Estimating a measure of dependence between two compositions

Jakob Bergman

2010:1

LUND
UNIVERSITY

DEPARTMENT OF STATISTICS

S-220 07 LUND
SWEDEN

Estimating a measure of dependence between two compositions

Jakob Bergman*†

February 19, 2010

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

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

*Department of Statistics, Lund University, Box 743, SE-220 07 Lund, Sweden
†e-mail: jakob.bergman@stat.lu.se

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

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

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 $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)^{\mathrm{T}}$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_D)^{\mathrm{T}}$ and γ, and the probability density function is

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

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 $A(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$ 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

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

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 $I(\theta_1 : \theta_0)$ may be estimated by

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

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

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

and the log likelihood function is

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

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

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

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

with

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

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

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

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

An estimator of the general measure of correlation is thus

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

2.2 Confidence intervals

Kent (1983) gives two proposals concerning confidence intervals for $\Gamma(\theta_1 : \theta_0)$: when the value of $\Gamma(\theta_1 : \theta_0)$ 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"

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

and the "small"

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

$1 - \alpha$ 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_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, 2009b), 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 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

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

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 (α, β, γ) 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	
α	β	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

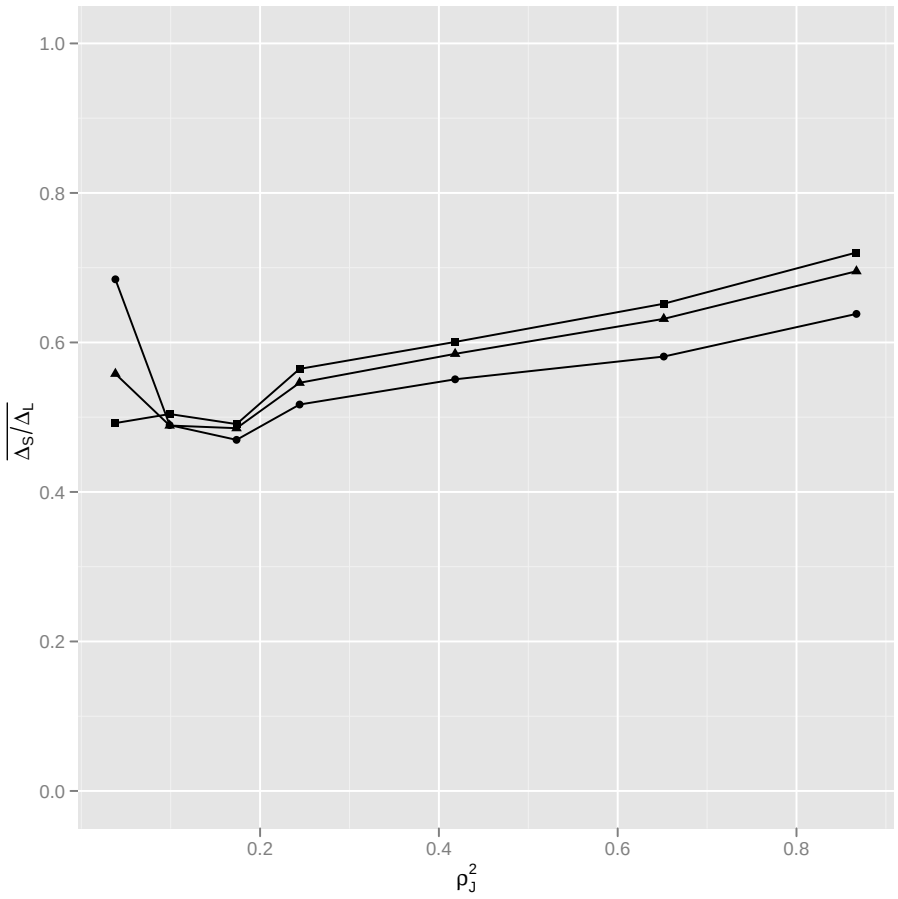

Figure 1. The average ratio between the widths of the "small" (Δ_S) and "large" (Δ_L) confidence intervals plotted for the seven different models (ρ_J^2) in Table 1 and for 50 (●), 100 (▲) and 250 (■) observations.

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 $\alpha_\cdot = \alpha_1 + \alpha_2$ and $\beta_\cdot = \beta_1 + \beta_2$. 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

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

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

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

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

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

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

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

where S_α is given in (8) and S_β is given in (9).

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

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

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.

A Derivatives of binomial coefficients

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

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

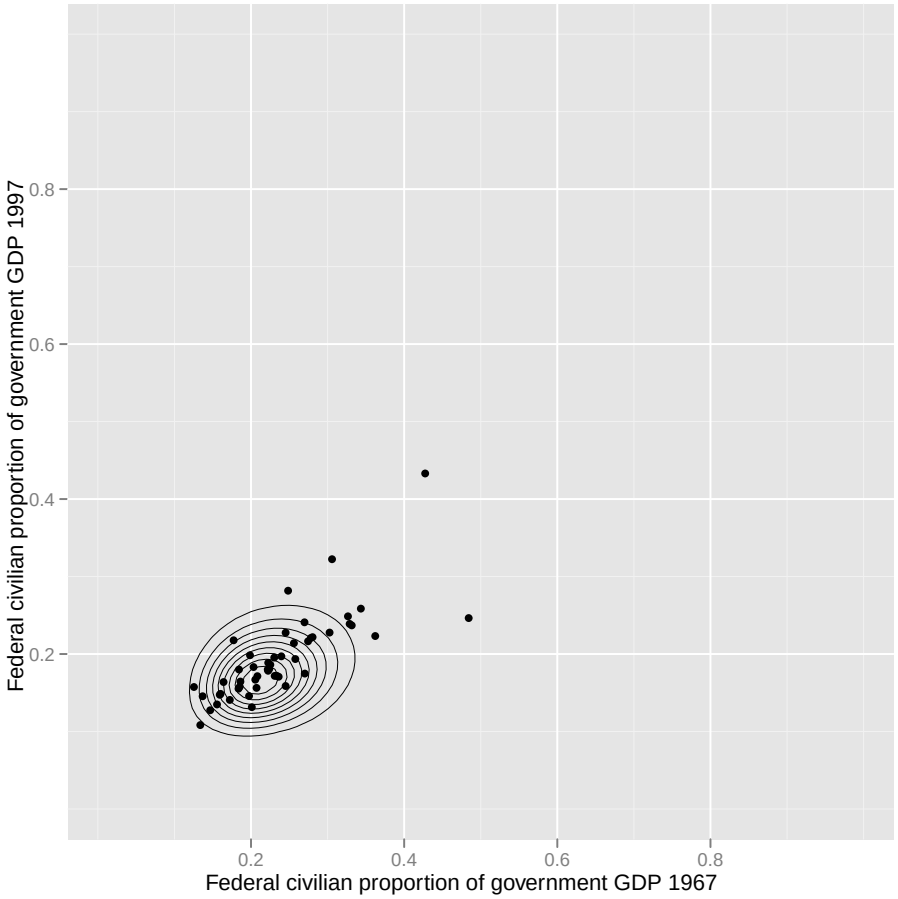

Figure 2. The federal civilian proportion of the government part of GDP for the 50 U.S. states and District of Columbia plotted for 1997 versus 1967 and the contour curves of the estimated bicompositional Dirichlet distribution.

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

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 $r-k$. □

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.