Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Pearson Residuals in Multi-way Contingency Tables

Shusaku Tsumoto, Shoji Hirano and Hidenao Abe
Department of Medical Informatics, Faculty of Medicine,
Shimane University
89-1 Enya-cho Izumo 693-8501 Japan
Email: {tsumoto,hirano,abe}@med.shimane-u.ac.jp

***Abstract*—This paper shows the meaning of Pearson residuals as an indicator of statistical indepndence in a multi-way contingency table. While information granules of statistical independence of two variables can be viewed as determinants of 2×2- submatrices, those of multivariate cases consist of linear sum of residuals for odds ratios, which can be viewed as an extention of determinants in 2×2 matrices.**

***Index Terms*—Granular Computing, Contingency Table, Pearson Residuals, Data Mining**

I. 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, which is continuation of studies on contingency matrix theory in [1]–[3].

The first important observation is that a contingency table compares two attributes with respect to information granularity. It is shown from the definition that statistifcal independence in a contingency table is a special form of linear depedence 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. Also, the results also show that partial statistical independence can be observed.

The second important observation is that 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.

The paper is organized as follows: Section 2 discusses the characteristics of contingency tables. Section 3 shows the conditions on statistical independence for a 2×2 table. Section 4 gives those for a $2 \times n$ table. Section 5 extends these results into a multi-way contingency table. Section 6 discusses statistical independence from matrix theory. Section 7 and 8 show pseudo statistical independence. Finally, Section 9 concludes this paper.

II. Contingency Matrix

Definition 1: Let R_1 and R_2 denote multinominal attributes in an attribute space A which have m and n values. A contingency tables $T(R_1, R_2)$ is a table of a set of the meaning of the following formulas: $|[R_1 = A_j]_A|$, $|[R_2 = B_i]_A|$, $|[R_1 = A_j \wedge R_2 = B_i]_A|$, $|U|$ ($i = 1, 2, 3, \cdots, n$ and $j = 1, 2, 3, \cdots, m$). This table is arranged into the form shown in Table I, where: $|[R_1 = A_j]_A| = \sum_{i=1}^{m} x_{1i} = x_{\cdot j}$, $|[R_2 = B_i]_A| = \sum_{j=1}^{n} x_{ji} = x_{i\cdot}$, $|[R_1 = A_j \wedge R_2 = B_i]_A| = x_{ij}$, $|U| = N = x_{\cdot\cdot}$ ($i = 1, 2, 3, \cdots, n$ and $j = 1, 2, 3, \cdots, m$).

TABLE I
Contingency Table ($m \times n$)

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

Definition 2: A contigency matrix $M_{R_1,R_2}(m, n, N)$ is defined as a matrix, which is composed of $x_{ij} = |[R_1 = A_j \wedge R_2 = B_i]_A|$, extracted from a contigency table defined in definition 1.

That is,

$$M_{R_1,R_2}(m, n, N) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}.$$

□

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 relations between two attributes with respect to intersection of their supporting sets. When two attributes have different number 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 the attribute which has a smaller number of equivalence classes. and the surplus number of equivalence classes of the attributes with larger number of equivalnce classes can be projected into other partitions. In other words, a $m \times n$ matrix or contingency table includes a projection from one attributes to the other one.

III. Statistical Independence in 2 × 2 Contingency Table

Let us consider a contingency table shown in Table I ($m = n = 2$). Statistical independence between R_1 and R_2 gives:

$$\begin{aligned}
P([R_1=0],[R_2=0]) &= P([R_1=0]) \times P([R_2=0]) \\
P([R_1=0],[R_2=1]) &= P([R_1=0]) \times P([R_2=1]) \\
P([R_1=1],[R_2=0]) &= P([R_1=1]) \times P([R_2=0]) \\
P([R_1=1],[R_2=1]) &= P([R_1=1]) \times P([R_2=1])
\end{aligned}$$

