A measure of dependence between two compositions

Jakob Bergman and Björn Holmquist

2009:4

LUND UNIVERSITY

DEPARTMENT OF STATISTICS

S-220 07 LUND
SWEDEN

A measure of dependence between two compositions

Jakob Bergman[*†] and Björn Holmquist[*]

November 18, 2009

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

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

[*]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. 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 similiarity 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.

A joint correlation coefficient between $\mathbf{X}$ and $\mathbf{Y}$ is proposed in Kent (1983) by

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

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

Independence between $\mathbf{X}$ and $\mathbf{Y}$ implies zero correlation if $g(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}; \theta)$ for some θ, 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

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

where $\mathbf{x}, \mathbf{y} \in \mathscr{S}^D$, $\alpha_j, \beta_j \in \mathscr{R}_+ (j = 1, \ldots, D)$ and $\gamma \in \mathscr{R}$. 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 α and β respectively, and hence $\mathbf{X}$ and $\mathbf{Y}$ are independent.

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 $\theta = (\gamma, \tilde{\alpha}, \tilde{\beta})^{\mathrm{T}}$, 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 $\theta_1 = (\gamma^{(1)}, \tilde{\alpha}^{(1)}, \tilde{\beta}^{(1)})^{\mathrm{T}}$ and $\theta_0 = (\gamma^{(0)}, \tilde{\alpha}^{(0)}, \tilde{\beta}^{(0)})^{\mathrm{T}}$, it can be shown, through the information inequality, that $\gamma^{(1)} = \gamma$, $\tilde{\alpha}^{(1)} = \tilde{\alpha}$, and $\tilde{\beta}^{(1)} = \tilde{\beta}$, but when $\gamma^{(0)} = 0$, in general $\tilde{\alpha}^{(0)} \neq \tilde{\alpha}$ and $\tilde{\beta}^{(0)} \neq \tilde{\beta}$.

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 $\boldsymbol{\theta} = (\gamma, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)^{\mathrm{T}}$ derive the following expressions:

$$\begin{aligned} c(\boldsymbol{\theta}) &= \log\left\{ 2^{-\gamma} \sum_{i=0}^{\infty} \binom{\gamma}{i} S_{\tilde{\alpha}} S_{\tilde{\beta}} \right\}, && (5) \\ \mathbf{b}(\boldsymbol{\theta}) &= \left(\frac{\partial c}{\partial \gamma}, \frac{\partial c}{\partial \tilde{\alpha}_1}, \frac{\partial c}{\partial \tilde{\alpha}_2}, \frac{\partial c}{\partial \tilde{\beta}_1}, \frac{\partial c}{\partial \tilde{\beta}_2} \right)^{\mathrm{T}}. && (6) \end{aligned}$$

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

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

where

$$\begin{aligned} \binom{\gamma}{i}' &= \frac{\mathrm{d}}{\mathrm{d}\gamma} \binom{\gamma}{i}, && (7) \\ B_{ij}(\tilde{\boldsymbol{\alpha}}) &= B(\tilde{\alpha}_1 + j + 1; \tilde{\alpha}_2 + i - j + 1), && (8) \\ \psi_{ij}^{(1)}(\tilde{\alpha}_1, \tilde{\alpha}_2) &= \psi(\tilde{\alpha}_1 + j + 1) - \psi(\tilde{\alpha}_1 + \tilde{\alpha}_2 + i + 2), && (9) \\ \psi_{ij}^{(2)}(\tilde{\alpha}_1, \tilde{\alpha}_2) &= \psi(\tilde{\alpha}_2 + i - j + 1) - \psi(\tilde{\alpha}_1 + \tilde{\alpha}_2 + i + 2). && (10) \end{aligned}$$

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 $\psi(x) = \mathrm{d}\log\Gamma(x)/\mathrm{d}x$.

