

# Residual as Linear Sum of Matrix Determinants in Multiway Contingency Tables

Shusaku Tsumoto and Shoji Hirano

*Department of Medical Informatics, Faculty of Medicine,  
Shimane University  
89-1 Enya-cho Izumo 693-8501 Japan  
E-mail: {tsumoto,hirano}@med.shimane-u.ac.jp*

## Abstract

A Pearson residual is defined as a residual between an observed value and expected one of each cell in a contingency table, which measures the degree of statistical dependence of two attribute-value pairs corresponding to the cell. This paper shows that this residual is decomposed into a linear sum of determinants of  $2 \times 2$  subtables, which means that the geometrical nature of the residuals can be viewed from grassmannian algebra.

*Keywords:* Pearson Residual, Determinants, Multiway Contingency Table, Information Granules

## 1. Introduction

Statistical independence between two attributes is a very important concept in data mining and statistics. The definition  $P(A, B) = P(A)P(B)$  show that the joint probability of  $A$  and  $B$  is the product of both probabilities. This gives several useful formula, such as  $P(A|B) = P(A)$ ,  $P(B|A) = P(B)$ . In a data mining context, these formulae show that these two attributes may not be correlated with each other. Thus, when  $A$  or  $B$  is a classification target, the other attribute may not play an important role in its classification.

Although independence is a very important concept, it has not been fully and formally investigated as a relation between two attributes.

In this paper, a statistical independence in a contingency table is focused on from the viewpoint of granular computing and linear algebra, which is continuation of studies on contingency matrix

theory<sup>2,4,6,5</sup>. Tsumoto<sup>2,5</sup> discusses that a contingency table compares two attributes with respect to information granularity and shows that statistical independence in a contingency table is a special form of linear dependence of two attributes. Especially, when the table is viewed as a matrix, the above discussion shows that the rank of the matrix is equal to 1.0. Tsumoto<sup>4</sup> matrix algebra is a key point of analysis of this table. A contingency table can be viewed as a matrix and several operations and ideas of matrix theory are introduced into the analysis of the contingency table. On the other hand, in another paper<sup>6</sup>, Tsumoto shows that the Pearson residual of a contingency matrix is represented as a linear sum of  $2 \times 2$  submatrices, whose number is equal to the degree of freedom.

In this paper, we extend residual analysis to multidimensional cases. The results show that multidimensional residuals are also represented as a linear

sum of determinants of  $2 \times 2$  submatrices, the number of which is equal to the degree of freedom in a given contingency table. Furthermore, the sum includes several kinds of units for statistical independence/dependence, such as total independence and partial independence. This also suggests that  $2 \times 2$  submatrices in a multidimensional data cube can be viewed as information granules for measuring statistical dependence.

The paper is organized as follows: Section 2 discusses the characteristics of contingency tables. Section 3 shows the conditions on statistical independence for contingency tables. Section 4 gives the results of residual analysis of multiway contingency tables. Section 5 illustrates the nature of residuals of multiway tables by using several samples. Section 6 shows that the number of the determinants of  $2 \times 2$  submatrices is exactly the same as the degree of freedom in a three-way contingency table. Finally, Section 7 concludes this paper.

## 2. Multiway Contingency Table

**Definition 1.** Let  $R_1, R_2, \dots, R_n$  denote  $n (\in N)$  multinominal attributes in an attribute space  $A$  which have  $m_1, m_2, \dots, m_n$  values. Let  $|R_j = A_{j_i}|$  denote the set of data whose  $j$ th-attribute is equal to  $A_{j_i}$  ( $i$ th-partition of  $j$ ). A contingency table  $T(R_1, R_2, \dots, R_n)$  is a table, each of whose cells can be defined as:

$$x_{i_1 i_2 \dots i_n} = \#\{x \in |R_1 = A_{i_1}| \wedge |R_2 = A_{i_2}| \dots \wedge |R_n = A_{i_n}|\},$$

with these marginal sums.  $\square$

For example, in the two dimensional case, this table is arranged into the form shown in Table 1, where:  $|[R_1 = A_j]_A| = \sum_{i=1}^m x_{i,j}$ ,  $|[R_2 = B_i]_A| = \sum_{j=1}^n x_{j,i}$ ,  $|[R_1 = A_j \wedge R_2 = B_i]_A| = x_{i,j}$ ,  $|U| = N = x..$  ( $i = 1, 2, 3, \dots, n$  and  $j = 1, 2, 3, \dots, m$ ).

**Definition 2.** A multiway contingency matrix  $M_{R_1, R_2, \dots, R_n}(N)$  is defined as: which is composed of

$$x_{i_1 i_2 \dots i_n} = \#\{x \in |R_1 = A_{i_1}| \wedge |R_2 = A_{i_2}| \dots \wedge |R_n = A_{i_n}|\},$$

where their marginal sums are not included as elements.  $\square$

For simplicity, if we do not need to specify  $R_1$  and  $R_2$ , we use  $M(m, n, N)$  as a contingency matrix with  $m$  rows,  $n$  columns and  $N$  samples.

