Austrian Journal of Statistics
September 2016, Volume 45, 45–56.
http://www.ajs.or.at/
doi:10.17713/ajs.v45i4.136

Bayesian Estimation of the Orthogonal Decomposition of a Contingency Table

Maria Isabel Ortego
Universitat Politècnica de Catalunya,
Spain

Juan José Egozcue
Universitat Politècnica de Catalunya,
Spain

Abstract

In a multinomial sampling, contingency tables can be parametrized by probabilities of each cell. These probabilities constitute the joint probability function of two or more discrete random variables. These probability tables have been previously studied from a compositional point of view. The compositional analysis of probability tables ensures coherence when analysing sub-tables. The main results are: (1) given a probability table, the closest independent probability table is the product of their geometric marginals; (2) the probability table can be orthogonally decomposed into an independent table and an interaction table; (3) the departure of independence can be measured using simplicial deviance, which is the Aitchison square norm of the interaction table.

In previous works, the analysis has been performed from a frequentist point of view. This contribution is aimed at providing a Bayesian assessment of the decomposition. The resulting model is a log-linear one, which parameters are the centered log-ratio transformations of the geometric marginals and the interaction table. Using a Dirichlet prior distribution of multinomial probabilities, the posterior distribution of multinomial probabilities is again a Dirichlet distribution. Simulation of this posterior allows to study the distribution of marginal and interaction parameters, checking the independence of the observed contingency table and cell interactions.

The results corresponding to a two-way contingency table example are presented.

Keywords: interaction, independence, simplicial deviance, multinomial sampling, Aitchison geometry of the simplex, orthogonal decomposition, R.

1. Introduction

Contingency tables have been studied for a long time. There are many examples, dating from the beginning of the XX-th century which afforded elementary, but relevant, questions about such kind of data (e.g. Yule 1912). Along the XX-th century many advances have been achieved. The introduction of log-linear models (Nelder 1974) and generalized linear models (McCullagh and Nelder 1983; Nelder and Wedderburn 1972) were important milestones in the study of contingency tables. From the seventies up to now many extensions, improvement of methods and generalisations have been presented, for instance, see Everitt (1977), Darroch, Lauritzen, and Speed (1980), Chambers and Welsh (1993), or Goodman (1996). However, challenges are still pendent for a straightforward solution, specially for the study of n-way contingency tables.

From the compositional point of view, the contingency tables have been studied only recently, with the early precedent of Kenett (1983). In the workshop CoDaWork 2008 (Girona, Spain) Egozcue, Díaz-Barrero, and Pawlowsky-Glahn (2008) introduced a perturbation-decomposition model for tables of multinomial parameters, thus opening new possibilities of analysis. This contribution was followed by other compositional attempts and applications (Gallo 2015; Fačevicová and Hron 2013). The approach proposed in Egozcue, Pawlowsky-Glahn, Templ, and Hron (2015) is a kind of log-linear model but it has some differences with the standard ones. The main differences are the way in which marginals are found and the definition of interactions.

Here, the model based on the orthogonal decomposition of multinomial contingency tables is used to carry out a Bayesian estimation of both close independent multinomial parameters and the subsequent interactions. The present goal is to show that orthogonal decomposition of multinomial contingency tables can be addressed using Bayesian estimation techniques. The zero problem, typical in compositional data analysis, is here overcome by estimating the model parameters (probabilities) underlying the contingency table, which are considered compositional parameters. Zeros in the observations do no produce any problem as their likelihood is well defined. This is a traditional way of dealing with zeroes in generalized linear models as multinomial logistic regression models (Nelder 1974) or Bayesian estimation of multinomial probabilities (e.g. Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015).

In Section 2, the main features of the model based on orthogonal decomposition of contingency tables are recalled. Some definitions of the Bayesian framework are introduced in Section 3. Examples are presented in Section 4.

2. Orthogonal decomposition model

Two-way contingency tables coming from a multinomial sampling are considered. They are generated by a row-classification into I classes, and a column-classification made of J classes. The total number, N, of classified individuals is then distributed on the $I \times J$ cells of the contingency table (CT) according to the classification. The number of individuals pertaining to the ij-cell is denoted n_{ij} for $i = 1, 2, \ldots, I$, and $j = 1, 2, \ldots, J$. The whole contingency table containing these counts is denoted $\mathbf{N}$. The table $\mathbf{N}$, as an array of counting random variables, is assumed to be multinomial distributed and its corresponding probability parameters denoted by p_{ij} $i = 1, 2, \ldots, I$, and $j = 1, 2, \ldots, J$. When arranged in a table, these probability parameters are called probability table (PT). The sample space of a CT, like $\mathbf{N}$, is $I \times J$ times the non-negative integers restricted to add to N. They are not conceived as compositional data, even when the frequencies $\mathbf{N}/N$ are computed. In fact, they can contain zero-counts and only can correspond to fractions with N as denominator. Alternatively, the probability parameters $\mathbf{P}$ are considered compositional. This can be summarized as (a) $\mathbf{P}$ is in $\mathcal{S}^D$, $D = I \cdot J$; (b) perturbation and powering, denoted $\oplus$, $\odot$ respectively, are vector space operations, and the dimension of $\mathcal{S}^D$ is $D - 1$; (c) the centered log-ratio (clr) transformation is defined and inner product, norm and distances in $\mathcal{S}^D$ are the ordinary Euclidean inner product, norm and distance of the clr transformed PT's. As well-known for $\mathcal{S}^D$ (Pawlowsky-Glahn and Egozcue 2001), the simplex endowed with $\oplus$, $\odot$, and the Aitchison inner product is a $D-1$-dimensional Euclidean space. More explicitly, consider two PT's, $\mathbf{P}$ and $\mathbf{Q}$ and a real number α. The perturbation $\mathbf{W} = \mathbf{P} \oplus \mathbf{Q}$ is a PT with entries $w_{ij} = p_{ij}q_{ij} / \sum_{km} p_{km}q_{km}$, $k = 1, \ldots, I$ and $m = 1, \ldots, J$. The α-powering $\mathbf{W} = \alpha \odot \mathbf{P}$ is a PT with entries $w_{ij} = p_{ij}^{\alpha} / \sum_{km} p_{km}^{\alpha}$. The clr of a PT is an $(I \times J)$-array, $\mathbf{V} = \mathrm{clr}(\mathbf{P})$ which entries are

