On Bicompositional Correlation

Bergman, Jakob

2010

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

On Bicompositional Correlation

Jakob Bergman

Department of Statistics
Lund University

Lund 2010

Department of Statistics
Lund University
Box 743
SE-220 07 Lund
Sweden
www.stat.lu.se

Doctoral Theses in Statistics
ISSN 1651-7938

LUSADG/SAST-1012/1-93
ISBN 978-91-628-8028-6

Printed in Sweden by Media-Tryck, Lund 2010

Contents

Acknowledgements

This thesis marks the end of some 20 years of training in the Swedish education system. Naturally, in the course of a training of such length, you will as a student come across a multitude of professors, teachers and tutors; some of them you will always remember, and some of them should be consigned to oblivion. Three teachers that have been important to me during my school and undergraduate years are Werner Persson, Kerstin Elwin and Per Lindblad. The late Werner Persson was the first to introduce me to descriptive statistics such as the median, but above all he taught me the importance of knowing where you come from. It was during Kerstin Elwin's history classes that I was sent to the city archives and there for the first time understood what research is. Per Lindblad showed me the importance of being passionate about your subject and made me love phonetics. (Had not the job market for phoneticians been so completely hopeless, this thesis might just as well been about some interesting phonetic problem.)

Now to the people who deserve recognition for their efforts during my postgraduate years.

First and foremost I would like to thank my supervisor Björn Holmquist for all his help and encouragement. I have especially appreciated Björn's patience with with all my questions (and some of them have not been very clever), and his ability to always make me regain hope and enthusiasm when I have been discouraged.

I am greatly indebted to Jan Lanke for helping me improve Paper I. Jan has also corrected my English and pointed out several other errors in this thesis, all for which I am very grateful. (Whatever errors remain, are of course my own.)

Krzysztof Nowicki has been very encouraging and given me advice on various issues, for which I extend my thanks.

I have enjoyed my conversations with Sven Berg, Peter Gustafsson and Johannes Nilsson. All three have in their own special way helped me shape my view on research, teaching and statistics in general during our discussions. I also want to thank my other colleagues at the Department of Statistics for making my time as a postgraduate student such a pleasant one.

Finally I want thank my family, my wonderful wife Karin, my parents Claes and Susanne, and my brother Carl, for always being there and for their endless love, support and confidence in me.

JAKOB BERGMAN

List of papers

This thesis is based on the following papers:

I Bergman, J. (2009). A bicompositional Dirichlet distribution. Technical Report 3, Department of Statistics, Lund University.

II Bergman, J. and B. Holmquist (2009). A measure of dependence between two compositions. Technical report 4, Department of Statistics, Lund University.

III Bergman, J. (2009). Generating random variates from a bicompositional Dirichlet distribution. Technical Report 5, Department of Statistics, Lund University.

IV Bergman, J. (2010). Estimating a measure of dependence between two compositions. Technical Report 1, Department of Statistics, Lund University.

Introduction

1 Compositions

An essential part of statistics is analysing measurements of various entities. Normally these values make perfect sense; we may be interested in the number of cars, the velocity of each car, or the weight of each car. There are however situations when we are not interested in the absolute values of our measurements, but the relative ones; the absolute values may not even be available to us. The absolute amount of a certain oxide in a rock sample or the absolute number of respondents who would vote for a certain party in a party preference survey are seldom of interest, whereas the relative amount of a certain oxide and the relative number of respondents are usually more interesting. We often refer to these relative values as proportions. The proportions of all the different outcomes must of course sum to 1 (or 100 %). A vector of these proportions is known as a *composition*, or put more mathematically: a composition is a vector of positive components summing to a constant, usually taken to be 1. As indicated above, compositions arise in many different areas; the geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all examples of compositions from three different scientific areas.

The sample space of a composition is the simplex. Without loss of generality we will always take the summation constant to be 1, and we define the D-dimensional simplex $\mathscr{S}^D$ as

$$\mathscr{S}^D = \left\{ (x_1, \ldots, x_D)^{\mathrm{T}} \in \mathscr{R}_+^D : \sum_{j=1}^{D} x_j = 1 \right\},$$

where $\mathscr{R}_+$ is the positive real space.

In this thesis we will refer to compositions with two components (or parts), i.e. $D = 2$, as *bicomponent*, with three components, i.e. $D = 3$, as *tricomponent*, and with more than two components, i.e. $D > 2$, as *multicomponent*. Please note the difference between *bicompositional* referring to two compositions and *bicomponent* referring to a composition with two components. The two notions will be used together as in "a bicomponent bicompositional distribution," i.e. a joint distribution of two compositions each with two components.

2 A short historical review

Compositions have been studied almost as long as the subject of modern statistics has existed. Pearson (1897) was the first to realize that if you divide two independent random variates with a third random variate, independent of the first two, the two quotients will be correlated. Pearson called this "spurious correlation" and warned researchers for this phenomenon. This "spurious correlation" of course applies to compositions, since compositions are usually made up of a number of measurements divided by their sum; in fact for compositions the denominator is not even independent of the measurements. Since then it should have been known that compositions have to be treated with care. During the following 60 years this was however usually not the case.

In 1986 Aitchison published his pivotal book *The Statistical Analysis of Compositional Data* (reprinted 2003). In this book he argues for the concept of logratio transformations as a way to resolve the problems caused by the compositional summation constraint. Aitchison presented two logratio transformations: the additive logratio transformation (ALR) and the centred logratio transformation (CLR). Later Egozcue et al. (2003) introduced the isometric logratio transformation (ILR). The ALR transformation consists of the logarithms of the components, omitting one, divided by the omitted reference component; the CLR transformation consists of the logarithms of the components divided by the geometric mean of the components. The ILR transformation is a much more complex transformation. If for example $\mathbf{x} = (x_1, x_2, x_3, x_4)^{\mathrm{T}} \in \mathscr{S}^4$, then the resulting vectors of the different transformations are the following:

$$\begin{aligned}
\operatorname{alr}(\mathbf{x}) &= \left(\log \frac{x_1}{x_4},\ \log \frac{x_2}{x_4},\ \log \frac{x_3}{x_4}\right)^{\mathrm{T}} \\
\operatorname{clr}(\mathbf{x}) &= \left(\log \frac{x_1}{g(\mathbf{x})},\ \log \frac{x_2}{g(\mathbf{x})},\ \log \frac{x_3}{g(\mathbf{x})},\ \log \frac{x_4}{g(\mathbf{x})}\right)^{\mathrm{T}} \\
\operatorname{ilr}(\mathbf{x}) &= \left(\frac{1}{\sqrt{2}} \log \frac{x_1}{x_2},\ \frac{1}{\sqrt{6}} \log \frac{x_1 x_2}{x_3^2},\ \frac{1}{\sqrt{12}} \log \frac{x_1 x_2 x_3}{x_4^3}\right)^{\mathrm{T}}
\end{aligned}$$

where $g(\mathbf{x}) = (x_1 \cdots x_D)^{1/D}$, i.e. the geometric mean. The three transformations are related, see for instance Barceló-Vidal et al. (2007).

Aitchison and Egozcue (2005) distinguish four phases in the evolution of compositional analysis, the first one being the phase until 1960s when the complications with compositional data were ignored, and the second being the phase from the 1960s until the 1980s when different ideas were tried to resolve the problems of the multivariate methods not working for compositional data. The third phase is that when the logratio methodology gains acceptance. The fourth phase started some ten years ago, with the realization that the simplex is a Hilbert space (see e.g. Pawlowsky-Glahn and Egozcue, 2001, 2002). This has given rise to a “stay-in-the-simplex” approach. This approach basically provides a way of modelling the operations done on the logratio transformed data, then usually referred to as *coordinates*, in the simplex.

3 Compositional time series

The interest for bicompositional correlation resulting in this thesis originally began as an interest in compositional time series (CTS), i.e. time series of compositions. Compositional time series arise in many different situations, for instance party preference surveys, labour force surveys or pollution measurements.

Even though there have only been relatively few papers published on CTS, there have been several approaches to CTS; these have been reviewed by Larrosa (2005) and Aguilar Zuil et al. (2007).

The first to discuss and use an ALR approach to CTS seem to be Aitchison (1986) and Brunsdon (1987), which were followed by Smith and Brunsdon (1989) and Brunsdon and Smith (1998). In that approach the CTS is transformed with an ALR, and the transformed series is then analysed with standard models, e.g. VAR or VARMA. Bergman (2008) and Aguilar and Barceló-Vidal (2008) have also used ILR to model the data. The choice of logratio transformation is of course arbitrary.

There have also been some ideas on how to model the time series on the simplex. Apart from Aitchison and Brunsdon, Billheimer and Guttorp (1995) and Billheimer et al. (1997) have used autoregressive and conditional autoregressive models. Barceló-Vidal et al. (2007) introduced a compositional ARIMA model, defined using the "stay-in-the-simplex" approach.

As an illustration of CTS we present a figure from Bergman (2008), where a time series from the Swedish labour force survey (AKU) was modelled. Figure 1 gives three views of the analysed time series; the top plot shows the time series in a ternary time series plot (sometimes referred to as a "Toblerone plot"), the middle plot shows the three components of the time series in a standard time series plot, and the bottom plot shows a standard time series plot of the ILR-transformed time series. In all three plots the structural change in the series due to the Swedish fiscal crisis during the early 1990s is clearly visible, as well as a seasonal pattern.

Figure 1 (Next page) Three views of a compositional time series. The top plot shows the time series in a ternary time series plot, where the top corner of the Simplex represents 100 % Unemployment, the bottom left corner 100 % Employment, and the bottom right corner that 100 % of the population are Not belonging to the labour force. The middle plot shows the three components of the time series in a standard time series plot. (Note that the vertical axis has been cut and has different scales in the different parts.) The bottom plot shows the ILR-transformed series. (The second component of the transformed series is plotted with a dotted line.) In all three plots the structural change in the series during the early 1990s is clearly visible, as well as the seasonal pattern.

Source: Statistics Sweden.

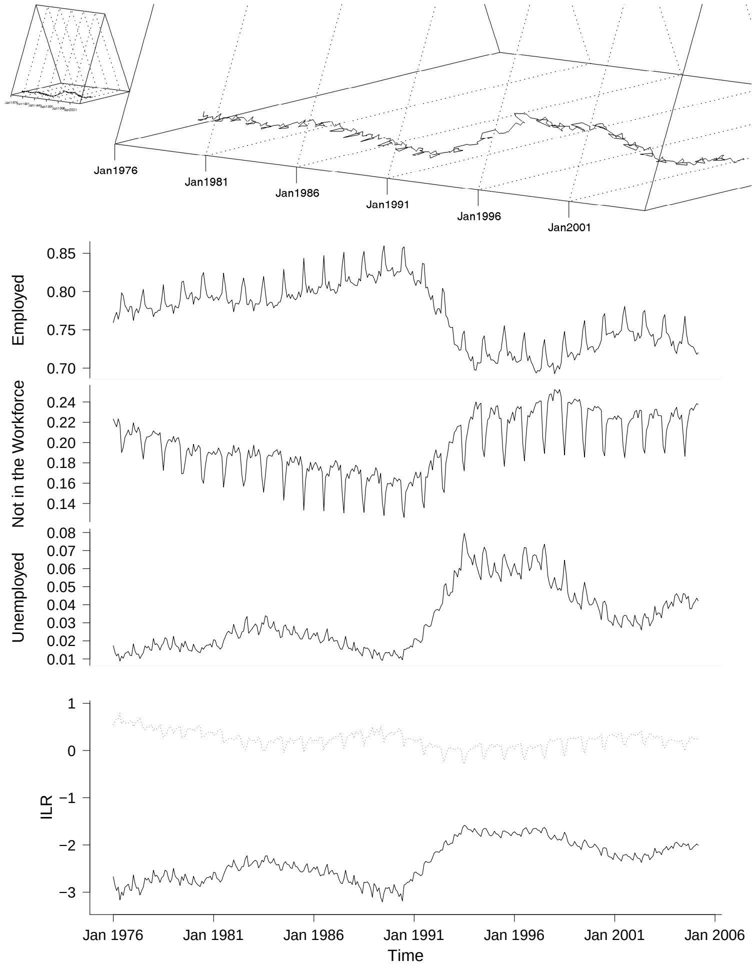
4 Correlation

Unlike the observations in cross-sectional data, the observations in time series are usually not independent. A not entirely unintuitive starting point for describing this dependence is to consider the concept of correlation. This thesis tries to target the question: "How do we model, measure and compare similarity or dissimilarity between two compositions?"

When hearing the word "correlation" most people would probability think of the product moment correlation coefficient

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}},$$

which measures the linear relationship between two variables. This is also how correlation is defined in *Encyclopedia of Statistical Sciences* (Rodriguez, 1982). However, correlation does not have be restricted to linear relationships or univariate variables. Dodge (2003) for instance states that it can be "used broadly to mean some kind of statistical relation between variables." This wider approach includes correlation coefficients that need not measure linear relationships, for instance the rank correlation coefficient Spearman's ρ_S. It is this wider approach we will utilize. We thus consider correlation as a measure



Jan1976
Jan1981
Jan1986
Jan1991
Jan1996
Jan2001
Employed
0.85
0.80
0.75
0.70
Not in the Workforce
0.24
0.22
0.20
0.18
0.16
0.14
Unemployed
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
ILR
1
0
−1
−2
−3
Jan 1976
Jan 1981
Jan 1986
Jan 1991
Jan 1996
Jan 2001
Jan 2006
Time

of similarity.

A good measure of correlation (or similarity) should also be able to compare not just two observations of the same composition at different time points, but also of two different compositions at the same time point. These two compositions might not even have equal numbers of components. We could for instance consider the correlation between some composition of the labour force and some composition of the gross domestic product. In this thesis we will however restrict our analysis to the correlation between two observations of the same composition, but with the introduction of suitable distributions, the result of this thesis is easily generalized to the above situations.

5 Bicompositions

In order to parametrically quantify the correlation between two compositions one needs to consider the joint distribution of the compositions. As stated above, the sample space of a D-component composition is the simplex $\mathscr{S}^D$. The sample space of two compositions $\mathbf{X}$, $\mathbf{Y}$, defined on $\mathscr{S}^D$, is consequently the Cartesian product $\mathscr{S}^D \times \mathscr{S}^D$. This is however not a simplex, but a manifold with two constraints, a *bisimplex*. We note that whereas the Cartesian product of two random vectors on the real space $\mathscr{R}^p$ will form a new random vector on the real space $\mathscr{R}^{p+p}$, this does not hold for two simplices: $\mathscr{S}^D \times \mathscr{S}^D \neq \mathscr{S}^{D+D}$.

The Cartesian product of two D-component compositions could have been denoted

$$\mathbf{Z} = (Z_1, \dots, Z_D, Z_{D+1}, \dots, Z_{D+D})^{\mathrm{T}},$$

where $\sum_{j=1}^{D} Z_j = \sum_{j=D+1}^{D+D} Z_j = 1$. However, throughout this thesis we choose to denote it

$$(\mathbf{X}, \mathbf{Y}) = (X_1, \dots, X_D, Y_1, \dots, Y_D)^{\mathrm{T}},$$

to stress the fact that we regard it primarily as two compositions and not as one *bicomposition*.

We will in this thesis base our modelling of correlation on an extension of the Dirichlet distribution. Following Aitchison (1986), we define the Dirichlet probability density function with parameter $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_D) \in \mathscr{R}_+^D$ as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_D)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_D)} x_1^{\alpha_1 - 1} \cdots x_D^{\alpha_D - 1},$$

where $\mathbf{x} = (x_1, \dots, x_D)^{\mathrm{T}} \in \mathscr{S}^D$ and $\Gamma(\cdot)$ is the Gamma function. We will present a bicompositional generalization of the Dirichlet distribution, defined on the Cartesian product of two simplices, i.e. a bisimplex. The notation $(\mathbf{X}, \mathbf{Y})$ will also allow us to emphasize the relationship between the new distribution and the product of two Dirichlet distributions.

In accordance with the Dirichlet integral, the new distribution is defined with respect to the Lebesgue measure. It remains as future work to reformulate it using the Aitchison (or simplicial) measure (Pawlowsky-Glahn, 2003) along the lines of Mateu-Figueras and Pawlowsky-Glahn (2005).

6 Outline of the thesis

This thesis is based on four papers concerning bicompositions and modelling the correlation between compositions. The contents of the papers are presented briefly below.

6.1 Paper I

We search the literature for distributions defined on the Cartesian product $\mathscr{S}^D \times \mathscr{S}^D$, and find a few bivariate Beta distributions for the bicomponent case, but no distributions defined on $\mathscr{S}^D \times \mathscr{S}^D$ when $D > 2$.

We introduce a bicompositional Dirichlet distribution. The distribution is defined on the Cartesian product $\mathscr{S}^D \times \mathscr{S}^D$ and is based on the product of two Dirichlet distributions. The probability density function is

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = A\left(\prod_{j=1}^{D} x_j^{\alpha_j-1} y_j^{\beta_j-1}\right)\left(\mathbf{x}^{\mathrm{T}}\mathbf{y}\right)^{\gamma},$$

where $\mathbf{x} = (x_1,\ldots,x_D)^{\mathrm{T}} \in \mathscr{S}^D$, $\mathbf{y} = (y_1,\ldots,y_D)^{\mathrm{T}} \in \mathscr{S}^D$, and $\alpha_j, \beta_j \in \mathscr{R}_+$ ($j = 1,\ldots,D$). The parameter space of γ depends on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; however, all non-negative values are always included. The parameter γ models the degree of covariation between $\mathbf{X}$ and $\mathbf{Y}$. When $\gamma = 0$, the distribution is the product of two independent Dirichlet distributions.

We prove that the distribution exists in the bicomponent case if and only if $\gamma > -\min(\alpha_1+\beta_2, \alpha_2+\beta_1)$ and at least for $\gamma \geq 0$ in the multicomponent case. We also give expressions for the normalization constant A for all γ in the bicomponent case and for integers γ in multicomponent case.

In the bicomponent case we present expressions for the cumulative distribution function and the product moment. In both the bicomponent and the multicomponent case, we derive expressions for the marginal probability density functions and the marginal moments, and for the conditional probability density distribution and conditional moments.