Since each probability is given as a ratio of each cell to N, the above equations are calculated as:

$$\begin{aligned}
\frac{x_{11}}{N} &= \frac{x_{11}+x_{12}}{N} \times \frac{x_{11}+x_{21}}{N} \\
\frac{x_{12}}{N} &= \frac{x_{11}+x_{12}}{N} \times \frac{x_{12}+x_{22}}{N} \\
\frac{x_{21}}{N} &= \frac{x_{21}+x_{22}}{N} \times \frac{x_{11}+x_{21}}{N} \\
\frac{x_{22}}{N} &= \frac{x_{21}+x_{22}}{N} \times \frac{x_{12}+x_{22}}{N}
\end{aligned}$$

Since $N = \sum_{i,j} x_{ij}$, the following formula will be obtained from these four formulae.

$$x_{11}x_{22} = x_{12}x_{21} \; or \; x_{11}x_{22} - x_{12}x_{21} = 0$$

Thus,

Theorem 1: If two attributes in a contingency table shown in a 2×2 contingency table are statistical indepedent, the following equation holds:

$$x_{11}x_{22} - x_{12}x_{21} = 0 \qquad (1)$$

□

It is notable that the above equation corresponds to the fact that the determinant of a matrix corresponding to this table is equal to 0. Also, when these four values are not equal to 0, the equation 1 can be transformed into:

$$\frac{x_{11}}{x_{21}} = \frac{x_{12}}{x_{22}}.$$

Let us assume that the above ratio is equal to $C(constant)$. Then, since $x_{11} = Cx_{21}$ and $x_{12} = Cx_{22}$, the following equation is obtained.

$$\frac{x_{11}+x_{12}}{x_{21}+x_{22}} = \frac{C(x_{21}+x_{22})}{x_{21}+x_{22}} = C = \frac{x_{11}}{x_{21}} = \frac{x_{12}}{x_{22}}. \qquad (2)$$

This equation also holds when we extend this discussion into a general case. Before getting into it, let us cosndier a 2×3 contingency table.

IV. Statistical Independence in 2 × 3 Contingency Table

Let us consider a 2×3 contingency table II. Statistical

TABLE II
Contingency Table (2 × 3)

	$R_1 = 0$	$R_1 = 1$	$R_1 = 2$	
$R_2 = 0$	x_{11}	x_{12}	x_{13}	$x_{1\cdot}$
$R_2 = 1$	x_{21}	x_{22}	x_{23}	$x_{2\cdot}$
	$x_{\cdot 1}$	$x_{\cdot 2}$	$x_{\dots 3}$	$x_{\cdot\cdot}$
			$(= \lvert U \rvert = N)$	

independence between R_1 and R_2 gives:

$$\begin{aligned}
P([R_1=0],[R_2=0]) &= P([R_1=0]) \times P([R_2=0]) \\
P([R_1=0],[R_2=1]) &= P([R_1=0]) \times P([R_2=1]) \\
P([R_1=0],[R_2=2]) &= P([R_1=0]) \times P([R_2=2]) \\
P([R_1=1],[R_2=0]) &= P([R_1=1]) \times P([R_2=0]) \\
P([R_1=1],[R_2=1]) &= P([R_1=1]) \times P([R_2=1]) \\
P([R_1=1],[R_2=2]) &= P([R_1=1]) \times P([R_2=2])
\end{aligned}$$

Since each probability is given as a ratio of each cell to N, the above equations are calculated as:

$$\begin{aligned}
\frac{x_{11}}{N} &= \frac{x_{11}+x_{12}+x_{13}}{N} \times \frac{x_{11}+x_{21}}{N} &(3)\\
\frac{x_{12}}{N} &= \frac{x_{11}+x_{12}+x_{13}}{N} \times \frac{x_{12}+x_{22}}{N} &(4)\\
\frac{x_{13}}{N} &= \frac{x_{11}+x_{12}+x_{13}}{N} \times \frac{x_{13}+x_{23}}{N} &(5)\\
\frac{x_{21}}{N} &= \frac{x_{21}+x_{22}+x_{23}}{N} \times \frac{x_{11}+x_{21}}{N} &(6)\\
\frac{x_{22}}{N} &= \frac{x_{21}+x_{22}+x_{23}}{N} \times \frac{x_{12}+x_{22}}{N} &(7)\\
\frac{x_{23}}{N} &= \frac{x_{21}+x_{22}+x_{23}}{N} \times \frac{x_{13}+x_{23}}{N} &(8)
\end{aligned}$$

