

Compositional tables analysis in coordinates

KAMILA FAČEVIČOVÁ

Department of Mathematical Analysis and Applications of Mathematics,
Palacký University Olomouc

KAREL HRON

Department of Mathematical Analysis and Applications of Mathematics,
Palacký University Olomouc

VALENTIN TODOROV

United Nations Industrial Development Organisation

MATTHIAS TEMPL

Institute of Statistics and Probability Theory, Vienna University of
Technology

Running headline: Fačevicová et al.: CoDa tables analysis

Abstract Compositional tables – a continuous counterpart to the contingency tables – carry relative information about relationships between row and column factors, thus for their analysis only ratios between cells of a table are informative. Consequently, the standard Euclidean geometry should be replaced by the Aitchison geometry on the simplex that enables decomposition of the table into its independent and interactive parts. The aim of the paper is to find interpretable coordinate representation for independent and interaction tables (in sense of balances and odds ratios of cells, respectively), where further statistical processing of compositional tables can be performed. Theoretical results are applied to real-world problems from a health survey and in macroeconomics.

Keywords: Aitchison geometry on the simplex; balances; compositional data; contingency tables; isometric logratio transformation

1 Introduction

The statistical analysis of a relation between two discrete random variables (factors) is popularly performed using contingency tables. Their cells contain counts resulting from a random sample. An extension of this basic problem setting is to analyze a sample of tables, where each table is a realization of a discrete bivariate distribution. Thus they represent a particular case of distribution-valued variables (?), where individual observations are summarized by distributions to preserve their intrinsic variability and enable to analyze statistically groups of individuals in a meaningful way. As an example we can mention a sample of tables containing the joint distribution of animals and vegetation in different plots of a survey area (see ?, p. 387) or the social vs. the economic structure of population in EU countries. Such a sample of tables is now subsequentially expressed as the term compositional tables (??). A compositional table itself represents one observation in a sample of distribution-valued variables with some (not necessary known) continuous multivariate distribution. On the other hand, a contingency table already collects results from n independent observations. Its cells contain counts as realization of discrete multinomial distribution. Obviously, compositional tables frequently appear in practice. However, up to now a concise methodology for their statistical treatment was not available. Although one possible approach to treat a sample of contingency tables statistically is to consider three-way contingency tables (?), where the third factor would be used to construct the sample of tables. However, this approach does not inherently contain the case of tables with continuous origin of entries (like for the first mentioned example) as well as a possibility of a random sample of tables. Correspondence analysis is another approach to analyse contingency tables (see, e.g., ?, for details) But again this method is not primarily designed for a sample of tables.

As a way out, compositional tables can be considered as a special case of D -part compositional data (??). These are multivariate observations consisting of relative information that express a relationship between two (row and column) factors. Thus they form a continuous analogy to the well-known contingency tables. As in the case of D -part compositions, cells (parts) of compositional tables are all strictly positive and represent their contributions to the whole. Consequently, for the analysis of compositional tables the magnitudes of their parts are not taken into account and the whole information is contained in ratios between the parts. This information will not change if all parts are multiplied by some positive constant, and thus compositions could be represented in proportions or in percentages (exactly as in the case of distributional data). For example, the total population in the European countries varies, but this is not relevant if we are interested only in the relative information on social vs. political structure of the population. The proportional representation is useful for comparison purposes of compositional data (compositional tables). It should not alter the results of any meaningful statistical analysis. However, this does not hold when standard methods are applied for statistical analysis of compositional tables. The reason is that contrary to standard multivariate observations (and the corresponding statistical methods) that rely on the Euclidean geometry in real space, compositional tables are characterized by their specific geometric nature, represented by the Aitchison geometry on the simplex (???)

As in the case of contingency tables, the aim of the analysis of compositional tables usually is to study relations between factors. Unfortunately, standard tests of independence, like the well-known Pearson χ^2 test as well as other popular confirmatory and exploration tools (?), are not suitable. They follow the general assumption that the table of independent classifications for the observed table is built up by multiplying the standard marginals of the observed table (arithmetic marginals). This assumption can be inter-

preted in the sense that there is an optimal (minimum) distance between the observed table and the product of arithmetic marginals. Nevertheless, there is no geometrical background that would enable us to proceed with further reasoning in this direction using metrical concepts as well as orthogonal projections and linear subspaces that would assist for the purpose (?).

Taking into account the relative character and the specific geometry of compositional tables (together with replacing the arithmetic marginals by the geometric ones), this analysis can be performed advantageously through a decomposition of the original table into its independent and interactive parts in the above described optimal sense (??). In particular, the interaction table conveys the key information for understanding the sources of association between both factors. In order to apply standard statistical methods like principal component analysis or discriminant analysis to the sample of decomposed compositional tables and extract possible sources of association between the row and column factors, it is convenient to express them in orthonormal coordinates with respect to the Aitchison geometry. As there is no canonical standard basis of the simplex, the crucial point (and the main goal of the paper) is to introduce an interpretable set of coordinates that could be used for representation of the interaction tables. The new coordinates can be interpreted in terms of odds ratios, forming an inherent part of statistical analysis of contingency tables.

The paper is organized as follows: Section 2 presents the general properties of D -part compositional data, which are applied to the special case of $I \times J$ compositional tables. The next two sections contain methodology for compositional tables analysis, based on the Aitchison geometry of compositional data. This section includes the decomposition of compositional tables into independent and interactive parts and representation of the interaction tables in interpretable coordinates. Finally, theoretical results are applied to a real-world problem in Section 5. Another example with an economic

motivation is included as supporting information.

2 Compositional tables and the Aitchison geometry

Since $I \times J$ compositional tables represent a special case of D -part compositional data, the general concepts can be easily adapted for compositional tables and used to derive the corresponding specific issues. A (random) D -part composition is defined as a row vector $\mathbf{x} = (x_1, x_2, \dots, x_D)$, where $x_i > 0$ for all i , and each part describes quantitatively its relative contribution to the whole (?). Thus the absolute values of the parts are not important, since all the relevant information in the composition is contained in the ratios between parts. Consequently, the composition could be rescaled (closed) to a prescribed constant sum representation $\kappa > 0$ (i.e. to 1 in case of proportions and 100 for percentages); formally, we refer to a closure operation and denote

$$\mathcal{C}(\mathbf{x}) = \left(\frac{\kappa \cdot x_1}{\sum_{i=1}^D x_i}, \frac{\kappa \cdot x_2}{\sum_{i=1}^D x_i}, \dots, \frac{\kappa \cdot x_D}{\sum_{i=1}^D x_i} \right).$$

The sample space of representations of D -part compositional data to a prescribed constant sum constraint κ is the simplex, a $(D-1)$ -dimensional subset of \mathbf{R}^D , defined as

$$S^D = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_D) \mid x_i > 0, i = 1, 2, \dots, D; \sum_{i=1}^D x_i = \kappa \right\}.$$

The definition of compositional data induces four main conditions that should be fulfilled to ensure their meaningful (not just statistical) analysis (?). The first of them is the *scale invariance*, which means that the results of the analysis should not depend on the particular sum κ of compositional parts. Scale invariance is also related to the property of *relative scale* of

compositions, since ratios should express the differences between observations rather than absolute distances. Another condition is *subcompositional coherence*. As in standard statistics the results obtained from a composition with D parts should not be in contradiction with results that are obtained from a subcomposition containing d parts, $d < D$. The final basic principle of compositional analysis is *permutation invariance*, which means that the analysis output should not depend on the order of parts in the composition. For the purpose of compositional tables this principle could be restated as follows: The result of the analysis should not depend on the order of rows and columns in the table.