$$v_{ij} = \log(p_{ij}) - \frac{1}{D} \sum_{k=1}^{I} \sum_{m=1}^{J} \log p_{km} \ .$$

The inverse clr-transformation is $\mathbf{P} = \mathcal{C} \exp(\mathbf{V})$, where exp operates componentwise and $\mathcal{C}$ is the closure operator. For any vector of D strictly positive real components,

$$\mathbf{z} = (z_1, z_2, \ldots, z_D) \in \mathbb{R}_+^D, \quad z_i > 0 \quad \text{for all } i = 1, 2, \ldots, D \ ,$$

the closure of $\mathbf{z}$ to $\kappa > 0$ is defined as

$$\mathcal{C}(\mathbf{z}) = \left[\frac{\kappa \cdot z_1}{\sum_{i=1}^{D} z_i}, \frac{\kappa \cdot z_2}{\sum_{i=1}^{D} z_i}, \ldots, \frac{\kappa \cdot z_D}{\sum_{i=1}^{D} z_i} \right] \ .$$

Denoting $\mathrm{clr}(\mathbf{P}) = \mathbf{V}$ and $\mathrm{clr}(\mathbf{Q}) = \mathbf{W}$, the Aitchison inner product, $\langle \cdot, \cdot \rangle_a$, and distance, $\mathrm{d}_a(\cdot, \cdot)$, of PT's is

$$\langle \mathbf{P}, \mathbf{Q} \rangle_a = \langle \mathbf{V}, \mathbf{W} \rangle = \sum_{i=1}^{I} \sum_{j=1}^{J} v_{ij} w_{ij} \quad , \quad \mathrm{d}_a^2(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^{I} \sum_{j=1}^{J} (v_{ij} - w_{ij})^2 \ ,$$

where $\langle \cdot, \cdot \rangle$ denotes the ordinary Euclidean inner product of arrays.

In these definitions, commonly used in compositional data analysis, there are, at least, two key points. The first one is the interpretability of the perturbation. Perturbation of PT's correspond to apply the Bayes formula to a PT, containing prior probabilities, using a likelihood arranged as a PT, up to the closure operation. The second point is that the subcompositional coherence (Pawlowsky-Glahn *et al.* 2015; Egozcue 2009), is guaranteed. In the case of PT's, subcompositional coherence assures that distance between two sub-tables have Aitchison distance smaller than or equal to the distance between the parent PT's.

The main result in Egozcue *et al.* (2008, 2015) is that, independent PTs constitute an $(I-1)(J-1)$-dimensional linear subspace of $\mathcal{S}^D$. This means that any PT can be projected orthogonally on this subspace. The consequence is that $\mathbf{P}$ is decomposed in a unique way as

$$\mathbf{P} = \mathbf{P}_{ind} \oplus \mathbf{P}_{int} \quad , \quad \mathbf{P}_{ind} \perp \mathbf{P}_{int} \ , \tag{1}$$

where $\mathbf{P}_{ind}$ is the projection of $\mathbf{P}$ on the independent subspace, and $\mathbf{P}_{int}$ is in the orthogonal complement. The PT $\mathbf{P}_{int}$ is called interaction PT. The independent PT is on its turn decomposed into two new PT's, called marginal PT's, which have equal rows and equal columns respectively. The independent PT is then decomposed as

$$\mathbf{P}_{ind} = (\mathbf{1}_I \mathbf{r}^\top) \oplus (\mathbf{c} \mathbf{1}_J^\top) \ , \tag{2}$$

where $\mathbf{r}$, $\mathbf{c}$ are compositions in $\mathcal{S}^J$ and $\mathcal{S}^I$ respectively, and they are treated as column vectors for matrix notation. The symbols $\mathbf{1}_I$ and $\mathbf{1}_J$ are column-vectors, with I and J components respectively, all of them equal to 1.

Equations 1 and 2 can be transformed by taking clr, which yields

$$\mathrm{clr}(\mathbf{P}) = \mathrm{clr}(\mathbf{P}_{ind}) + \mathrm{clr}(\mathbf{P}_{int}) = \mathbf{1}_I (\mathrm{clr}(\mathbf{r}))^\top + \mathrm{clr}(\mathbf{c}) \mathbf{1}_J^\top + \mathrm{clr}(\mathbf{P}_{int}) \ . \tag{3}$$

It should be remarked that $\mathrm{clr}(\mathbf{r})$ and $\mathrm{clr}(\mathbf{c})$ are clr transformations of compositions in $\mathcal{S}^J$ and $\mathcal{S}^I$ respectively and they are not PT's.

