A measure of dependence between two compositions

Bergman, Jakob; Holmquist, Björn

Published in:
Australian & New Zealand Journal of Statistics

DOI:
10.1111/j.1467-842X.2012.00688.x

2012

Document Version:
Peer reviewed version (aka post-print)

A measure of dependence between two compositions

(Running Title: Dependence between two compositions)

JAKOB BERGMAN*AND BJÖRN HOLMQUIST

Lund University

Summary

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. We also present an estimator of the joint correlation coefficient 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 joint correlation between the 1967 and 1997, and the 1977 and 1997 compositions of the government

[1]Author to whom correspondence should be addressed.
[0]Department of Statistics, Box 743, SE-220 07 Lund, Sweden.
[0]e-mail: jakob.bergman@stat.lu.se

gross domestic product for the 50 U.S. states and District of Columbia.

Key words: correlation; Dirichlet distribution; empirical confidence coefficient; Fraser information; joint correlation coefficient; simplex

1 Introduction

A composition is a vector of positive components summing to a constant, usually taken to be 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. We will refer to compositions with two components as *bicomponent*, to compositions with three components as *tricomponent* and to compositions with more than two components as *multicomponent*.

Due to the summation constraint the components of a composition are not independent. Much research have been concerned with describing how the components of a composition correlate, i.e. the intra-compositional dependence. A review of different independence concepts pertaining to partitions of a composition is presented by Aitchison (1986, Chap. 10).

Our concern is correlations between compositions, a topic which has previously not been given equally much attention. We believe that a measure of inter-compositional dependence is needed in order to describe various naturally occurring phenomena, 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. There exist many proposals of measures for comparing the similarity (or dissimilarity) between two compositions based on intuitive meanings of similarity in different contexts. For example in ecology the abundance of species in two communities (or the percentage cover of different species in two communities) have been proposed to be compared using the Bray-Curtis measure of 'percent similarity' (Bray & Curtis 1957), or using chi-square distance for quantifying resemblance in abundance data.

Some such techniques adapt to situations where many pairs of compositions are available and the general tendency of similarity between the pairs is to be quantified. Again the approach is to consider measures that intuitively would strengthen the idea of a general similarity between pairs.

The approach taken here is somewhat the opposite. We describe, using a joint distribution, the simultaneous outcome of pairs of compositions. We characterize in this joint distribution what describes this dependence between the composition pairs. This will in effect be through a very general definition of correlation in terms of information gain in the dependence model relative a model of independence (Kent 1983). We also describe how this h property, in terms of a particular parameter, could be estimated using a sample of paired observations of compositions. The statistic derived for use in a sample of paired observations of compositions will hence be an estimator of the corresponding theoretical parameter describing the correlation. This makes inference about this property possible.

2 The bicompositional Dirichlet distribution

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

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

where $\mathscr{R}_+$ is the positive real space. The joint sample space of two compositions is the Cartesian product of two simplices $\mathscr{S}^D \times \mathscr{S}^D$. Unfortunately very few distributions with dependence structures defined on $\mathscr{S}^D \times \mathscr{S}^D$ are available. For the bicomponent case there have been proposed a few bivariate Beta distributions (though usually not in a compositional context). See for instance Olkin & Liu 2003, Nadarajah & Kotz 2005, Nadarajah 2006 or Nadarajah 2007. Apart from the fact that these distributions do not enable modelling of multicomponent compositions, some are unable to model independence between the compositions. Jørgensen & Lauritzen (2000) have introduced a distribution defined on the Cartesian product of p bicomponent simplices $(\mathscr{S}^2)^p$, i.e. a bicomponent multicompositional distribution, however still only a bicomponent distribution.

The logistic-normal distribution on the simplex (see e.g. Aitchison & Shen 1980; Aitchison 1986) can be generalized to bicompositional distribution on $\mathscr{S}^D \times \mathscr{S}^D$ through two inverse additive log-ratio transformations of a multivariate normal distribution on $\mathscr{R}^{D-1} \times \mathscr{R}^{D-1}$. The actual dependence structure between the compositions is then determined through the cross covariance matrix between the two normal sub-vectors on $\mathscr{R}^{D-1}$. A similar technique applies when considering logistic skew-normal distributions (see e.g. Mateu-Figueras *et al.* 2005) where the dependence structure between the compositions will be even more complex.