Specifically, if a IJ -part composition

$$\mathbf{x} = \mathcal{C}(x_{11}, \dots, x_{1J}, \dots, x_{I1}, \dots, x_{IJ}) \quad (1)$$

carries primarily information about a relation between two factors, then it is convenient to reorder it into a form of a $I \times J$ compositional table

$$\mathbf{x} = \mathcal{C} \begin{pmatrix} x_{11} & \cdots & x_{1J} \\ \vdots & \ddots & \vdots \\ x_{I1} & \cdots & x_{IJ} \end{pmatrix},$$

which represents relationship between these two factors with I and J values, respectively, displayed now in rows and columns. Contrary to contingency tables, a compositional table could be considered as a result of a single observation in a sample of tables with multivariate continuous distribution, and its analysis thus requires a different approach. Analogously for a subcomposition of D -part compositional data, a subtable (or partial table) of $I \times J$ compositional table can be defined as resulting from omitting the whole row(s) and/or column(s) of the original table. Similarly as in the general case of compositional data, $I \times J$ compositional tables can be assumed to follow the Aitchison geometry on the simplex that takes the above requirements

into account. The sample space of compositional tables is a subset of IJ part simplex, but formed only by compositions of type (??). The algebraic-geometrical structure of the Aitchison geometry is formed by operations of perturbation and power transformation. According to ?, these operations result for $I \times J$ compositional tables \mathbf{x} and \mathbf{y} and a real number α in compositional tables

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C} \begin{pmatrix} x_{11}y_{11} & \cdots & x_{1J}y_{1J} \\ \vdots & \ddots & \vdots \\ x_{I1}y_{I1} & \cdots & x_{IJ}y_{IJ} \end{pmatrix}, \quad \alpha \odot \mathbf{x} = \mathcal{C} \begin{pmatrix} x_{11}^\alpha & \cdots & x_{1J}^\alpha \\ \vdots & \ddots & \vdots \\ x_{I1}^\alpha & \cdots & x_{IJ}^\alpha \end{pmatrix},$$

respectively. Here $\mathbf{n} = \mathcal{C} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$ denotes the neutral element in the $(D - 1)$ -dimensional vector space $(\mathcal{S}^D, \oplus, \odot)$. The Aitchison inner product of two compositional tables \mathbf{x} and \mathbf{y} is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \frac{1}{2IJ} \sum_{i,j} \sum_{k,l} \ln \frac{x_{ij}}{x_{kl}} \ln \frac{y_{ij}}{y_{kl}}. \quad (2)$$

Then, from the Euclidean vector space properties of the Aitchison geometry,

$$\|\mathbf{x}\|_A = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_A} \quad \text{and} \quad d_A(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_A$$

$(\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus [(-1) \odot \mathbf{y}])$ represent the Aitchison norm of a table \mathbf{x} and the distance between two compositional tables \mathbf{x} and \mathbf{y} , respectively.

Standard statistical methods that rely on the Euclidean geometry in real space (?, Chapter 2) should not be applied to raw compositional data (compositional tables). It is convenient to assign the orthonormal coordinates to a composition $\mathbf{x} = \mathcal{C}(x_1, \dots, x_D) \in \mathcal{S}^D$ through an isometric (ilr) logratio transformation (??). This results in a $(D - 1)$ -dimensional real vector $\mathbf{z} = h(\mathbf{x}) = (\langle \mathbf{x}, \mathbf{e}_1 \rangle_A, \dots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_A) = (z_1, z_2, \dots, z_{D-1})$, where

$\mathbf{e}_i = \mathcal{C}(e_{i1}, \dots, e_{iD})$, $i = 1, \dots, D - 1$, form an orthonormal basis on the simplex. The real vector \mathbf{z} represents orthonormal coordinates in the real space \mathbf{R}^{D-1} . For compositions $\mathbf{x}_1, \mathbf{x}_2, \in S^D$ and $\alpha, \beta \in \mathbf{R}$ the following properties hold,

$$h(\alpha \odot \mathbf{x}_1 \oplus \beta \odot \mathbf{x}_2) = \alpha \cdot \mathbf{z}_1 + \beta \cdot \mathbf{z}_2, \quad \langle \mathbf{x}_1, \mathbf{x}_2 \rangle_A = \langle \mathbf{z}_1, \mathbf{z}_2 \rangle \quad . \quad (3)$$

Thus the ilr transformation forms an isometric isomorphism from \mathcal{S}^D to \mathbf{R}^{D-1} . Obviously, orthonormal (ilr) coordinates \mathbf{z} depend on a particularly chosen orthonormal basis of \mathcal{S}^D . A sequential binary partition (SBP) (?) represents one possibility to construct interpretable orthonormal coordinates. The procedure of SBP includes $D - 1$ steps as follows. In the first step of the partition all compositional parts are divided into two groups and in the following steps, each group is splitted again into two new subgroups, one with u parts marked by $+$, and the other with v parts marked by $-$. This process continues until all groups of parts consist of only one single part. The result of each step is a basis vector \mathbf{e}_i with elements $e_{ij} = \exp\left(\frac{1}{u}\sqrt{uv/u+v}\right)$, if j -th unit of \mathbf{x} was marked as $+$, $e_{ij} = \exp\left(-\frac{1}{v}\sqrt{uv/u+v}\right)$, if j -th unit was marked as $-$, and $e_{ij} = 1$, if the j -th part was not involved in the partition in the i -th step, $i = 1, \dots, D - 1$, $j = 1, \dots, D$. The basis vectors are used for construction of the resulting coordinates

$$z_i = \langle \mathbf{x}, \mathbf{e}_i \rangle_A = \sqrt{\frac{uv}{u+v}} \ln \frac{(x_{j_1} x_{j_2} \dots x_{j_u})^{1/u}}{(x_{k_1} x_{k_2} \dots x_{k_v})^{1/v}} \quad , \quad (4)$$

where $\{j_1, j_2, \dots, j_u\}$ and $\{k_1, k_2, \dots, k_v\}$ are indices of parts of \mathbf{x} which were splitted in the i -th step of the partition into the first or second subgroup, respectively. The orthonormal coordinates, obtained using SBP, are also called balances, and the entire partition is usually recorded in a table. Note that different ilr coordinate systems are linked through an orthogonal transformation (?).

Balances can be easily interpreted in the general case of D -part compositional data (??). This is also true for 2×2 compositional tables, see ? for details. Compositional tables require another approach to provided interpretation. Here it would be preferable to have coordinates that follow odds ratio interpretation of contingency tables (?, p.44). Nevertheless, to construct such coordinates a deeper insight into the geometrical properties of compositional tables is necessary and will be provided in the next section.

3 Geometrical properties of $I \times J$ compositional tables and their decomposition

Since the analysis of compositional tables is based on projections of the table onto subspaces with specific interpretation (?), such projections shall be introduced before we proceed to construction of any reasonable orthonormal coordinates.