The marginal row and column, $\mathbf{r}$ and $\mathbf{c}$ respectively, are obtained from $\mathbf{P}$ as the closed geometric means by columns and rows of $\mathbf{P}$ respectively. This feature indicates that the nearest independent PT, in the sense of Aitchison geometry in $\mathcal{S}^D$, is not obtained from the traditional (arithmetic) marginals. This is an important difference from common analysis of contingency tables. As a consequence, $\mathrm{clr}(\mathbf{P}_{ind})$ has the property that its arithmetic and geometric marginals are equal up to a closure; and the geometric marginals of $\mathbf{P}_{int}$ are neutral in the simplex (i.e. all their elements are equal).

The decomposition in Equation 3 implicitly defines a log-linear model which is revealed after taking clr^{-1} in Equation 3. The log-linear model is then

$$\mathbf{P} = \mathcal{C} \exp[\mathrm{clr}(\mathbf{P}_{ind}) + \mathrm{clr}(\mathbf{P}_{int})] = \mathcal{C} \exp[\mathbf{1}_I (\mathrm{clr}(\mathbf{r}))^\top + \mathrm{clr}(\mathbf{c}) \mathbf{1}_J^\top + \mathrm{clr}(\mathbf{P}_{int})] \ , \tag{4}$$

where the parameters are the J-coefficients in $\mathrm{clr}(\mathbf{r})$, the I coefficients in $\mathrm{clr}(\mathbf{c})$ and the $D = I \cdot J$ coefficients in $\mathrm{clr}(\mathbf{P}_{int})$. However, coefficients of any clr add to zero, and the number of free parameters is $(J-1)+(I-1)+(IJ-1) = IJ+I+J-3$. The number of clr-parameters in Equation 4 can be reduced to $IJ+I+J-3$ using ilr-coordinates, but this strategy is not used here as the clr-parameters can be interpreted directly.

In order to interpret the results when the log-linear model is fitted to a CT, some derived parameters may be useful. When the norm $\|\mathbf{P}_{int}\|_a$ is null, $\mathbf{P}_{int}$ is the neutral element in $\mathcal{S}^D$ and $\mathbf{P}$ is an independent PT. Therefore, $\|\mathbf{P}_{int}\|_a^2$ is an overall measure of dependence which was named simplicial deviance. When considered relative to the Aitchison square norm of $\mathbf{P}$, it can be called relative simplicial deviance. The corresponding definitions are

$$\Delta^2(\mathbf{P}) = \|\mathbf{P}_{int}\|_a^2 \quad , \quad R^2_\Delta(\mathbf{P}) = \frac{\|\mathbf{P}_{int}\|_a^2}{\|\mathbf{P}_{ind}\|_a^2 + \|\mathbf{P}_{int}\|_a^2} \ , \tag{5}$$

where $\|\mathbf{P}_{ind}\|_a^2 + \|\mathbf{P}_{int}\|_a^2 = \|\mathbf{P}\|_a^2$ due to the orthogonal decomposition (Equation 1). Remarkably, $\Delta^2(\mathbf{P})$ does not depend on the marginals of $\mathbf{P}$; such a property is not shared by $R^2_\Delta(\mathbf{P})$. However $R^2_\Delta(\mathbf{P})$ has clear interpretation based on the facts of $0 \le R^2_\Delta(\mathbf{P}) \le 1$, $R^2_\Delta(\mathbf{P}) = 0$ implies independence of $\mathbf{P}$, whereas $R^2_\Delta(\mathbf{P}) = 1$ indicates that the nearest independent PT to $\mathbf{P}$ is the neutral (uniform) PT, and it can be considered as a pure interaction PT.

In order to interpret the coefficients of $\mathbf{V} = \mathrm{clr}(\mathbf{P}_{int})$ it should be taken into account that the simplicial deviance is decomposed

$$\Delta^2(\mathbf{P}) = \|\mathbf{P}_{int}\|_a^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} v_{ij}^2 \ , \tag{6}$$

so that each cell contributes to the simplicial deviance with v_{ij}^2 thus deserving the name of cell interaction. A way of presenting these cell interactions is computing their relative value to the simplicial deviance or expressing them as percent of contribution. However, the signs of v_{ij} are important as they indicate whether the probability in the cell p_{ij} is smaller than the predicted probability using $\mathbf{P}_{ind}$ (negative v_{ij}) or it is larger than this predicted probability (positive v_{ij}). It has been proposed to use an interaction array reporting in each cell the value $\mathrm{sign}(v_{ij})(v_{ij}^2/\Delta^2(\mathbf{P}))$. Unfortunately, the values of v_{ij} cannot be interpreted separately as they add to zero. The analyst should look for large absolute values in the interaction array coupled by positive-negative interactions. Cells interactions are then interpreted jointly as the sources of interaction are frequently coupled.