One of the important observations from granular computing is that a contingency table shows the counting relations between two attributes with respect to intersection of their supporting sets. When two attributes have different numbers of equivalence classes, the situation may be a little complicated. But, in this case, due to knowledge about linear algebra, we only have to consider an attribute which has a smaller number of equivalence classes. and the surplus number of equivalence classes of the attributes with larger number of equivalence classes can be projected into the other partitions. In other words, a  $m \times n$  matrix or contingency table includes a projection from one attributes to the other one.

## 3. Statistical Independence in Multi-way Contingency Tables

Let us consider a  $m \times n$  contingency table shown in Table 1. Statistical independence of  $R_1$  and  $R_2$  gives the following formulae:

$$P([R_1 = A_i, R_2 = B_j]) = P([R_1 = A_i])P([R_2 = B_j]) \\ (i = 1, \dots, m, j = 1, \dots, n).$$

According to the definition of the table,

$$\frac{x_{ij}}{N} = \frac{\sum_{k=1}^n x_{ik}}{N} \times \frac{\sum_{l=1}^m x_{lj}}{N}. \quad (1)$$

Thus, we have obtained:

$$x_{ij} = \frac{\sum_{k=1}^n x_{ik} \times \sum_{l=1}^m x_{lj}}{N}. \quad (2)$$

Thus, for a fixed  $j$ ,

$$\frac{x_{iaj}}{x_{ibj}} = \frac{\sum_{k=1}^n x_{iak}}{\sum_{k=1}^n x_{ibk}}$$

In the same way, for a fixed  $i$ ,

$$\frac{x_{ija}}{x_{ijb}} = \frac{\sum_{l=1}^m x_{lja}}{\sum_{l=1}^m x_{ljb}}$$

Table 1: Contingency Table ( $m \times n$ )

	$A_1$	$A_2$	$\dots$	$A_n$	Sum
$B_1$	$x_{11}$	$x_{12}$	$\dots$	$x_{1n}$	$x_{1\cdot}$
$B_2$	$x_{21}$	$x_{22}$	$\dots$	$x_{2n}$	$x_{2\cdot}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$B_m$	$x_{m1}$	$x_{m2}$	$\dots$	$x_{mn}$	$x_{m\cdot}$
Sum	$x_{\cdot 1}$	$x_{\cdot 2}$	$\dots$	$x_{\cdot n}$	$x_{\cdot\cdot} = N$

Since this relation will hold for any  $j$ , the following equation is obtained:

$$\frac{x_{i_a 1}}{x_{i_b 1}} = \frac{x_{i_a 2}}{x_{i_b 2}} \dots = \frac{x_{i_a n}}{x_{i_b n}} = \frac{\sum_{k=1}^n x_{i_a k}}{\sum_{k=1}^n x_{i_b k}} \quad (3)$$

Since the right hand side of the above equation will be constant, thus all the ratios are constant. Thus,

**Theorem 1.** *If two attributes in a contingency table shown in Table 1 are statistical independent, the following equations hold:*

$$\frac{x_{i_a 1}}{x_{i_b 1}} = \frac{x_{i_a 2}}{x_{i_b 2}} \dots = \frac{x_{i_a n}}{x_{i_b n}} = const. \quad (4)$$

for all rows:  $i_a$  and  $i_b$  ( $i_a, i_b = 1, 2, \dots, m$ ).

□

### 3.1. Three-way Table

Let “ $\bullet$ ” denote as the sum over the row or column of a contingency matrix. That is ,

$$x_{i\bullet} = \sum_{j=1}^n x_{ij} \quad (5)$$

$$x_{\bullet j} = \sum_{i=1}^m x_{ij}, \quad (6)$$

where (5) and (6) shows marginal column and row sums. Then, it is easy to see that

$$x_{\bullet\bullet} = N,$$

where  $N$  denotes the sample size.

Then, Equation (2) is reformulated as:

$$\frac{x_{ij}}{x_{\bullet\bullet}} = \frac{x_{i\bullet}}{x_{\bullet\bullet}} \times \frac{x_{\bullet j}}{x_{\bullet\bullet}} \quad (7)$$

That is,

$$x_{ij} = \frac{x_{i\bullet} \times x_{\bullet j}}{x_{\bullet\bullet}}$$

Or

$$x_{ij} x_{\bullet\bullet} = x_{i\bullet} x_{\bullet j}$$

Thus, statistical independence can be viewed as the specific relations between assignments of  $i, j$  and “ $\cdot$ ”. By use of the above relation, Equation (4) can be rewritten as:

$$\frac{x_{i_1 j}}{x_{i_2 j}} = \frac{x_{i_1 \bullet}}{x_{i_2 \bullet}},$$

where the right hand side gives the ratio of marginal column sums.

Equation (7) can be extended into multivariate cases. Let us consider a three attribute case.

Statistical independence with three attributes is defined as:

$$\frac{x_{ijk}}{x_{\bullet\bullet\bullet}} = \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \times \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \times \frac{x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}}, \quad (8)$$

Thus,

$$x_{ijk} x_{\bullet\bullet\bullet}^2 = x_{i\bullet\bullet} x_{\bullet j\bullet} x_{\bullet\bullet k}, \quad (9)$$

which corresponds to:

$$P(A = a, B = b, C = c) = P(A = a)P(B = b)P(C = c), \quad (10)$$

where A,B,C correspond to the names of attributes for  $i, j, k$ , respectively.