For our purpose we shall here consider the bicompositional Dirichlet distribution proposed by Bergman (2009). The proposed distribution has the probability density function

$$f(\boldsymbol{x}, \boldsymbol{y}) = A \left(\prod_{j=1}^{D} x_j^{\alpha_j - 1} y_j^{\beta_j - 1} \right) \left(\boldsymbol{x}^\top \boldsymbol{y} \right)^\gamma, \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_D)^\top \in \mathscr{S}^D$, $\boldsymbol{y} = (y_1, \ldots, y_D)^\top \in \mathscr{S}^D$, 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 under various parameter settings are given in Bergman (2009). 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 $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent in that case. The bicompositional Dirichlet distribution forms an exponential family with natural parameters $\boldsymbol{\theta} = (\gamma, \tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})^\top$, where $\tilde{\alpha}_j = \alpha_j - 1$ and $\tilde{\beta}_j = \beta_j - 1$.

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

$$F(\boldsymbol{\theta}) = \int \log f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) g(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y}, \tag{2}$$

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

By choosing $\boldsymbol{\theta}_i$ to maximize $F(\boldsymbol{\theta})$ in the parameter space Θ_i, '$\boldsymbol{\theta}_i$ is the theoretical analogue of the maximum likelihood estimate of $\boldsymbol{\theta}$ over the parameter space Θ_i' (Kent 1983). We divide $\boldsymbol{\theta}$ into two parts $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})$, where $\boldsymbol{\psi}$ is the parameter of interest and $\boldsymbol{\lambda}$ is a nuisance parameter.

If the model forms an exponential family

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

the maximized Fraser information may be calculated as

$$F(\boldsymbol{\theta}_i) = \boldsymbol{\theta}_i{}^\top \boldsymbol{b}(\boldsymbol{\theta}_i) - c(\boldsymbol{\theta}_i), \tag{3}$$

for the two parameter spaces $\Theta_i, i = 1, 2$, where $\boldsymbol{b}(\boldsymbol{\theta})$ is the vector of partial derivatives of $c(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

If for $\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{\psi} = \mathbf{0}\}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are modelled as independent, the information gain of allowing for dependence between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in the model is $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0) = 2\{F(\boldsymbol{\theta}_1) - F(\boldsymbol{\theta}_0)\}$. Since $F(\boldsymbol{\theta}_i)$ is the maximized expected log likelihood, $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ is the theoretical analogue of -2 times the log likelihood ratio statistic.

Kent (1983) proposed a joint correlation coefficient between $\boldsymbol{X}$ and $\boldsymbol{Y}$ defined as $\rho_J^2 = 1 - \exp\{-\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)\}$. As is easily seen, $0 \leq \rho_J^2 < 1$ and tending to one as $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ tends to infinity. Independence between $\boldsymbol{X}$ and $\boldsymbol{Y}$ implies zero ρ_J^2-correlation if $g(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$ for some $\boldsymbol{\theta}$, or 'the model Θ_1 forms a regular exponential family' (Inaba & Shirahata 1986, p. 346). The subscript J indicates that the information gain is with respect to the joint distribution rather than to conditional distributions which in general would give a different kind of correlation measure, see Kent (1983).

We shall assume that the true density function $g(\boldsymbol{x}, \boldsymbol{y})$ is the bicompositional Dirichlet probability density function (1) and that the two families of parametric models $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}_i)$ also are bicompositional Dirichlet distributions. The Fraser information now equals the Kullback-Leibler information (Kullback & Leibler 1951). The parameter of interest ψ in these models is γ. Denoting $\boldsymbol{\theta}_1 = (\gamma^{(1)}, \tilde{\boldsymbol{\alpha}}^{(1)}, \tilde{\boldsymbol{\beta}}^{(1)})^\top$ and $\boldsymbol{\theta}_0 = (\gamma^{(0)}, \tilde{\boldsymbol{\alpha}}^{(0)}, \tilde{\boldsymbol{\beta}}^{(0)})^\top$, 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 $\tilde{\boldsymbol{\beta}}^{(0)} \neq \tilde{\boldsymbol{\beta}}$. Hence calculation of $F(\boldsymbol{\theta}_0)$ requires maximization, usually numerically, with respect to $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$.

2.1 The bicomponent case

For two components, the calculation of the joint correlation coefficient ρ_J^2 has been done through the information gain as described earlier for some different situations. Figure 1 depicts the joint correlation coefficient ρ_J^2, calculated for five different sets of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ values and 49 values of γ ranging from -1.25 to 2.0.