3. Bayesian analysis

Assume that an $I \times J$ contingency table $\mathbf{N}$ has been observed as the result of a multinomial sampling. After adopting the log-linear model (Equation 4), the multinomial probabilities p_{ij} can be expressed as functions of the clr's of the geometric marginals $\mathrm{clr}(\mathbf{r}) = \mathbf{z}^{(r)} = (z_1^{(r)}, z_2^{(r)}, \ldots, z_J^{(r)})$, $\mathrm{clr}(\mathbf{c}) = \mathbf{z}^{(c)} = (z_1^{(c)}, z_2^{(c)}, \ldots, z_I^{(c)})$, and the entries of $\mathbf{V} = \mathrm{clr}(\mathbf{P}_{int})$ denoted v_{ij}. Hence, the likelihood of these parameters, given the observation has the form

$$L(\mathbf{z}^{(r)}, \mathbf{z}^{(c)}, \mathbf{V} \ |\mathbf{N}) = K \cdot \prod_{i=1}^{I}\prod_{j=1}^{J} p_{ij}^{n_{ij}} \ ,$$

where all p_{ij} are functions of $\mathbf{z}^{(r)}, \mathbf{z}^{(c)}, \mathbf{V}$ and K the normalizing constant corresponding to the multinomial density. In order to simplify the estimation procedure, a Dirichlet distribution (e.g. Aitchison 1986) can be chosen as initial joint distribution of the p_{ij}. If the chosen parameters of the Dirichlet distribution are $a_{ij} > 0$, the final or posterior distribution of the parameters is again a Dirichlet distribution with parameters $p_{ij} + a_{ij}$ and, therefore, the posterior distribution is

$$f(\mathbf{z}^{(r)}, \mathbf{z}^{(c)}, \mathbf{V} \mid \mathbf{N}) = \frac{\Gamma\left(\sum_k \sum_m a_{km}\right)}{\prod_k \prod_m \Gamma(a_{km})} \prod_{i=1}^{I} \prod_{j=1}^{J} p_{ij}^{n_{ij}+a_{ij}-1} \quad , \quad \sum_k \sum_m p_{ij} = 1 \ , \tag{7}$$

The goals of the Bayesian procedure are, at least, three: (a) estimation of posterior distribution of parameters $\mathbf{z}^{(r)}$, $\mathbf{z}^{(c)}$, $\mathbf{V}$ and their marginal distributions; (b) checking the hypothesis of independence of the observed CT; (c) study the distribution of the cell interactions v_{ij} and checking whether they can be considered null or not. These three tasks are hardly carried out using the explicit distribution (Equation 7). A way out consists of drawing independent realisations from Equation 7, and then, studying the simulated sample of parameters thus accomplishing goal (a).

Checking independence of the observed CT is performed through a predictive p-value (Bayarri and Berger 2000; Meng 1994) as proposed in goal (a). Assume that for each possible set of posterior parameters, $\mathbf{z}_0^{(r)}$, $\mathbf{z}_0^{(c)}$, $\mathbf{V}_0$, a likelihood ratio test is carried out on the hypothesis

$$H_0: \ \mathbf{z}^{(r)} = \mathbf{z}_0^{(r)} \ , \ \mathbf{z}^{(c)} = \mathbf{z}_0^{(c)} \ , \ \mathbf{V} = \mathbf{0} \ , \tag{8}$$

using the statistic

$$\Lambda = -2 \log \left(\frac{L(\mathbf{z}_0^{(r)}, \mathbf{z}_0^{(c)}, \mathbf{V} = \mathbf{0} \mid \mathbf{N})}{L(\hat{\mathbf{z}}^{(r)}, \hat{\mathbf{z}}^{(c)}, \hat{\mathbf{V}} \mid \mathbf{N})} \right) \ , \tag{9}$$

where $\hat{\mathbf{z}}^{(r)}$, $\hat{\mathbf{z}}^{(c)}$, $\hat{\mathbf{V}}$ denote the maximum likelihood estimators based on the sample CT. Asymptotically with N, the statistic Λ has distribution χ^2 with degrees of freedom $IJ + I + J - 3$. This corresponds to the number of estimated parameters, compared with no free parameter in H_0. For each set of values $\mathbf{z}_0^{(r)}$, $\mathbf{z}_0^{(c)}$, $\mathbf{V}_0$, one p-value α_{p0} is obtained. The p-value α_{p0}, as a function of the observed CT, has uniform distribution under asymptotic conditions (Robins, van der Vaart, and Ventura 2000). A predictive p-value, α, with asymptotic uniform distribution, is obtained using

$$\alpha = \Phi\left(\frac{1}{m} \sum_{k=1}^{m} \Phi^{-1}\left(\alpha_{p0}^{(k)}\right) \right) \ , \tag{10}$$

where the sum goes through the set of p-values corresponding to the m-simulated sample of parameters $\mathbf{z}_0^{(r)}$, $\mathbf{z}_0^{(c)}$; and Φ denotes the standard normal distribution function (Ortego 2015). Small values of α suggest rejection of the independence H_0.

The assessment of the hypothesis that a single cell interaction v_{ij} is null is performed using Bayesian discrepancy p-values (Gelman, Meng, and Stern 1996), that is, computing the posterior probability of $v_{ij} \leq 0$ across the simulated sample. When this p-value is small (near to zero), or large (near to 1), rejection of v_{ij} is suggested. This accomplishes goal (c).