In the statistical context, statistical independence requires a hierarchical model. That is, statistical independence of three attributes requires that all the

two pairs of three attributes should satisfy the equations of statistical independence. Thus, for Equation (10), the following equations should satisfy:

$$\begin{aligned} P(A = a, B = b) &= P(A = a)P(B = b), \\ P(B = b, C = c) &= P(B = b)P(C = c), \text{ and} \\ P(A = a, C = c) &= P(A = a)P(C = c). \end{aligned}$$

Thus,

$$x_{ij\bullet}x_{\bullet\bullet\bullet} = x_{i\bullet\bullet}x_{\bullet j\bullet} \quad (11)$$

$$x_{i\bullet k}x_{\bullet\bullet\bullet} = x_{i\bullet\bullet}x_{\bullet\bullet k} \quad (12)$$

$$x_{\bullet jk}x_{\bullet\bullet\bullet} = x_{\bullet j\bullet}x_{\bullet\bullet k} \quad (13)$$

From Equation (9) and Equation (11),

$$x_{ijk}x_{\bullet\bullet\bullet} = x_{ij\bullet}x_{\bullet\bullet k},$$

Therefore,

$$\frac{x_{ijk}}{x_{ij\bullet}} = \frac{x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}} \quad (14)$$

In the same way, the following equations are obtained:

$$\frac{x_{ijk}}{x_{i\bullet k}} = \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \quad (15)$$

$$\frac{x_{ijk}}{x_{\bullet jk}} = \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \quad (16)$$

In summary, the following theorem is obtained.

**Theorem 2.** *If a three-way contingency table satisfy statistical independence, then the following three equations should be satisfied:*

$$\begin{aligned} \frac{x_{ijk}}{x_{ij\bullet}} &= \frac{x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}} \\ \frac{x_{ijk}}{x_{i\bullet k}} &= \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \\ \frac{x_{ijk}}{x_{\bullet jk}} &= \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \end{aligned}$$

□

Thus, the equations corresponding to Theorem 1 are obtained as follows.

**Corollary 3.** *If three attributes in a contingency table shown in Table 1 are statistical independent, the*

following equations hold:

$$\begin{aligned} \frac{x_{ijk_a}}{x_{ijk_b}} &= \frac{x_{\bullet\bullet k_a}}{x_{\bullet\bullet k_b}} \\ \frac{x_{i_j a k}}{x_{i_j b k}} &= \frac{x_{\bullet j a \bullet}}{x_{\bullet j b \bullet}} \\ \frac{x_{i_a j k}}{x_{i_b j k}} &= \frac{x_{i_a \bullet \bullet}}{x_{i_b \bullet \bullet}} \end{aligned}$$

for all  $i, j$ , and  $k$ .

□

### 3.2. Multi-way Table

The above discussion can be easily extended into a multi-way contingency table.

**Theorem 4.** *If a  $m$ -way contingency table satisfy statistical independence, then the following equation should be satisfied for any  $k$ -th attribute  $i_k$  and  $j_k$  ( $k = 1, 2, \dots, n$ ) where  $n$  is the number of attributes.*

$$\frac{x_{i_1 i_2 \dots i_k \dots i_n}}{x_{i_1 i_2 \dots j_k \dots i_n}} = \frac{x_{\bullet \dots i_k \dots \bullet}}{x_{\bullet \dots j_k \dots \bullet}}$$

Also, the following equation should be satisfied for any  $i_k$ :

$$\begin{aligned} x_{i_1 i_2 \dots i_n} \times x_{\bullet \dots \bullet}^{n-1} \\ = x_{i_1 \bullet \dots \bullet} x_{\bullet i_2 \dots \bullet} \times \dots \times x_{\bullet \dots i_k \dots \bullet} \times \dots \times x_{\bullet \dots \bullet i_n} \end{aligned}$$

□

Thus, a Pearson residual, a difference between an observed value for each cell in a contingency table and an expected value, is defined as:

$$\begin{aligned} \sigma_{i_1 i_2 \dots i_n} &= x_{i_1 i_2 \dots i_n} \\ &- \frac{x_{i_1 \bullet \dots \bullet} \times \dots \times x_{\bullet \dots i_k \dots \bullet} \times \dots \times x_{\bullet \dots \bullet i_n}}{x_{\bullet \dots \bullet}^{n-1}} \end{aligned} \quad (17)$$

## 4. Information Granule for Contingency Matrices

### 4.1. Residual of Contingency Matrices

Tsumoto and Hirano<sup>3,6</sup> discusses the meaning of pearson residuals from the viewpoint of linear algebra.

From Equation (17), the residual is defined as:

$$\sigma_{ij} = x_{ij} - \frac{x_{i\bullet} \times x_{\bullet j}}{x_{\bullet\bullet}}$$

And simple calculation leads to the following theorem.