6.2 Paper II

We consider two families of parametric models $\{f(\mathbf{x},\mathbf{y};\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_i\}$ ($i = 0, 1$) with $\Theta_0 \subset \Theta_1$ when modelling $(\mathbf{X},\mathbf{Y})$ and assume that the true joint density function is $g(\mathbf{x},\mathbf{y})$. Kent (1983) defines the Fraser information as

$$F(\boldsymbol{\theta}) = \int \log f(\mathbf{x},\mathbf{y};\boldsymbol{\theta}) g(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}$$

and the information gain as

$$\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0) = 2\{F(\boldsymbol{\theta}_1) - F(\boldsymbol{\theta}_0)\},$$

where θ_i is the parameter value that maximizes $F(\theta)$ under the parameter space Θ_i ($i = 0, 1$). Using $\Gamma(\theta_1 : \theta_0)$, Kent (1983) proposes a general measure of correlation, or joint correlation coefficient, between $(\mathbf{X}, \mathbf{Y})$ defined as

$$\rho_J^2 = 1 - \exp\{-\Gamma(\theta_1 : \theta_0)\},$$

where $\mathbf{X}$ and $\mathbf{Y}$ are modelled as independent quantities under Θ_0.

We use the bicompositional Dirichlet distribution presented in Paper I to model two compositions $\mathbf{X}$ and $\mathbf{Y}$. We let $\theta = (\alpha, \beta, \gamma)$ and $\Theta_0 = \{\theta : \gamma = 0\}$, while Θ_1 is the unrestricted parameter space.

The joint correlation coefficient is calculated, utilizing that the bicompositional Dirichlet distribution constitutes an exponential family of distributions, and it is presented graphically for a large number of bicomponent bicompositional models. We note that ρ_J^2 as a function of γ is not symmetric around 0.

We also calculate the joint correlation coefficient for nine tricomponent bicompositional models.

In the Appendices we present and examine expressions for the first derivative of the binomial coefficient

$$\frac{\mathrm{d}}{\mathrm{d}r}\binom{r}{n},$$

and we also give a suggestion for numerical integration over $\mathscr{S}^3 \times \mathscr{S}^3$.

6.3 Paper III

We use the rejection method to generate random variates with a bicompositional Dirichlet density f. Given a dominating density g and a constant $c \geq 1$ such that $f(\mathbf{x}, \mathbf{y}) \leq cg(\mathbf{x}, \mathbf{y})$, and a random number U uniformly distributed on the unit interval, a generated variate $(\mathbf{x}, \mathbf{y})$ is accepted if

$$U \leq \frac{f(\mathbf{x}, \mathbf{y})}{cg(\mathbf{x}, \mathbf{y})},$$

otherwise it is rejected and new $(\mathbf{x}, \mathbf{y})$ and U are generated until acceptance. We hence need to find dominating densities g and constants c. We examine three cases.

First we look at the (trivial) case when $\gamma = 0$, i.e. the product of two independent Dirichlet distributions. Dirichlet distributed variates are easily generated using Gamma distributed variates, and thus we need not use the rejection method.

Secondly we examine the case when $\gamma > 0$. We use a bicompositional Dirichlet distribution with $\gamma = 0$, i.e. the product of two independent Dirichlet distribution, as dominating density. We find that the random variate is accepted if $U \leq (\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma}$. Evidently, we need not calculate the normalization constant $A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$, and hence we can generate random numbers from bicompositional Dirichlet distributions whose probability density functions we cannot calculate. When γ is very large, the method will be slow, as the acceptance probability $\Pr\{U \leq (\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma}\} = (\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma}$ will be very low. We note that we can always use a uniform density as g, with $c = \max_{\mathbf{x},\mathbf{y}} f(\mathbf{x}, \mathbf{y})$. This is though only applicable for non-negative integers γ, since it is necessary to calculate $A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$.

Thirdly we examine the bicomponent case when $\gamma < 0$. We partition the sample space into four quadrants Q1-Q4, and choose a quadrant Qk ($k = 1, 2, 3, 4$) randomly with probability

$$\iint_{\mathrm{Q}k} f(x, y)\,\mathrm{d}x\,\mathrm{d}y \qquad (k = 1, 2, 3, 4),$$

where $f(x, y)$ is the bicomponent bicompositional Dirichlet probability density function viewed as a function of x and y. For each of the quadrants we find a dominating density based on the product of two Dirichlet distributions and a constant c, and generate a random variate using the rejection method. A slight problem with the method is to find effective ways of generating random Dirichlet distributed variates that are restricted to a particular quadrant.

We compare the efficiencies of the two suggestions for dominating densities, Dirichlet and uniform, with a Monte Carlo study.

6.4 Paper IV

We present maximum likelihood estimates of the parameter $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ of the bicompositional Dirichlet distribution presented in Paper I. Following Kent (1983) we also present an estimator of the general measure of correlation, or joint correlation coefficient, presented in Paper II, assuming that the data follow a bicompositional Dirichlet distribution,

$$\hat{\rho}_J^2 = 1 - \exp\{-\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0)\},$$

where $\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0)$ is an estimator of the information gain when allowing for dependence,

$$\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) = \frac{2}{n}\left(\sum_{k=1}^{n} \log f(\mathbf{x}_k, \mathbf{y}_k; \hat{\boldsymbol{\theta}}_1) - \sum_{k=1}^{n} \log f(\mathbf{x}_k, \mathbf{y}_k; \hat{\boldsymbol{\theta}}_0)\right),$$

and $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_0$ are the maximum likelihood estimates under the parameter spaces Θ_1 and Θ_0, respectively.

We also present two confidence intervals for the joint correlation coefficient: one when $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ is large,

$$\left[1 - \exp\left\{-\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) + \sqrt{s^2\chi^2_{1;\alpha}/n}\right\},\right.$$
$$\left. 1 - \exp\left\{-\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) - \sqrt{s^2\chi^2_{1;\alpha}/n}\right\}\right],$$

where s^2 is the sample variance of $2\log\{f(\mathbf{x}_j, \mathbf{y}_j; \hat{\boldsymbol{\theta}}_1)/f(\mathbf{x}_j, \mathbf{y}_j; \hat{\boldsymbol{\theta}}_0)\}$ and $\chi^2_{1;\alpha}$ is the upper α quantile of the χ^2_1 distribution; and one when $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ is small,

$$\left[1 - \exp\left\{-\frac{\varkappa_{1;\alpha/2}(\hat{a})}{n}\right\}, 1 - \exp\left\{-\frac{\delta_{1;\alpha/2}(\hat{a})}{n}\right\}\right],$$

where $\varkappa_{1;\alpha/2}$ and $\delta_{1;\alpha/2}$ are non-centrality parameters of certain χ_1^2 distributions and $\hat{a} = n\widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0)$.

Using a Monte Carlo study, we compare the empirical confidence coefficients of the two intervals for a number of models. The random variates are generated by means of the method described in Paper III. It is apparent for the models that we have examined that the "small" confidence interval (based on non-central χ^2-distributions) will produce the smaller intervals, yielding an empirical confidence coefficient for almost all models of approximately 95 %, when the nominal confidence coefficient is 95 %. The "large" confidence intervals will in general be wider.

We also examine a bias correction, suggested by Kent (1983), of the information gain estimator. This correction involves the second derivative of the binomial coefficient

$$\frac{\mathrm{d}^2}{\mathrm{d}r^2}\binom{r}{n},$$

and an expression for this is given in the appendix of that paper. In our examples, however, the suggested correction actually yields estimates that are more biased than the uncorrected ones. We believe that this might be due to numerical issues, as the correction involves a large number of infinite sums. Due to this lack of improvement we have not used this bias correction in our estimations.

As an example we have also estimated the general measure of correlation for GDP data from the 50 U.S. states and District of Columbia. The estimate of the general measure of correlation is

$$\hat{\rho}_J^2 = 0.3027,$$

with a "small" confidence interval of

$$(0.0993, 0.5371)$$

thus indicating that composition of the government GDP in 1967 is correlated with that in 1997.

References

Aguilar, L. and C. Barceló-Vidal (2008, May). Multivariate ARIMA compositional time series analysis. In J. Daunis i Estadella and J. Martín-Fernández (Eds.), *Proceedings of CoDaWork'08, The 3rd Compositional Data Analysis Workshop, CD-ROM*. Univeristy of Girona, Girona (Spain).

Aguilar Zuil, L., C. Barceló-Vidal, and J. M. Larrosa (2007). Compositional time series analysis: A review. In *Proceedings of the 56th Session of the ISI (ISI 2007)*, Lisboa, August 22-29.

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

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.

Aitchison, J. and J. J. Egozcue (2005). Compositional data analysis: Where are we and where should we be heading? *Mathematical Geology 37*(7), 829–850.

Barceló-Vidal, C., L. Aguilar, and J. Martín-Fernández (2007). Time series of compostional data: A first approach. In *Proceedings of the 22nd International Workshop of Statistical Modelling (IWSM 2007), Barcelona, July 2-6*, pp. 81–86.

Bergman, J. (2008, May). Compositional time series: An application. In J. Daunis i Estadella and J. Martín-Fernández (Eds.), *Proceedings of CoDaWork'08, The 3rd Compositional Data Analysis Workshop, CD-ROM*. Univeristy of Girona, Girona (Spain).

Billheimer, D. and P. Guttorp (1995, Nov). Spatial models for discrete compositional data. Technical report, Dept. of Statistics, University of Washington, Seattle.

Billheimer, D., P. Guttorp, and W. F. Fagan (1997). Statistical analysis and interpretation of discrete compositional data. Technical Report Series 11, NRSCE.

Brunsdon, T. M. (1987). *Time series analysis of compositional data*. Ph. D. thesis, Dept. of Mathematics, University of Southampton.

Brunsdon, T. M. and T. M. F. Smith (1998). The time series analysis of compositional data. *Journal of Official Statistics 14*(3), 237–253.

Dodge, Y. (Ed.) (2003). *The Oxford Dictionary of Statistical Terms* (6th ed.). Oxford: Oxford University Press.

Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology 35*(3), 279–300.

Kent, J. T. (1983). Information gain and a general measure of correlation. *Biometrika 70*(1), 163–173.

Larrosa, J. M. (2005, Oct). Compositional time series: Past and present. Technical Report Econometrics 0510002, EconWPA.

Mateu-Figueras, G. and V. Pawlowsky-Glahn (2005, October). The Dirichlet distribution with respect to the Aitchison measure on the simplex - a first approach. In G. Mateu-Figueras and C. Barceló-Vidal (Eds.), *Proceedings of CoDaWork'05, The 2nd Compositional Data Analysis Workshop*. Universitat de Girona.

Pawlowsky-Glahn, V. (2003). Statistical modelling on coordinates. In S. Thió-Henestrosa and J. A. Martín-Fernández (Eds.), *Proceedings of CoDaWork'03, Compositional Data Analysis Workshop*. Universitat de Girona.

Pawlowsky-Glahn, V. and J. J. Egozcue (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment 15*(5), 384–398.

Pawlowsky-Glahn, V. and J. J. Egozcue (2002). BLU estimators and compositional data. *Mathematical Geology 34*(3), 259–274.

Pearson, K. (1897). Mathematical contributions to the theory of evolution.—on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London LX*, 489–498.

Rodriguez, R. N. (1982). Correlation. In S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Volume 2. New York: John Wiley & Sons.

Smith, T. M. F. and T. M. Brunsdon (1989). The time series analysis of compositional data. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 26–32.

I

A bicompositional Dirichlet distribution

Jakob Bergman

Abstract

The simplex $\mathscr{S}^D$ is the sample space of a D-part composition. There are only a few distributions defined on the simplex and even fewer defined on the Cartesian product $\mathscr{S}^D \times \mathscr{S}^D$. Based on the Dirichlet distribution, defined on $\mathscr{S}^D$, we propose a new bicompositional Dirichlet distribution defined on $\mathscr{S}^D \times \mathscr{S}^D$, and examine some of its properties, such as moments as well as marginal and conditional distributions. The proposed distribution allows for modelling of covariation between compositions without leaving $\mathscr{S}^D \times \mathscr{S}^D$.

Keywords

Cartesian product · Compositional data · Dirichlet distribution · Simplex

1 Introduction

1.1 Compositions

A composition is a vector of positive components summing to a constant, usually 1. The components of a composition are what we usually think of as proportions (at least when the vector sums to 1). Compositions arise in many different areas; the geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all three examples of compositions from different scientific areas. For more examples of compositions, see for instance Aitchison (1986) or the reprint Aitchison (2003).

Compositions differ from other multivariate random vectors on the real space due the summation constraint. Whereas the Cartesian product of two random vectors on the real space $\mathscr{R}^p$ will form a new random vector on the real space $\mathscr{R}^{p+p}$, this does not hold for two simplices: $\mathscr{S}^D \times \mathscr{S}^D \neq \mathscr{S}^{D+D}$.

When describing dependency structures, compositional analysis has been primarily concerned with describing the dependency structures within a composition, i.e. the relation between the components of a composition. Aitchison (1986, Ch. 10) for instance devotes an entire chapter to this, and as a recent example Ongaro et al. (2008) construct a new distribution for modelling such relations. In this paper we will not be considering the relation between the components of a composition, but the relation between two compositions. We will use the term *bicompositional* when referring to two compositions (with same number of components) and the term *unicompositional* when referring to one composition. A composition with two components will be referred to as a *bicomponent* composition, as opposed to a *multicomponent* composition.

1.2 The simplex

The sample space of a composition is the simplex. (For simplicity we will always take the summing constant to be 1.) We define the D-dimensional simplex space $\mathscr{S}^D$ as

$$\mathscr{S}^D = \left\{ (x_1, \ldots, x_D)^{\mathrm{T}} \in \mathscr{R}_+^D : \sum_{j=1}^{D} x_j = 1 \right\}$$

where $\mathscr{R}_+$ is the positive real space. As noted above, it is a sample space that occurs in a wide variety of applications. There are however only a limited number of distributions defined on the simplex; the two most notable are the

Dirichlet distribution and the logistic normal distribution class described by Aitchison and Shen (1980). There are also a number of generalizations of these two distributions.

The sample space of the joint distribution of two compositions is the Cartesian product $\mathscr{S}^D \times \mathscr{S}^D$, which is not a simplex, but a manifold with two constraints. $\mathscr{S}^D \times \mathscr{S}^D$ is the subspace of $\mathscr{R}_+^{D+D}$, where

$$\sum_{j=1}^{D} x_j = \sum_{j=D+1}^{D+D} x_j = 1.$$

There are almost no distributions defined on $\mathscr{S}^D \times \mathscr{S}^D$. For the case $D = 2$, the bicomponent case, there have been proposed a few bivariate Beta distributions (though usually not in a compositional context). See for instance in recent years Olkin and Liu (2003), Nadarajah and Kotz (2005), Nadarajah (2006) or Nadarajah (2007). For $D > 2$, the multicomponent case, we have not found any distributions at all in the literature.

2 An example

As an example we look at the proportion x of the Gross Domestic Product (GDP) due to agriculture etc., mining, utilities, construction and manufacturing for the 50 U.S. states and District of Columbia. Every observation hence is a composition $(x, 1-x)^{\mathrm{T}}$. (Data come from the Bureau of Economic Analysis, U.S. Department of Commerce.) The observations for 2007 are plotted versus the observations for 1997 in Figure 1. We see that for almost all of the states the proportion x is less for 2007 than for 1997. For 1997 most states vary between $(0.2, 0.8)$ and $(0.4, 0.6)$, but for 2007 most are between $(0.1, 0.9)$ and $(0.2, 0.8)$. To be able to model this relation we need distributions defined on $\mathscr{S}^2 \times \mathscr{S}^2$.

We could of course also have divided the GDP into three parts, e.g. (Agriculture, mining etc.; Other private industries; Government). This would

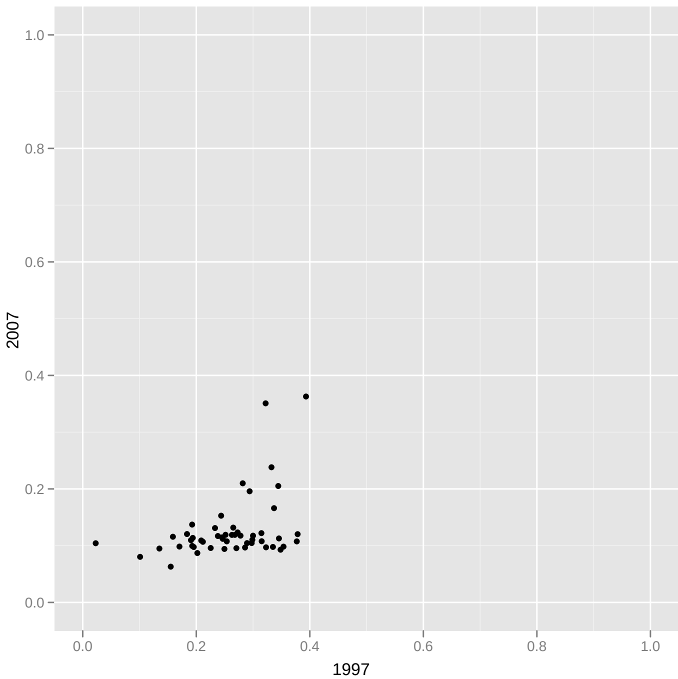


Figure 1 The combined agriculture, mining, utilities, construction, and manufacturing proportion of the GDP for the 50 U.S. states and District of Columbia for 2007 plotted versus the same proportion for 1997.

Source: Bureau of Economic Analysis, U.S. Department of Commerce.

however require a four-dimensional plot. We acknowledge though the need for modelling the covariation between two compositions with D components in the space $\mathscr{S}^D \times \mathscr{S}^D$.

As well as studying the joint probability density of two compositions, we could study the probability density for a composition conditioned on the values of another composition. In our example we might study a composition in 2007, conditioned on the value in 1997. Also for this context we need suitable distributions to be able to create models.

3 A bicompositional Dirichlet distribution

Following Aitchison (1986), we define $d = D - 1$ and $\mathbf{x} = (x_1, \ldots, x_d, x_D)^{\mathrm{T}}$, and let $\mathbf{x} \in \mathscr{S}^D$. The well-known (unicompositional) Dirichlet distribution with parameter $\boldsymbol{\alpha} \in \mathscr{R}_+^D$ is defined as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_D)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_D)} x_1^{\alpha_1 - 1} \cdots x_D^{\alpha_D - 1}$$

where $\Gamma(\cdot)$ is the Gamma function.