4. A simple example

4.1. Marks in a subject

The marks obtained by $N = 104$ students in a exam of a college-level statistics subject are considered. Theoretical and practical (mostly problems) questions in the exams are marked separately. In this context, we want to know if the performance in theoretical questions can be considered independent from the performance in practical questions.

The results of the exam may be classified into four groups: A, B, C, D, corresponding to the numeric interval of Spanish marks over 10 points. The results corresponding to this group of students have been organized in a two-way table T (Table 1). We assume that these marks

Table 1: Two-way contingency table containing the marks of the May examination of 104 students. Mark of the theoretical part of the exam (rows) vs. mark of the practical part (columns). The equivalence between marks A, B, C, D and traditional Spanish scores is indicated in the first column.

mark (theory)\ mark (prob)	A [8.5,10]	B [7,8.5)	C [5,7)	D [0,5)
A [8.5,10]	1	0	4	4
B [7,8.5)	2	4	6	13
C [5,7)	0	3	11	25
D [0,5)	1	1	5	24

have been observed as a result of a multinomial sampling with probabilities p_{ij}. A Bayesian framewok is chosen for the estimation of the table parameters p_{ij}. For simplicity, a joint Dirichlet distribution has been assumed for these probabilities.

A Dirichlet prior has been set for the multinomial probabilities. Then, the posterior distribution of these parameters corresponds again to a Dirichlet distribution (Equation 7). A large sample of the posterior distribution has been drawn. This sample is used to describe the uncertainty of parameter estimates and other quantities of interest derived from them. For this data set, a sample of the posterior of length 10,000 has been obtained (e.g. Table 2).

Table 2: Example of a sample PT drawn from the posterior Dirichlet distribution

t\p	A	B	C	D
A	0.01	0.00	0.06	0.08
B	0.01	0.03	0.04	0.17
C	0.00	0.04	0.10	0.24
D	0.02	0.02	0.02	0.16

The tables sampled PT's from the posterior Dirichlet distribution should be properly treated, due to their compositional character. The clr coordinates of the cells for each table have been computed (e.g. Table 3). The row and column geometric marginals of the clr coordinates have also been obtained for each of the tables of the posterior sample. Also, each of these tables has been decomposed in its independent (e.g. Table 4) and interaction table (e.g. Table 5) following Equation 1. That is, a sample of independent and interaction tables has been obtained from the sample of posterior tables. This allows to describe the uncertainty of quantities of interest derived from them, such as deviance, $\Delta^2(\mathbf{P})$, relative deviance, $R^2_\Delta(\mathbf{P})$, among others.

Table 3: Example of clr-coordinates of a sample PT drawn from the posterior Dirichlet distribution. Row and column geometric marginals.

t\p	A	B	C	D	rmarg
A	-1.36	-5.39	1.02	1.30	-1.105
B	-0.57	0.35	0.68	2.06	0.631
C	-4.19	0.52	1.48	2.38	0.048
D	-0.26	0.05	-0.08	2.00	0.426
cmarg	-1.594	-1.117	0.776	1.936	

The departure from independence for the two sets of marks of interest may be measured observing the simplicial deviance (squared Aitchison norm) of the interaction component of drawn posterior tables (Equation 5). Figure 1 shows the histogram of simplicial deviances corresponding to the obtained sample of interaction tables. As the deviance is a measure of dependence, $0 \leq \Delta^2(\mathbf{P}) < +\infty$, a visual comparison with the zero value (red line) is included. For the marks in the example, although the median value (blue line) is low, the amount of variability in the deviance values points to lack of independence between the theoretical and

Table 4: Example of clr-coordinates of the independent component of a sample PT drawn from the posterior Dirichlet distribution

t\p	A	B	C	D
A	-2.70	-2.22	-0.33	0.83
B	-0.96	-0.49	1.41	2.57
C	-1.55	-1.07	0.82	1.98
D	-1.17	-0.69	1.20	2.36

Table 5: Example of clr-coordinates of the interaction component of a sample PT drawn from the posterior Dirichlet distribution

t\p	A	B	C	D
A	1.34	-3.17	1.35	0.47
B	0.39	0.83	-0.72	-0.50
C	-2.64	1.59	0.65	0.40
D	0.90	0.74	-1.28	-0.36

practical marks.

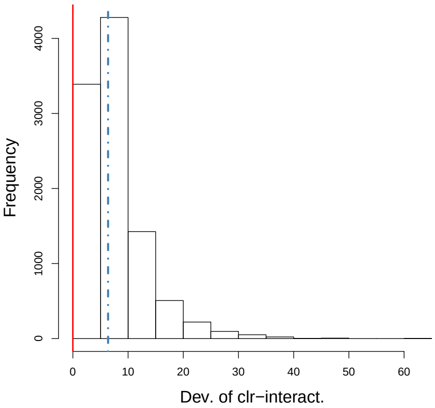

Figure 1: Histogram of posterior simplicial deviance (square norm of the clr-interaction) for final marks. Red line (solid): null value. Blue line (dashed): median.

The relative simplicial deviance $R^2_\triangle(\mathbf{P})$ may seem easier to interpret than the deviance, as $0 \leq R^2_\triangle(\mathbf{P}) \leq 1$, but this interpretation should be taken with caution as this parameter is not marginal invariant. Figure 2 shows the histogram corresponding to the relative simplicial deviance of the posterior sample. A zero-line is also included for a visual comparison. In this case, the majority of the relative deviance values are around 0.3, being near to its median value, reassuring the interpretation of lack of independence.