**Theorem 5.** *The residual  $\sigma_{ij}$  of two-dimensional contingency matrix is obtained as:*

$$\begin{aligned} \sigma_{ij} &= \frac{1}{x_{\bullet\bullet}} \{x_{ij}x_{\bullet\bullet} - x_{i\bullet} \times x_{\bullet j}\} \\ &= \frac{1}{x_{\bullet\bullet}} \left\{ x_{ij} \sum_{k \neq i} \sum_{l \neq j} x_{kl} - \left( \sum_{l \neq j} x_{il} \right) \left( \sum_{k \neq i} x_{kj} \right) \right\} \\ &= \frac{1}{x_{\bullet\bullet}} \sum_{\substack{k \neq i \\ l \neq j}} (x_{ij}x_{kl} - x_{kj}x_{il}) \\ &= \frac{1}{x_{\bullet\bullet}} \sum_{\substack{k \neq i \\ l \neq j}} \Delta_{k,l}^{i,j} \end{aligned}$$

where  $\Delta_{k,l}^{i,j}$  is the determinant of a  $2 \times 2$  submatrix of the original contingency matrix with selection of  $i$  and  $k$  rows and  $j$  and  $l$  columns. Also, the sum takes over  $k = m$  and  $l = n$ . Equivalently, the above formula can be represented as:

$$\sigma_{ij}x_{\bullet\bullet} = \sum_{\substack{k \neq i \\ l \neq j}} \Delta_{k,l}^{i,j}$$

where the sum takes over  $k = m$  and  $l = n$ .  
□

Especially, for  $m = n = 2$ , the residual is equal to the determinant of the original matrix.

$$\sigma_{ij}x_{\bullet\bullet} = \Delta_{k,l}^{i,j}$$

where  $k \neq i$  and  $l \neq j$ . Thus,  $\sigma_{11}x_{\bullet\bullet} = \Delta_{2,2}^{1,1}$ ,

In the case of  $3 \times 3$  tables, the following formulas

are obtained.

$$\begin{aligned} \sigma_{11} &= \frac{1}{x_{\bullet\bullet}} \left( \Delta_{2,2}^{1,1} + \Delta_{2,3}^{1,1} + \Delta_{3,2}^{1,1} + \Delta_{3,3}^{1,1} \right) \\ \sigma_{12} &= \frac{1}{x_{\bullet\bullet}} \left( \Delta_{2,1}^{1,2} + \Delta_{2,3}^{1,2} + \Delta_{3,1}^{1,2} + \Delta_{3,3}^{1,2} \right) \\ \sigma_{21} &= \frac{1}{x_{\bullet\bullet}} \left( \Delta_{1,2}^{2,1} + \Delta_{1,3}^{2,1} + \Delta_{3,2}^{2,1} + \Delta_{3,3}^{2,1} \right) \\ \sigma_{22} &= \frac{1}{x_{\bullet\bullet}} \left( \Delta_{1,1}^{2,2} + \Delta_{1,3}^{2,2} + \Delta_{3,1}^{2,2} + \Delta_{3,3}^{2,2} \right) \end{aligned}$$

As shown in the above formulae, the number of the subdeterminants in  $\sigma_{ij}$  is equal to  $(3 - 1) * (3 - 1) = 4$ , the degree of freedom of  $3 \times 3$  contingency table.

Thus, a  $2 \times 2$  submatrix in a contingency table can be viewed as a information granule for statistical (in)dependence. Furthermore, the number of granules is equal to the degree of freedom of the table.

Can we generalize these results into statistical independence of multivariate cases? This question is answered in this paper as follows.

#### 4.2. Information Granule for Three-way Contingency Tables

The residual for  $x_{ijk}$  is defined as:

$$\sigma_{ijk} = x_{ijk} - \frac{x_{i\bullet\bullet} \times x_{\bullet j\bullet} \times x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}^2} \quad (18)$$

Here, we define “partial residuals” in which one of three attributes are summarized (marginalized) as follows:

$$\sigma_{\bullet jk} = x_{\bullet jk} - \frac{x_{\bullet j\bullet} \times x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}} \quad (19)$$

$$\sigma_{i\bullet k} = x_{i\bullet k} - \frac{x_{i\bullet\bullet} \times x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}} \quad (20)$$

$$\sigma_{ij\bullet} = x_{ij\bullet} - \frac{x_{i\bullet\bullet} \times x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \quad (21)$$

Then, by using (19), the residual (18) is refor-

mualted as:

$$\begin{aligned}
 \sigma_{ijk} &= x_{ijk} - \frac{x_{i\bullet\bullet} \times x_{\bullet j\bullet} \times x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}^2} \\
 &= x_{ijk} - \frac{x_{i\bullet\bullet} x_{\bullet j\bullet} x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet} x_{\bullet\bullet\bullet}} \\
 &= x_{ijk} - \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} (x_{\bullet jk} - \sigma_{\bullet jk}) \\
 &= \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \sigma_{\bullet jk} + x_{ijk} - \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} x_{\bullet jk} \\
 &= \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \sigma_{\bullet jk} \\
 &\quad + \frac{1}{x_{\bullet\bullet\bullet}} (x_{ijk} x_{\bullet\bullet\bullet} - x_{i\bullet\bullet} x_{\bullet jk}) \quad (22)
 \end{aligned}$$

Since  $\sigma_{\bullet jk}$  is equivalent to the residual of two-dimensional contingency matrix, it can be represented as a linear sum of  $2 \times 2$  submatrices. Thus, the latter part  $x_{ijk} x_{\bullet\bullet\bullet} - x_{i\bullet\bullet} x_{\bullet jk}$  should be examined. Let us denote this part by  $\sigma_{jk}^i$ .