From equation (3) and (6),

$$\frac{x_{11}}{x_{21}} = \frac{x_{11}+x_{12}+x_{13}}{x_{21}+x_{22}+x_{23}}$$

In the same way, the following equation will be obtained:

$$\frac{x_{11}}{x_{21}} = \frac{x_{12}}{x_{22}} = \frac{x_{13}}{x_{23}} = \frac{x_{11}+x_{12}+x_{13}}{x_{21}+x_{22}+x_{23}} \qquad (9)$$

Thus, we obtain the following theorem:

Theorem 2: If two attributes in a contingency table shown in Table II are statistical indepedent, the following equations hold:

$$\begin{aligned}
x_{11}x_{22} - x_{12}x_{21} &= x_{12}x_{23} - x_{13}x_{22} \\
&= x_{13}x_{21} - x_{11}x_{23} = 0 \qquad (10)
\end{aligned}$$

□

It is notable that this discussion can be easily extended into a $2xn$ contingency table where $n > 3$. The important equation 9 will be extended into

$$\begin{aligned}
\frac{x_{11}}{x_{21}} &= \frac{x_{12}}{x_{22}} = \cdots = \frac{x_{1n}}{x_{2n}} \\
&= \frac{x_{11}+x_{12}+\cdots+x_{1n}}{x_{21}+x_{22}+\cdots+x_{2n}} = \frac{\sum_{k=1}^{n} x_{1k}}{\sum_{k=1}^{n} x_{2k}} \qquad (11)
\end{aligned}$$

Thus,

Theorem 3: If two attributes in a contingency table ($2 \times k(k = 2, \cdots, n)$) are statistical indepedent, the following equations hold:

$$\begin{aligned} x_{11}x_{22} - x_{12}x_{21} &= x_{12}x_{23} - x_{13}x_{22} = \cdots \\ &= x_{1n}x_{21} - x_{11}x_{n3} = 0 \end{aligned} \quad (12)$$

□

It is also notable that this equation is the same as the equation on collinearity of projective geometry [4].

V. Statistical Independence in $m \times n$ Contingency Table

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

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

According to the definition of the table,

$$\frac{x_{ij}}{N} = \frac{\sum_{k=1}^{n} x_{ik}}{N} \times \frac{\sum_{l=1}^{m} x_{lj}}{N}. \quad (13)$$

Thus, we have obtained:

$$x_{ij} = \frac{\sum_{k=1}^{n} x_{ik} \times \sum_{l=1}^{m} x_{lj}}{N}. \quad (14)$$

Thus, for a fixed j,

$$\frac{x_{i_a j}}{x_{i_b j}} = \frac{\sum_{k=1}^{n} x_{i_a k}}{\sum_{k=1}^{n} x_{i_b k}}$$

In the same way, for a fixed i,

$$\frac{x_{i j_a}}{x_{i j_b}} = \frac{\sum_{l=1}^{m} x_{l j_a}}{\sum_{l=1}^{m} x_{l j_b}}$$

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

$$\frac{x_{i_a 1}}{x_{i_b 1}} = \frac{x_{i_a 2}}{x_{i_b 2}} \cdots = \frac{x_{i_a n}}{x_{i_b n}} = \frac{\sum_{k=1}^{n} x_{i_a k}}{\sum_{k=1}^{n} x_{i_b k}}. \quad (15)$$

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

Theorem 4: If two attributes in a contingency table shown in Table I are statistical indepedent, the following equations hold:

$$\frac{x_{i_a 1}}{x_{i_b 1}} = \frac{x_{i_a 2}}{x_{i_b 2}} \cdots = \frac{x_{i_a n}}{x_{i_b n}} = const. \quad (16)$$