Based on the Dirichlet distribution we will now define a bicompositional generalization for $\mathbf{X}$ and $\mathbf{Y}$.

Definition 1 (Probability Density Function)**.** Let $\mathbf{X}, \mathbf{Y} \in \mathscr{S}^D$ and let

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = A \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) \left(\mathbf{x}^{\mathrm{T}} \mathbf{y} \right)^{\gamma} \tag{1}$$

where $\alpha_j, \beta_j \in \mathscr{R}_+$ for $j = 1, \ldots, D$ and γ is a real number.

The parameter γ models the degree of covariation between $\mathbf{X}$ and $\mathbf{Y}$. If $\gamma = 0$, (1) reduces to the product of two independent Dirichlet probability density functions

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = A \prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1}$$

where A is known:

$$A = \frac{\Gamma(\alpha_1 + \cdots + \alpha_D)\,\Gamma\left(\beta_1 + \cdots + \beta_D\right)}{\prod_{j=1}^{D} \Gamma(\alpha_j)\Gamma(\beta_j)}.$$

This motivates that (1) is referred to as the probability density function of a bicompositional Dirichlet distribution. We let $\mathfrak{D}_1^D(\boldsymbol{\alpha})$ denote the unicompositional Dirichlet distribution with D components and parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)^{\mathrm{T}}$ and we let $\mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ denote the bicompositional Dirichlet distribution with D components and parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and γ, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_D)^{\mathrm{T}}$.

Next we examine some of the properties of this distribution: first in the special bicomponent case, and then in the general multicomponent case.

4 The bicomponent case

For two bicomponent compositions, $\mathbf{X} = (X, 1 - X)^{\mathrm{T}}$ and $\mathbf{Y} = (Y, 1 - Y)^{\mathrm{T}}$, the probability density function is a function of $\mathbf{x} = (x, 1 - x)^{\mathrm{T}}$ and $\mathbf{y} = (y, 1 - y)^{\mathrm{T}}$, but as it is completely determined by (x, y) we will for simplicity treat it as function of (x, y). Hence the probability density function (1) is reduced to

$$f_{X,Y}(x, y) = A x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} y^{\beta_1 - 1}(1-y)^{\beta_2 - 1} \left\{xy + (1 - x)(1 - y)\right\}^{\gamma}. \quad (2)$$

This distribution could of course also be considered a bivariate Beta distribution, if we regard it as a function of (x, y).

We begin our investigation of the distribution by stating for what values of γ the distribution exists.

Theorem 1. *The distribution $\mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ exists if and only if $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$.*

Proof. If γ is a non-negative then the last factor in

$$x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1} y^{\beta_1 - 1}(1 - y)^{\beta_2 - 1} \left\{xy + (1 - x)(1 - y)\right\}^{\gamma} \quad (3)$$

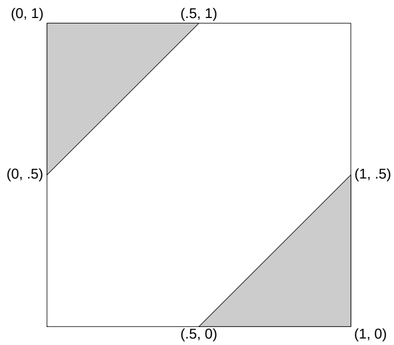


Figure 2 The integration area of (4). The denominator is close to 0 only when (x, y) is close to either $(0, 1)$ or $(1, 0)$.

is bounded for $x, y \in [0, 1]$ and the integral of (3) exists. If γ is negative then the last factor is unbounded.

Let γ be negative and $g = -\gamma$. Then the integral of (3) may be written as

$$\int_0^1 \int_0^1 \frac{x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} y^{\beta_1 - 1}(1-y)^{\beta_2 - 1}}{\{xy + (1-x)(1-y)\}^g} \,\mathrm{d}x\,\mathrm{d}y. \tag{4}$$

The denominator is close to 0 only when (x, y) is close to either $(0, 1)$ or $(1, 0)$. For the rest of the integration area the denominator is bounded away from 0. In the triangle $(0, 0.5)$, $(0, 1)$, $(0.5, 1)$, i.e. the top left triangle, $\{xy + (1-x)(1-y)\}$ may be estimated from below by $(x + 1 - y)/2$ and from above by $(x + 1 - y)$. In the bottom right triangle, $\{xy + (1-x)(1-y)\}$ may be estimated from below by $(y + 1 - x)/2$ and from above by $(y + 1 - x)$. The integration area, with the top left and bottom right triangles shaded, is shown in Figure 2.

The integral (4) thus exists if and only if the integral

$$\int_{\frac{1}{2}}^{1} \left(\int_{0}^{y-\frac{1}{2}} \frac{x^{\alpha_1 - 1}(1-y)^{\beta_2 - 1}}{(x+1-y)^g} \, \mathrm{d}x \right) \mathrm{d}y \tag{5}$$

and the corresponding integral over the bottom triangle both exist. Introducing $u = x$ and $v = x + 1 - y$, (5) turns into

$$\int_{0}^{\frac{1}{2}} \left(\int_{0}^{v} \frac{u^{\alpha_1 - 1}(v-u)^{\beta_2 - 1}}{v^g} \, \mathrm{d}u \right) \mathrm{d}v. \tag{6}$$

If we replace u by vt, (6) may be written as

$$\int_{0}^{\frac{1}{2}} v^{\alpha_1 + \beta_2 - g - 1} \, \mathrm{d}v \int_{0}^{1} t^{\alpha_1 - 1}(1-t)^{\beta_2 - 1} \, \mathrm{d}t.$$

The second integral of this product always exists, but the first one exists if and only if $\alpha_1 + \beta_2 > g$. With an analogous argument for the integral over the bottom right triangle, we can show that this integral exists if and only if $\alpha_2 + \beta_1 > g$. Hence the density (2) exists if and only if $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$. □

We next determine the normalization constant A.

Theorem 2. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, the normalization constant A is determined by*

$$\begin{aligned} \frac{1}{A} = \frac{1}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} & \left(\sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B(\alpha_1 + j, \alpha_2 + i - j) \right) \\ & \cdot \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right) \end{aligned} \tag{7}$$

where $B(p, q)$ is the Beta function:

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,\mathrm{d}t = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Proof. Let ξ and η be defined by $x = \frac{1}{2}(1+\xi)$ and $y = \frac{1}{2}(1+\eta)$. Then $1-x = \frac{1}{2}(1-\xi)$, $1-y = \frac{1}{2}(1-\eta)$ and $xy + (1-x)(1-y) = \frac{1}{2}(1+\xi\eta)$ and the Binomial expansion yields

$$\{xy + (1-x)(1-y)\}^{\gamma} = \frac{1}{2^{\gamma}} \sum_{i=0}^{\infty} \binom{\gamma}{i} (\xi\eta)^i. \tag{8}$$

We note that, since $0 < x < 1$ and $0 < y < 1$, and thus $-1 < \xi < 1$ and $-1 < \eta < 1$, the series on the right-hand side of (8) converges. The normalization constant is then determined by

$$\begin{aligned}
\frac{1}{A} &= \frac{1}{2^{\gamma}} \sum_{i=0}^{\infty} \binom{\gamma}{i} \iint x^{\alpha_1-1}(1-x)^{\alpha_2-1} y^{\beta_1-1}(1-y)^{\beta_2-1}(\xi\eta)^i\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{2^{\gamma}} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\int x^{\alpha_1-1}(1-x)^{\alpha_2-1}\xi^i\,\mathrm{d}x\right)\left(\int y^{\beta_1-1}(1-y)^{\beta_2-1}\eta^i\,\mathrm{d}y\right) \\
&= \frac{1}{2^{\gamma}} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\int x^{\alpha_1-1}(1-x)^{\alpha_2-1}\{x-(1-x)\}^i\,\mathrm{d}x\right) \\
&\qquad\qquad \cdot \left(\int y^{\beta_1-1}(1-y)^{\beta_2-1}\{y-(1-y)\}^i\,\mathrm{d}y\right) \\
&= \frac{1}{2^{\gamma}} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^{i} \binom{i}{j}(-1)^{i-j} B(\alpha_1+j, \alpha_2+i-j)\right) \\
&\qquad\qquad \cdot \left(\sum_{k=0}^{i} \binom{i}{k}(-1)^{i-k} B(\beta_1+k, \beta_2+i-k)\right)
\end{aligned}$$

where all integrals are over the unit interval. $\square$

Empirical trials show that the series (7) converges quickly for most examples. If all the parameters are close to 0, the series may converge very slowly. We have also found that convergence can be slow when γ is negative and close to $-\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$. These situations however mean that the probability is concentrated near the edges of the sample space and hence perhaps of little practical importance.

If γ is a non-negative integer the results are simplified as shown next.

Theorem 3. *If $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the normalization constant A is determined by*

$$\frac{1}{A} = \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + j, \alpha_2 + \gamma - j) B(\beta_1 + j, \beta_2 + \gamma - j) \tag{9}$$

where $B(\cdot, \cdot)$ is the Beta function.

Proof. The result follows from expanding the last factor of (2) using the Binomial theorem

$$\{xy + (1-x)(1-y)\}^{\gamma} = \sum_{j=0}^{\gamma} \binom{\gamma}{j} x^j y^j (1-x)^{\gamma-j} (1-y)^{\gamma-j} \tag{10}$$

and then integrating

$$\sum_{j=0}^{\gamma} \binom{\gamma}{j} x^{\alpha_1-1}(1-x)^{\alpha_2-1} y^{\beta_1-1}(1-y)^{\beta_2-1} x^j y^j (1-x)^{\gamma-j}(1-y)^{\gamma-j}$$

using the definition of the Beta function. $\square$

We proceed by also stating the cumulative distribution function.

Theorem 4. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$*, the cumulative distribution function is*

$$\begin{aligned} F_{X,Y}(x, y) = \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} &\left(\sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_x(\alpha_1 + j, \alpha_2 + i - j) \right) \\ &\cdot \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B_y(\beta_1 + k, \beta_2 + i - k) \right) \end{aligned} \tag{11}$$

where A *is the constant given in Theorem 2 and* $B_x(p, q)$ *is the incomplete Beta function defined as*

$$B_x(p, q) = \int_0^x t^{p-1}(1 - t)^{q-1} \, \mathrm{d}t.$$

Proof. The proof is similar to the proof of Theorem 2, but uses the definition of the incomplete Beta function instead of the Beta function. □

If γ is a non-negative integer the cumulative distribution function may be expressed more simply.

Theorem 5. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* γ *is a non-negative integer, the cumulative distribution function is*

$$F_{X,Y}(x, y) = A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B_x(\alpha_1 + j, \alpha_2 + \gamma - j) B_y(\beta_1 + j, \beta_2 + \gamma - j) \tag{12}$$

where A *is the constant given in Theorem 3 and* $B_x(\cdot, \cdot)$ *is the incomplete Beta function.*

Proof. The proof is similar to the proof of Theorem 3, but uses the definition of the incomplete Beta function instead of the Beta function. □

Next we give the product moments of the distribution.

Theorem 6. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$, *the product moment* $E(X^n Y^m)$ *is*

$$\begin{aligned}
E(X^n Y^m) &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B(\alpha_1 + n + j, \alpha_2 + i - j) \right) \\
&\qquad \cdot \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B(\beta_1 + m + k, \beta_2 + i - k) \right) \\
&= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B(\alpha_1 + n + j, \alpha_2 + i - j) G(\boldsymbol{\alpha}, j, \gamma, n) \right) \\
&\qquad \cdot \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B(\beta_1 + m + k, \beta_2 + i - k) G(\boldsymbol{\beta}, k, \gamma, m) \right)
\end{aligned}$$

where

$$G(\boldsymbol{\alpha}, j, \gamma, n) = \prod_{l=0}^{n-1} \frac{\alpha_1 + j + l}{\alpha_1 + \alpha_2 + \gamma + l} \tag{13}$$

and A is the constant given in Theorem 2.

Proof. The proof of the first equality is similar to that of Theorem 2. The second equality follows from repeated use of the identity

$$B(p+1, q) = \frac{p}{p+q} B(p, q). \tag{14}$$

□

As before, if γ is a non-negative integer, the calculations are simplified.

Theorem 7. *If* $(X,Y) \in \mathfrak{D}_2^2(\alpha_1,\alpha_2,\beta_1,\beta_2;\gamma)$ *and* γ *is a non-negative integer, the product moment* $E(X^nY^m)$ *is*

$$\begin{aligned} E(X^nY^m) &= A\sum_{j=0}^{\gamma}\binom{\gamma}{j}B(\alpha_1+n+j,\alpha_2+\gamma-j)B(\beta_1+m+j,\beta_2+\gamma-j) \\ &= A\sum_{j=0}^{\gamma}\binom{\gamma}{j}B(\alpha_1+j,\alpha_2+\gamma-j)B(\beta_1+j,\beta_2+\gamma-j) \\ &\qquad\qquad\qquad\qquad \cdot G(\boldsymbol{\alpha},j,\gamma,n)G(\boldsymbol{\beta},j,\gamma,m) \end{aligned}$$

where G *is given in (13) and* A *is the constant given in Theorem 3.*

Proof. The proof of the first equality is similar to the proof of Theorem 3. The second equality follows from repeated use of the identity (14). □

4.1 Marginal distributions

In the example in Section 2 we noted that not only the joint distribution, but also the conditional distributions may be of interest when modelling bi-compositional data. In order to determine the properties of the conditional distributions, we first need to derive some of the properties of the marginal distributions.

Theorem 8. *If* $(X,Y) \in \mathfrak{D}_2^2(\alpha_1,\alpha_2,\beta_1,\beta_2;\gamma)$ *and* $\gamma > -\min(\alpha_1+\beta_2,\alpha_2+\beta_1)$*, the marginal probability density function of* X *is*

$$\begin{aligned} f_X(x) = \frac{A}{2^\gamma}\sum_{i=0}^{\infty}\binom{\gamma}{i}x^{\alpha_1-1}(1-x)^{\alpha_2-1}\{x-(1-x)\}^i \\ \cdot\left(\sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}B(\beta_1+k,\beta_2+i-k)\right) \end{aligned} \quad (15)$$

where A *is the constant given in Theorem 2.*

Proof. The result follows by expanding (2) as in the proof of Theorem 2 to get

$$\frac{A}{2^\gamma}\sum_{i=0}^{\infty}\binom{\gamma}{i}x^{\alpha_1-1}(1-x)^{\alpha_2-1}\{x-(1-x)\}^i y^{\beta_1-1}(1-y)^{\beta_2-1}\{y-(1-y)\}^i,$$

then using the Binomial Theorem to expand

$$\{y-(1-y)\}^i=\sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}y^k(1-y)^{i-k}, \tag{16}$$

and finally integrating

$$\frac{A}{2^\gamma}\sum_{i=0}^{\infty}\binom{\gamma}{i}x^{\alpha_1-1}(1-x)^{\alpha_2-1}\{x-(1-x)\}^i \cdot y^{\beta_1-1}(1-y)^{\beta_2-1}\sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}y^k(1-y)^{i-k}$$

with respect to y using the definition of the Beta function. □

As previously the results are simplified for non-negative integer values on γ.

Theorem 9. *If $(X,Y)\in\mathfrak{D}_2^2(\alpha_1,\alpha_2,\beta_1,\beta_2;\gamma)$ and γ is a non-negative integer, the marginal probability density function of X is*

$$f_X(x)=A\sum_{j=0}^{\gamma}\binom{\gamma}{j}B(\beta_1+j,\beta_2+\gamma-j)x^{\alpha_1+j-1}(1-x)^{\alpha_2+\gamma-j-1} \tag{17}$$

where A is the constant given in Theorem 3.

Proof. The result follows by expanding the last factor of (2) using the identity (10) and then integrating with respect to y using the definition of the Beta function. $\square$

When $\gamma = 0$, we see that (17) is reduced to the common unicompositional Dirichlet distribution.

For completeness we also state the moments of the marginal distributions.

Theorem 10. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$*, the nth moment of* X *is*

$$\begin{aligned} E(X^n) &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B(\alpha_1 + n + j, \alpha_2 + i - j) \right) \\ &\qquad \cdot \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right) \\ &= \frac{A}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \left(\sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B(\alpha_1 + j, \alpha_2 + i - j) \right) \\ &\qquad \cdot \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right) \\ &\qquad \cdot \left(\prod_{l=0}^{n-1} \frac{\alpha_1 + i + l}{\alpha_1 + \alpha_2 + \gamma + l} \right) \end{aligned}$$

where A *is the constant given in Theorem 2.*

Proof. The result follows by expanding $\{x - (1-x)\}^i$ in (15) using the identity (16), integrating the product of (15) and x^n, and repeatedly using the identity (14). $\square$

Again, the calculations are simplified for integer values.

Theorem 11. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* γ *is a non-negative integer, the nth moment of* X *is*

$$
\begin{aligned}
E(X^n) &= A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + n + j, \alpha_2 + \gamma - j) B(\beta_1 + j, \beta_2 + \gamma - j) \\
&= A \sum_{j=0}^{\gamma} \binom{\gamma}{j} B(\alpha_1 + j, \alpha_2 + \gamma - j) B(\beta_1 + j, \beta_2 + \gamma - j) G(\boldsymbol{\alpha}, j, \gamma, n)
\end{aligned}
$$

where A *is the constant given in Theorem 3 and* $G(\boldsymbol{\alpha}, j, \gamma, n)$ *is given in (13).*

Proof. The result follows from direct calculation and repeated use of the identity (14). □

We note that if $\gamma = 0$, then $E(X) = \frac{\alpha_1}{\alpha_1 + \alpha_2}$, i.e. precisely the expectation of a Dirichlet distribution, as one would expect.

Due to the symmetry of the bicompositional Dirichlet distribution, all of the above results of course also apply to Y (with appropriate changes of α_i to β_i and vice versa).

4.2 Conditional distributions

We now proceed to the bicomponent conditional distributions, first stating the conditional probability density function.