$$\begin{aligned}
 \sigma_{jk}^i &= x_{ijk} x_{\bullet\bullet\bullet} - x_{i\bullet\bullet} x_{\bullet jk} \\
 &= x_{ijk} \left( x_{i\bullet\bullet} + \sum_{l \neq i} x_{l\bullet\bullet} \right) \\
 &\quad - x_{i\bullet\bullet} \left( x_{ijk} + \sum_{m \neq i} x_{mjk} \right) \\
 &= x_{ijk} \sum_{l \neq i} x_{l\bullet\bullet} - x_{i\bullet\bullet} \sum_{m \neq i} x_{mjk} \\
 &= \sum_{l \neq i} (x_{ijk} x_{l\bullet\bullet} - x_{i\bullet\bullet} x_{ljk}) \\
 &= \sum_{\substack{l \neq i \\ m \neq j \text{ or } n \neq k}} (x_{ijk} x_{lmn} - x_{imn} x_{ljk}) \\
 &= \sum_{\substack{l \neq i \\ n \neq k}} \Delta(j)_{ln}^{ik} + \sum_{\substack{l \neq i \\ m \neq j}} \Delta(k)_{lm}^{ij} \\
 &\quad + \sum_{\substack{l \neq i \\ m \neq j \\ n \neq k}} (x_{ijk} x_{lmn} - x_{imn} x_{ljk}), \quad (23)
 \end{aligned}$$

where  $\Delta(j)_{ln}^{ik}$  denotes the determinant of  $2 \times 2$  matrix in which  $i$  and  $l$  rows and  $k$  and  $n$  columns are selected with  $m = j$  fixed and  $\Delta(k)_{lm}^{ij}$  denotes the determinant of  $2 \times 2$  matrix in which  $i$  and  $l$  rows and  $j$  and  $m$  columns are selected with  $n = k$  fixed. On

the other hand, the last part is different from the former two sums. This sum is zero if the conditional independence are satisfied, that is, if  $x_{ijk} = x_{i\bullet\bullet} x_{\bullet jk}$ , whose meaning is  $i$  and  $j$  are statistical independence with  $k$  fixed. Thus, the following theorem is obtained.

**Theorem 6.** *The residual of three-dimensional contingency table is decomposed as:*

$$\begin{aligned}
 \sigma_{ijk} &= \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \sigma_{\bullet jk} + \sigma_{jk}^i \\
 &= \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \sigma_{\bullet jk} \\
 &\quad + \frac{1}{x_{\bullet\bullet\bullet}} \left( \sum_{\substack{l \neq i \\ n \neq k}} \Delta(j)_{ln}^{ik} + \sum_{\substack{l \neq i \\ m \neq j}} \Delta(k)_{lm}^{ij} \right. \\
 &\quad \left. + \sum_{\substack{l \neq i \\ m \neq j \\ n \neq k}} (x_{ijk} x_{lmn} - x_{imn} x_{ljk}) \right). \quad (24)
 \end{aligned}$$

□

### 4.3. Information Granule for Multi-way Contingency Table

The structure of Equation (24) suggests that the residual can be defined in a recursive way.

**Theorem 7.** *Let  $n$  be a number of attributes in a contingency table ( i.e.  $n$ -way contingency table. The residual of the cell  $(i_1, i_2, \dots, i_n)$ ,  $\sigma_{i_1 i_2 \dots i_n}$  is decomposed as follows.*

$$\begin{aligned}
 \sigma_{i_1 i_2 \dots i_n} &= \frac{x_{i_1 \dots i_n}}{x_{\bullet \dots \bullet}} \sigma_{\bullet i_2 \dots i_n} + \frac{1}{x_{\bullet \dots \bullet}} \sigma_{i_1}^{i_2 \dots i_n} \\
 &= \frac{x_{i_1 \dots i_n}}{x_{\bullet \dots \bullet}} \sigma_{\bullet i_2 \dots i_n} \\
 &\quad + \frac{1}{x_{\bullet \dots \bullet}} (x_{i_1 i_2 \dots i_n} x_{\bullet \dots \bullet} - x_{i_1 \dots \bullet} x_{\bullet i_2 \dots i_n}), \quad (25)
 \end{aligned}$$

□

For example, in the case of a four-dimensional contingency table, a residual  $\sigma_{i_1 i_2 i_3 i_4}$  is decomposed

as:

$$\begin{aligned}\sigma_{i_1 i_2 i_3 i_4} &= \frac{x_{i_1 \bullet \bullet \bullet}}{x_{\bullet \bullet \bullet \bullet}} \sigma_{\bullet i_2 i_3 i_4} + \frac{1}{x_{\bullet \bullet \bullet \bullet}} \sigma_{i_2 i_3 i_4}^{i_1} \\ &= \frac{x_{i_1 \bullet \bullet \bullet}}{x_{\bullet \bullet \bullet \bullet}} \sigma_{\bullet i_2 i_3 i_4} \\ &+ \frac{1}{x_{\bullet \bullet \bullet \bullet}} (x_{i_1 i_2 i_3 i_4} x_{\bullet \bullet \bullet \bullet} - x_{i_1 \bullet \bullet \bullet} x_{\bullet i_2 i_3 i_4}),\end{aligned}$$

where  $\sigma_{\bullet i_2 i_3 i_4}$  can be decomposed as shown in Equation (24):