Various projections are used for different purposes in the case of compositional tables. At first, projections of a compositional table \mathbf{x} onto row subspaces $\mathcal{S}^{IJ}(\text{row}_i)$, for $i = 1, \dots, I$, each with dimension $J - 1$, are considered. These projections are denoted by $\text{row}_i(\mathbf{x})$ and according to ? they result in

$$\text{row}_i(\mathbf{x}) = \mathcal{C} \begin{pmatrix} g(\text{row}_i[\mathbf{x}]) & g(\text{row}_i[\mathbf{x}]) & \cdots & g(\text{row}_i[\mathbf{x}]) \\ \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{iJ} \\ \cdots & \cdots & \cdots & \cdots \\ g(\text{row}_i[\mathbf{x}]) & g(\text{row}_i[\mathbf{x}]) & \cdots & g(\text{row}_i[\mathbf{x}]) \end{pmatrix},$$

where $g(\text{row}_i[\mathbf{x}])$ denotes the geometric mean of elements in the i -th row of \mathbf{x} . The projection onto the subspace, formed by the i -th row of the compositional table \mathbf{x} , $\text{row}_i[\mathbf{x}] = \mathcal{C}(x_{i1}, \dots, x_{iJ}) \in \mathcal{S}^J, i = 1, \dots, I$, is thus a $I \times J$ table

$\text{row}_i(\mathbf{x})$ whose entries consist of the i -th row itself and the rest elements are equal to geometric mean of $\text{row}_i[\mathbf{x}]$.

Analogously, also projections of the compositional table \mathbf{x} onto its columns, $\text{col}_j[\mathbf{x}] = \mathcal{C}(x_{1j}, \dots, x_{Ij}) \in \mathcal{S}^I, j = 1, \dots, J$, forming subspaces $\mathcal{S}^{IJ}(\text{col}_j)$ with dimension $I - 1$, can be constructed. Similarly to the case of projections onto rows, the resulting projected compositional tables $\text{col}_j(\mathbf{x})$ are given by the j -th column of \mathbf{x} and its geometric mean in the other parts of the table.

Orthogonality between $\text{row}_i(\mathbf{x})$ and $\text{row}_{i'}(\mathbf{x}), i \neq i'$, or between $\text{col}_j(\mathbf{x})$ and $\text{col}_{j'}(\mathbf{x}), j \neq j'$, can be proven directly using the Aitchison inner product or the isometric properties of the clr transformation (??) (?).

The projection onto the subspace of the i -th row results in a compositional table $\text{row}_i(\mathbf{x})$ that explains the relative information (ratios) exclusively for this row. In order to complete the information about the original compositional table \mathbf{x} , it is necessary to introduce a projection that explains the remaining ratios between parts in different rows (?). In other words, a projection onto the subspace of dimension $I - 1$ that forms the orthogonal complement to row subspaces $\mathcal{S}^{IJ}(\text{row}_i), i = 1, \dots, I$, needs to be constructed. This subspace will be denoted as $\mathcal{S}^{IJ}(\text{row}^\perp)$ and projection onto this subspace as row^\perp . Analogously as before (??) we get a compositional table

$$\text{row}^\perp(\mathbf{x}) = \mathcal{C} \begin{pmatrix} g(\text{row}_1[\mathbf{x}]) & g(\text{row}_1[\mathbf{x}]) & \dots & g(\text{row}_1[\mathbf{x}]) \\ g(\text{row}_2[\mathbf{x}]) & g(\text{row}_2[\mathbf{x}]) & \dots & g(\text{row}_2[\mathbf{x}]) \\ \dots & \dots & \dots & \dots \\ g(\text{row}_I[\mathbf{x}]) & g(\text{row}_I[\mathbf{x}]) & \dots & g(\text{row}_I[\mathbf{x}]) \end{pmatrix}, \quad (5)$$

formed by row geometric means of the original table. Similarly, projection of \mathbf{x} onto subspace orthogonal to column subspaces, $\mathcal{S}^{IJ}(\text{col}^\perp)$, of dimension $J - 1$ that carries information about ratios between different columns of the

original compositional table, results in

$$\text{col}^\perp(\mathbf{x}) = \mathcal{C} \begin{pmatrix} g(\text{col}_1[\mathbf{x}]) & g(\text{col}_2[\mathbf{x}]) & \dots & g(\text{col}_J[\mathbf{x}]) \\ g(\text{col}_1[\mathbf{x}]) & g(\text{col}_2[\mathbf{x}]) & \dots & g(\text{col}_J[\mathbf{x}]) \\ \dots & \dots & \dots & \dots \\ g(\text{col}_1[\mathbf{x}]) & g(\text{col}_2[\mathbf{x}]) & \dots & g(\text{col}_J[\mathbf{x}]) \end{pmatrix}. \quad (6)$$

From their construction, projections $\text{row}^\perp(\mathbf{x})$ and $\text{col}^\perp(\mathbf{x})$ are orthogonal to all row or column projections, respectively, and even to each other (see ? for proof). This fact is crucial for compositional tables analysis as will be shown later.

Orthogonality of all row/column subspaces allows to reconstruct the original compositional table \mathbf{x} using decompositions

$$\mathbf{x} = \text{row}^\perp(\mathbf{x}) \oplus \left(\bigoplus_{i=1}^I \text{row}_i(\mathbf{x}) \right) = \text{col}^\perp(\mathbf{x}) \oplus \left(\bigoplus_{j=1}^J \text{col}_j(\mathbf{x}) \right).$$

As mentioned above, projections $\text{row}^\perp(\mathbf{x})$ and $\text{col}^\perp(\mathbf{x})$ carry information exclusively about ratios between parts of different rows and columns, respectively. This information is sufficient for the reconstruction of the compositional table, when row and column factors are independent (motivated by the probabilistic sense of the formulation). This corresponds to the case when the original table can be expressed as a product of row and column (geometric) marginals of \mathbf{x} (??), similarly as for contingency tables (?). The resulting $I \times J$ compositional table $\mathbf{x}_{ind} = \text{row}^\perp(\mathbf{x}) \oplus \text{col}^\perp(\mathbf{x})$, obtained as a perturbation of these two projections, is called *independence table* with related parts

$$x_{ij}^{ind} = \left(\prod_{k=1}^I \prod_{l=1}^J x_{kl} \right)^{\frac{1}{IJ}}. \quad (7)$$

x_{ij} denote parts of the original compositional table \mathbf{x} . Since the dimensions of subspaces $\mathcal{S}^{IJ}(\text{row}^\perp)$ and $\mathcal{S}^{IJ}(\text{col}^\perp)$ are $I - 1$ and $J - 1$, respectively, dimension of the subspace of independence tables \mathcal{S}_{ind}^{IJ} equals $I + J - 2$.

The remaining information about the original table, i.e. about the relations between row and column factors, is contained in the *interaction table* \mathbf{x}_{int} , which is orthogonal to \mathbf{x}_{ind} and results in the decomposition

$$\mathbf{x} = \mathbf{x}_{ind} \oplus \mathbf{x}_{int} \quad . \quad (8)$$

The interaction table can be obtained from (??) as $\mathbf{x}_{int} = \mathbf{x} \ominus \mathbf{x}_{ind}$. It also forms an $I \times J$ compositional table and its parts can be computed from the original table \mathbf{x} by

$$x_{ij}^{int} = \left(\prod_{k=1}^I \prod_{l=1}^J \frac{x_{ij}}{x_{kj}x_{il}} \right)^{\frac{1}{IJ}} . \quad (9)$$

From Equation (??) and orthogonality between \mathbf{x}_{ind} and \mathbf{x}_{int} it follows that the dimension of the subspace of interaction tables, \mathcal{S}_{int}^{IJ} , equals $I \cdot J - 1 - (I + J - 2) = (I - 1)(J - 1)$. In the following section, interpretable orthonormal coordinates for interaction tables will be of particular interest.