Theorem 12. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$*, the conditional probability density function for* Y *conditioned on* $X = x$ *is*

$$
f_{Y|X=x}(y) = C y^{\beta_1 - 1} (1 - y)^{\beta_2 - 1} \left\{ xy + (1 - x)(1 - y) \right\}^{\gamma} \tag{18}
$$

where

$$
\frac{1}{C} = \frac{1}{2^{\gamma}} \sum_{i=0}^{\infty} \binom{\gamma}{i} \{x - (1 - x)\}^i \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B(\beta_1 + k, \beta_2 + i - k) \right).
$$

Proof. The result follows directly from (2) and Theorem 8. □

Theorem 13. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* γ *is a non-negative integer, the conditional probability density function for* Y *conditioned on* $X = x$ *is*

$$f_{Y|X=x}(y) = D \sum_{k=0}^{\gamma} \binom{\gamma}{k} x^k (1-x)^{\gamma-k} y^{\beta_1+k-1}(1-y)^{\beta_2+\gamma-k-1} \tag{19}$$

where

$$\frac{1}{D} = \sum_{j=0}^{\gamma} \binom{\gamma}{j} x^j (1-x)^{\gamma-j} B(\beta_1 + j, \beta_2 + \gamma - j).$$

Proof. The result follows from directly from (2) and Theorem 9. □

We note that when $\gamma = 0$, (19) simplifies to the probability density function of the unicompositional Dirichlet distribution with parameters β_1 and β_2.

We also derive the moments for the conditional distributions.

Theorem 14. *If* $(X, Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ *and* $\gamma > -\min(\alpha_1+\beta_2, \alpha_2+\beta_1)$*, the nth moment of* $Y|X = x$ *is*

$$E(Y^n|X = x) =$$
$$\frac{C}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \{x - (1-x)\}^i \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} B(\beta_1 + n + k, \beta_2 + i - k) \right) \tag{20}$$

where C *was given in Theorem 12*

Proof. The result follows by using the identity (8) on (19) to get

$$\frac{C}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \{x - (1-x)\}^i y^{\beta_1-1}(1-y)^{\beta_2-1}\{y - (1-y)\}^i,$$

than expanding $\{y-(1-y)\}^i$ using (16), and finally integrating

$$y^n \frac{C}{2^\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} \{x-(1-x)\}^i y^{\beta_1-1}(1-y)^{\beta_2-1} \left(\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} y^k (1-y)^{i-k} \right)$$

with respect to y using the definition of the Beta function. $\square$

Theorem 15. *If $(X,Y) \in \mathfrak{D}_2^2(\alpha_1, \alpha_2, \beta_1, \beta_2; \gamma)$ and γ is a non-negative integer, the nth moment of $Y|X=x$ is*

$$E(Y^n|X=x) = D \sum_{j=0}^{\gamma} \binom{\gamma}{j} x^j (1-x)^{\gamma-j} B(\beta_1+n+j, \beta_2+\gamma-j) \quad (21)$$

where D was given in Theorem 13.

Proof. The result follows by integrating the product of y^n and (19) using the definition of the Beta function. $\square$

In both (20) and (21), the Beta function may be expanded to a product similar to the ones in Theorems 10 and 11.

5 The multicomponent case

We now turn to the case when there are $D > 2$ components. Unfortunately, due to the increased complexity, the results for this case are less elaborate than for the bicomponent case.

The distribution exists if γ is a non-negative real number. If γ furthermore is a non-negative integer, the normalization constant may easily be determined. To simplify the expressions, we let $\alpha_{.} = \alpha_1 + \cdots + \alpha_D$, $\beta_{.} = \beta_1 + \cdots + \beta_D$, $k_{.} = k_1 + \cdots + k_D$ and $\mathbf{k} = (k_1, \ldots, k_D)$. We denote the multinomial coefficient

$$\binom{\gamma}{\mathbf{k}} = \frac{\gamma!}{k_1! \cdots k_D!}.$$

Theorem 16. *If* $(\mathbf{X},\mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha},\boldsymbol{\beta};\gamma)$ *and* γ *is a non-negative integer, then the normalization constant is determined by*

$$\frac{1}{A} = \sum_{\substack{\mathbf{k}\geq 0\\ k_.=\gamma}} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^D \Gamma(\alpha_i + k_i)}{\Gamma(\alpha_. + \gamma)} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)}. \tag{22}$$

Proof. The result follows from expanding the factor $(\mathbf{x}^{\mathrm{T}}\mathbf{y})^\gamma$ by using the Multinomial Theorem and from the properties of the Dirichlet integral:

$$\begin{aligned}
\frac{1}{A} &= \int_{\mathbf{y}}\int_{\mathbf{x}} \left(\prod_{j=1}^D x_j^{\alpha_j-1} y_j^{\beta_j-1}\right) \left(\mathbf{x}^{\mathrm{T}}\mathbf{y}\right)^\gamma \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \\
&= \int_{\mathbf{y}}\int_{\mathbf{x}} \left(\prod_{j=1}^D x_j^{\alpha_j-1} y_j^{\beta_j-1}\right) \sum_{\substack{\mathbf{k}\geq 0\\ k_.=\gamma}} \binom{\gamma}{\mathbf{k}} x_1^{k_1} y_1^{k_1} \cdots x_D^{k_D} y_D^{k_D} \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \\
&= \sum_{\substack{\mathbf{k}\geq 0\\ k_.=\gamma}} \binom{\gamma}{\mathbf{k}} \int_{\mathbf{x}} \prod_{j=1}^D x_j^{\alpha_j+k_j-1} \,\mathrm{d}\mathbf{x} \int_{\mathbf{y}} \prod_{j=1}^D y_j^{\beta_j+k_j-1} \,\mathrm{d}\mathbf{y} \qquad \square
\end{aligned}$$

For computational purposes, we note that (22) can for instance be written as

$$\frac{1}{A} = \sum_{k_1=0}^{\gamma} \sum_{k_2=0}^{\gamma-k_1} \cdots \sum_{k_d=0}^{\gamma-\sum_{i=1}^{d-1} k_i} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^D \Gamma(\alpha_i + k_i)}{\Gamma(\alpha_. + \gamma)} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)}$$

if we define $k_D = \gamma - k_1 - \cdots - k_d$.

5.1 Marginal distributions

Before we examine the conditional distributions we give the marginal distributions for the multicomponent case.

First we give the probability density function.

Theorem 17. *If* $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ *and* γ *is a non-negative integer, then the marginal probability density function for* $\mathbf{X}$ *is*

$$f_{\mathbf{X}}(\mathbf{x}) = A \sum_{\substack{\mathbf{k} \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{\alpha_1 + k_1 - 1} \cdots x_D^{\alpha_D + k_D - 1} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)}. \tag{23}$$

where A *is the constant given in Theorem 16.*

Proof. The proof is similar to that of Theorem 16, but only integrating over $\mathbf{y}$. □

Just as for the bicomponent case, we note that when $\gamma = 0$, (23) simplifies to a unicompositional Dirichlet distribution with parameter $\boldsymbol{\alpha}$. When γ is a positive integer, (23) just like (1) is a mixture of unicompositional Dirichlet distributions.

Next we determine the moments of the components of the marginal distributions.

Theorem 18. *If* $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ *and* γ *is a non-negative integer, then*

$$E(X_j^n) = A \sum_{\substack{\mathbf{k} \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^D \Gamma(\alpha_i + k_i)}{\Gamma(\alpha_. + \gamma)} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)} \prod_{l=0}^{n-1} \frac{\alpha_j + k_j + l}{\alpha_. + \gamma + l} \tag{24}$$

for $j = 1, \ldots, D$ *and* $n = 1, 2, \ldots$.

Proof. The result follows by using the Dirichlet integral and the identity $\Gamma(x+1) = x\Gamma(x)$ repeatedly. □

5.2 Conditional distributions

Having determined the marginal distributions we continue by determining the multicomponent conditional distributions. We begin with the conditional probability density function of the composition $\mathbf{Y}$ conditioned on the composition $\mathbf{X} = \mathbf{x}$.

Theorem 19. *If* $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ *and* γ *is a non-negative integer, then the conditional probability density function for* $\mathbf{Y}$ *given* $\mathbf{X} = \mathbf{x}$ *is*

$$f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \frac{y_1^{\beta_1 - 1} \cdots y_D^{\beta_D - 1} (x_1 y_1 + \cdots + x_D y_D)^\gamma}{\sum_{\substack{\mathbf{k} \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{k_1} \cdots x_D^{k_D} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)}}. \tag{25}$$

Proof. The result follows directly from (2) and Theorem 17. □

As before, we note that if $\gamma = 0$, (25) reduces to a unicompositional Dirichlet distribution.

Using (25) we next determine the moments of the multicomponent conditional distribution.

Theorem 20. *If* $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{D}_2^D(\boldsymbol{\alpha}, \boldsymbol{\beta}; \gamma)$ *and* γ *is a non-negative integer, then*

$$E(Y_j^n | \mathbf{X} = \mathbf{x}) = B \sum_{\substack{\mathbf{k} \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{k_1} \cdots x_D^{k_D} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)} \prod_{l=0}^{n-1} \frac{\beta_j + k_j + l}{\beta_. + \gamma + l} \tag{26}$$

for $j = 1, \ldots, D$ *and* $n = 1, 2, \ldots$*; here*

$$\frac{1}{B} = \sum_{\substack{\mathbf{k} \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} x_1^{k_1} \cdots x_D^{k_D} \frac{\prod_{i=1}^D \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)}.$$

Proof. The proof is similar to the proof of Theorem 18. □

6 An example, cont'd

To illustrate the bicompositional Dirichlet distribution, we fit a bicompositional Dirichlet model to the data presented in Section 2 using the method of maximum likelihood. The contour lines of the model are shown in Figure 3 together with the 51 observations.

7 Discussion

There are not many distributions defined on the sample space consisting of the Cartesian product of two Simplices. In this paper we have proposed a bicompositional generalization of the unicompositional Dirichlet distribution. The proposed bicompositional Dirichlet distribution allows for modelling of covariation between compositions without leaving $\mathscr{S}^D \times \mathscr{S}^D$. We have determined in the bicomponent case for what parameter values the distribution exists. We have also derived the marginal and conditional distributions and their moments. For the bicomponent case we have also derived the cumulative distribution function and the product moment.

The distribution is defined with respect to the Lebesgue measure in accordance with the Dirichlet integral. It remains as future work to reformulate it using the Aitchison (or simplicial) measure (Pawlowsky-Glahn, 2003) along the lines of Mateu-Figueras and Pawlowsky-Glahn (2005).

The proposed distribution is meant to be a first suggestion for modelling bicompositional data and it is developed primarily to possess properties that will exploited in future work. Among these properties are, apart from simplicity, the ability to model dependence and independence between compositions and also the fact that the distributions constitute an exponential family of distributions.

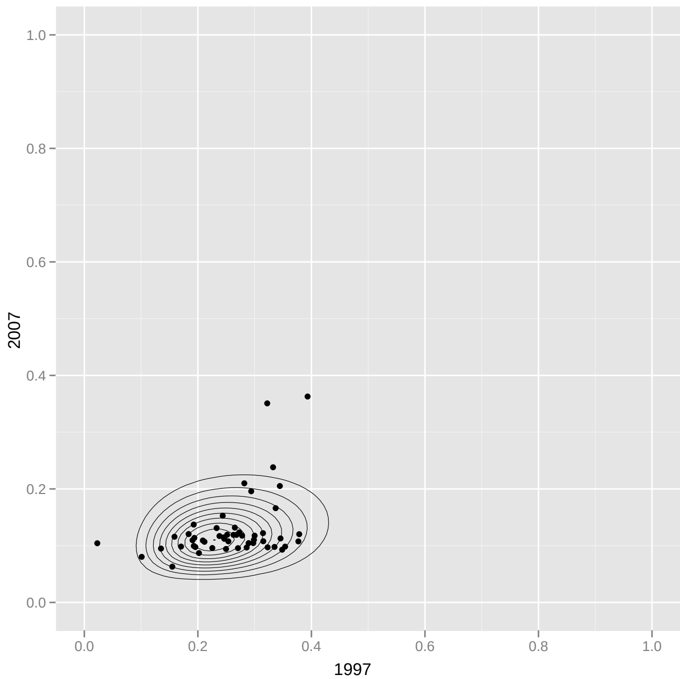


Figure 3 The data from Figure 1 plotted together with the contour lines of the density with the parameter values estimated with maximum likelihood estimators from the data.

Source: Bureau of Economic Analysis, U.S. Department of Commerce.

8 Acknowledgements

The author wishes to thank Professor Jan Lanke for all his time and effort in helping improving this article.

References

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

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.

Aitchison, J. and S. M. Shen (1980). Logistic-normal distributions: Some properties and uses. *Biometrika 67*(2), 261–272.

Mateu-Figueras, G. and V. Pawlowsky-Glahn (2005, October). The Dirichlet distribution with respect to the Aitchison measure on the simplex - a first approach. In G. Mateu-Figueras and C. Barceló-Vidal (Eds.), *Proceedings of CoDaWork'05, The 2nd Compositional Data Analysis Workshop*. Universitat de Girona.

Nadarajah, S. (2006). The bivariate F_3-beta distribution. *Communications of the Korean Mathematical Society 21*(2), 363–374.

Nadarajah, S. (2007). A new bivariate beta distribution with application to drought data. *Metron LXV*(2), 153–174.

Nadarajah, S. and S. Kotz (2005). Some bivariate beta distributions. *Statistics 39*(5), 457–466.

Olkin, I. and R. Liu (2003). A bivariate beta distribution. *Statistics and Probability Letters 62*(4), 407–412.

Ongaro, A., S. Migliorati, and G. Monti (2008). A new distribution on the simplex containing the Dirichlet family. In J. Daunis i Estadella and J. Martín-Fernández (Eds.), *Proceedings of CoDaWork'08, The 3rd Compositional Data Analysis Workshop, CD-ROM*. Universitat de Girona.

Pawlowsky-Glahn, V. (2003). Statistical modelling on coordinates. In S. Thió-Henestrosa and J. A. Martín-Fernández (Eds.), *Proceedings of CoDaWork'03, Compositional Data Analysis Workshop*. Universitat de Girona.

II

A measure of dependence between two compositions

Jakob Bergman and Björn Holmquist

Abstract

A composition is a vector of positive components summing to a constant. We consider the problem of describing the correlation between two compositions. Using a bicompositional Dirichlet distribution, we calculate a joint correlation coefficient, based on the concept of information gain, between two compositions. Numerical values of the joint correlation coefficient are calculated for compositions of two and three components.

Keywords

Binomial coefficient differentiation · Composition · Correlation · Dirichlet distribution · Simplex

1 Introduction

A composition is a vector of positive components summing to a constant, usually taken to be 1. The components of a composition are what we usually think of as proportions (at least when the vector sums to 1). Compositions arise in many different areas; the geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all examples of compositions from three different scientific areas. For more examples of compositions, see for instance Aitchison (2003).

The sample space of a composition is the simplex. Without loss of generality we will always take the summing constant to be 1, and we define the D-dimensional simplex $\mathscr{S}^D$ as

$$\mathscr{S}^D = \left\{ (x_1, \ldots, x_D)^{\mathrm{T}} \in \mathscr{R}_+^D : \sum_{j=1}^{D} x_j = 1 \right\},$$

where $\mathscr{R}_+$ is the positive real space. The joint sample space of two compositions is the Cartesian product of two simplices $\mathscr{S}^D \times \mathscr{S}^D$. It should be noted that, unlike the case for real spaces, $\mathscr{S}^D \times \mathscr{S}^D \neq \mathscr{S}^{D+D}$ and that $\mathscr{S}^D \times \mathscr{S}^D$ is not a even simplex, but a manifold with two constraints.

The compositional data analysis has previously been concerned with describing how the components of a composition correlate, i.e. the intra-compositional dependence. The components of a composition are not independent due to the summation constraint. A review of different independence concepts pertaining to partitions of a composition is presented in (Aitchison, 2003, Chap. 10).

Correlation between compositions has however previously not been given very much attention. We investigate the dependence between two compositions, i.e. the inter-compositional dependence, using a measure of dependence suggested in Kent (1983) based on the concept of information gain. We believe that a measure of inter-compositional dependence is needed in order to describe, for instance, the spatial similarity between two geochemical compositions measured at different locations, or the temporal similarity between party preference surveys conducted at different times.

2 Information gain and Kent's general measure of correlation

If we consider two families of parametric models $\{f(\mathbf{x}, \mathbf{y}; \theta), \theta \in \Theta_i\}$ $(i = 0, 1)$ with $\Theta_0 \subset \Theta_1$ and the true joint density function is $g(\mathbf{x}, \mathbf{y})$, the Fraser information is defined in Kent (1983) as

$$F(\theta) = \int \log f(\mathbf{x}, \mathbf{y}; \theta) g(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y}, \tag{1}$$

that is, $F(\theta)$ is the expected log-likelihood.

By choosing θ_i to maximize $F(\theta)$ in the parameter space Θ_i, "θ_i is the theoretical analogue of the maximum likelihood estimate of θ over the parameter space Θ_i" (Kent, 1983). We will in the following partition θ into two parts $\theta = (\psi, \lambda)$, where ψ is the parameter of interest and λ is a nuisance parameter.

If the model forms a canonical exponential family, that is

$$f(\mathbf{x}, \mathbf{y}; \theta) = \exp\{\psi^{\mathrm{T}} \mathbf{v}(\mathbf{x}, \mathbf{y}) + \lambda^{\mathrm{T}} \mathbf{w}(\mathbf{x}, \mathbf{y}) - c(\theta)\},$$

the Fraser information may be calculated as

$$F(\theta_i) = \theta_i^{\mathrm{T}} \mathbf{b}(\theta_i) - c(\theta_i), \tag{2}$$

where $\mathbf{b}(\theta)$ is the vector of partial derivatives of $c(\theta)$ with respect to θ.

If for $\Theta_0 = \{\theta : \psi = \mathbf{0}\}$, $\mathbf{X}$ and $\mathbf{Y}$ are modelled as independent, the information gain of allowing for dependence between $\mathbf{X}$ and $\mathbf{Y}$ in the model is

$$\Gamma(\theta_1 : \theta_0) = 2\{F(\theta_1) - F(\theta_0)\}.$$

Since $F(\theta_i)$ is the maximized expected log likelihood, $\Gamma(\theta_1 : \theta_0)$ is the theoretical analogue of -2 times the log likelihood ratio statistic.