$$\begin{aligned}\sigma_{\bullet i_2 i_3 i_4} &= \frac{x_{\bullet i_2 \bullet \bullet}}{x_{\bullet \bullet \bullet \bullet}} \sigma_{\bullet \bullet i_3 i_4} + \frac{1}{x_{\bullet \bullet \bullet \bullet}} \sigma_{\bullet i_3 i_4}^{i_2} \\ &= \frac{x_{\bullet i_2 \bullet \bullet}}{x_{\bullet \bullet \bullet \bullet}} \sigma_{\bullet \bullet i_3 i_4} \\ &+ \frac{1}{x_{\bullet \bullet \bullet \bullet}} (x_{\bullet i_2 i_3 i_4} x_{\bullet \bullet \bullet \bullet} - x_{\bullet i_2 \bullet \bullet} x_{\bullet \bullet i_3 i_4}).\end{aligned}$$

Then,  $\sigma_{\bullet i_3 i_4}^{i_2}$  is reformulated as shown in Equation (23).

## 5. Examples

### 5.1. $2 \times 2 \times 2$ Contingency Table

The residual of a cell  $(1, 1, 1)$  is decomposed as:

$$\begin{aligned}\sigma_{111} &= \frac{x_{1 \bullet \bullet}}{x_{\bullet \bullet \bullet}} \sigma_{\bullet 11} + \frac{1}{x_{\bullet \bullet \bullet}} (x_{111} x_{\bullet \bullet \bullet} - x_{1 \bullet \bullet} x_{\bullet 11}) \\ &= \frac{x_{1 \bullet \bullet}}{x_{\bullet \bullet \bullet}} \sigma_{\bullet 11} + \frac{1}{x_{\bullet \bullet \bullet}} \sigma_{11}^1\end{aligned}$$

and

$$\begin{aligned}\sigma_{11}^1 &= (x_{111} x_{212} - x_{211} x_{112}) \\ &+ (x_{111} x_{221} - x_{211} x_{121}) \\ &+ (x_{111} x_{222} - x_{211} x_{112}),\end{aligned}$$

which consists of three components. Since the number of determinants of  $\sigma_{\bullet 11}$  is equal to 1 and that of  $\sigma_{11}^1$  is equal to 3, the total number of the determinants is equal to 4. In the case of conditional independence  $x_{ijk} = x_{i \bullet \bullet} x_{\bullet jk}$ , the last part of  $\sigma_{11}^1$  is equal to 0, the total number of the determinants in the residual is equal to 3. It is notable that these numbers on the degree of freedom are exactly the same as those of the degree of freedom in contingency table

analysis<sup>1</sup>. This result is not accidental, as shown in Section 6.

The geometrical structure of  $\sigma_{11}^1$  can be depicted as Figure 1. The shadowed parts correspond to  $x_{111} x_{212} - x_{211} x_{112}$ ,  $x_{111} x_{221} - x_{211} x_{121}$ , and  $x_{111} x_{222} - x_{211} x_{112}$ . Thus, these three parts give three possible combinations of parallelograms which has  $x_{111} x_{211}$  as the fixed edge.

### 5.2. $3 \times 2 \times 3$ Contingency Table

The residual of a cell  $(1, 1, 1)$  is decomposed as:

$$\begin{aligned}\sigma_{111} &= \frac{x_{1 \bullet \bullet}}{x_{\bullet \bullet \bullet}} \sigma_{\bullet 11} + \frac{1}{x_{\bullet \bullet \bullet}} (x_{111} x_{\bullet \bullet \bullet} - x_{1 \bullet \bullet} x_{\bullet 11}) \\ &= \frac{x_{1 \bullet \bullet}}{x_{\bullet \bullet \bullet}} \sigma_{\bullet 11} + \frac{1}{x_{\bullet \bullet \bullet}} \sigma_{11}^1\end{aligned}$$

and

$$\begin{aligned}\sigma_{11}^1 &= (x_{111} x_{212} - x_{211} x_{112}) \\ &+ (x_{111} x_{213} - x_{211} x_{113}) \\ &+ (x_{111} x_{221} - x_{211} x_{121}) \\ &+ (x_{111} x_{222} - x_{211} x_{122}) \\ &+ (x_{111} x_{223} - x_{211} x_{123}) \\ &+ (x_{111} x_{231} - x_{211} x_{131}) \\ &+ (x_{111} x_{232} - x_{211} x_{132}) \\ &+ (x_{111} x_{233} - x_{211} x_{133})\end{aligned}$$

which consists of eight components. The geometrical structure of  $\sigma_{11}^1$  can be depicted as Figure 2. The shadowed parts correspond to  $x_{111} x_{212} - x_{211} x_{112}$ ,  $x_{111} x_{213} - x_{211} x_{113}$ ,  $x_{111} x_{221} - x_{211} x_{121}$ ,  $x_{111} x_{222} - x_{211} x_{122}$ ,  $x_{111} x_{223} - x_{211} x_{123}$ ,  $x_{111} x_{231} - x_{211} x_{131}$ ,  $x_{111} x_{232} - x_{211} x_{132}$ , and  $x_{111} x_{233} - x_{211} x_{133}$ . Thus, these eight parts give eight possible combinations of parallelograms which has  $x_{111} x_{211}$  as the fixed edge.