4 Compositional tables analysis

As in the case of contingency tables, the goal of compositional tables analysis is primarily to study the relationship between row and column factors and to answer the question whether these factors are independent. The analysis of independence is simplified a lot through the decomposition of the original table into its independent and interaction parts (??). Note that this decomposition has no counterpart with similar geometrical properties in the standard analysis of independence in contingency tables, provided by Pearson's χ^2 -statistic. In case of independence between row and column factors in a compositional table, the whole information about the original table \mathbf{x} is concentrated in the independence table \mathbf{x}_{ind} , while the interaction table \mathbf{x}_{int} just equals the neutral element \mathbf{n} . On the other hand, the importance of the interaction table increases if the relationship between factors becomes

stronger. A deeper investigation of the interaction table can yield information about the sources of association between row and column factors. Thus, interpretable coordinates are useful also for the exploratory analysis of a sample of compositional tables. In order to provide additional statistical inference about independence and to apply standard statistical methods for structural analysis of compositional tables (like principal component analysis or correlation analysis), it is also convenient to express compositional tables in orthonormal coordinates. Although the coordinates of an interaction table seem to be of primary interest, the general task is to express all compositional tables $(\mathbf{x}, \mathbf{x}_{ind}, \mathbf{x}_{int})$ in their ilr coordinates $(\mathbf{z}, \mathbf{z}_{ind}, \mathbf{z}_{int})$ and then proceed with statistical analysis.

As mentioned in Section 2, the interpretation of orthonormal coordinates depends on the methodology used for their construction. Since different systems of orthonormal coordinates are just orthogonal rotations of each other, for the well-known Hotelling test about mean value of a random sample (in case of normality), or discriminant analysis, interpretation is not of primary interest. Nevertheless, this is not the case of correlation analysis or principal component analysis, where interpretable orthonormal coordinates are crucial. For compositional data, sequential binary partition (SBP) seems to be a proper way to obtain coordinates with interpretation in terms of ratios of the original compositional parts (??). However, SBP becomes quite complex for representation of compositional tables in coordinates, in particular for the interaction table. A naive approach would be to derive coordinates for the interaction table using expression (??) first and then proceed to SBP using formula (??). Obviously, this would lead to quite tedious computations and when the size of the compositional table grows, it becomes difficult to find an easily interpretable set of coordinates. Furthermore, using SBP we will obtain usually more nonzero coordinates than the dimension of the subspace of the interaction table $((I - 1)(J - 1))$, which will lead to singularity problems

for most of standard statistical methods.

In a second, preferable, approach, the goal is to compute the set of coordinates of the interaction table directly from the original compositional table \mathbf{x} . For deriving the main formula for coordinates of the interaction table, it is first necessary to construct $(I - 1)$ coordinates of $\mathcal{S}^{IJ}(\text{row}^\perp)$ (to explain ratios between different rows of \mathbf{x}) and $(J - 1)$ coordinates of $\mathcal{S}^{IJ}(\text{col}^\perp)$ (for columns). From SBP it is easy to see that one such choice of balances would lead to the generating vectors $\mathbf{e}_i^{\text{row}}, i = 1, \dots, I - 1$, and $\mathbf{e}_j^{\text{col}}, j = 1, \dots, J - 1$, with parts

$$\begin{aligned} e_{kj}^{\text{row}} &= \exp(0) && \text{for } k = 1, \dots, i - 1, \quad j = 1, \dots, J, \\ e_{kj}^{\text{row}} &= \exp\left(\sqrt{\frac{I-i}{J(I-i+1)}}\right) && \text{for } k = i, \quad j = 1, \dots, J, \\ e_{kj}^{\text{row}} &= \exp\left(-\sqrt{\frac{1}{(I-i+1)(IJ-iJ)}}\right) && \text{for } k = i + 1, \dots, I, \quad j = 1, \dots, J, \end{aligned}$$

and

$$\begin{aligned} e_{il}^{\text{col}} &= \exp(0) && \text{for } i = 1, \dots, I, \quad l = 1, \dots, j - 1, \\ e_{il}^{\text{col}} &= \exp\left(\sqrt{\frac{J-j}{I(J-j+1)}}\right) && \text{for } i = 1, \dots, I, \quad l = j, \\ e_{il}^{\text{col}} &= \exp\left(-\sqrt{\frac{1}{(J-j+1)(IJ-jI)}}\right) && \text{for } i = 1, \dots, I, \quad l = j + 1, \dots, J. \end{aligned}$$

These vectors lead to coordinates

$$z_i^{\text{row}} = \sqrt{\frac{(I-i)J}{I-i+1}} \ln \frac{(x_{i1} \dots x_{iJ})^{1/J}}{(x_{i+1,1} \dots x_{iJ})^{1/(IJ-iJ)}}, \quad i = 1, \dots, I - 1$$

(for rows), and

$$z_j^{\text{col}} = \sqrt{\frac{I(J-j)}{J-j+1}} \ln \frac{(x_{1j} \dots x_{Ij})^{1/I}}{(x_{1,j+1} \dots x_{Ij})^{1/(IJ-Ij)}}, \quad j = 1, \dots, J - 1$$

(for columns), respectively. These $I + J - 2$ orthonormal coordinates express nonzero coordinate representation for the independence table and their number reflects the dimension of $\mathcal{S}_{\text{ind}}^{IJ}$. Because of the mutual orthogonality of the subspaces corresponding to tables $\text{row}^\perp(\mathbf{x})$, $\text{col}^\perp(\mathbf{x})$ and \mathbf{x}_{int} and decomposition (??), the remaining $(I - 1)(J - 1)$ coordinates of \mathbf{x}_{ind} are equal to

zero. Conversely, a coordinate representation of the interaction table results in zero coordinates of the corresponding independence table.