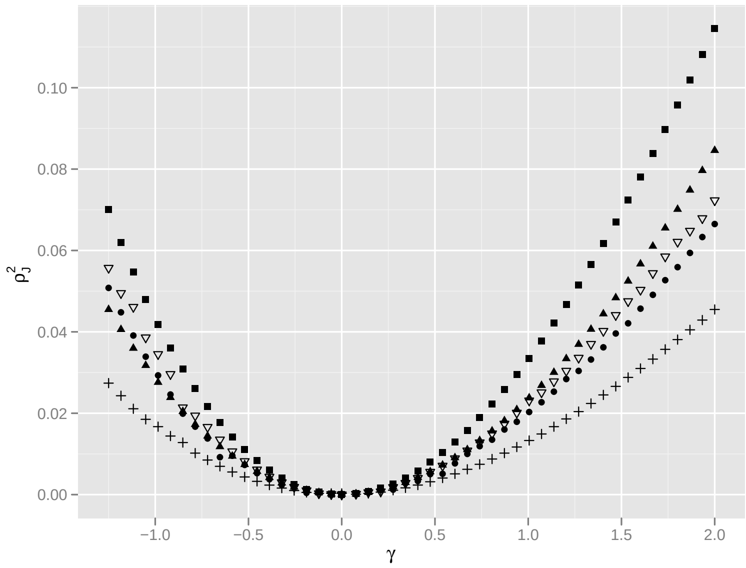


Figure 1. The joint correlation coefficient ρ_J^2 calculated for γ ranging from -1.25 to 2.0 for the $(\boldsymbol{\alpha}; \boldsymbol{\beta})$ parameter values (2.1, 2.4; 2.2, 2.3) (■), (2.1, 2.2; 3.6, 3.5) (▲), (5.2, 2.0; 2.0, 2.0) (▽), (1.9, 6.4; 3.2, 2.1) (●) and (4.1, 2.4; 4.1, 2.4) (+).

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

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 numerically, 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 the ρ_J^2 is not symmetric around 0; the rate at which ρ_J^2 increases 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 non-negative integers of γ. 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:

$$\begin{aligned}
F(\boldsymbol{\theta}_i) &= \int \log\{f(\mathbf{x},\mathbf{y};\boldsymbol{\theta}_i)\}g(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= \int \log\left\{Ax_1^{\alpha_1^{(i)}-1}x_2^{\alpha_2^{(i)}-1}x_3^{\alpha_3^{(i)}-1}y_1^{\beta_1^{(i)}-1}y_2^{\beta_2^{(i)}-1}y_3^{\beta_3^{(i)}-1}(\mathbf{x}^{\mathrm{T}}\mathbf{y})^{\gamma^{(i)}}\right\}g(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= \log A + (\alpha_1^{(i)}-1)\int \log(x_1)g(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&\qquad + \cdots \\
&\qquad + (\beta_3^{(i)}-1)\int \log(y_3)g(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&\qquad + \gamma^{(i)}\int \log(x_1y_1+x_2y_2+x_3y_3)g(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}
\end{aligned}$$

Thus $F(\boldsymbol{\theta}_i)$ 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 $\boldsymbol{\alpha}$ belongs to the 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 members of the 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 seven first terms of $F(\theta_i)$ exactly. For example:

$$
\begin{aligned}
&\int \log(x_j) g(\mathbf{x},\mathbf{y}) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= \int \log(x_j) A x_1^{\alpha_1-1} x_2^{\alpha_2-1} x_3^{\alpha_3-1} y_1^{\beta_1-1} y_2^{\beta_2-1} y_3^{\beta_3-1} (x_1y_1 + x_2y_2 + x_3y_3)^\gamma \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= \int \log(x_j) \sum_{\substack{k_i \geq 0 \\ k_\cdot = \gamma}} \binom{\gamma}{\mathbf{k}} A x_1^{\alpha_1+k_1-1} x_2^{\alpha_2+k_2-1} x_3^{\alpha_3+k_3-1} y_1^{\beta_1+k_1-1} y_2^{\beta_2+k_2-1} y_3^{\beta_3+k_3-1} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= A \sum_{\substack{k_i \geq 0 \\ k_\cdot = \gamma}} \binom{\gamma}{\mathbf{k}} \int \log(x_j) x_1^{\alpha_1+k_1-1} x_2^{\alpha_2+k_2-1} x_3^{\alpha_3+k_3-1} \mathrm{d}\mathbf{x} \int y_1^{\beta_1+k_1-1} y_2^{\beta_2+k_2-1} y_3^{\beta_3+k_3-1} \mathrm{d}\mathbf{y} \\
&= A \sum_{\substack{k_i \geq 0 \\ k_\cdot = \gamma}} \binom{\gamma}{\mathbf{k}} \frac{\prod_{i=1}^3 \Gamma(\alpha_1 + k_i)}{\Gamma(\alpha_\cdot + \gamma)} \frac{\prod_{i=1}^3 \Gamma(\beta_1 + k_i)}{\Gamma(\beta_\cdot + \gamma)} \left\{ \psi(\alpha_j + k_j) - \psi(\alpha_\cdot + \gamma) \right\}
\end{aligned}
$$

In the above expression

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

and thus denotes the multinomial coefficient.