### 5.3. $3 \times 3 \times 3$ Contingency Table

The residual of a cell  $(1, 1, 1)$  is decomposed as:

$$\begin{aligned}\sigma_{111} &= \frac{x_{1 \bullet \bullet}}{x_{\bullet \bullet \bullet}} \sigma_{\bullet 11} + \frac{1}{x_{\bullet \bullet \bullet}} (x_{111} x_{\bullet \bullet \bullet} - x_{1 \bullet \bullet} x_{\bullet 11}) \\ &= \frac{x_{1 \bullet \bullet}}{x_{\bullet \bullet \bullet}} \sigma_{\bullet 11} + \frac{1}{x_{\bullet \bullet \bullet}} \sigma_{11}^1\end{aligned}$$

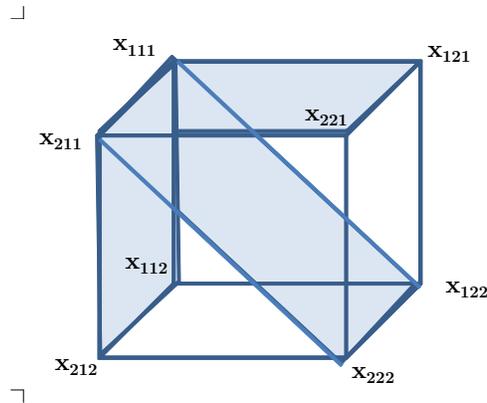


Figure 1: Geometrical Structure of  $2 \times 2 \times 2$  Contingency Table

and

$$\begin{aligned}
 \sigma_{11}^1 = & (x_{111}x_{212} - x_{211}x_{112}) \\
 & + (x_{111}x_{213} - x_{211}x_{113}) \\
 & + (x_{111}x_{221} - x_{211}x_{121}) \\
 & + (x_{111}x_{222} - x_{211}x_{122}) \\
 & + (x_{111}x_{223} - x_{211}x_{123}) \\
 & + (x_{111}x_{231} - x_{211}x_{131}) \\
 & + (x_{111}x_{232} - x_{211}x_{132}) \\
 & + (x_{111}x_{233} - x_{211}x_{133}) \\
 & + (x_{111}x_{312} - x_{311}x_{112}) \\
 & + (x_{111}x_{313} - x_{311}x_{113}) \\
 & + (x_{111}x_{321} - x_{311}x_{121}) \\
 & + (x_{111}x_{322} - x_{311}x_{122}) \\
 & + (x_{111}x_{323} - x_{311}x_{123}) \\
 & + (x_{111}x_{331} - x_{311}x_{131}) \\
 & + (x_{111}x_{332} - x_{311}x_{132}) \\
 & + (x_{111}x_{333} - x_{311}x_{133})
 \end{aligned}$$

which consists of 16 components. The geometrical structure of  $\sigma_{11}^1$  can be depicted as Figure 3. In the same way as  $2 \times 3 \times 3$  table, these 16 parts give all the combinations of parallelograms which has  $x_{111}x_{211}$  and  $x_{111}x_{311}$  as the fixed edges.

## 6. Discussion

### 6.1. #Degree of Freedom = #Information Granules

In the above sections, we mentioned that the number of determinants of  $2 \times 2$  submatrices in a three-way contingency table is equal to the degree of freedom of  $\chi^2$  test statistic in contingency table analysis. This section shows that it is not accidental.

Everitt<sup>1</sup> shows that the degree of freedom of  $\chi^2$  test statistic is given as:

$$\begin{aligned}
 d.f. &= rcl - (r-1) - (c-1) - (l-1) - 1 \\
 &= rcl - r - c - l + 2
 \end{aligned} \tag{26}$$

On the other hand, with the hypothesis of conditional independence on  $(r-1)$  row probabilities and column  $\times$  layer probabilities  $(cl-1)$ , the degree of freedom is given as:

$$\begin{aligned}
 d.f. &= rcl - (cl-1) - (r-1) - 1 \\
 &= rcl - cl - r + 1
 \end{aligned} \tag{27}$$

Let us go back to Equation 25 in Theorem 6 and assume that indices  $i, j$ , and  $k$  correspond to row, column, and layer, respectively. The numbers of the determinants are summarized into:

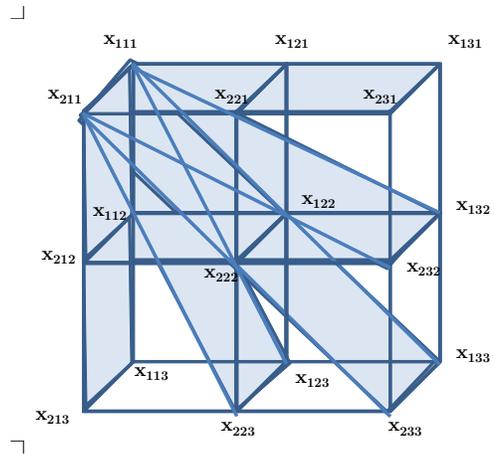


Figure 2: Geometrical Structure of  $2 \times 3 \times 3$  Contingency Table

$$\begin{aligned} \sigma_{\bullet jk} & (c-1)(l-1) \\ \Delta(j)_{ln}^{ik} & (r-1)(l-1) \\ \Delta(k)_{lm}^{ij} & (r-1)(c-1) \\ x_{ijk}x_{lmn} - x_{imn}x_{ljk} & (r-1)(c-1)(l-1) \end{aligned}$$

The total number of the determinants sum is:

$$\begin{aligned} \text{sum} &= (c-1)(l-1) + (r-1)(l-1) \\ &+ (r-1)(c-1) + (r-1)(c-1)(l-1) \\ &= rcl - r - c - l + 2, \end{aligned}$$

which is equal to Equation 26. Thus, the following theorem is obtained.