Following Section 3, the next step is to form a new basis, orthonormal to the basis corresponding to the first $I + J - 2$ nonzero coordinates of independence table. Moreover, this basis should determine the non-zero coordinates of the interaction table. Among other options (and after tedious calculations), a stepwise procedure to assess the basis looks most convenient. Accordingly, the method firstly assigns a basis compositional vector to table given only by the parts x_{11}, x_{12}, x_{21} and x_{22} . This basis element compares parts at the main diagonal x_{11}, x_{22} with parts at the minor diagonal x_{12}, x_{21} of the 2×2 partial table and thus the first basis composition has the form

$$\mathbf{e}^{22} = \exp\left(\frac{1}{2}, -\frac{1}{2}, 0, \dots, -\frac{1}{2}, \frac{1}{2}, 0, \dots\right) \quad ,$$

where the upper index expresses the dimension of the current partial table. Obviously, an odds-ratio interpretation of the resulting coordinate is possible. In the next step the third column is added to the previous partial table and the basis vector \mathbf{e}^{23} deals with the new partial table with $r = 2$ rows and $s = 3$ columns and parts $x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}$. The corresponding basis element compares again parts at the main diagonal of a virtual 2×2 table with parts at the minor diagonal, when these diagonals are formed by geometric mean of x_{11} and x_{12} (that thus merges information on the employed components together) and part x_{23} , and by geometric mean of x_{21} and x_{22} , and part x_{13} , respectively. This results in

$$\mathbf{e}^{23} = \exp\left(\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, -\frac{1}{\sqrt{3}}, 0, \dots, -\frac{1}{2\sqrt{3}}, -\frac{1}{2\sqrt{3}}, \frac{1}{\sqrt{3}}, 0, \dots\right).$$

In general, in each step the method extends the partial table by one row or column. The basis composition \mathbf{e}^{rs} compares parts at the main diagonal (formed by geometric mean of all parts at rows of order smaller than r and column of order smaller than s and by part x_{rs}) and parts at the minor

diagonal (formed by geometric mean of first $s - 1$ parts of the r -th row and by geometric mean of first $r - 1$ parts of the s -th column). This resulting basis vector has the elements

$$\begin{aligned}
e_{ij}^{rs} &= \exp\left(\sqrt{\frac{1}{rs(r-1)(s-1)}}\right) & \text{for} & & i = 1, \dots, r-1, & j = 1, \dots, s-1, \\
e_{ij}^{rs} &= \exp\left(-\sqrt{\frac{r-1}{rs(s-1)}}\right) & \text{for} & & i = r, & j = 1, \dots, s-1, \\
e_{ij}^{rs} &= \exp\left(-\sqrt{\frac{s-1}{rs(r-1)}}\right) & \text{for} & & i = 1, \dots, r-1, & j = s, \\
e_{ij}^{rs} &= \exp\left(\sqrt{\frac{(r-1)(s-1)}{rs}}\right) & \text{for} & & i = r, & j = s, \\
e_{ij}^{rs} &= \exp(0), & & & & \text{otherwise,}
\end{aligned}$$

where the upper index represents the particular choice of $r = 2, 3, \dots, I$ and $s = 2, 3, \dots, J$. This procedure continues until $r = I$ and $s = J$ and so we get a system of $(I - 1)(J - 1)$ basis vectors, orthogonal to basis of coordinates z_i^{row} and z_j^{col} , for $i = 1, \dots, I - 1, j = 1, \dots, J - 1$. The basis of $I \times J$ compositional tables is formed by these three systems of basis vectors.

For example, the basis of 2×3 compositional tables contains compositions

$$\begin{aligned}
\mathbf{e}^{22} &= \exp(1/2, -1/2, 0, -1/2, 1/2, 0), \\
\mathbf{e}^{23} &= \exp\left(1/2\sqrt{3}, 1/2\sqrt{3}, -1/\sqrt{3}, -1/2\sqrt{3}, -1/2\sqrt{3}, 1/\sqrt{3}\right), \\
\mathbf{e}_1^{\text{row}} &= \exp\left(1/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}, -1/\sqrt{6}, -1/\sqrt{6}, -1/\sqrt{6}\right), \\
\mathbf{e}_1^{\text{col}} &= \exp\left(1/\sqrt{3}, -1/2\sqrt{3}, -1/2\sqrt{3}, 1/\sqrt{3}, -1/2\sqrt{3}, -1/2\sqrt{3}\right), \\
\mathbf{e}_2^{\text{col}} &= \exp(0, 1/2, -1/2, 0, 1/2, -1/2).
\end{aligned}$$

Basis vectors \mathbf{e}^{rs} lead to the following nonzero coordinates of the interaction table (out of $IJ - 1$)

$$z_{rs}^{\text{int}} = \frac{1}{\sqrt{r \cdot s \cdot (r-1) \cdot (s-1)}} \ln \prod_{i=1}^{r-1} \prod_{j=1}^{s-1} \frac{x_{ij} x_{rs}}{x_{is} x_{rj}} \quad (10)$$

for $r = 2, 3, \dots, I$ and $s = 2, 3, \dots, J$, since coordinates z_i^{row} and z_j^{col} , for $i = 1, \dots, I - 1, j = 1, \dots, J - 1$, applied on \mathbf{x}_{int} are always zero. Although

the above formula is advantageous for interpretation purposes, in practice it is easier to compute coordinates of the interaction table from the following modified formula with expanded products

$$\frac{1}{\sqrt{r \cdot s \cdot (r-1) \cdot (s-1)}} \ln \frac{x_{11}x_{12} \cdots x_{1,s-1} \cdots x_{r-1,1} \cdots x_{r-1,s-1}x_{rs}^{(r-1)(s-1)}}{x_{r1}^{r-1} \cdots x_{r,s-1}^{r-1}x_{1s}^{s-1} \cdots x_{r-1,s}^{s-1}} \quad (11)$$

for $r = 2, 3, \dots, I$ and $s = 2, 3, \dots, J$. Even though x_{ij} 's in both formulas stand for parts of the original table \mathbf{x} , the result would not change if they are replaced by parts of the interaction table \mathbf{x}_{int} .

In fact, this coordinate representation contains also the nonzero coordinates of the interaction tables of all tables with sizes smaller than the considered $I \times J$ table. For example, the set of four nonzero coordinates of 3×3 interaction table contains two nonzero coordinates of the 2×3 table as well as of the 3×2 table, and in turn both (as well as 3×3 table) contain the only nonzero ilr coordinate of the 2×2 interaction table.

The interpretability of new coordinates (??) is supported by their relation to odds ratios of parts in the original table (?, p.44). This fact is obvious directly from the form of (??). The $I \times J$ table contains $\binom{I}{2} \cdot \binom{J}{2}$ odds ratios of type $(x_{ac}x_{bd})/(x_{bc}x_{ad})$, where a and b are indices of rows and c and d are indices of columns. All these odds ratios are contained in the set of coordinates since each coordinate of the interaction table is formed by the sum of logarithms of odds ratios which compare the cell of original table with the position (r, s) with all cells that are north-west from the r -th row and s -th column (this feature will be thoroughly analyzed in the first example). Of course, permutation of rows and/or columns in the original compositional table leads to a new set of coordinates. Table ?? compares features of coordinate representations of the interaction table, computed using both SBP (corresponding to the first approach) and the new presented formula (??). Finally, nonzero coordinates of \mathbf{x}_{ind} and \mathbf{x}_{int} (z_i^{row} , z_j^{col} and z_{rs}^{int}) can be used

for coordinate representation of the original compositional table.

Table 1 about here.

5 Example - relationship between age and BMI index

The above methodology for coordinate representation of compositional tables is illustrated in an example analyzing the relationship between age and BMI index in 18 European countries (??). For this purpose a sample of 3×4 compositional tables was collected. Each of the tables records the population structure of a country in 2008 according to age and BMI index ((weight in kg)/(height in m)²). The two factors to be considered correspond to the age classes 25 – 44, 45 – 64, 65 – 84 and their BMI index in the categories underweight, normal, overweight, and obesity, respectively. Note that finer categories of age are available, but the chosen classes lead to better interpretability. Table ?? shows an example of a compositional table from the sample from Czech Republic.

Table 2 about here.