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

□

A. Three-way Table

Let "•" denote as the sum over the row or column of a contingency matrix. That is ,

$$x_{i\bullet} = \sum_{j=1}^{n} x_{ij} \quad (17)$$

$$x_{\bullet j} = \sum_{i=1}^{m} x_{ij}, \quad (18)$$

where (17) and (18) 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 (14) is reformulated as:

$$\frac{x_{ij}}{x_{\bullet\bullet}} = \frac{x_{i\bullet}}{x_{\bullet\bullet}} \times \frac{x_{\bullet j}}{x_{\bullet\bullet}} \quad (19)$$

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

Statistical independence with three attributes is defined as:

$$\frac{x_{ijk}}{x_{\bullet\bullet\bullet}} = \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \times \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \times \frac{x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}}, \quad (20)$$

Thus,

$$x_{ijk}x_{\bullet\bullet\bullet}^2 = x_{i\bullet\bullet}x_{\bullet j\bullet}x_{\bullet\bullet k}, \quad (21)$$

which corresponds to:

$$P(A = a, B = b, C = c) = P(A = a)P(B = b)P(C = c), \quad (22)$$

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

In statistical context, statistical independence requires hiearchical 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 (22), the following equations should satisfy:

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

Thus,

$$x_{ij\bullet}x_{\bullet\bullet\bullet} = x_{i\bullet\bullet}x_{\bullet j\bullet} \quad (23)$$

$$x_{i\bullet k}x_{\bullet\bullet\bullet} = x_{i\bullet\bullet}x_{\bullet\bullet k} \quad (24)$$

$$x_{\bullet jk}x_{\bullet\bullet\bullet} = x_{\bullet j\bullet}x_{\bullet\bullet k} \quad (25)$$

From Equation (21) and Equation (23),

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

Therefore,

$$\frac{x_{ijk}}{x_{ij\bullet}} = \frac{x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}} \quad (26)$$

In the same way, the following equations are obtained:

$$\frac{x_{ijk}}{x_{i\bullet k}} = \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \quad (27)$$

$$\frac{x_{ijk}}{x_{\bullet jk}} = \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \quad (28)$$

In summary, the following theorem is obtained.

Theorem 5: If a three-way contingency table satisfy statistical independence, then the following three equations should be satisfied:

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

$$\frac{x_{ijk}}{x_{i\bullet k}} = \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}}$$

$$\frac{x_{ijk}}{x_{\bullet jk}} = \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}}$$

□

Thus, the equations corresponding to Theorem 4 are obtained as follows.

Corollary 1: If three attributes in a contingency table shown in Table I are statistical indepedent, the following equations hold:

$$\frac{x_{ijk_a}}{x_{ijk_b}} = \frac{x_{\bullet\bullet k_a}}{x_{\bullet\bullet k_b}}$$

$$\frac{x_{ij_ak}}{x_{ij_bk}} = \frac{x_{\bullet j_a\bullet}}{x_{\bullet j_b\bullet}}$$

$$\frac{x_{i_ajk}}{x_{i_bjk}} = \frac{x_{i_a\bullet\bullet}}{x_{i_b\bullet\bullet}}$$

for all i,j, and k.

□

B. Multi-way Table

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

Theorem 6: 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 $(k = 1, 2, \cdots, n)$ where n is the number of attributes.

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

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

$$x_{i_1 i_2 \cdots i_n} \times x^{n-1}_{\bullet\bullet\cdots\bullet} = x_{i_1\bullet\cdots\bullet}x_{\bullet i_2\cdots\bullet} \times \cdots \times x_{\bullet\bullet\cdots i_k\cdots\bullet} \times \cdots \times x_{\bullet\bullet\cdots\bullet i_n}$$

□

VI. Information Granule for Contingency Matrix

A. Residual of Contingency Matrix

Tsumoto and Hirano [5] discusses the meaning of pearson residuals from the viewpoint of linear algebra.

The residual is defined as a difference between an observed value for each cell in a contingency matrix and an expected value:

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

And simple calculation leads to the following theorem.

Theorem 7: The residual of $M_{R_1,R_2}(m,n,N)$ is obtained as:

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

where $\Delta^{i,j}_{k,l}$ is the determinant of a 2×2 submatrix of $M_{R_1,R_2}(m,n,N)$ 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:

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

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

□

For example, for $M_{R_1,R_2}(2,2,N)$ $(m = n = 2)$,

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

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

In the case of $M_{R_1,R_2}(3,3,N)$ $(m = n = 3)$, the following formulas are obtained.