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 $\alpha_\cdot$ are replaced by β_j and $\beta_\cdot$ respectively.

The last term of $F(\theta_1)$ must be integrated numerically. (See Appendix B for integration over $\mathscr{S}^3 \times \mathscr{S}^3$.) This is not the case for $F(\theta_0)$, as $\gamma^{(0)} = 0$, but instead, in order to obtain θ_0, $F(\theta)$ 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 leveling off towards 1 as γ increases, something that is not really visible in Figure 1.

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.

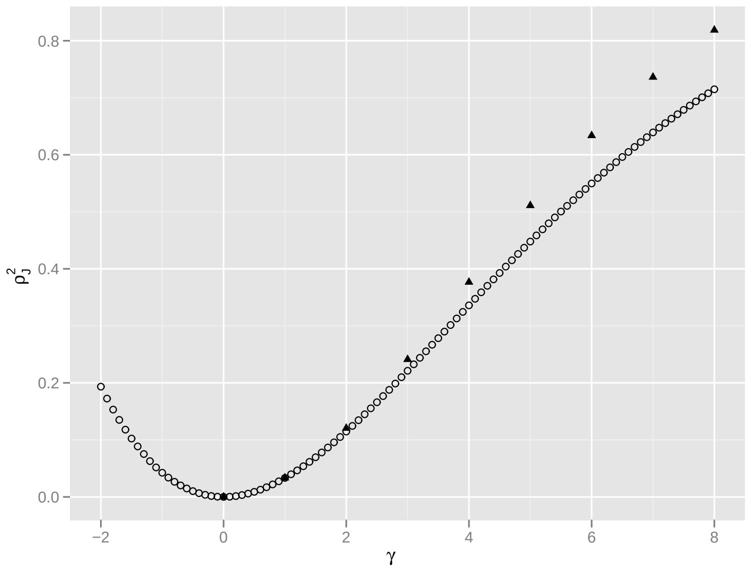


Figure 2. The joint correlation coefficient ρ_J^2 calculated for γ ranging from -2 to 8 for bicomponent models with $(\boldsymbol{\alpha}; \boldsymbol{\beta})$ parameter values (2.1, 2.4; 2.2, 2.3) (○) and tricomponent models with $(\boldsymbol{\alpha}; \boldsymbol{\beta})$ parameter values (2.1, 2.4,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

The binomial coefficient may also be expressed as

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

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

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

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. Hence 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 nominator 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

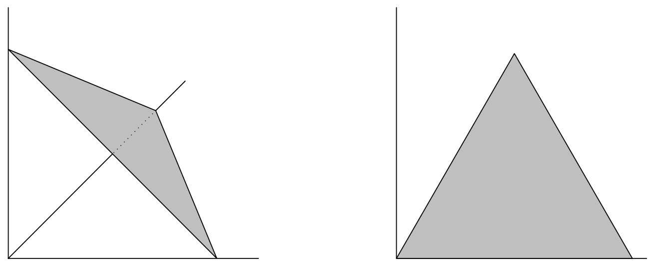
$$f(r) = r(r-1)\cdots(r-(k-1)),$$

we may write

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

where

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



(a) (b)

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

and hence

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

□

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

The simplex $\mathscr{S}^3$ is the triangle in the $\mathscr{R}^3_+$ 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

$$\mathbf{x}(s,t) = \begin{pmatrix} t\left(\tfrac{2}{3}\right)^{1/2} \\ s2^{-1/2} - t6^{-1/2} \\ 1 - s6^{-1/2} - t2^{-1/2} \end{pmatrix}, \qquad \mathbf{y}(u,v) = \begin{pmatrix} v\left(\tfrac{2}{3}\right)^{1/2} \\ u2^{-1/2} - v6^{-1/2} \\ 1 - u6^{-1/2} - v2^{-1/2} \end{pmatrix}$$

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, December). 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.