FIGURE 1 ABOUT HERE

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. This may seem peculiar but is a natural consequence of that the intra-dependence between the parts of a composition implies that correlations between two compositions are in general not only depending of the corresponding components of the two compositions, but also have to be influenced by the remaining parts of the composition, and thus also of the parameters that rule their relative magnitudes.

It may 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 likely due to numerical issues in the maximization.

2.2 The tricomponent case

The normalization constant in the multicomponent case of the bicompositional Dirichlet distribution is hitherto calculated only for γ being non-negative integers (Bergman 2009). We will hence calculate the joint correlation coefficient only for such γ values. Also, since differentiation with respect to γ may not be meaningful, we will not use (3) in the calculation, but will instead utilize the definition given in (2).

The Fraser information for the tricomponent bicompositional Dirichlet distribution is

$$
\begin{aligned}
F(\boldsymbol{\theta}_i) = \int \log\{A x_1^{\alpha_1^{(i)}-1} x_2^{\alpha_2^{(i)}-1} x_3^{\alpha_3^{(i)}-1} y_1^{\beta_1^{(i)}-1} y_2^{\beta_2^{(i)}-1} y_3^{\beta_3^{(i)}-1} (\boldsymbol{x}^\top \boldsymbol{y})^{\gamma^{(i)}}\} g(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \\
= \log A + (\alpha_1^{(i)} - 1) \int \log(x_1) g(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \\
+ \cdots \\
+ (\beta_3^{(i)} - 1) \int \log(y_3) g(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \\
+ \gamma^{(i)} \int \log(x_1 y_1 + x_2 y_2 + x_3 y_3) g(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y},
\end{aligned}
$$

for $i = 0, 1$. Thus $F(\boldsymbol{\theta}_i)$ equals the sum of a constant, six log expectations, and the expectation $\mathrm{E}\{\log(\boldsymbol{x}^\top \boldsymbol{y})\}$.

Using the Multinomial Theorem and also that for a Dirichlet distributed random variable $\boldsymbol{X} = (X_1, \ldots, X_D)^\top$, $\mathrm{E}\{\log(X_j)\} = \Psi(\alpha_j) - \Psi(\alpha_.)$, where Ψ denotes the digamma function (Abramowitz & Stegun 1964; Aitchison 1986) and for generic a, $a_. = a_1 + \cdots + a_D$, we may calculate the first seven terms of $F(\boldsymbol{\theta}_i)$ exactly. For example:

$$
\begin{aligned}
\int \log(x_j) g(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \\
= A \sum_{\substack{k_i \geq 0 \\ k_. = \gamma}} \binom{\gamma}{\boldsymbol{k}} \frac{\prod_{i=1}^3 \Gamma(\alpha_i + k_i)}{\Gamma(\alpha_. + \gamma)} \frac{\prod_{i=1}^3 \Gamma(\beta_i + k_i)}{\Gamma(\beta_. + \gamma)} \left\{\Psi(\alpha_j + k_j) - \Psi(\alpha_. + \gamma)\right\}.
\end{aligned}
$$

where

$$
\binom{\gamma}{\boldsymbol{k}} = \frac{\gamma!}{k_1! k_2! k_3!},
$$

i.e. the multinomial coefficient.

The integral $\int \log(y_j) g(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}$ 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 I 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 $\boldsymbol{\beta}^{(0)}$.

Using these approaches the joint correlation coefficient has been calculated for bicomponent models with parameters $\boldsymbol{\alpha} = (2.1, 2.4)^\top$ and $\boldsymbol{\beta} = (2.2, 2.3)^\top$ for γ ranging from -2 to 8, and for tricomponent models with parameters $\boldsymbol{\alpha} = (2.1, 2.4, 2.3)^\top$ and $\boldsymbol{\beta} = (2.2, 2.3, 2.1)^\top$ for non-negative integer values of γ upto 8.

FIGURE 2 ABOUT HERE

In Figure 2 these values are plotted and we see how the joint correlation coefficient is levelling off towards 1 as γ increases, something that is not really visible in Figure 1. We can also see again that the joint correlation not only depend on γ but to some extent on the full set of parameters in the distribution.

In both the bicomponent and tricomponent case the Fraser information, and hence the joint correlation coefficient, is calculated 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).

3 Estimation

We will now focus on estimating the joint correlation coefficient in a bicompositional Dirichlet probability density function under the restriction of a bicomponent model.