Applying Equation (??), the values of the independence table are

$$\mathbf{x}_{ind} = \begin{pmatrix} 0.0061 & 0.1716 & 0.2218 & 0.1090 \\ 0.0039 & 0.1090 & 0.1409 & 0.0692 \\ 0.0020 & 0.0569 & 0.0736 & 0.0361 \end{pmatrix} .$$

Using Equation (??) the interaction table can be obtained,

$$\mathbf{x}_{int} = \begin{pmatrix} 0.1813 & 0.0973 & 0.0483 & 0.0387 \\ 0.0444 & 0.0707 & 0.0967 & 0.1085 \\ 0.0541 & 0.0632 & 0.0930 & 0.1037 \end{pmatrix} .$$

Note that these tables follow the condition $\mathbf{x}_{ind} \oplus \mathbf{x}_{int} = \mathbf{x}$. If the factors were independent, the interaction table would equal to the neutral element on the simplex, i.e. all parts would be approximately $1/(IJ) = 1/12 = 0.0833$. In case of the Czech Republic it is easy to see that this condition does not hold as well as in the case of the other countries. This feature is clearly visible also from the mean interaction table (in sense of the Aitchison geometry)

$$\bar{\mathbf{x}}_{int} = \frac{1}{n} \odot \bigoplus_{k=1}^n \mathbf{x}_{int,k} = \begin{pmatrix} 0.1483 & 0.0967 & 0.0589 & 0.0465 \\ 0.0554 & 0.0753 & 0.0917 & 0.1031 \\ 0.0604 & 0.0682 & 0.0922 & 0.1035 \end{pmatrix} .$$

The above findings lead to a preliminary conclusion that age and BMI index are not independent, nevertheless, further verification using statistical analysis in coordinates is necessary in order to search for possible sources of association between these two factors.

In order to express the independence table in coordinates, two SBPs according to Table ?? were introduced.

Table 3 about here.

The steps of SBP1 result in the first two nonzero coordinates of the independence table that contain relative information (ratios) between different rows of \mathbf{x} . The next three coordinates result from SBP2 and separate columns of the original compositional table. For example, the coordinates of the independence table in case of the Czech Republic equal to

$$\mathbf{z}_{ind} = (1.2713, 0.9189, -4.9160, 0.1399, 0.8703, 0, 0, 0, 0, 0).$$

When both SBPs from Table ?? are applied to \mathbf{x}_{int} , the resulting coordinates are equal to zero, as well as coordinates of SBP1 and SBP2 applied to $\text{col}^\perp(\mathbf{x})$ and $\text{row}^\perp(\mathbf{x})$, respectively. Thus, because of decomposition (??), the same coordinates would be obtained if SBPs from Table ?? were applied

directly to the independence table \mathbf{x}_{ind} from (??), or if SBP1 was applied to $\text{row}^\perp(\mathbf{x})$ and SBP2 to $\text{col}^\perp(\mathbf{x})$, respectively. As a consequence of (??) and (??), the coordinates of the independence table also form coordinates of the original table \mathbf{x} . The remaining coordinates of \mathbf{x} equal to $(I - 1)(J - 1) = 6$ nonzero coordinates of the interaction table, and can be expressed using formula (??). In case of the Czech Republic, these coordinates are

$$\mathbf{z}_{int} = (0, 0, 0, 0, 0, 0.5439, 0.8988, 0.8428, 0.1354, 0.4648, 0.4441) \quad ,$$

where the first five zero coordinates refer to SBP1 and SBP2 applied to \mathbf{x}_{int} . The relation of the coordinates of the interaction table to the partial tables and odds ratios within them is illustrated in Figure ??. Their basic descriptive statistics for the given data set are summarized in Table ??.

Figure 1 about here.

Table 4 about here.

The first nonzero coordinate z_{22}^{int} is computed for $r = s = 2$ for all 18 European countries. From Table ?? it could be seen that the sample mean equals to 0.3674, and the standard deviation is 0.1488. This coordinate could be interpreted as a ratio of the chance that people with age between 25 and 44 years will be underweight rather than normal weight and the same chance for people between 45 and 64. From (??) the mean odds ratio $e^{2 \cdot 0.3674} \approx 2$ is obtained. Consequently, the chance that younger people are underweight is about twice as high as for people with age between 45 and 64. Considering also the small standard deviation of this coordinate, it can be concluded that age and BMI index are not independent for this particular table (see also ?, for further reasoning).

The next coordinate z_{23}^{int} corresponds to a table for people aged between 25 – 44 or 45 – 64 being under-, normal, or overweight, respectively. From

the estimates of the sample mean and standard deviation (Table ??) it can be concluded that age and BMI index are again far from independence. This coordinate could be also interpreted as sum of logarithms of two odds ratios, divided by $2\sqrt{3}$. The first odds ratio compares chances that people underweight against overweight for age ranges 25 – 44 and 45 – 64 years, respectively. The second odds ratio compares almost the same with the only difference of taking normal weight instead of underweight. The sum of logarithms of these odds ratios is $2\sqrt{3} \cdot 0.6096 = 2.1117 > 0$ on average. Consequently, at least one of the chances that one is underweight against overweight, or the normal weight against overweight, respectively, is higher for people between 25 and 44 years. This fact contradicts once again the preliminary hypothesis that BMI index and age in this partial table are independent factors. Coordinate z_{24}^{int} , which adds the column for obese people, has almost the same interpretation.

The fourth coordinate z_{32}^{int} corresponds to a partial table with three age ranges (25–44, 45–64 and 65–84) and two weight possibilities (underweight and normal weight). This coordinate is the only one whose true value could be seriously considered zero since its sample mean equals 0.1057 and standard deviation 0.2412. This might point to a conclusion that in this partial table age and BMI are independent. In such a case also the other coordinates of tables with lower sizes would need to have similar features, however, it does not hold for the corresponding coordinate z_{22}^{int} . The interpretation of z_{32}^{int} is analogous to the previous cases.

Since the remaining coordinates of the interaction table could be interpreted analogously as in the previous cases, they are only described using Figure ?? and Table ?. Their values rather suggest a relation between both factors than their independence. Furthermore, to summarize, the first three nonzero coordinates of the interaction table carry information about odds ratios, which compare chances of lower weight ranges to a higher one for age

group between 25 and 44 years and group between 45 and 64 years. The first coordinate compares underweight with normal weight. In the next coordinate, these two groups are both compared with overweight. Finally, the third coordinate compares groups with underweight, normal weight and overweight with the group of obese people. The last three coordinates compare the same chances, but now the first age group contains age ranges 25 – 44 and 45 – 64 together and the second group covers exclusively age range 65 – 84 years. Quite interesting is the absence of negative values in the sample means of all coordinates, lower weight categories are thus typical for younger population.

Figure 2 about here.

In order to extend the univariate conclusions to a multivariate one, the coordinates of the interaction table as well as of the original compositional table and the independence table are also analyzed using the well-known biplot (?) of the first two principal components of the corresponding coordinates. In Figure ?? biplots of the original, independence and interaction tables are collected. The biplot of the original compositional table seems to be dominated by high variability of the coordinates of the independence table, thus here mainly the data structure (with Romania and Slovakia as outlying observations) can be observed. The other two biplots provide further information on the relations leading to independence and interaction between the age and BMI factors.