The simplicial deviance is an overall measure of dependence, but often more detail is needed. The cell values of the interaction table (e.g. Table 6) provide this detail, but the direct interpretation of the values may be confusing due to its compositional character. In order to obtain a detailed description of interaction using the appropriate scale, the clr-coordinates of interaction PT in the posterior sample have been computed. Figure 3 shows the histograms of the cell interactions as a summary of the obtained results. For visual comparison, a zero line has been added to each histogram. Visually, a zero line near the median of the histogram indicates no interaction added by that cell (e.g. histogram corresponding to cell 4). If the zero line is *far* from the center of the histogram (e.g. histogram corresponding to cell 1), that cell may be adding interaction to the deviance (Equation 6), and should be studied.

However, for an easier understanding of the importance of each cell, the interaction array of

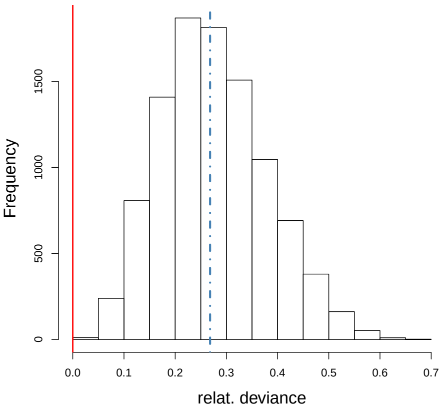

Figure 2: Histogram of relative simplicial deviance (square norm of clr-interac / total) corresponding to the posterior sample. Final marks. Red line (solid): null value. Blue line (dashed): median.

the cells has also been computed (e.g. Table 6), measuring the percentage of interaction added to the deviance by each cell, and including the sign of this interaction. The histogram of the signed interaction array of the posterior sample is shown in Figure 4. Visually, cells with interaction arrays clearly different from zero should be studied, as they are the influential ones. It seems that the most influential cells for the departure of independence are cells number 1, namely 'A in theory' vs 'A in practical' marks and number 3, 'A in theory' vs 'C in practical' marks, with more or less the same weight and opposite signs (see Figure 4, first row). The positive sign of the interaction array for cell number 1 means that the predicted probability for the cell is larger than the predicted by the independent table, while the predicted probabilities for cell 3 (negative sign) are lower than the probabilities predicted by the independent table. Other cells, as cell 5, are also influential, but with a lower weight. The hypothesis of null interaction has also been assessed by means of a Bayesian p-value based on a discrepancy (posterior probability of $v_{ij} \leq 0$ across the sample) (Table 7). If the zero value is central in the sample, i.e. the proportion of $v_{ij} \leq 0$ is near 0.5, the hypothesis is not rejected. Otherwise, small or large proportions, lead to the rejection of the null interaction hypothesis. For instance, for cell number 3, the Bayesian p-value is 0.956, and therefore the null hypothesis is clearly rejected. For cell number 1, the p-value is 0.101 and, although the value is low, the decision of rejection of null cell interaction is not so straightforward.

Table 6: Example of interaction array from the sample

t\p	A	B	C	D
A	6.27	0.54	-24.25	2.84
B	-34.89	2.42	8.79	1.92
C	6.37	-1.82	1.49	-5.74
D	0.77	-0.89	0.55	-0.46

Table 7: Assessment of null interaction hypothesis. Bayesian p-value based on discrepancy (posterior probability of $v_{ij} \leq 0$ across the sample)

t\p	A	B	C	D
A	0.101	0.297	0.956	0.455
B	0.773	0.131	0.223	0.697
C	0.261	0.873	0.175	0.746
D	0.880	0.914	0.115	0.052

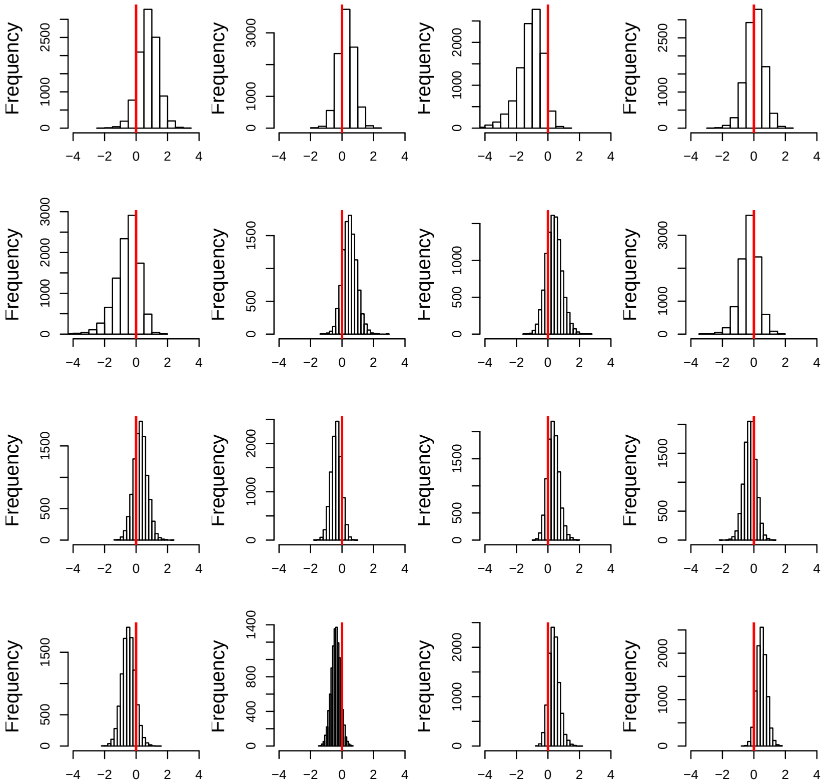

Figure 3: Histograms of clr-cell interactions for the posterior sample. Red lines (solid): null interaction