The parameter space for models allowing dependence between the compositions, i.e. unrestricted models, is $\Theta_1 = \{\alpha_1 > 0, \alpha_2 > 0, \beta_1 > 0, \beta_2 > 0, \gamma > -\min(\alpha_1 + \beta_2, \alpha_2 + \beta_1)\}$ while the parameter for models not allowing dependence is $\Theta_0 = \{\alpha_1 > 0, \alpha_2 > 0, \beta_1 > 0, \beta_2 > 0, \gamma = 0\}$. The information gained by allowing dependence $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ may be estimated by

$$\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) = \frac{2}{n}\{\sum_{k=1}^{n} \log f(\boldsymbol{x}_k, \boldsymbol{y}_k; \hat{\boldsymbol{\theta}}_1) - \sum_{k=1}^{n} \log f(\boldsymbol{x}_k, \boldsymbol{y}_k; \hat{\boldsymbol{\theta}}_0)\}, \tag{4}$$

where $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_0$ are the maximum likelihood estimates under the parameter spaces Θ_1 and Θ_0, respectively (Kent 1983).

3.1 Maximum likelihood estimates

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

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

where

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

Here $S_{\boldsymbol{\alpha}} = \sum_{j=0}^{i} [i!/\{j!(i-j)!\}](-1)^{i-j} B(\alpha_1 + j, \alpha_2 + i - j)$, where $B(\cdot, \cdot)$ denotes the Beta function.

Finding the maximum likelihood estimates will in general require numerical methods. We stress the fact that the parameter space for γ depends on the values of the other parameters.

The maximum likelihood estimate of $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ under the parameter space Θ_i $(i = 0, 1)$ is denoted $\hat{\boldsymbol{\theta}}_i$. An estimator of the joint correlation coefficient is thus $\hat{\rho}_J^2 = 1 - \exp\{-\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0)\}$.

3.2 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 $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ 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' $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ are likely to prove most useful.'

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

$$\left\{\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) - (s^2 \chi^2_{1;\alpha} n^{-1})^{1/2}, \widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) + (s^2 \chi^2_{1;\alpha} n^{-1})^{1/2}\right\} \tag{5}$$

where s^2 is the sample variance of $2\log\{f(\boldsymbol{x}_j, \boldsymbol{y}_j; \hat{\boldsymbol{\theta}}_1)/f(\boldsymbol{x}_j, \boldsymbol{y}_j; \hat{\boldsymbol{\theta}}_0)\}$ for $j =$

$1, \ldots, n$ 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 $\hat{a}$)

$$\left\{ \frac{\mu\kappa_{1;\alpha/2}(\hat{a}/\mu)}{n}, \frac{\mu\delta_{1;\alpha/2}(\hat{a}/\mu)}{n} \right\}, \tag{6}$$

where $\hat{a} = n\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0)$ and $\kappa_{1;\alpha}(a)$ and $\delta_{1;\alpha}(a)$ are the values of the non-centrality parameters of a non-central chi square distribution defined by $\Pr[\chi^2_1\{\kappa_{1;\alpha}(a)\} \geq a] = \alpha$ and $\Pr[\chi^2_1\{\delta_{1;\alpha}(a)\} \leq a] = \alpha$, respectively, except that $\kappa_{1;\alpha}(a) \equiv 0$ if $\Pr[\chi^2_1\{0\} \geq a] > \alpha$ (Kent 1983, p. 169). The constant μ is equal to 1, as we are convinced that the true density function belongs to $\{f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) | \boldsymbol{\theta} \in \Theta_1\}$. The α in (5) and (6) is one minus the confidence coefficient, not to be confused with the parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$ of the bicomponent bicompositional Dirichlet distribution.

We thus transform the confidence intervals of $\Gamma(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_0)$ yielding the 'large'

$$\begin{aligned} \Bigg[1 - \exp\left\{ -\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) + (s^2\chi^2_{1;\alpha}n^{-1})^{1/2} \right\}, \\ 1 - \exp\left\{ -\widehat{\Gamma}(\hat{\boldsymbol{\theta}}_1 : \hat{\boldsymbol{\theta}}_0) - (s^2\chi^2_{1;\alpha}n^{-1})^{1/2} \right\} \Bigg] \end{aligned} \tag{7}$$

and the 'small'

$$\left[1 - \exp\left\{ -\frac{\kappa_{1;\alpha/2}(\hat{a})}{n} \right\}, 1 - \exp\left\{ -\frac{\delta_{1;\alpha/2}(\hat{a})}{n} \right\} \right] \tag{8}$$

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

Asymptotically the likelihood ratio test for $\boldsymbol{\theta} \in \Theta_0$ (i.e. $\gamma = 0$, or $\rho_J^2 = 0$) against $\boldsymbol{\theta} \in \Theta_1$ rejects Θ_0 if and only if the one-sided upper confidence interval analogous to (6) does not contain 0.