Kent (1983) proposed a joint correlation coefficient between $\mathbf{X}$ and $\mathbf{Y}$ defined as

$$\rho_J^2 = 1 - \exp\{-\Gamma(\theta_1 : \theta_0)\}.$$

As is easily seen, $0 \leq \rho_J^2 \leq 1$.

Independence between $\mathbf{X}$ and $\mathbf{Y}$ implies zero correlation if $g(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ for some $\boldsymbol{\theta}$, or "the model Θ_1 forms a regular exponential family" Inaba and Shirahata (1986).

3 The bicompositional Dirichlet distribution

In order to calculate a joint correlation coefficient between two compositions a suitable distribution is needed. Unfortunately very few distributions with dependence structures defined on $\mathscr{S}^D \times \mathscr{S}^D$ are available. One distribution for modelling random vectors on $\mathscr{S}^D \times \mathscr{S}^D$ is proposed in Bergman (2009). The proposed distribution, called the bicompositional Dirichlet distribution, has the probability density function

$$f(\mathbf{x}, \mathbf{y}) = A \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) \left(\mathbf{x}^{\mathrm{T}} \mathbf{y} \right)^{\gamma}, \qquad (3)$$

where $\mathbf{x} = (x_1, \ldots, x_D)^{\mathrm{T}} \in \mathscr{S}^D$, $\mathbf{y} = (y_1, \ldots, y_D)^{\mathrm{T}} \in \mathscr{S}^D$, and $\alpha_j, \beta_j \in \mathscr{R}_+$ ($j = 1, \ldots, D$). The parameter space of γ depends on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; however, all non-negative values are always included. Expressions for the normalization constant A are given in Bergman (2009). If $\gamma = 0$, the probability density function (3) is equal to the product of two Dirichlet probability density functions with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively, and hence $\mathbf{X}$ and $\mathbf{Y}$ are independent in that case.

When $\mathbf{X}, \mathbf{Y} \in \mathscr{S}^2$ we shall refer to this as the *bicomponent* case, and similarly to $\mathscr{S}^3$ as the *tricomponent* case and to $\mathscr{S}^D (D > 2)$ as the *multicomponent* case.

The bicompositional Dirichlet distribution forms a canonical exponential family with parameters $\boldsymbol{\theta} = (\gamma, \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})^{\mathrm{T}}$, where

$$\begin{aligned} \tilde{\alpha}_j &= \alpha_j - 1, \\ \tilde{\beta}_j &= \beta_j - 1. \end{aligned}$$

We shall assume that the true density function $g(\mathbf{x}, \mathbf{y})$ is the bicompositional Dirichlet probability density function (3) and that the two families of parametric models $f(\mathbf{x}, \mathbf{y}; \theta_i)$ also are bicompositional Dirichlet distributions. The parameter of interest in these models is $\psi = \gamma$. Denoting $\theta_1 = (\gamma^{(1)}, \tilde{\boldsymbol{\alpha}}^{(1)}, \tilde{\boldsymbol{\beta}}^{(1)})^{\mathrm{T}}$ and $\theta_0 = (\gamma^{(0)}, \tilde{\boldsymbol{\alpha}}^{(0)}, \tilde{\boldsymbol{\beta}}^{(0)})^{\mathrm{T}}$, it can be shown, through the information inequality, that $\gamma^{(1)} = \gamma$, $\tilde{\boldsymbol{\alpha}}^{(1)} = \tilde{\boldsymbol{\alpha}}$, and $\tilde{\boldsymbol{\beta}}^{(1)} = \tilde{\boldsymbol{\beta}}$, but when $\gamma^{(0)} = 0$, in general $\tilde{\boldsymbol{\alpha}}^{(0)} \neq \tilde{\boldsymbol{\alpha}}$ and $\tilde{\boldsymbol{\beta}}^{(0)} \neq \tilde{\boldsymbol{\beta}}$.

3.1 The bicomponent case

If we define

$$S_{\tilde{\alpha}} = \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B(\tilde{\alpha}_1 + j + 1; \tilde{\alpha}_2 + i - j + 2), \tag{4}$$

where $B(a, b)$ is the Beta function, and $S_{\tilde{\beta}}$ in the same way as equation (4) but with $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ replaced with $\tilde{\beta}_1$ and $\tilde{\beta}_2$, we may for the bicomponent bicompositional Dirichlet distribution with parameters $\theta = (\gamma, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)^{\mathrm{T}}$ derive the following expressions:

$$c(\theta) = \log\left\{2^{-\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} S_{\tilde{\beta}}\right\}, \tag{5}$$

$$\mathbf{b}(\theta) = \left(\frac{\partial c}{\partial \gamma}, \quad \frac{\partial c}{\partial \tilde{\alpha}_1}, \quad \frac{\partial c}{\partial \tilde{\alpha}_2}, \quad \frac{\partial c}{\partial \tilde{\beta}_1}, \quad \frac{\partial c}{\partial \tilde{\beta}_2}\right)^{\mathrm{T}}. \tag{6}$$

The partial derivatives of (5) in (6) are

$$\frac{\partial c}{\partial \gamma} = \frac{\sum_{i=0}^{\infty} \binom{\gamma}{i}' S_{\tilde{\alpha}} S_{\tilde{\beta}}}{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} S_{\tilde{\beta}}} - \log 2,$$

$$\frac{\partial c}{\partial \tilde{\alpha}_1} = \frac{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\beta}} \left\{ \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\tilde{\boldsymbol{\alpha}}) \Psi_{ij}^{(1)}(\tilde{\alpha}_1, \tilde{\alpha}_2) \right\}}{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} S_{\tilde{\beta}}},$$

$$\frac{\partial c}{\partial \tilde{\alpha}_2} = \frac{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\beta}} \left\{ \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\tilde{\boldsymbol{\alpha}}) \Psi_{ij}^{(2)}(\tilde{\alpha}_1, \tilde{\alpha}_2) \right\}}{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} S_{\tilde{\beta}}},$$

$$\frac{\partial c}{\partial \tilde{\beta}_1} = \frac{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} \left\{ \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\tilde{\boldsymbol{\beta}}) \Psi_{ij}^{(1)}(\tilde{\beta}_1, \tilde{\beta}_2) \right\}}{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} S_{\tilde{\beta}}},$$

$$\frac{\partial c}{\partial \tilde{\beta}_2} = \frac{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} \left\{ \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\tilde{\boldsymbol{\beta}}) \Psi_{ij}^{(2)}(\tilde{\beta}_1, \tilde{\beta}_2) \right\}}{\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} S_{\tilde{\beta}}},$$

where

$$\binom{\gamma}{i}' = \frac{\mathrm{d}}{\mathrm{d}\gamma} \binom{\gamma}{i}, \tag{7}$$

$$B_{ij}(\tilde{\boldsymbol{\alpha}}) = B(\tilde{\alpha}_1 + j + 1; \tilde{\alpha}_2 + i - j + 1), \tag{8}$$

$$\Psi_{ij}^{(1)}(\tilde{\alpha}_1, \tilde{\alpha}_2) = \Psi(\tilde{\alpha}_1 + j + 1) - \Psi(\tilde{\alpha}_1 + \tilde{\alpha}_2 + i + 2), \tag{9}$$

$$\Psi_{ij}^{(2)}(\tilde{\alpha}_1, \tilde{\alpha}_2) = \Psi(\tilde{\alpha}_2 + i - j + 1) - \Psi(\tilde{\alpha}_1 + \tilde{\alpha}_2 + i + 2). \tag{10}$$

Analogous expressions of equations (8)-(10) for $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are implied. The function in equations (9) and (10) is the digamma function

$$\Psi(x) = \frac{\mathrm{d} \log \Gamma(x)}{\mathrm{d}x}.$$

Expressions for the derivative of the binomial coefficient (7) are discussed in Appendix A.

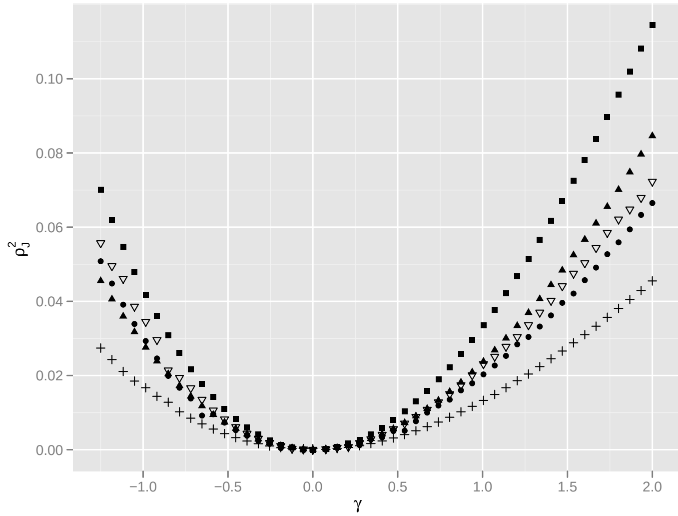


Figure 1 The joint correlation coefficient ρ_J^2 calculated for γ ranging from -1.25 to 2.0 for the $(\boldsymbol{\alpha};\boldsymbol{\beta})$ parameter values (2.1, 2.4; 2.2, 2.3) (■), (2.1, 2.2; 3.6, 3.5) (▲), (5.2, 2.0; 2.0, 2.0) (▽), (1.9, 6.4; 3.2, 2.1) (●) and (4.1, 2.4; 4.1, 2.4) (+).

If $\Theta_0 = \{\boldsymbol{\theta} : \gamma = 0\}$, the joint correlation coefficient ρ_J^2 may be calculated through the information gain as described earlier. However $F(\boldsymbol{\theta}_0)$ requires maximization, usually numerical, with respect to $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$.

Figure 1 depicts the joint correlation coefficient ρ_J^2, calculated for five different sets of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ values and 49 of values of γ ranging from -1.25 to 2.0. As can be seen in the figure, the joint correlation coefficient depends primarily on the value of γ but also to some extent on the rest of the parameters. It should be noted that ρ_J^2 is not symmetric around 0; the rate at which

ρ_J^2 changes differs for negative and positive γ and we note that the vertical order of the five graphs in the figure are different for negative and positive γ. The small deviations in the curvature of the graphs, e.g. at -0.65, are due numerical issues.

3.2 The tricomponent case

Since the normalization constant of the bicompositional Dirichlet distribution in the multicomponent case hitherto is only calculated for non-negative integers γ, we may only calculate the joint correlation coefficient for such γ values. This also disables us from using equation (2) in the calculations, as differentiation with respect to γ is not meaningful. We will instead utilize the definition given in equation (1).

The Fraser information for the tricomponent bicompositional Dirichlet distribution is the following:

$$
\begin{aligned}
F(\theta_i) &= \int \log\{f(\mathbf{x},\mathbf{y};\theta_i)\} g(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \\
&= \int \log\left\{ A x_1^{\alpha_1^{(i)}-1} x_2^{\alpha_2^{(i)}-1} x_3^{\alpha_3^{(i)}-1} y_1^{\beta_1^{(i)}-1} y_2^{\beta_2^{(i)}-1} y_3^{\beta_3^{(i)}-1} (\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma^{(i)}} \right\} g(\mathbf{x},\mathbf{y}) d\mathbf{x} d\mathbf{y} \\
&= \log A + (\alpha_1^{(i)} - 1) \int \log(x_1) g(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \\
&\qquad + \cdots \\
&\qquad + (\beta_3^{(i)} - 1) \int \log(y_3) g(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \\
&\qquad + \gamma^{(i)} \int \log(x_1 y_1 + x_2 y_2 + x_3 y_3) g(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}
\end{aligned}
$$

Thus $F(\theta_i)$ equals the sum of a constant, six log expectations, and the expectation $E\{\log(\mathbf{x}^{\mathrm{T}}\mathbf{y})\}$. (For the sake of brevity we will use the notation $a_. = a_1 + \cdots + a_D$ for the rest of this section.)

Before proceeding, we note that the Dirichlet distribution with parameter $\boldsymbol{\alpha}$ constitutes an exponential family of distributions, with sufficient statistic $\mathbf{T}(\mathbf{x}) = (\log x_1, \ldots, \log x_D)^{\mathrm{T}}$ and normalization constant

$$A(\boldsymbol{\alpha}) = \sum_{j=1}^{D} \log \Gamma(\alpha_j) - \log \Gamma(\alpha_.).$$

Since $E\{\mathbf{T}(\mathbf{X})\} = \partial A(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}$ for distributions that constitute an exponential family, we conclude that the log expectation of a Dirichlet distribution with parameter $\boldsymbol{\alpha}$ is

$$E\{\log(X_j)\} = \Psi(\alpha_j) - \Psi(\alpha_.). \tag{11}$$

Using the Multinomial Theorem and equation (11), we may calculate the first seven terms of $F(\boldsymbol{\theta}_i)$ exactly. For example:

$$\begin{aligned}
&\int \log(x_j) g(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \\
&= \int \log(x_j) A x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} x_3^{\alpha_3 - 1} y_1^{\beta_1 - 1} y_2^{\beta_2 - 1} y_3^{\beta_3 - 1} (x_1 y_1 + x_2 y_2 + x_3 y_3)^{\gamma}\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \\
&= \int \log(x_j) \sum_{\substack{k_i \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} A\, \mathbf{x}^{\boldsymbol{\alpha} + \mathbf{k} - 1} \mathbf{y}^{\boldsymbol{\beta} + \mathbf{k} - 1}\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \\
&= A \sum_{\substack{k_i \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} \int \log(x_j)\, \mathbf{x}^{\boldsymbol{\alpha} + \mathbf{k} - 1}\, \mathrm{d}\mathbf{x} \int \mathbf{y}^{\boldsymbol{\beta} + \mathbf{k} - 1}\, \mathrm{d}\mathbf{y} \\
&= A \sum_{\substack{k_i \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^{3} \Gamma(\alpha_1 + k_i)}{\Gamma(\alpha_. + \gamma)} \frac{\prod_{i=1}^{3} \Gamma(\beta_1 + k_i)}{\Gamma(\beta_. + \gamma)} \left\{ \Psi(\alpha_j + k_j) - \Psi(\alpha_. + \gamma) \right\}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{x}^{\boldsymbol{\alpha} + \mathbf{k} - 1} &= x_1^{\alpha_1 + k_1 - 1} x_2^{\alpha_2 + k_2 - 1} x_3^{\alpha_3 + k_3 - 1}, \\
\mathbf{y}^{\boldsymbol{\beta} + \mathbf{k} - 1} &= y_1^{\beta_1 + k_1 - 1} y_2^{\beta_2 + k_2 - 1} y_3^{\beta_3 + k_3 - 1}
\end{aligned}$$

and the multinomial coefficient is denoted

$$\binom{\gamma}{\mathbf{k}} = \frac{\gamma!}{k_1!k_2!k_3!}.$$

The integral $\int \log(y_j)g(\mathbf{x}, \mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}$ analogously yields the same result except for the last factor, where α_j and $\alpha_.$ are replaced by β_j and $\beta_.$ respectively.

The last term of $F(\boldsymbol{\theta}_1)$ must be integrated numerically. (See Appendix B for integration over $\mathscr{S}^3 \times \mathscr{S}^3$.) This is not the case for $F(\boldsymbol{\theta}_0)$, as $\gamma^{(0)} = 0$, but instead, in order to obtain $\boldsymbol{\theta}_0$, $F(\boldsymbol{\theta})$ must be maximized with respect to $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$.

In Figure 2 the joint correlation coefficient is plotted for γ ranging from -2 to 8 for bicomponent models with parameters $\boldsymbol{\alpha} = (2.1, 2.4)^{\mathrm{T}}$ and $\boldsymbol{\beta} = (2.2, 2.3)^{\mathrm{T}}$, and for tricomponent models with parameters $\boldsymbol{\alpha} = (2.1, 2.4, 2.3)^{\mathrm{T}}$ and $\boldsymbol{\beta} = (2.2, 2.3, 2.1)^{\mathrm{T}}$. In this figure we see how the joint correlation coefficient is levelling off towards 1 as γ increases, something that is not really visible in Figure 1.

Appendix A Differentiating binomial coefficients

We first define the binomial coefficient.

Definition 1. The binomial coefficient is defined as

$$\binom{r}{k} = \frac{r(r-1)\cdots(r-(k-1))}{k!} \tag{12}$$

where r is a real number and k is a non-negative integer.

The binomial coefficient may also be expressed as

$$\binom{r}{k} = \frac{\Gamma(r+1)}{k!\Gamma(r-k+1)}. \tag{13}$$

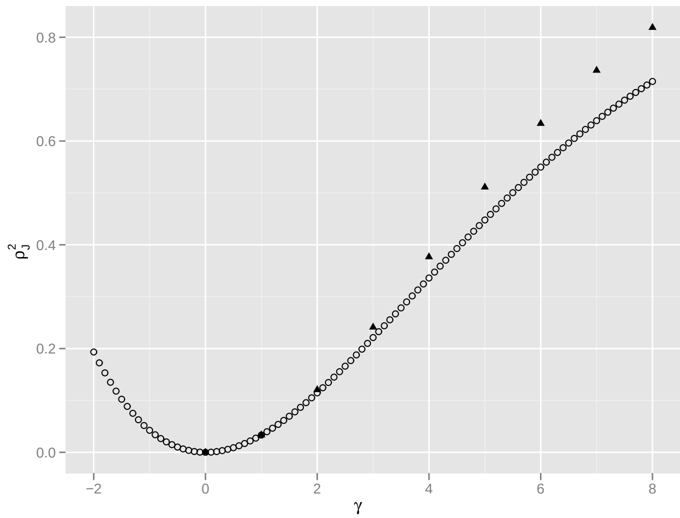


Figure 2 The joint correlation coefficient ρ_J^2 calculated for γ ranging from -2 to 8 for bicomponent models with $(\boldsymbol{\alpha}; \boldsymbol{\beta})$ parameter values (2.1, 2.4; 2.2, 2.3) (○) and tricomponent models with $(\boldsymbol{\alpha}; \boldsymbol{\beta})$ parameter values (2.1, 2.4, 2.3; 2.2, 2.3, 2.1) (▲).