The biplot of the independence table shows that its first two nonzero coordinates (that carry relative information on relations between the rows of the original table, i.e. age ranges) are strongly positively correlated, and also negatively correlated with the remaining three nonzero coordinates (explaining the relations between the columns representing BMI categories). From the directions of arrows (loadings) we can observe that moving from the left to the right side of the biplot, the values of the first two coordinates decrease

and the next three coordinates increase. Also the locations of countries resulting from the principal component scores provide interesting information about the data structure, like cluster of countries Belgium, Denmark and France with quite high values of coordinates z_1^c, z_2^c and z_3^c . It means that these countries contribute substantially to the independence between age and BMI index, in particular due to the high relative contributions of underweight people over all other age categories. Similarly, we can observe Romania as an outlying observation, with particular importance of the positive ratio between overweight and obese people. Contrary, Poland and Lithuania lay in the centre of the biplot. The centre represents an average behaviour of both factors.

The interaction biplot shows some interesting features as well. In particular, the first three nonzero coordinates are strongly correlated and the last three ones as well, but no correlation between these two groups is visible. This means that odds ratios with the third row of the interaction table (age range 65–84 years) yield results different from those within the younger categories. Also in this biplot, Belgium, Denmark and France are placed quite near to each other and these western European countries together with Switzerland and Austria represent states with lower values of all coordinates, thus with rather smaller BMI (weight) growth for increasing age. On the other hand, countries like Czech Republic and Estonia with high values of coordinates $z_{22}^{int}, z_{23}^{int}$ and z_{24}^{int} indicate a substantial weight growth from the younger to middle age generation, and thus contribute substantially to interaction between the factors. An interesting outlier is represented by Turkey with small values of coordinates $z_{32}^{int}, z_{33}^{int}$ and z_{34}^{int} . This testifies that the weight growth tends to be particularly small from 25–44 and 45–64 to 65–84 age group, just conversely to Slovakia. Nearest to the origin are placed Poland and Lithuania again, i.e. these countries do not contribute neither to interaction nor independence between the age and BMI factors.

Interestingly, small correlation between coordinates z_{22}^{int} , z_{23}^{int} , z_{24}^{int} and z_{32}^{int} , z_{33}^{int} , z_{34}^{int} remains unaltered when rows of the original compositional tables are permuted, see Figure ???. This result indicates an independence behaviour of single row factor values (age groups) with respect to BMI categories.

Figure 3 about here.

6 Conclusions

As far as we know, a concise methodology for statistical analysis of a sample of compositional tables was not available before. This paper takes the very first steps in this direction. Its main aim was to introduce a coordinate representation for decomposition of a two-way compositional table into independence and interaction tables and to show how they can be used for a reasonable statistical analysis. In particular, coordinates for the interaction table were constructed that can be interpreted in sense of odds ratios of elements of the table. This interpretation is similar as for contingency tables, and thus provides a natural generalization to the sample of compositional tables. Moreover, the introduced coordinates can be also considered as a starting point for development of more general representations of two-way compositional tables. Finally, the way how the coordinates were obtained motivates also the possibility of considering some higher dimensional generalization of odds ratios for coordinate representation of k -way compositional tables, $k \geq 2$.

Coordinate representation of both independent and interaction tables also enables their graphical visualization using a biplot of principal component scores and loadings that can be used for further analysis of the data structure and to observe relations between coordinates. This paper opens also further questions, e.g., concerning interpretation of coordinates or the corresponding

statistical inference. The coordinate representation of compositional tables seems to provide counterparts to standard problems in contingency tables, as represented by the test of symmetry or Stuart’s test. We hope that they will be introduced in a near future.

Supporting information. Additional information for this article is available online including an analysis of technology intensity and resource efficiency in manufacturing.

References

- AGRESTI, A. (2002). *Categorical data analysis*, 2nd ed. J. Wiley & Sons, New York.
- AITCHISON, J. (1986). *The statistical analysis of compositional data*. Chapman and Hall, London.
- BILLARD, L. & DIDAY, E. (2007). *Symbolic data analysis: Conceptual statistics and data mining*. Wiley, Chichester.
- EATON, M. L. (1983). *Multivariate statistics. A vector space approach*. John Wiley & Sons, New York.
- EGOZCUE, J. J. (2009). Reply to “On the Harker Variation Diagrams” by J.A. Cortés. *Mathematical Geosciences* **41**, 829–834.
- EGOZCUE, J. J., DÍAZ-BARRERO, J. L. & PAWLOWSKY-GLAHN, V. (2008). Compositional analysis of bivariate discrete probabilities. In *Proceedings of CODAWORK’08, The 3rd Compositional Data Analysis Workshop*. (eds J. Daunis-i-Estadella & J. A. Martín-Fernández) University of Girona, Spain.

- EGOZCUE, J. J. & PAWLOWSKY-GLAHN, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology* **37**, 795–828.
- EGOZCUE, J. J., PAWLOWSKY-GLAHN, V., MATEU-FIGUERAS, G. & BARCELÓ-VIDAL, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology* **35**, 279–300.
- EGOZCUE, J. J., PAWLOWSKY-GLAHN, V., TEMPL, M. & HRON, K. (2015). Independence in contingency tables using simplicial geometry. *Communications in Statistics - Theory and Methods*, 44:18, 3978–3996.
- EUROSTAT (2013). Body mass index (BMI) by sex, age and educational level - collection round 2008. (Available from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_ehis_de1&lang=en. Accessed September 12, 2013.)
- EUROSTAT (2013). Population on 1 January: Structure indicators. (Available from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjanind&lang=en. Accessed September 12, 2013.)
- FAČEVIČOVÁ, K., HRON, K., TODOROV, V., GUO, D. & TEMPL, M. (2014). Logratio approach to statistical analysis of 2×2 compositional tables. *Journal of Applied Statistics* **41**, 944–958.
- FIŠEROVÁ, E. & HRON, K. (2011). On interpretation of orthonormal coordinates for compositional data. *Mathematical Geosciences* **43**, 455–468.
- GABRIEL, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **58**, 453–467.
- GREENACRE, M. (2007). *Correspondence analysis in practice*, 2nd edition. Chapman and Hall/CRC Press, London (UK).

MATEU-FIGUERAS, G. & PAWLOWSKY-GLAHN, V. (2008). A critical approach to probability laws in geochemistry. *Mathematical Geosciences* **40**, 489–502.

PAWLOWSKY-GLAHN, V. & EGOZCUE, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment (SEERRA)* **15**, 384–398.

PAWLOWSKY-GLAHN, V., EGOZCUE, J.J. & TOLOSANA-DELGADO, R. (2015). *Modeling and analysis of compositional data*. Wiley, Chichester.

Address Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University, 17. listopadu 12, CZ-77146 Olomouc, Czech Republic.

E-mail kamila.facevicova@gmail.com

Disclaimer The views expressed herein are those of the authors and do not necessarily reflect the views of the United Nations Industrial Development Organization.

Acknowledgements The paper was supported by the grant COST Action CRoNoS IC1408 and the grant IGA_PrF_2015_013 Mathematical Models of the Internal Grant Agency of the Palacký University in Olomouc.

Table 1: Comparison of approaches for coordinate representation of \mathbf{x}_{int} .