4 Comparison of the confidence intervals

In order to examine the properties of the two confidence intervals (7) and (8), we conducted a Monte Carlo study for six models with different ρ_J^2 and for different numbers of observations ($n = 50, 100, 250$). For every combination of model and number of observations we generate random variates (Bergman 2012), estimate $\hat{\rho}_J^2$, compute the two confidence intervals, and record in how many cases the true value of ρ_J^2 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 empirical confidence coefficients are close to this; they vary between 0.90 and 1.00. We note that especially the 'large' confidence intervals seem to have empirical confidence coefficients that are too high, indicating overly wide confidence intervals. 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 lower limit less than zero would in practice be replaced by zero as both the information gain and the joint correlation coefficient are non-negative. On the other hand, a confidence limit that is not restricted to the appropriate parameter space is of less practical use.

TABLE 1 ABOUT HERE

Kent (1983) notes that the estimator (4) is biased and suggests a less biased estimator. However, in our case numerical examples indicated 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 calculations, which consist of large numbers of infinite sums. Due to this lack of improvement we have not used this bias correction in our estimations. It can however be concluded from this limited study that the 'small' interval gives a fairly reasonable (correct) coverage probability in general, with a tendency to be slightly too conservative (too wide) when ρ_J^2 is large.

5 An example

We illustrate the estimation of the joint correlation coefficient presented in Section 3 with an example. The data consist of the composition of the government gross domestic product for the 50 U.S. states and District of Columbia, for the years 1967, 1977 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.

We estimate the correlation between the GDP composition in 1967 and that in 1997. The maximum likelihood estimates of the parameters under Θ_1 are $\hat{\boldsymbol{\alpha}} = (16.32, 14.41)^\top$, $\hat{\boldsymbol{\beta}} = (17.31, 43.20)^\top$, and $\hat{\gamma} = 57.41$. The estimate of the joint correlation coefficient is $\hat{\rho}_J^2 = 0.3027$, with a 'small' confidence interval of $(0.0993, 0.5371)$. An analogous one-sided upper confidence interval will have lower bound at least 0.0993, thus indicating that composition of the government GDP in 1967 is correlated with the composition of the government GDP in 1997.

For comparison, we also calculated the corresponding values for the years 1977 and 1997. This yielded the following maximum likelihood estimates: $\hat{\boldsymbol{\alpha}} = (19.02, 31.14)^\top$, $\hat{\boldsymbol{\beta}} = (18.33, 45.98)^\top$, and $\hat{\gamma} = 51.02$. The estimated joint correlation coefficient is $\hat{\rho}_J^2 = 0.3538$, with a 'small' 95 % confidence interval of $(0.1369, 0.5852)$. We note that the estimated correlation is slightly higher between 1977 and 1997 than between 1967 and 1997. This could be explained by the greater temporal proximity. The confidence intervals are however to a large extent overlapping.

6 Discussion

The compositional data analysis has through history primarily been concerned with modelling the dependence between the components of one composition, the intra-compositional dependence. This is a very natural focus due to the intrinsic correlation of compositions. However, understanding and modelling the dependence between two compositions, the inter-compositional dependence, is also of interest.

We have considered correlation as a measure of similarity in accordance with e.g. Dodge (2003). We have demonstrated that the joint correlation coefficient measuring the similarity between two compositions may be found under the assumption of a bicompositional Dirichlet distribution. As the joint correlation coefficient is in the range 0 to 1, this also enables the possibility to compare the dependence between two D-part compositions with the dependence between two N-part compositions for $D \neq N$.

The approach taken here also allows for the possibility of calculating correlation between one D-part composition and another N-part composition when a suitable family of joint distributions on $\mathscr{S}^D \times \mathscr{S}^N$ is defined.