The derivative of equation (13) with respect to r is

$$\frac{\mathrm{d}}{\mathrm{d}r}\binom{r}{k} = \frac{\Gamma(r+1)}{k!\Gamma(r-k+1)}\left\{\Psi(r+1) - \Psi(r-k+1)\right\} \tag{14}$$

where $\Psi(x)$ is the digamma function $\mathrm{d}\log\Gamma(x)/\,\mathrm{d}x$. However, equation (14) is not defined if r is an integer less than k. In order to give an expression that is useful for all alternatives, the derivative of the binomial coefficient must be based on the expression given in Definition 1.

Theorem 1. *The derivative of the binomial coefficient with respect to r is*

$$\frac{\mathrm{d}}{\mathrm{d}r}\binom{r}{k} = \frac{1}{k!}\sum_{i=0}^{k-1}\prod_{j=0}^{k-1} I(i,j)$$

where

$$I(i,j) = \begin{cases} 1 & (i=j), \\ r-j & (i \neq j). \end{cases}$$

Proof. Differentiating equation (12) with respect to r means differentiating the numerator consisting of a product of k factors:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}r}\prod_{i=0}^{k-1}(r-i) &= 1\cdot\prod_{i=1}^{k-1}(r-i) + (r-0)\frac{\mathrm{d}}{\mathrm{d}r}\prod_{i=1}^{k-1}(r-i) \\
&= 1\cdot(r-1)\cdots(r-k+1) \\
&\quad + (r-0)\cdot 1\cdot(r-2)\cdots(r-k+1) \\
&\quad + \cdots \\
&\quad + (r-0)(r-1)\cdots(r-k+2)\cdot 1
\end{aligned}$$

The derivative is hence a sum of k terms, each consisting of the product $r(r-1)\cdots(r-(k-1))$, where ith factor of the ith term is is replaced by 1. If we define

$$f(r) = r(r-1)\cdots(r-k+1),$$

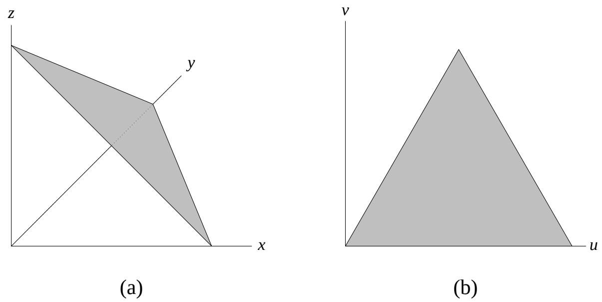


Figure 3 Two views of $\mathscr{S}^3$: (a) as seen in the normal coordinate system and (b) as seen in the plane.

we may write

$$f'(r) = \sum_{i=0}^{k-1} \prod_{j=0}^{k-1} I(i,j)$$

where

$$I(i,j) = \begin{cases} 1 & (i = j), \\ r - j & (i \neq j), \end{cases}$$

and hence

$$\frac{\mathrm{d}}{\mathrm{d}r} \binom{r}{k} = \frac{f'(r)}{k!}. \qquad \square$$

Appendix B Integration over $\mathscr{S}^3 \times \mathscr{S}^3$

The simplex $\mathscr{S}^3$ is the triangle in $\mathscr{R}_+^3$ where $x + y + z = 1$; it is depicted in Figure 3(a). Obviously, this triangle lies in a plane and may be viewed that way as shown in Figure 3(b). Integrating over $\mathscr{S}^3$ in $\mathscr{R}^3$ is thus equivalent to integrating over the triangle defined by

$$\begin{aligned} &0 < u < 2^{1/2}, \\ &0 < v < \left(\tfrac{3}{2}\right)^{1/2} - 3^{1/2}\left|2^{-1/2} - u\right|, \end{aligned}$$

in $\mathscr{R}^2$. Integration over $\mathscr{S}^3 \times \mathscr{S}^3$ analogously becomes a quadruple integral. However, since the tricomponent bicompositional Dirichlet distribution is defined on $\mathscr{S}^3 \times \mathscr{S}^3$, the $\mathscr{R}^2 \times \mathscr{R}^2$ coordinates must be transformed into compositions to get the density. Using

$$\mathbf{x}(s,t) = \mathbf{y}(s,t) = \begin{pmatrix} t\left(\frac{2}{3}\right)^{1/2} \\ s2^{-1/2} - t6^{-1/2} \\ 1 - s6^{-1/2} - t2^{-1/2} \end{pmatrix},$$

the integral of $g(\mathbf{x},\mathbf{y})$ over $\mathscr{S}^3 \times \mathscr{S}^3$ becomes

$$\int g(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} = \int_{s=0}^{2^{1/2}} \int_{t=0}^{\left(\frac{3}{2}\right)^{1/2} - 3^{1/2}\left|2^{-1/2}-s\right|} \int_{u=0}^{2^{1/2}} \int_{v=0}^{\left(\frac{3}{2}\right)^{1/2} - 3^{1/2}\left|2^{-1/2}-u\right|} g(\mathbf{x}(s,t), \mathbf{y}(u,v))\,\mathrm{d}v\,\mathrm{d}u\,\mathrm{d}t\,\mathrm{d}s.$$

References

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.

Bergman, J. (2009). A bicompositional Dirichlet distribution. Technical Report 3, Department of Statistics, Lund University.

Inaba, T. and S. Shirahata (1986). Measures of dependence in normal models and exponential models by information gain. *Biometrika 73*(2), 345–352.

Kent, J. T. (1983). Information gain and a general measure of correlation. *Biometrika 70*(1), 163–173.

III

Generating random variates from a bicompositional Dirichlet distribution

Jakob Bergman

Abstract

A composition is a vector of positive components summing to a constant. The sample space of a composition is the simplex and the sample space of two compositions, a bicomposition, is a Cartesian product of two simplices. We present a way of generating random variates from a bicompositional Dirichlet distribution defined on the Cartesian product of two simplices using the rejection method. We derive a general solution for finding a dominating density function and a rejection constant, and also compare this solution to using a uniform dominating density function. Finally some examples of generated bicompositional random variates, with varying number of components, are presented.

Keywords

Bicompositional Dirichlet distribution · Composition · Dirichlet distribution · Random variate generation · Rejection method · Simplex

1 Introduction

A composition is a vector of positive components summing to a constant. The components of a composition are what we usually think of as proportions (at least when the vector sums to 1). Compositions arise in many different areas; the geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all examples of compositions from three different scientific areas. For more examples of compositions, see for instance Aitchison (2003).

The sample space of a composition is the simplex. Without loss of generality we will always take the summing constant to be 1, and we define the D-dimensional simplex $\mathscr{S}^D$ as

$$\mathscr{S}^D = \left\{ \mathbf{x} = (x_1, \ldots, x_D)^{\mathrm{T}} \in \mathscr{R}_+^D : \sum_{j=1}^{D} x_j = 1 \right\}, \tag{1}$$

where $\mathscr{R}_+$ is the positive real space. The joint sample space of two compositions is the Cartesian product of two simplices $\mathscr{S}^D \times \mathscr{S}^D$. It should be noted that, unlike the case for real Cartesian product spaces, $\mathscr{S}^D \times \mathscr{S}^D \neq \mathscr{S}^{D+D}$ and that $\mathscr{S}^D \times \mathscr{S}^D$ is not even a simplex, but a manifold with two constraints.

2 The rejection method

Leydold (1998) notes that apart from the multinormal and Wishart distributions, papers on generating bivariate and multivariate random variates are rare and most suggested general methods have disadvantages. The only universal algorithm for generating multivariate random variates is the algorithm presented by Leydold and Hörmann (1998), which is a generalization of algorithms for the univariate and bivariate case given in different versions by Gilks and Wild (1992) and Hörmann (1995). However, Leydold (1998) concludes that this algorithm is very slow and suggests an alternative algorithm which requires a function of the density to be concave. The class of distributions that will be utilized in this paper is very versatile and it is therefore hard to find a function that fulfils the requirements. Hence we will use the rejection method to construct a specialized method for generating bicompositional random variates.

Let f be the density from which we wish to generate random variates. Let $c \geq 1$ be a constant and g be a density such that

$$f(\mathbf{z}) \leq cg(\mathbf{z}) \tag{2}$$

for all $\mathbf{z}$. We now generate a random variate $\mathbf{Z}$ with density g and a random number U uniformly distributed on the unit interval. The variate $\mathbf{Z}$ is accepted if

$$U \leq \frac{f(\mathbf{Z})}{cg(\mathbf{Z})}, \tag{3}$$

otherwise we reject $\mathbf{Z}$ and generate new $\mathbf{Z}$ and U until acceptance.

We thus need to find a dominating density g and constant c, and preferably such choices of g and c that will give high probabilities of acceptance and hence make the random variate generation efficient.

3 The bicompositional Dirichlet distribution

Bergman (2009) proposed a distribution, called the bicompositional Dirichlet distribution, for modelling random vectors on $\mathscr{S}^D \times \mathscr{S}^D$. The proposed distribution has the probability density function

$$f(\mathbf{x}, \mathbf{y}) = A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) \left(\mathbf{x}^{\mathrm{T}} \mathbf{y} \right)^{\gamma}, \tag{4}$$

where $\mathbf{x}, \mathbf{y} \in \mathscr{S}^D$ and $\alpha_j, \beta_j \in \mathscr{R}_+ (j = 1, \ldots, D)$. The parameter space of γ depends on $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)^{\mathrm{T}}$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_D)^{\mathrm{T}}$; however, all non-negative values are always included. Expressions for the normalization constant A are given in Bergman (2009). If $\gamma = 0$, the probability density function (4) is the product of two Dirichlet probability density functions with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively, and hence $\mathbf{X}$ and $\mathbf{Y}$ are independent in that case.

When $\mathbf{X}, \mathbf{Y} \in \mathscr{S}^2$ we shall refer to this as the *bicomponent* case, and similarly to $\mathscr{S}^3$ as the *tricomponent* case, and to $\mathscr{S}^D(D > 2)$ as the *multicomponent* case.

4 Generating random bicompositions

Here, the bicomposition $(\mathbf{X}, \mathbf{Y})$ will take the role of $\mathbf{Z}$ in Section 2.

4.1 The case when $\gamma = 0$

A Dirichlet distributed random variate is easily generated using Gamma distributed variates. Let V_i is a Gamma distributed variate with parameter $(\alpha_i, 1)$ and $X_i = V_i / \sum_{j=1}^{D} V_j$ $(i = 1, \ldots, D)$, then $\mathbf{X} = (X_1, \ldots, X_D)$ is Dirichlet distributed with parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)$ (Devroye, 1986, pp. 593–596).

Hence, to generate a random bicompositional Dirichlet distributed variate $(\mathbf{x}, \mathbf{y})$ with parameter $(\boldsymbol{\alpha}, \boldsymbol{\beta}, 0)$, we need only to generate a Dirichlet distributed variate $\mathbf{x}$ with parameter $\boldsymbol{\alpha}$ and a Dirichlet distributed variate $\mathbf{y}$ with parameter $\boldsymbol{\beta}$.

4.2 The case when $\gamma > 0$

When $\gamma > 0$, we may use the product of two Dirichlet distributions, i.e. a bicompositional Dirichlet distribution with $\gamma = 0$, as a dominating density, since $0 < \mathbf{x}^{\mathrm{T}}\mathbf{y} < 1$ and thus

$$A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) (\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma} \leq A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right).$$

The inequality (2) now becomes

$$A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) (\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma} \leq cA(\boldsymbol{\alpha}, \boldsymbol{\beta}, 0) \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right), \quad (5)$$

which holds if we choose

$$c = \frac{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)}{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, 0)} > 1. \tag{6}$$

The last inequality is true, since for fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ is a non-negative increasing function of γ.

A random variate $(\mathbf{x}, \mathbf{y})$ with a bicompositional Dirichlet distribution with parameter $(\boldsymbol{\alpha}, \boldsymbol{\beta}, 0)$ is generated as described in Section 4.1. We accept the variate $(\mathbf{x}, \mathbf{y})$ if

$$U \leq \frac{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)\left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1}\right)(\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma}}{\frac{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)}{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, 0)} A(\boldsymbol{\alpha}, \boldsymbol{\beta}, 0)\left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1}\right)}, \tag{7}$$

i.e. if

$$U \leq (\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma}; \tag{8}$$

otherwise it is rejected and new $(\mathbf{x}, \mathbf{y})$ and U are generated until acceptance.

We note that this procedure does not require the calculation of $A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ and hence is applicable for all non-negative γ. We thus have the slightly surprising situation that we may generate random variates from distributions whose densities we cannot calculate.

Using a product of two Dirichlet distributions as dominating density is however not always very efficient, as $(\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma}$ may be close to 0 when γ is large. When $\gamma \geq 0$, and $\alpha_j, \beta_j \geq 1$ ($j = 1, \ldots, D$), it is easily seen that the density (4) will have an upper bound. We may therefore use a uniform density as g, with $c = \max_{\mathbf{x},\mathbf{y}} f(\mathbf{x}, \mathbf{y})$. This is though only applicable for non-negative integers γ, since it is necessary to calculate $A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$.

4.3 The case when $\gamma < 0$ and $D = 2$

The bicomponent case is simpler as $\mathbf{x} = (x, 1 - x)^{\mathrm{T}}$ and $\mathbf{y} = (y, 1 - y)^{\mathrm{T}}$. This has enabled the distribution to be defined also for $\gamma < 0$. We will in this section view the density as a function of x and y.

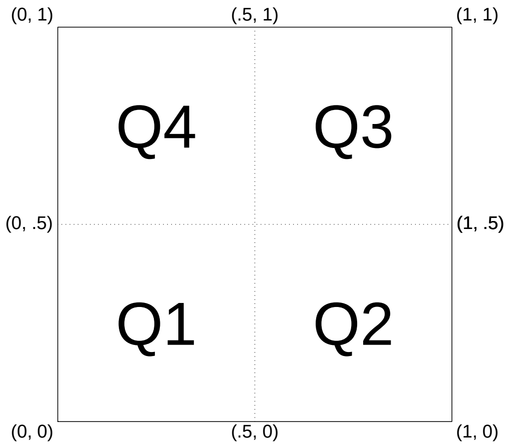


Figure 1 The four quadrants Q1-Q4 of the sample space $\mathscr{S}^2 \times \mathscr{S}^2$; the horizontal axis represents *x* and the vertical axis represents *y*.

Bergman (2009) showed that the bicomponent bicompositional Dirichlet density exists if and only if $\gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)$. If $\gamma < 0$, the factor $(\mathbf{x}^{\mathsf{T}}\mathbf{y})^{\gamma}$ will tend to infinity when x is close to 0 and y is close to 1, and also when x is close to 1 and y is close to 0. We therefore divide the sample space $\mathscr{S}^2 \times \mathscr{S}^2$ into four quadrants, denoted Q1-Q4 counter-clockwise from the origin. Figure 1 shows the $\mathscr{S}^2 \times \mathscr{S}^2$ with the four quadrants.

To generate a random variate from a bicomponent bicompositional Dirichlet distribution with parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and

$$-\min(\alpha_2, \beta_2) < \gamma < 0,$$

we first randomly choose a quadrant Qk ($k = 1, 2, 3, 4$) with probability

$$p_k = \iint_{\mathrm{Q}k} f(x, y)\,\mathrm{d}x\,\mathrm{d}y \qquad (k = 1, 2, 3, 4), \tag{9}$$

where $f(x, y)$ is the bicomponent bicompositional Dirichlet probability density function (4) viewed as a function of x and y. Expressions for the cumulative distribution function have been given by Bergman (2009), which may be used in calculating p_k. Depending on which quadrant is chosen, we then choose a dominating density g and a constant c in the following manner.

Q1 & Q3. In quadrants Q1 and Q3, $\mathbf{x}^{\mathrm{T}}\mathbf{y} > 1/2$ and we may hence use a product of two Dirichlet (or equivalently Beta) distributions with parameters $\boldsymbol{\alpha}$ respectively $\boldsymbol{\beta}$ as g and a constant

$$c = \frac{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)}{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, 0)2^{\gamma}}. \tag{10}$$

Q2. In quadrant Q2, $\mathbf{x}^{\mathrm{T}}\mathbf{y}$ is bounded from below by $(1 - x)/2$, and hence

$$(\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma} \leq 2^{-\gamma}(1 - x)^{\gamma}$$

as $\gamma < 0$. We therefore use a product of two Dirichlet distributions with parameters $(\alpha_1, \alpha_2 + \gamma)$ respectively $\boldsymbol{\beta}$ as the density g and the constant c given by

$$c = \frac{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)}{A(\alpha_1, \alpha_2 + \gamma, \boldsymbol{\beta}, 0)2^{\gamma}}. \tag{11}$$

Q4. Analogously, in quadrant Q4, $\mathbf{x}^{\mathrm{T}}\mathbf{y} > (1 - y)/2$ and we hence use a product of two Dirichlet distributions with parameters $\boldsymbol{\alpha}$ and $(\beta_1, \beta_2 + \gamma)$, respectively, as g and c given in (12).

$$c = \frac{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)}{A(\boldsymbol{\alpha}, \beta_1, \beta_2 + \gamma, 0)2^{\gamma}} \tag{12}$$

Table 1 Comparisons of the estimated acceptance probabilities depending on choice of dominating density. We clearly see that the product of two Dirichlet densities can be very inefficient for large values of γ, but also that it may be much more efficient than a uniform density for some distributions.

Parameter values					Dominating density	
α_1	α_2	β_1	β_2	γ	Dirichlet	Uniform
2.1	3.1	5.5	2.3	0.3	0.769	0.222
2.1	3.1	5.5	2.3	3.2	0.103	0.200
2.1	3.1	5.5	2.3	7.7	0.007	0.110
2.1	3.1	5.5	2.3	−1.2	0.208	0.208
2.1	3.1	0.7	2.3	3.2	0.185	NA
7.1	4.2	6.3	8.5	0.3	0.769	0.119
7.1	4.2	6.3	8.5	3.2	0.100	0.125
7.1	4.2	6.3	8.5	7.7	0.005	0.135
7.1	1.2	12.5	3.1	3.2	0.357	0.031

In all four cases, we must though assure that the generated variates with density g are restricted to that particular quadrant.

5 Comparison of the two dominating densities

The efficiency of the generation process will usually depend on the choice of dominating density. In most cases we have a possibility to choose between two different dominating densities: a product of two independent Dirichlet densities or a uniform density. In general, the product of two Dirichlet distributions will often be more efficient when γ is close to 0, but may however be highly inefficient when γ is large.