**Theorem 8.** *The total number of determinants of  $2 \times 2$  submatrices in a three-way contingency table is equal to the degree of freedom of  $\chi^2$  test statistic in contingency table analysis.*

Furthermore, since the total number of the determinants of two-way contingency tables with  $c$  rows and  $l$  columns is  $(c-1)(l-1)$ , the difference between the number of the determinants of three-way tables and that of two tables,  $\delta$  is equal to:

$$\begin{aligned} \delta &= (c-1)(l-1) + (r-1)(l-1) \\ &+ (r-1)(c-1) + (r-1)(c-1)(l-1) \\ &- (c-1)(l-1) \\ &= rcl - cl - r + 1, \end{aligned}$$

which is equal to Equation 27. Thus,

**Theorem 9.** *The total number of determinants of  $2 \times 2$  submatrices in a three-way contingency table can be decomposed into the sum of the total number of determinants in a two-way contingency table and that in a three-way contingency table with conditional independence. It will be our future work to generalize these results into multi-way contingency tables, whose dimension is larger than 4.*

## 6.2. Towards Applications

Since the determinants of  $2 \times 2$  subtables (submatrices) can be viewed as information granules, these values measure the local nature of a given table. On the other hand, a  $\chi^2$ -test statistic gives the global nature of a given table. If we only need to evaluate the global degree of statistical independence, a  $\chi^2$ -test statistic is sufficient. However, when we want to evaluate the precise nature of a given table, or the local degree of statistical independence, the values for each subtable will play important roles in further investigation of the local relations between attributes.

In other words, information granules give local natures of statistical independence and those behavior is integrated into the global nature of a given ta-

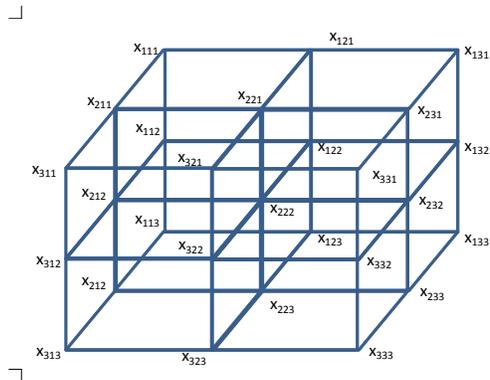


Figure 3: Geometrical Structure of  $2 \times 3 \times 3$  Contingency Table

ble: integration of local into global behavior can also be viewed as a study on granular computing. Then, do such a multi-way contingency table have intermediate granules between the determinant of a  $2 \times 2$  subtables and a Pearson residual or  $\chi^2$ -test statistics? This will be the next target of our research.

## 7. Conclusion

This paper focuses on statistical independence of three variables from the viewpoint of linear algebra and show that information granules of statistical independence of three or four variables are decomposed into linear sum of determinants of  $2 \times 2$ -submatrices. The analysis also shows that the residuals of a multiway contingency table can be defined in a recursive way. For example, the residuals of four-way tables are described as those of three-way and two-way tables.

Then, we focus on this recursive nature in the case of three-way contingency tables. The geometric characteristics of selected determinants in a residual  $x_{ijk}$  show that the components of the sum give possible combinations of parallelograms which has one

fixed edge with  $x_{ijk}$  as a vertex. Furthermore, The total number of determinants of  $2 \times 2$  submatrices in a three-way contingency table is equal to the degree of freedom of  $\chi^2$  test statistic in contingency table analysis.

Thus, the determinants of  $2 \times 2$  matrices are principal information granules to measure the degree of statistical dependence in a given contingency table.

1. B.S. Everitt. *The Analysis of Contingency Tables*. Chapman & Hall/CRC, 2nd edition, 1992.
2. Shusaku Tsumoto. Contingency matrix theory: Statistical dependence in a contingency table. *Inf. Sci.*, 179(11):1615–1627, 2009.
3. Shusaku Tsumoto and Shoji Hirano. Meaning of pearson residuals - linear algebra view -. In *Proceedings of IEEE GrC 2007*. IEEE press, 2007.
4. Shusaku Tsumoto and Shoji Hirano. Contingency matrix theory ii: Degree of dependence as granularity. *Fundam. Inform.*, 90(4):427–442, 2009.
5. Shusaku Tsumoto and Shoji Hirano. Dependency and granularity in data. In Robert A. Meyers, editor, *Encyclopedia of Complexity and Systems Science*, pages 1864–1872. Springer, 2009.
6. Shusaku Tsumoto and Shoji Hirano. Statistical independence and determinants in a contingency table - interpretation of pearson residuals based on linear algebra -. *Fundam. Inform.*, 90(3):251–267, 2009.