The use of a single-valued measure for describing the inter-compositional dependence based on the information gain, is in line with the fact that ρ_J^2 corresponds, in a different model, to a combination of the squared canonical correlations between two multivariate dependent normal variables (Kent 1983, Sec 10), thus reflecting the dependence not only in the same coordinate but also the dependence due to all other coordinates of the two variables. This is furthermore closely related to what would be the correlation measure between compositions whose log-ratio transforms were dependent multivariate normally distributed, i.e. themselves logistic-normal, as outlined in Section 2.

We have also shown how to estimate this joint correlation coefficient with a point estimate and two confidence intervals. The two confidence intervals were compared 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 are in general wider.

We have demonstrated that the joint correlation coefficient between compositions can be calculated for real data using an example for American gross domestic product data.

Future work will include considering other families of bicompositional joint distributions allowing for the calculation of the joint correlation between the compositions.

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

Integrating over $\mathscr{S}^3 \subset \mathscr{R}^3$ is equivalent to integrating over a triangle in $\mathscr{R}^2$ defined by a coordinate along the basis of the triangle, $0 < u < 2^{1/2}$, and an orthogonal coordinate from the midpoint of this basis up to the edge of the triangle, $0 < v < (3/2)^{1/2} - 3^{1/2}\left|2^{-1/2} - u\right|$. Analogously, integration over $\mathscr{S}^3 \times \mathscr{S}^3$ 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

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

the integral of a function $h(\boldsymbol{x}, \boldsymbol{y})$ over $\mathscr{S}^3 \times \mathscr{S}^3$ becomes

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

where the Jacobian of the transformation is 1/3.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover, 9th ed.

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

AITCHISON, J. & SHEN, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika* **67**, 261–272.

BERGMAN, J. (2009). A bicompositional Dirichlet distribution. Tech. Rep. 3, Department of Statistics, Lund University.

BERGMAN, J. (2012). Generating random variates from a bicompositional Dirichlet distribution. *Journal of Statistical Computation and Simulation* **82**, 797–805.

BRAY, J. R. & CURTIS, J. T. (1957). An ordination of upland forest communities of southern Wisconsin. *Ecological Monographs* **27**, 325–349.

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

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

JØRGENSEN, B. & LAURITZEN, S. (2000). Multivariate dispersion models. *Journal of Multivariate Analysis* **74**, 267–281.

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

KULLBACK, S. & LEIBLER, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86.

MATEU-FIGUERAS, G., PAWLOWSKY-GLAHN, V. & BARCELÓ-VIDAL, C. (2005). The additive logistic skew-normal distribution on the simplex. *Stochastic Environmental Research and Risk Assessment* **19**, 205–214.

NADARAJAH, S. (2006). The bivariate F_3-beta distribution. *Commun. Korean Math. Soc.* **21**, 363–374.

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

NADARAJAH, S. & KOTZ, S. (2005). Some bivariate beta distributions. *Statistics* **39**, 457–466.

OLKIN, I. & LIU, R. (2003). A bivariate beta distribution. *Statist. Probab. Lett.* **62**, 407–412.

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

Table 1: The empirical confidence coefficient is presented for six 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 for nominal confidence level 95 %. 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	0.926	0.954
				100	0.930	0.958
				250	0.992	0.970
(9, 7)	(4, 2)	4.5	0.099	50	0.992	0.964
				100	0.978	0.904
				250	0.998	0.960
(4, 3)	(4, 2)	4.5	0.174	50	0.998	0.950
				100	0.978	0.918
				250	1.000	0.942
(4, 3)	(3, 4)	4.5	0.244	50	0.998	0.954
				100	1.000	0.968
				250	0.998	0.958
(4, 3)	(3, 4)	9.5	0.652	50	1.000	0.976
				100	1.000	0.944
				250	1.000	0.986
(4, 3)	(3, 4)	14.0	0.867	50	0.998	0.982
				100	1.000	0.980
				250	0.996	0.984