To compare the efficiencies of the two dominating densities we generated

Table 2 Comparisons of the estimated acceptance probabilities for some multicomponent bicompositional Dirichlet distributions using a Dirichlet and a uniform dominating density.

Parameter values			Density	
α	β	γ	Dir.	Unif.
(2, 2, 2)	(2, 2, 2)	1	0.333	0.145
(2, 2, 2)	(2, 2, 2)	7	0.001	0.085
(2.1, 1.2, 3.2, 4.1, 2.8)	(3.2, 2.2, 5.3, 1.8, 2.9)	1	0.204	0.000
(2.1, 1.2, 3.2, 4.1, 2.8)	(3.2, 2.2, 5.3, 1.8, 2.9)	3	0.009	0.000

25,000 random variates for each of the dominating densities from a number of different bicomponent bicompositional Dirichlet distributions, and calculated the average number of trials to generate one random variate. Table 1 shows the results presented as the estimated probability of acceptance (the reciprocal of the average number of trials) as well as the results for a distribution where only a Dirichlet product is available as dominating density as the distribution density function does not have an upper bound. We note that the probability of acceptance with a uniform density can be much (almost 30 times) larger than the probability of acceptance with a with a Dirichlet density. On the other hand we also see that there are distributions for which the probability of acceptance with a with a Dirichlet density is more than 10 times the probability of acceptance with a uniform density. As a graphical illustration of the differences between the distributions, 150 generated random variates from four of the distributions in Table 1 are plotted for each of the two dominating densities in Figure 2 together with contour curves of the density.

The differences in efficiency between the two dominating densities is even more obvious for the multicomponent bicompositional Dirichlet distribution

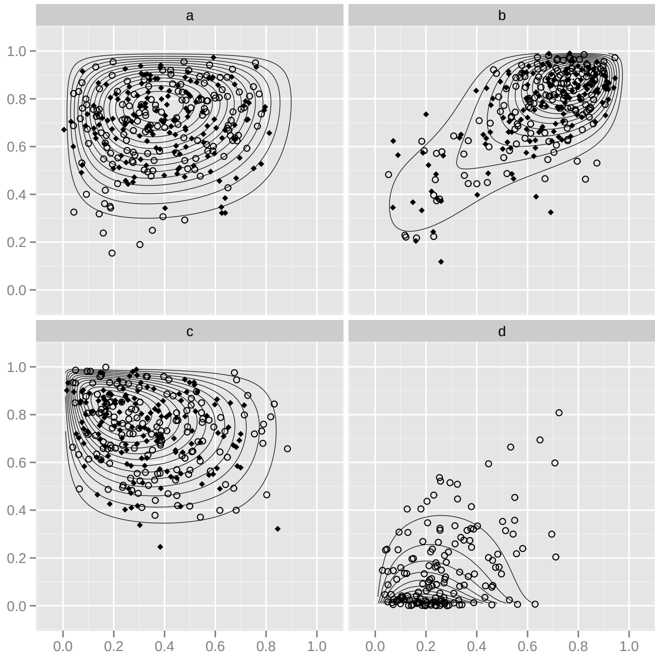


Figure 2 150 random variates generated from four different bicomponent bicompositional Dirichlet distributions with $(\boldsymbol{\alpha}; \boldsymbol{\beta}; \gamma)$ parameters $(2.1, 3.1;\ 5.5, 2.3;\ 0.3)$ (a), $(2.1, 3.1;\ 5.5, 2.3;\ 7.7)$ (b), $(2.1, 3.1;\ 5.5, 2.3;\ -1.2)$ (c), and $(2.1, 3.1;\ 0.7, 2.3;\ 3.2)$ (d), using the product of two Dirichlet densities (○) and a uniform density (◆) as dominating density. Since the distribution in (d) does not have an upper bound, a uniform density may not be utilized. As a reference, the contour curves of the true densities are also drawn.

examples presented in Table 2. Here again, we generated 25,000 random variates, this time from four different multicomponent bicompositional Dirichlet distributions using both of the two dominating densities. For the tricomponent distributions, when $\gamma = 1$, the Dirichlet density has a probability of acceptance of more than twice that of the uniform density, but when $\gamma = 7$ the probability of acceptance of the uniform density is more than 80 times that of the the Dirichlet density. For the two distributions with five components, we see that the Dirichlet density is much more effective for both cases. This is in accordance with Devroye (1986, p. 557), who notes that as the dimension D increases the rejection constant often deteriorates quickly when a uniform density is used.

6 Conclusions

The choice of the dominating density is evidently crucial to the efficiency of this random variate generation. When γ is close to 0 or the number of components is large, a product of two Dirichlet density functions seems the most efficient, otherwise a uniform density function (if possible) is recommended. What is meant by close is however dependent of the other parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, so when in doubt, the recommendation would be to generate a small number of variates with each dominating density and see which is the most efficient for the particular parameter values in question. We note that the efficiency of the method seems to degrade as the dimension (i.e. the number of components) increases, and that further research is needed to find more efficient dominating densities for distributions with a large number of components and for large γ values.

It remains yet to find a way of generating random numbers for the bicomponent case when $-\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1) < \gamma < -\min(\alpha_2, \beta_2)$ and the density function does not have an upper bound.

The random variate generation might further be made more efficient for at least the bicomponent case, by adopting the quadrant scheme also for positive γ; especially when the probability mass is concentrated in one or two of the quadrants, which is often the case for large γ, this might speed up the generation process considerably.

References

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.

Bergman, J. (2009). A bicompositional Dirichlet distribution. Technical Report 3, Department of Statistics, Lund University.

Devroye, L. (1986). *Non-uniform random variate generation*. New York: Springer-Verlag.

Gilks, W. R. and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics 41*(2), 337–348.

Hörmann, W. (1995). A rejection technique for sampling from T-concave distributions. *ACM Transactions on Mathematical Software 21*(2), 182–193.

Leydold, J. (1998). A rejection technique for sampling from log-concave multivariate distributions. *ACM Transactions on Modeling and Computer Simulation 8*(3), 254–280.

Leydold, J. and W. Hörmann (1998). A sweep-plane algorithm for generating random tuples in simple polytopes. *Mathematics of Computation 67*(224), 1617–1635.

IV

Estimating a measure of dependence between two compositions

Jakob Bergman

Abstract

We present an estimator of the general measure of correlation for bicompositional data for a sample from a bicompositional Dirichlet distribution. Two confidence intervals are also presented and we examine their empirical confidence coefficient using a Monte Carlo study. Finally we apply the estimator to a data set analysing the correlation between the 1967 and 1997 composition of the government GDP for the 50 U.S. states and District of Columbia.

Keywords

Composition · Correlation · Dirichlet distribution · Empirical confidence coefficient · Estimation · Joint correlation coefficient

1 Introduction

A composition is a vector of positive components summing to a constant, usually taken to be 1. The components of a composition are what we usually think of as proportions (at least when the vector sums to 1). Compositions arise in many different areas; the geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all examples of compositions from three different scientific areas. For more examples of compositions, see for instance Aitchison (2003).

The sample space of a composition is the simplex. Without loss of generality we will always take the summation constant to be 1, and we define the D-dimensional simplex $\mathscr{S}^D$ as

$$\mathscr{S}^D = \left\{ (x_1, \dots, x_D)^{\mathrm{T}} \in \mathscr{R}_+^D : \sum_{j=1}^{D} x_j = 1 \right\},$$

where $\mathscr{R}_+$ is the positive real space.

We will refer to compositions with two components, i.e. $D = 2$, as *bi-component*.

2 Estimation of the correlation

Following the ideas of Kent (1983), Bergman and Holmquist (2009) derived a general measure of correlation ρ_J^2 for data from a bicompositional Dirichlet distribution. The bicompositional Dirichlet distribution, defined on the Cartesian product $\mathscr{S}^D \times \mathscr{S}^D$, was introduced by Bergman (2009a). The distribution has three parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_D)^{\mathrm{T}}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_D)^{\mathrm{T}}$ and γ, and the probability density function is

$$f(\mathbf{x}, \mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) \left(\mathbf{x}^{\mathrm{T}} \mathbf{y} \right)^{\gamma}, \qquad (1)$$

where $\mathbf{x} = (x_1, \dots, x_D)^{\mathrm{T}} \in \mathscr{S}^D$, $\mathbf{y} = (y_1, \dots, y_D)^{\mathrm{T}} \in \mathscr{S}^D$, and $\alpha_j, \beta_j \in \mathscr{R}_+$ ($j = 1, \dots, D$). The parameter space of γ depends on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; however, all non-negative values are always included. Expressions for the normalization constant $A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ are given in Bergman (2009a). If $\gamma = 0$, the probability density function (1) is the product of two Dirichlet probability density functions with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively, and hence the two compositions are independent in that case.

The general measure of correlation (or joint correlation coefficient) is defined as

$$\rho_J^2 = 1 - \exp\{-\Gamma(\theta_1 : \theta_0)\}, \tag{2}$$

where $\Gamma(\theta_1 : \theta_0)$ is the information gain of modelling the data with $\theta_1 \in \Theta_1$ instead of $\theta_0 \in \Theta_0 \subset \Theta_1$ in the parametric model $f(\mathbf{x}, \mathbf{y}; \theta_i)$ $(i = 0, 1)$. The information gain is defined as

$$\Gamma(\theta_1 : \theta_0) = 2\{F(\theta_1) - F(\theta_0)\}, \tag{3}$$

where $F(\theta_i)$ $(i = 0, 1)$ is the maximized Fraser information

$$F(\theta_i) = \max_{\theta \in \Theta_i} \int \log f(\mathbf{x}, \mathbf{y}; \theta) g(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y}; \tag{4}$$

here $g(\mathbf{x}, \mathbf{y})$ is the true probability density function.

We assume that $g(\mathbf{x}, \mathbf{y})$ is a bicompositional Dirichlet probability density function and restrict our estimation to the bicomponent models. Since we are interested in modelling the correlation between two compositions (the intercompositional correlation), we want to calculate the information gained by allowing dependence between the compositions as compared to independent compositions. The parameter spaces are then

$$\Theta_1 = \{\alpha_1 > 0, \alpha_2 > 0, \beta_1 > 0, \beta_2 > 0, \gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)\}$$

and

$$\Theta_0 = \{\alpha_1 > 0, \alpha_2 > 0, \beta_1 > 0, \beta_2 > 0, \gamma = 0\}.$$

According to Kent (1983), the information gain $\Gamma(\theta_1 : \theta_0)$ may be estimated by

$$\widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0) = \frac{2}{n} \left(\sum_{k=1}^{n} \log f(\mathbf{x}_k, \mathbf{y}_k; \hat{\theta}_1) - \sum_{k=1}^{n} \log f(\mathbf{x}_k, \mathbf{y}_k; \hat{\theta}_0) \right), \tag{5}$$

where $\hat{\theta}_1$ and $\hat{\theta}_0$ are the maximum likelihood estimates under the parameter spaces Θ_1 and Θ_0, respectively.

2.1 Maximum likelihood estimates

If we assume a sample of n independent observations $(\mathbf{x}_j, \mathbf{y}_j)$ $(j = 1, \ldots, n)$ from a bicomponent bicompositional Dirichlet distribution with parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and γ, the likelihood function becomes

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = \{A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)\}^n \prod_{k=1}^{n} \left((\mathbf{x}_k^{\mathrm{T}} \mathbf{y}_k)^{\gamma} \prod_{j=1}^{2} x_{kj}^{\alpha_j - 1} y_{kj}^{\beta_j - 1} \right) \tag{6}$$

and the log likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) &= -nc(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) + \gamma \sum_{k=1}^{n} \log(\mathbf{x}_k^{\mathrm{T}} \mathbf{y}_k) \\ &\quad + \sum_{k=1}^{n} \sum_{j=1}^{2} \{(\alpha_j - 1) \log x_{kj} + (\beta_j - 1) \log y_{kj}\} \end{aligned} \tag{7}$$

where $c(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = -\log A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = \log \left(2^{-\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}} S_{\boldsymbol{\beta}} \right)$. Here

$$S_{\boldsymbol{\alpha}} = \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\boldsymbol{\alpha}), \tag{8}$$

$$S_{\boldsymbol{\beta}} = \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\boldsymbol{\beta}), \tag{9}$$

with

$$B_{ij}(\boldsymbol{\alpha}) = B(\alpha_1 + j, \alpha_2 + i - j), \tag{10}$$

where $B(\cdot, \cdot)$ denotes the Beta function.

The maximum likelihood estimates are of course the parameter values that yield the maximum value of (7). However, finding those values will in general require numerical methods. We have used the R function `constrOptim`, which also utilizes the score function

$$\mathbf{U}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = \begin{bmatrix} \frac{\partial \ell}{\partial \gamma} & \frac{\partial \ell}{\partial \alpha_1} & \frac{\partial \ell}{\partial \alpha_2} & \frac{\partial \ell}{\partial \beta_1} & \frac{\partial \ell}{\partial \beta_2} \end{bmatrix}^{\mathrm{T}}, \tag{11}$$

where

$$\frac{\partial \ell}{\partial \gamma} = -n \frac{\partial c}{\partial \gamma} + \sum_{j=1}^{n} \log(\mathbf{x}_j^{\mathrm{T}} \mathbf{y}_j), \tag{12}$$

$$\frac{\partial \ell}{\partial \alpha_1} = -n \frac{\partial c}{\partial \alpha_1} + \sum_{j=1}^{n} \log x_{j1}, \tag{13}$$

$$\frac{\partial \ell}{\partial \alpha_2} = -n \frac{\partial c}{\partial \alpha_2} + \sum_{j=1}^{n} \log x_{j2}, \tag{14}$$

$$\frac{\partial \ell}{\partial \beta_1} = -n \frac{\partial c}{\partial \beta_1} + \sum_{j=1}^{n} \log y_{j1}, \tag{15}$$

$$\frac{\partial \ell}{\partial \beta_2} = -n \frac{\partial c}{\partial \beta_2} + \sum_{j=1}^{n} \log y_{j2}. \tag{16}$$

The maximum likelihood estimate of $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ under the parameter space Θ_i $(i = 0, 1)$ is denoted $\hat{\boldsymbol{\theta}}_i$. Trivially, the estimate of γ under Θ_0 is $\hat{\gamma} = 0$.

An estimator of the general measure of correlation is thus

$$\hat{\rho}_J^2 = 1 - \exp\{-\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0)\}. \tag{17}$$

2.2 Confidence intervals

Kent (1983) gives two proposals concerning confidence intervals for $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$: when the value of $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ is “large” and when it is “small”. Kent does not indicate which values of $\Gamma(\theta_1 : \theta_0)$ that are to be considered "large" and which are to be considered "small," other than that it depends on the number of observations n. He notes though that "the asymptotics for 'small' $\Gamma(\theta_1 : \theta_0)$ are likely to prove most useful."

The first $1 - \alpha$ confidence interval ("large") is

$$\left(\widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0) - \sqrt{\frac{s^2 \chi^2_{1;\alpha}}{n}}, \widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0) + \sqrt{\frac{s^2 \chi^2_{1;\alpha}}{n}} \right) \tag{18}$$

where s^2 is the sample variance of

$$2 \log \frac{f(\mathbf{x}_j, \mathbf{y}_j; \hat{\theta}_1)}{f(\mathbf{x}_j, \mathbf{y}_j; \hat{\theta}_0)} \qquad (j = 1, \ldots, n)$$

and $\chi^2_{1;\alpha}$ is the upper α quantile of the χ^2_1 distribution.

The second $1 - \alpha$ confidence interval ("small") is (corrected for an apparently misprinted $\hat{\alpha}$ instead of $\hat{a}$)

$$\left(\frac{\mu \varkappa_{1;\alpha/2}(\hat{a}/\mu)}{n}, \frac{\mu \delta_{1;\alpha/2}(\hat{a}/\mu)}{n} \right), \tag{19}$$

where

$$\hat{a} = n\widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0)$$

and $\varkappa_{1;\alpha}(a)$ and $\delta_{1;\alpha}(a)$ are the values of the non-centrality parameters of a non-central chi square distribution defined as

$$\Pr[\chi^2_1\{\varkappa_{1;\alpha}(a)\} \geq a] = \alpha, \quad \Pr[\chi^2_1\{\delta_{1;\alpha}(a)\} \leq a] = \alpha.$$

The constant μ is the common value of the eigenvalues, which are assumed to be equal, of a rather complicated matrix. However, for our purposes μ is always equal to 1, as we are convinced that the true density function belongs to $\{f(\mathbf{x}, \mathbf{y}; \theta) | \theta \in \Theta_1\}$. (The α in (18) and (19) is one minus the confidence coefficient, not to be confused with the parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^{\mathrm{T}}$ of the bicomponent bicompositional Dirichlet distribution.)

We thus transform the confidence intervals of $\Gamma(\theta_1 : \theta_0)$ yielding the "large"

$$\left[1 - \exp\left\{-\widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0) + \sqrt{s^2\chi^2_{1;\alpha}n^{-1}}\right\}, \right.$$
$$\left. 1 - \exp\left\{-\widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0) - \sqrt{s^2\chi^2_{1;\alpha}n^{-1}}\right\}\right] \quad (20)$$

and the "small"

$$\left[1 - \exp\left\{-\frac{\varkappa_{1;\alpha/2}(\hat{a})}{n}\right\}, 1 - \exp\left\{-\frac{\delta_{1;\alpha/2}(\hat{a})}{n}\right\}\right] \quad (21)$$

$1 - \alpha$ confidence intervals of ρ^2_J.

3 Comparison of the confidence intervals

If the first confidence interval (18) "includes or nearly includes 0, then," according to Kent (1983), "provided n is large enough for the asymptotics to be valid, the confidence interval of the next section [(19)] is probably more reliable."

In order to examine the properties of the two confidence intervals (20) and (21), we conduct a Monte Carlo study for seven models with different ρ^2_J and for different numbers of observations ($n = 50, 100, 250$). For every combination of model and number of observations we generate random variates (Bergman, 2009b), estimate $\hat{\rho}^2_J$, compute the two confidence intervals, and record in how many cases the true value of ρ^2_J is covered by the two intervals (the empirical confidence coefficient). The results are presented in Table 1. The nominal confidence coefficient in the study is 0.95 and we see clearly from the table that most of the empirical confidence coefficients are close to this; the empirical confidence coefficients vary between 0.88 and 1.00. We note that especially the "large" confidence intervals seem to have a too high empirical confidence coefficients, indicating overly wide confidence intervals.