SBP to \mathbf{x}_{int}	Formula (??)
+ Interpretation in terms of balances.	+ Related with odds ratios in \mathbf{x} .
– Procedure with several steps.	+ Quick computation directly from \mathbf{x} .
– More nonzero coordinates than dimension of $\mathcal{S}^{I \times J}(\mathbf{x}_{int})$.	+ Number of nonzero coordinates equal to dimension of $\mathcal{S}^{I \times J}(\mathbf{x}_{int})$.
	+ Full-rank observations.

Table 2: Structure of population in the Czech Republic in 2008 according to age and BMI index (in proportions).

CZE	under	normal	over	obesity
25 – 44	0.0144	0.2196	0.1410	0.0554
45 – 64	0.0022	0.1014	0.1792	0.0988
65 – 84	0.0014	0.0473	0.0900	0.0493

Table 3: Sequential binary partitions used for expression of independence tables in coordinates

SBP1	x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{24}	x_{31}	x_{32}	x_{33}	x_{34}	u	v
Step 1	+	+	+	+	–	–	–	–	–	–	–	–	4	8
Step 2					+	+	+	+	–	–	–	–	4	4
SBP2	x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{24}	x_{31}	x_{32}	x_{33}	x_{34}	r	s
Step 1	+	–	–	–	+	–	–	–	+	–	–	–	3	9
Step 2		+	–	–		+	–	–		+	–	–	3	6
Step 3			+	–			+	–			+	–	3	3

Table 4: Sample means and standard deviations (according to the analyzed 18 European countries) of nonzero coordinates of the interaction table.

\mathbf{z}_{int}	z_{22}^{int}	z_{23}^{int}	z_{24}^{int}	z_{32}^{int}	z_{33}^{int}	z_{34}^{int}
Sample mean	0.3674	0.6096	0.6494	0.1057	0.3624	0.3783
Sample st. dev.	0.1488	0.1357	0.1426	0.2412	0.2175	0.1945

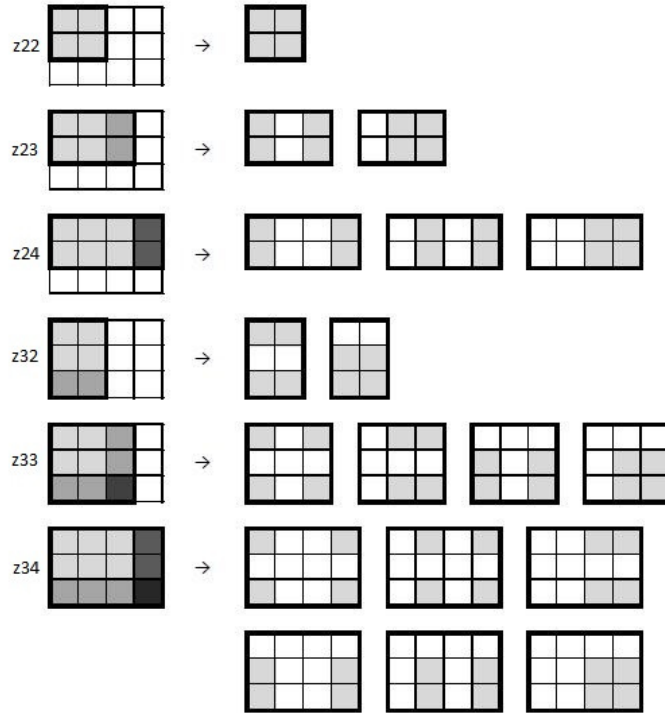


Figure 1: Relation of coordinates of the interaction table to the partial tables of \mathbf{x} and odds ratios within them. In the first column the shades of grey denote the parts of the compositional table \mathbf{x} used for computation of each coordinate; darker shade grades represent higher power of the corresponding parts in formula (??). The right part of the figure represents the odds ratios contained in each coordinate. This gives a visual interpretation of Equation (??) in case of 3×4 tables of age groups and BMI index in Example 1. Specifically, the second coordinate z_{23}^{int} is computed only from parts $x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}$ and this coordinate could be interpreted as $1/2\sqrt{3}$ of logarithm of two multiplied odds ratios, $(x_{11}x_{23})/(x_{13}x_{21})$ and $(x_{12}x_{23})/(x_{13}x_{22})$.

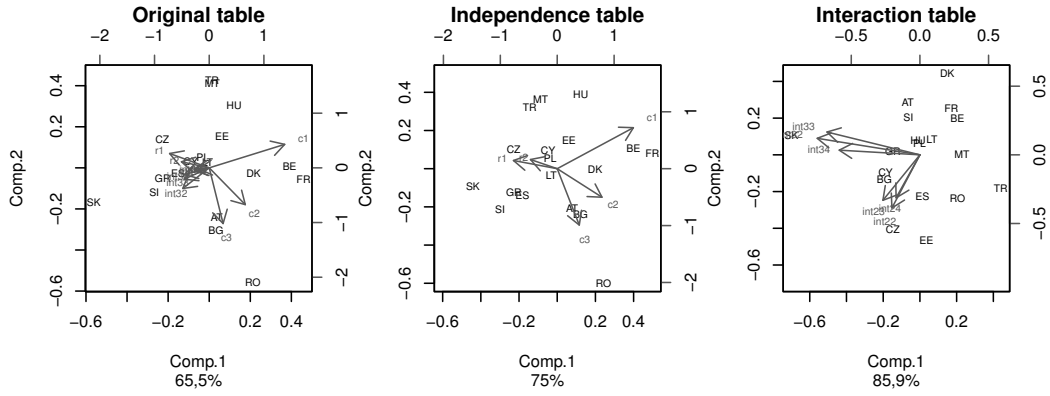


Figure 2: Biplots of coordinates of original, independence, and interaction tables.

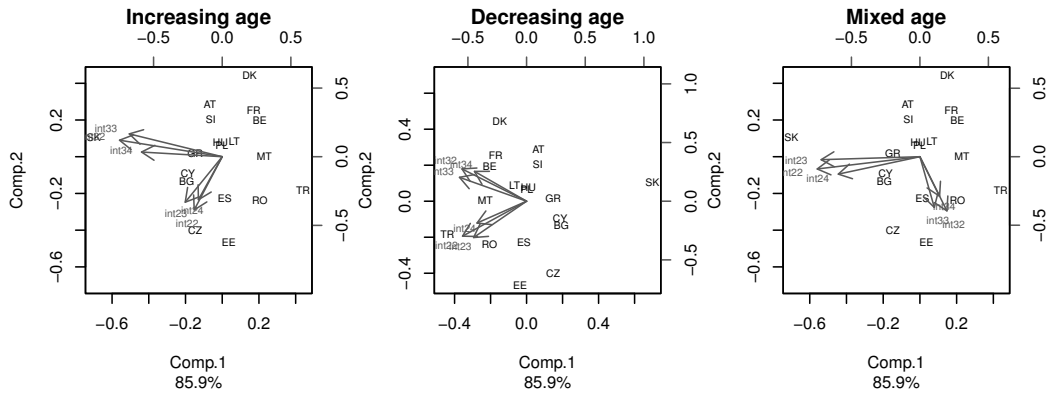


Figure 3: Biplots of coordinates of the interaction table with rows in increasing (25 – 44, 45 – 64 and 65 – 84), decreasing (65 – 84, 45 – 64 and 25 – 44) and mixed (25 – 44, 65 – 84 and 45 – 64) order of age categories.