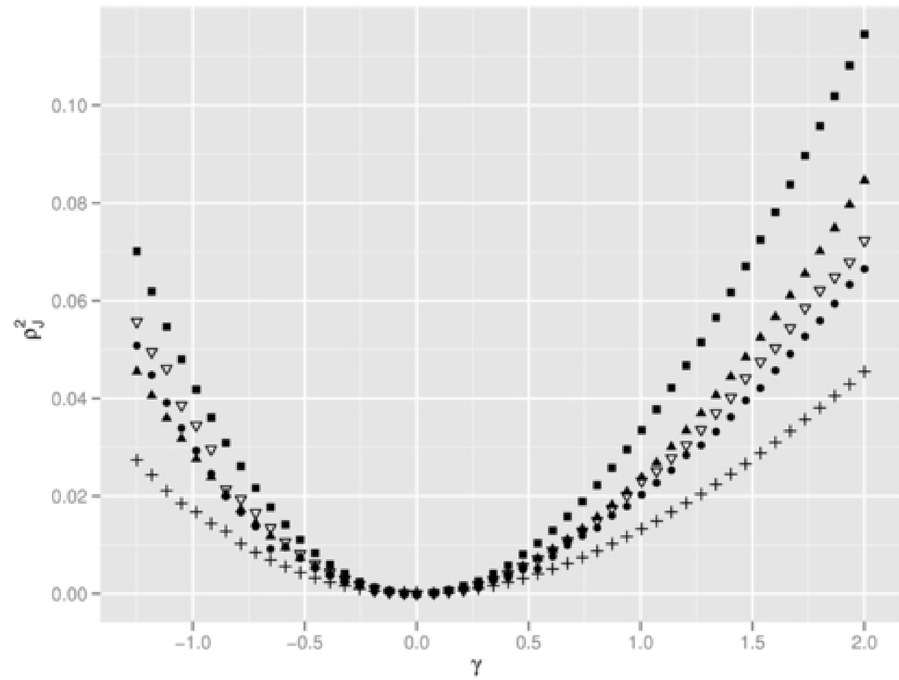

Figure 1: The joint correlation coefficient ρ_J^2 calculated for γ ranging from -1.25 to 2.0 for the $(\boldsymbol{\alpha};\boldsymbol{\beta})$ parameter values (2.1, 2.4; 2.2, 2.3) (■), (2.1, 2.2; 3.6, 3.5) (▲), (5.2, 2.0; 2.0, 2.0) (▽), (1.9, 6.4; 3.2, 2.1) (●) and (4.1, 2.4; 4.1, 2.4) (+).

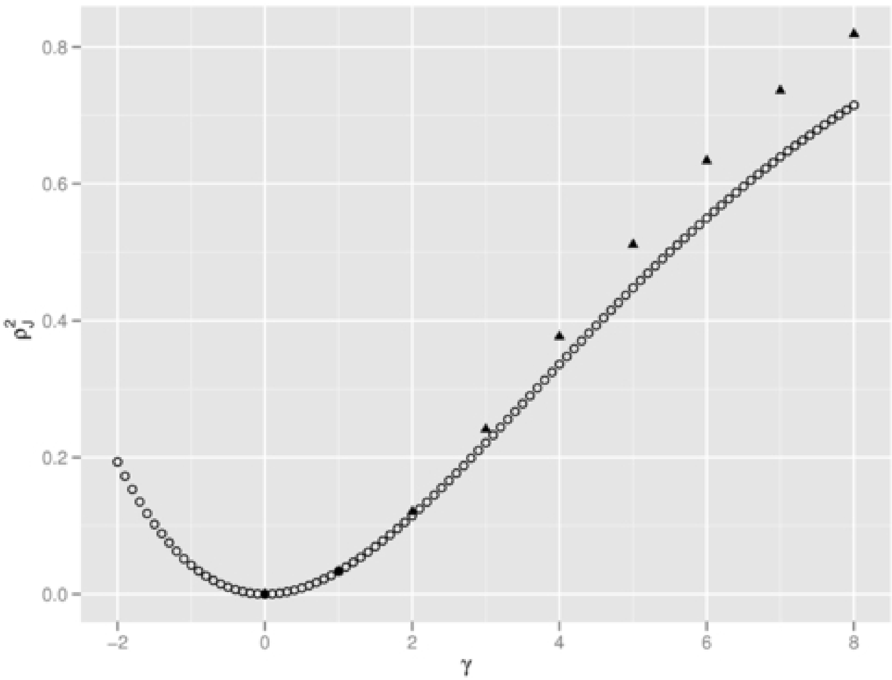

Figure 2: The joint correlation coefficient ρ_J^2 calculated for γ ranging from -2 to 8 for bicomponent models with $(\boldsymbol{\alpha};\boldsymbol{\beta})$ parameter values (2.1, 2.4; 2.2, 2.3) ($\circ$) and tricomponent models with $(\boldsymbol{\alpha};\boldsymbol{\beta})$ parameter values (2.1,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