The ratios between the average widths of the "small" and "large" confidence intervals are plotted in Figure 1. We see in the figure that the average ratio between the widths of the "small" and the "large" confidence intervals is about 0.5 when $\rho_J^2 < 0.2$, and around 0.6 for larger ρ_J^2. We also note that, perhaps not very surprisingly, the ratio increases as the sample size and the correlation coefficient are increased. However, for 250 observations and $\rho_J^2 = 0.867$, the average width of the "small" confidence interval is less than 0.75 of that of the "large" one. It should be noted though that as the "large" confidence intervals are not guaranteed to be non-negative, the comparisons are from a practical point of view not entirely fair; a confidence interval with a lower limit less than zero would in practice of course have it replaced by zero as both the information gain and the general measure of correlation are non-negative. On the other hand, a confidence limit that is not restricted to the appropriate parameter space is of course of less practical use.

4 Bias correction

Kent (1983) notes that the estimator (5) is biased and suggests a less biased estimator

$$\widehat{\Gamma}(\hat{\theta}_1 : \hat{\theta}_0) - \frac{\hat{B}}{n} \tag{22}$$

where

$$\hat{B} = \mathrm{tr}\{\widehat{\mathbf{H}}(\hat{\theta}_1)^{-1}\hat{\mathbf{J}}(\hat{\theta}_1)\} - \mathrm{tr}\{\widehat{\mathbf{H}}_{\lambda\lambda}(\hat{\theta}_0)^{-1}\hat{\mathbf{J}}_{\lambda\lambda}(\hat{\theta}_0)\}. \tag{23}$$

In (23), $\hat{\mathbf{J}}(\hat{\theta})$ is an estimate of the expected squared score matrix $\mathbf{J}(\theta) = E\{\mathbf{U}(\theta)\mathbf{U}(\theta)^{\mathrm{T}}\}$

$$\hat{\mathbf{J}}(\theta) = \frac{1}{n}\sum \mathbf{U}(\theta)\mathbf{U}(\theta)^{\mathrm{T}}, \tag{24}$$

Table 1 The empirical confidence coefficient is presented for seven different models $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ and three different numbers of observations n. For each model and number of observations, 500 samples of random variates are generated and the two confidence intervals ("large" and "small") for the correlation coefficient are calculated. We then calculate the proportion of the confidence intervals that cover the true value of the correlation coefficient ρ_J^2 for that model.

Parameter values					Interval	
$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	γ	ρ_J^2	n	"large"	"small"
(3, 2.3)	(4, 2)	1.5	0.038	50	0.932	0.970
				100	0.946	0.964
				250	0.980	0.958
(9, 7)	(4, 2)	4.5	0.099	50	0.964	0.950
				100	0.972	0.882
				250	0.988	0.948
(4, 3)	(4, 2)	4.5	0.174	50	0.968	0.954
				100	0.980	0.910
				250	1.000	0.946
(4, 3)	(3, 4)	4.5	0.244	50	0.992	0.954
				100	0.994	0.962
				250	0.998	0.954
(4, 3)	(3, 4)	6.5	0.418	50	0.996	0.968
				100	1.000	0.982
				250	1.000	0.968
(4, 3)	(3, 4)	9.5	0.652	50	1.000	0.992
				100	1.000	0.996
				250	1.000	0.984
(4, 3)	(3, 4)	14.0	0.867	50	0.998	0.996
				100	1.000	0.994
				250	1.000	0.994

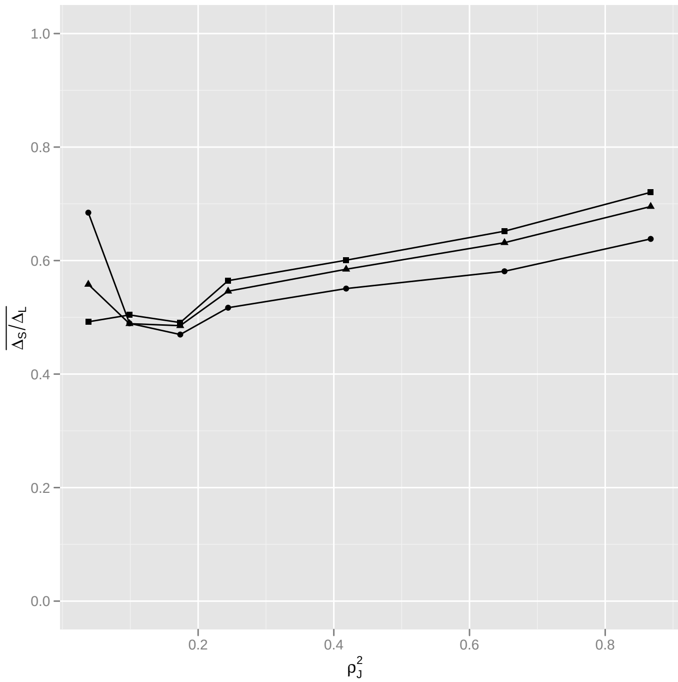


Figure 1 The average ratio between the widths of the "small" (Δ_S) and "large" (Δ_L) confidence intervals plotted for the seven different models (ρ_J^2) in Table 1 and for 50 (●), 100 (▲) and 250 (■) observations.

and $\widehat{\mathbf{H}}(\hat{\theta})$ is the estimate of minus the expected score derivative matrix $\mathbf{H}(\theta) = -E\{\partial \mathbf{U}(\theta)/\partial \theta^{\mathrm{T}}\}$

$$\widehat{\mathbf{H}}(\theta) = -\frac{1}{n}\sum \frac{\partial \mathbf{U}(\theta)}{\partial \theta}. \tag{25}$$

The matrices above with the subscript $_{\lambda\lambda}$ refers to the 4×4 part of the matrix not depending on γ, that is the top left part $\mathbf{H}_{\lambda\lambda}$ if the matrix is partitioned

$$\mathbf{H} = \left[\begin{array}{c|c} \mathbf{H}_{\lambda\lambda} & \mathbf{H}_{\lambda\gamma} \\ \hline \mathbf{H}_{\gamma\lambda} & \mathrm{H}_{\gamma\gamma} \end{array}\right]$$

and of course analogously for $\mathbf{J}_{\lambda\lambda}$.

Calculating $\widehat{\mathbf{H}}(\theta)$ requires calculating the second derivatives of the log likelihood $\partial^2 \ell(\theta)/\partial \theta^2 = \partial \mathbf{U}(\theta)/\partial \theta$. An expression for (25) may be found, but it is not presented here as it would require a large amount of space. However to give an example of the complexity of the calculations necessary we present five of the second derivatives. We first introduce some notation (to enable the expressions to fit into the page).

Let $\alpha_. = \alpha_1 + \alpha_2$ and $\beta_. = \beta_1 + \beta_2$. We use the digamma and trigamma functions

$$\Psi(z) = \frac{\mathrm{d}\log \Gamma(z)}{\mathrm{d}z}, \tag{26}$$

$$\psi_1(z) = \frac{\mathrm{d}^2 \log \Gamma(z)}{\mathrm{d}z^2} = \frac{\mathrm{d}\Psi(z)}{\mathrm{d}z}, \tag{27}$$

and define

$$\Psi_{ij}^{(1)}(\boldsymbol{\alpha}) = \Psi(\alpha_1 + j) - \Psi(\alpha_. + i), \tag{28}$$

$$\Psi_{ij}^{(2)}(\boldsymbol{\alpha}) = \Psi(\alpha_2 + i - j) - \Psi(\alpha_. + i). \tag{29}$$

The first and second derivatives of the binomial coefficient are denoted

$$\binom{\gamma}{i}' = \frac{\mathrm{d}}{\mathrm{d}\gamma}\binom{\gamma}{i}, \tag{30}$$

$$\binom{\gamma}{i}'' = \frac{\mathrm{d}^2}{\mathrm{d}\gamma^2}\binom{\gamma}{i}. \tag{31}$$

Calculation of the second derivative of the binomial coefficient (31) is discussed in Appendix A. We also define

$$S_{\boldsymbol{\alpha}}^{(k)} = \frac{\partial S_{\boldsymbol{\alpha}}}{\partial \alpha_k} = \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\boldsymbol{\alpha}) \Psi_{ij}^{(k)}(\boldsymbol{\alpha}) \tag{32}$$

$$S_{\boldsymbol{\alpha}}^{(kl)} = \frac{\partial^2 S_{\boldsymbol{\alpha}}}{\partial \alpha_k \partial \alpha_l} = \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} B_{ij}(\boldsymbol{\alpha}) \{ \Psi_{ij}^{(k)}(\boldsymbol{\alpha}) \Psi_{ij}^{(l)}(\boldsymbol{\alpha}) - \psi_1(\alpha_{.} + i) \}, \tag{33}$$

where $B_{ij}(\boldsymbol{\alpha})$ is given in (10), and we define $S_{\boldsymbol{\beta}}^{(k)}$ and $S_{\boldsymbol{\beta}}^{(kl)}$ analogously. We finally define

$$K = \sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}} S_{\boldsymbol{\beta}}, \tag{34}$$

where $S_{\boldsymbol{\alpha}}$ is given in (8) and $S_{\boldsymbol{\beta}}$ is given in (9).

Using this notation we present five of the elements in $\widehat{\mathbf{H}}(\boldsymbol{\theta})$:

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \gamma^2} &= -n \frac{\left(\sum_{i=0}^{\infty} \binom{\gamma}{i}'' S_{\boldsymbol{\alpha}} S_{\boldsymbol{\beta}}\right) K - \left(\sum_{i=0}^{\infty} \binom{\gamma}{i}' S_{\boldsymbol{\alpha}} S_{\boldsymbol{\beta}}\right)^2}{K^2} \\
\frac{\partial^2 \ell}{\partial \gamma \partial \alpha_1} &= -n \frac{\left(\sum_{i=0}^{\infty} \binom{\gamma}{i}' S_{\boldsymbol{\alpha}}^{(1)} S_{\boldsymbol{\beta}}\right) K - \left(\sum_{i=0}^{\infty} \binom{\gamma}{i}' S_{\boldsymbol{\alpha}} S_{\boldsymbol{\beta}}\right)\left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(1)} S_{\boldsymbol{\beta}}\right)}{K^2} \\
\frac{\partial^2 \ell}{\partial \alpha_1^2} &= -n \frac{\left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(11)} S_{\boldsymbol{\beta}}\right) K - \left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(1)} S_{\boldsymbol{\beta}}\right)^2}{K^2} \\
\frac{\partial^2 \ell}{\partial \alpha_1 \partial \alpha_2} &= -n \frac{\left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(12)} S_{\boldsymbol{\beta}}\right) K - \left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(1)} S_{\boldsymbol{\beta}}\right)\left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(2)} S_{\boldsymbol{\beta}}\right)}{K^2} \\
\frac{\partial^2 \ell}{\partial \alpha_1 \partial \beta_1} &= -n \frac{\left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(1)} S_{\boldsymbol{\beta}}^{(1)}\right) K - \left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}}^{(1)} S_{\boldsymbol{\beta}}\right)\left(\sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\boldsymbol{\alpha}} S_{\boldsymbol{\beta}}^{(1)}\right)}{K^2}
\end{aligned}$$

The remaining the elements of the matrix may be expressed in a similar fashion.

Assuming that the true density belongs to $\{f(\mathbf{x}, \mathbf{y}; \theta) | \theta \in \Theta_1\}$, then, according to Kent (1983), $\mathrm{tr}\{\widehat{\mathbf{H}}(\hat{\theta}_1)^{-1}\hat{\mathbf{J}}(\hat{\theta}_1)\}$ is equal to the number of parameters in the model, in our case five.

However, numerical examples indicate that the bias corrected estimates are, contrary to Kent's claim, actually more biased than the uncorrected ones, especially for models with large ρ_J^2. We believe that this increased bias might be due to numerical issues in calculating $\widehat{\mathbf{H}}(\hat{\theta})$, which, as demonstrated above, consists of a multitude of infinite sums. Due to this lack of improvement we have not used this bias correction in our estimations.

5 An application

We illustrate the estimation of the general measure of correlation presented in Section 2 with an example. The data consist of the composition of the government Gross Domestic Product (GDP) for the 50 U.S. states and District of Columbia, for the years 1967 and 1997. The composition is originally (Federal civilian, Federal military, State and local), but we have collapsed the Federal military and the State and local, to create a bicomponent composition. Data come from the Bureau of Economic Analysis, U.S. Department of Commerce.

The maximum likelihood estimates of the parameters under Θ_1 are

$$\hat{\alpha} = (16.32, 14.41)^{\mathrm{T}}, \qquad \hat{\beta} = (17.31, 43.20)^{\mathrm{T}}, \qquad \hat{\gamma} = 57.41.$$

The data and the contour curves of the bicompositional Dirichlet distribution with the above parameter estimates are shown in Figure 2. The estimate of the general measure of correlation is

$$\hat{\rho}_J^2 = 0.3027,$$

with a "small" confidence interval of

$$(0.0993, 0.5371)$$

thus indicating that composition of the government GDP in 1967 is correlated with the composition of the government GDP in 1997.

6 Discussion

The compositional data analysis has through history primarily been concerned with modelling the dependence between the components of a composition, the intra-compositional dependence. However, understanding and modelling the dependence between compositions, the inter-compositional dependence, is also of interest; this is of course especially evident when we are studying compositional processes.

Kent (1983) introduced a general measure of correlation and this was developed by Bergman and Holmquist (2009) for two compositions using the only known distribution on the Cartesian product $\mathscr{S}^D \times \mathscr{S}^D$ (Bergman, 2009a). In this paper we have shown how to estimate the general correlation coefficient ρ_J^2 with a point estimate and two confidence intervals. We have also compared the two confidence intervals and it is apparent for the models that we have examined that the so called "small" confidence interval (based on non-central χ^2-distributions) will produce the smaller intervals, yielding an empirical confidence coefficient for almost all models of approximately 95 %, when the nominal confidence coefficient is 95 %. The "large" confidence intervals will in general be wider.

As an example we have also estimated the general measure of correlation for GDP data from the 50 U.S. states and District of Columbia.

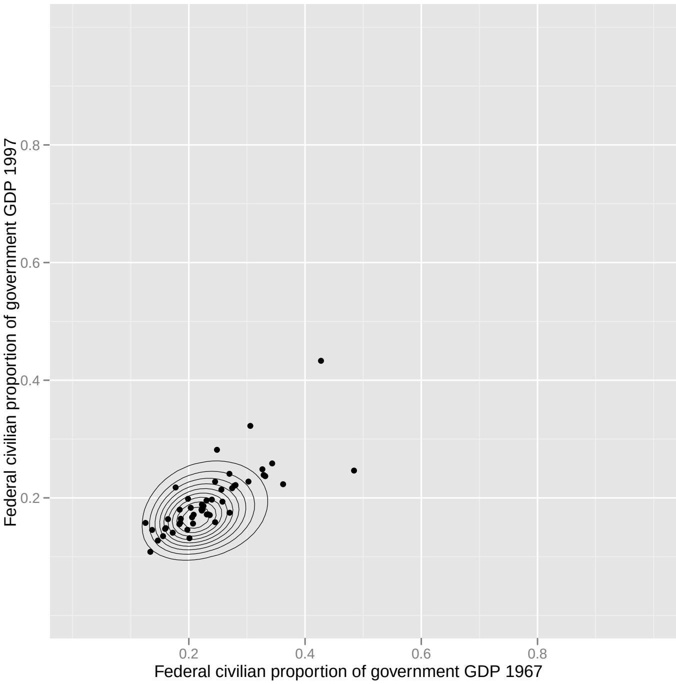


Figure 2 The federal civilian proportion of the government part of GDP for the 50 U.S. states and District of Columbia plotted for 1997 versus 1967 and the contour curves of the estimated bicompositional Dirichlet distribution.

Source: Bureau of Economic Analysis, U.S. Department of Commerce.

Appendix A Derivatives of binomial coefficients

Theorem 1. *The second derivative of the binomial coefficient with respect to r is*

$$\frac{\mathrm{d}^2}{\mathrm{d}r^2}\binom{r}{n} = \frac{1}{n!}\sum_{i=0}^{n-1}\sum_{\substack{j=0\\ j\neq i}}^{n-1}\prod_{\substack{k=0\\ k\neq i\\ k\neq j}}^{n-1}(r-k). \tag{35}$$

Proof. Bergman and Holmquist (2009, Theorem A.2) give an expression for the first derivative of the binomial coefficient:

$$\frac{\mathrm{d}}{\mathrm{d}r}\binom{r}{n} = \frac{1}{n!}\sum_{i=0}^{n-1}\prod_{j=0}^{n-1} I(i,j)$$

where

$$I(i,j) = \begin{cases} 1 & (i=j), \\ r-j & (i\neq j). \end{cases}$$

This is thus a sum of n terms each consisting of a product of n factors $r-j$, where the jth factor of the jth term is replaced by 1; hence each term in practice consists of a product of $n-1$ factors:

$$\begin{aligned} &(r-1)(r-2)\cdots(r-n+1) \\ +&(r-0)(r-2)\cdots(r-n+1) \\ +&\cdots \\ +&(r-0)(r-1)\cdots(r-n+2) \end{aligned}$$

Differentiating this expression yields a sum of n terms ($i=0,\ldots,n-1$), each consisting of a sum $n-1$ terms ($j=0,\ldots,n; j\neq i$), each in turn consisting of a product of $n-2$ factors ($k=0,\ldots,n; k\neq i, k\neq j$) where every factor is $r-k$. $\square$

References

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.

Bergman, J. (2009a). A bicompositional Dirichlet distribution. Technical Report 3, Department of Statistics, Lund University.

Bergman, J. (2009b). Generating random variates from a bicompositional Dirichlet distribution. Technical Report 5, Department of Statistics, Lund University.

Bergman, J. and B. Holmquist (2009). A measure of dependence between two compositions. Technical Report 4, Department of Statistics, Lund University.

Kent, J. T. (1983). Information gain and a general measure of correlation. *Biometrika 70*(1), 163–173.