4.2. An independence test

Simplicial deviance, relative deviance and the interaction array are useful quantities to study dependence in a contingency table. However, it is usual to discuss independence in contingency tables by means of a test (e.g. Equation 8). In our example, are the marks for theory and practice in the exam independent? In the established theoretical context, that can be rephrased as, does the contingency table of marks, T, belong to the subspace of independent tables?

$$H_0 : T = \mathbf{P}_{ind} \in \mathcal{S}^D_{ind} \quad ; \quad H_1 : T = \mathbf{P} \notin \mathcal{S}^D_{ind}$$

The selected likelihood ratio test statistic (Equation 9), is based on the sample of estimates of the independent component $\mathbf{P}_{ind}$, $\widehat{\mathbf{P}}_{ind}$. For each table of the sample of the posterior distribution, its decomposition into independent and interaction component has been obtained in section 4.1. The proposed test statistic and its corresponding predictive p-value have been computed for each of these decompositions. This sample of p-values can be used to measure the uncertainty of the decision of the independence test. Figure 5 shows the histogram of these predictive p-values for the posterior sample of tables. It can be observed that there is variability in the sample of p-values, with a majority of small values, leading to the rejection of the independence hypothesis. However, the lack of uniformity of p-values and their relative scale are problematic for their interpretation (Robins *et al.* 2000). Therefore, the predictive p-values of the sample have been suitably transformed and combined, in order to obtain a summary p-value, α, with asymptotic uniform distribution. In this case, α is nearly 0, and the independence hypothesis has been rejected, as already pointed out by the deviance values. That is, it cannot be considered that the theory and practical marks of this exam are independent.

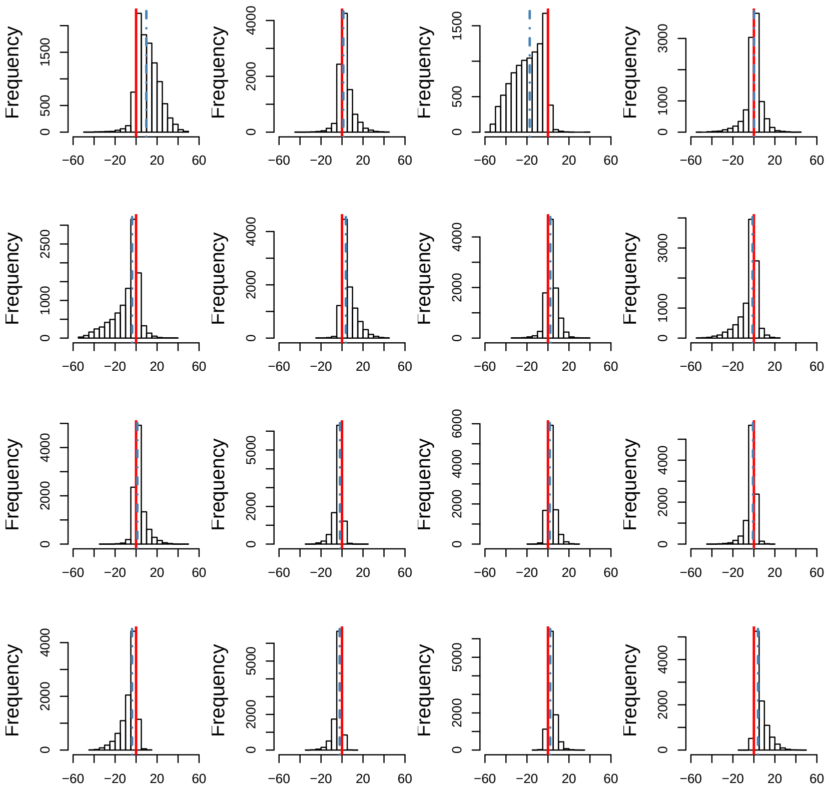

Figure 4: Histogram of the interaction array for each cell. Red line (solid): null value. Blue line (dashed): median

5. Conclusions

Contingency tables have been broadly studied, although only recently they have been treated from the compositional point of view. The orthogonal decomposition of multinomial contingency tables has been presented. Also, a Bayesian framework for the estimation of the parameters of contingency tables has been introduced as a novelty in the compositional treatment of these tables.

A two-way contingency table containing marks from an exam of a college-level statistics course has been studied as an example. Results show that theory and practical exam marks cannot be considered as independent as suggested by the table decomposition and their summary statistics. The Bayesian point of view allows considering uncertainty of estimators and summary statistics. Moreover, the Bayesian approach deals with small or null counts in the original table very efficiently. The multinomial probabilities of the table are assumed compositional. Contrarily, counts in the original contingency table are not reduced to frequencies thus avoiding zero replacements or imputations. This latter fact makes Bayesian estimation very useful in the context of compositional analysis of probability parameters.

Acknowledgements

This research has been partially funded by the Ministerio de Economía y Competividad under project "CODA-RETOS" (Ref. MTM2015-65016-C2-2-R); and by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) of the Generalitat de Catalunya under the project "Compositional and Spatial Analysis" (COSDA) (Ref: 2014SGR551;2014-2016).

References

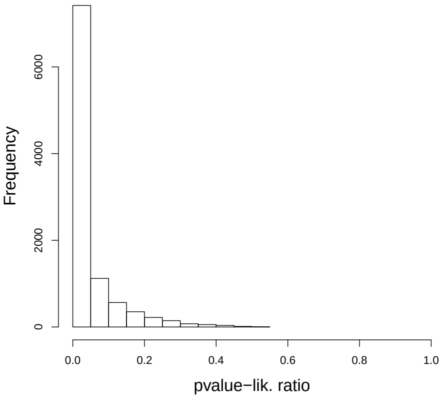

Figure 5: Predictive p-value of the multinomial likelihood ratio independence test. Final marks

Aitchison J (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). ISBN 0-412-28060-4. 416 p.

Bayarri MJ, Berger JO (2000). "P-values for composite null models." *Journal of the American Statistical Association*, **95**, 1127–1142.

Chambers RL, Welsh AH (1993). "Log-linear Models for Survey Data with Non-ignorable Non-response." *J. R. Statist. Soc. B*, **55**(1), 157–170.

Darroch JN, Lauritzen SL, Speed TP (1980). "Markov fields and Log-linear interaction models for contingency tables." *The Annals of Statistics*, **8**(3), 522–539.

Egozcue JJ (2009). "Reply to "On the Harker variation diagrams; ..." by J. A. Cortés." *Mathematical Geosciences*, **41**(7), 829–834.

Egozcue JJ, Díaz-Barrero JL, Pawlowsky-Glahn V (2008). "Compositional analysis of bivariate discrete probabilities." In J Daunis-i Estadella, JA Martín-Fernández (eds.), *Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop*, pp. 1–11. University of Girona, Girona. ISBN 978-84-8458-272-4, URL http://hdl.handle.net/10256/717.

Egozcue JJ, 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. doi:10.1080/03610926.2013.824980.

Everitt BS (1977). *The Analysis of Contingency Tables*. John Wiley & Sons, Inc. New York, New York, USA. ISBN 0-470-71135-3.

Fačevicová K, Hron K (2013). "Statistical analysis of compositional 2 X 2 tables." In K Hron, P Filzmoser, M Templ (eds.), *Proceedings of the 5th International Workshop on Compositional Data Analysis*. TU Wien, Wien. ISBN 978-3-200-03103-6.

Gallo M (2015). "Tucker3 Model for Compositional Data." *Communications in Statistics - Theory and Methods*, **44**(21), 4441–4453. doi:10.1080/03610926.2013.798664.

Gelman A, Meng XL, Stern H (1996). "Posterior predictive assessment of model fitness via realized discrepancies (with discussion)." *Statistica Sinica*, **6**, 733–807.

Goodman LA (1996). "A Single General Method for the Analysis of Cross-Classified Data: Reconciliation and Synthesis of Some Methods of Pearson, Yule, and Fisher, and Also Some Methods of Correspondence Analysis and Association Analysis." *Journal of the American Statistical Association*, **91**(433), 408–428.

Kenett RS (1983). “On an Exploratory Analysis of Contingency Tables.” *The Statistician*, **32**(3), 395–403.

McCullagh P, Nelder JA (1983). *Generalized Linear Models*. Chapman and Hall, London, UK. 522 p.

Meng XL (1994). “Posterior predictive p-values.” *Annals of Statistics*, **22**, 1142–1160.

Nelder JA (1974). “Log linear models for contingency tables: a generalization of classical least squares.” *Appl. Statist.*, **23**, 323–329.

Nelder JA, Wedderburn RWM (1972). “Generalized linear models.” *Journal of the Royal Statistical Society, series A*, **135**, 370–384.

Ortego MI (2015). *Estimación Bayesiana de cópulas extremales en procesos de Poisson*. Ph.D. thesis, Universitat Politècnica de Catalunya.

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

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015). *Modeling and analysis of compositional data*. Statistics in practice. John Wiley & Sons, Chichester UK. ISBN 9781118443064. 272 pp.

Robins JM, van der Vaart A, Ventura V (2000). “Asymptotic Distribution of p-Values in Composite Null Models.” *Journal of the American Statistical Association*, **95**(452), 1143–1156.

Yule GU (1912). “On the Methods of Measuring Association Between Two Attributes.” *Journal of the Royal Statistical Society*, **75**, 579–642.

Affiliation:

Maria Isabel Ortego
Civil and Environmental Engineering Department.
Campus Nord UPC. Edifici C2
Universitat Politècnica de Catalunya
08034 Barcelona, Spain
E-mail: ma.isabel.ortego@upc.edu

Juan José Egozcue
Civil and Environmental Engineering Department.
Campus Nord UPC. Edifici C2
E-mail: juan.jose.egozcue@upc.edu

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 45 *Submitted:* 2015-10-15
September 2016 *Accepted:* 2016-04-20