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

Thus, a 2×2 submatrix in a contingency table can be viewed as a information granule for statistical (in)dependence.

Can we generalize this results into statistical independence of three variables ? This is our main question to be partially answered in this paper.

B. Information Granule for $2 \times 2 \times 2$ Data Cube

Let us get back to Equation (21), (23), (24) and (25).

From Equation (21), the residual for x_{ijk} is obtained as:

$$\sigma_{ijk} = x_{ijk} - \frac{x_{i\bullet\bullet} \times x_{\bullet j\bullet} \times x_{\bullet\bullet k}}{x^2_{\bullet\bullet\bullet}}.$$

For simplicity, let us confine to $2 \times 2 \times 2$-data cube. Then, the above residual for x_{111} will be:

$$\begin{aligned}
\sigma_{111} &= x_{111} - \frac{x_{1\bullet\bullet} \times x_{\bullet 1\bullet} \times x_{\bullet\bullet 1}}{x^2_{\bullet\bullet\bullet}} \\
&= \frac{1}{x^2_{\bullet\bullet\bullet}} \left\{ x_{111} \left(x^2_{\bullet\bullet\bullet} - x_{\bullet 1\bullet} x_{\bullet\bullet 1} \right) \right. \\
&\quad \left. - \sum_{\substack{k \neq i \ or \\ l \neq j}} x_{1jk} x_{\bullet 1\bullet} x_{\bullet\bullet 1} \right\} \\
&= \frac{1}{x^2_{\bullet\bullet\bullet}} \left\{ x_{\bullet 1\bullet} (x_{111} x_{\bullet\bullet 2} - x_{112} x_{\bullet\bullet 1}) \right. \\
&\quad + x_{\bullet\bullet 1} (x_{111} x_{\bullet 2\bullet} - x_{121} x_{\bullet 1\bullet}) \\
&\quad \left. + x_{111} x_{\bullet 2\bullet} x_{\bullet\bullet 2} - x_{122} x_{\bullet 1\bullet} x_{\bullet\bullet 1} \right\}
\end{aligned}$$

Thus, the following proposition is obtained:

Proposition 1: The residual $sigma_{111}$ of $2 \times 2 \times 2$-data cube is obtained as:

$$\begin{aligned}
\sigma_{111} x^2_{\bullet\bullet\bullet} &= x_{\bullet 1\bullet} (x_{111} x_{\bullet\bullet 2} - x_{112} x_{\bullet\bullet 1}) \\
&\quad + x_{\bullet\bullet 1} (x_{111} x_{\bullet 2\bullet} - x_{121} x_{\bullet 1\bullet}) \\
&\quad + x_{111} x_{\bullet 2\bullet} x_{\bullet\bullet 2} - x_{122} x_{\bullet 1\bullet} x_{\bullet\bullet 1} \qquad (29)
\end{aligned}$$

□

The above formula shows that the products of the odds ratio and the marginal distribution play central roles in statistical independence.

Let us extend the notation of subdeterminants in two-dimensional cases into three-dimentional ones as:

$$\begin{aligned}
\Delta^{ijk}_{imn} &= x_{ijk} (\sum_{m \neq j} x_{\bullet m\bullet})(\sum_{n \neq k} x_{\bullet\bullet n}) \\
&\quad - (\sum_{\substack{m \neq j \\ n \neq k}} x_{imn}) x_{\bullet j\bullet} x_{\bullet\bullet k}. \qquad (30)
\end{aligned}$$

Then, the above formula 29 can be represented as:

$$\sigma_{111} x^2_{\bullet\bullet\bullet} = x_{\bullet 1\bullet} \Delta^{111}_{112} + + x_{\bullet\bullet 1} \Delta^{111}_{121} + \Delta^{111}_{122} \qquad (31)$$

C. Information Granule for $l \times m \times n$ Data Cube

The above results can be generalized, although the results are rather complicated [6].

Theorem 8: The residual σ_{ijk} of $l \times m \times n$-data cube is obtained as:

$$\begin{aligned}
\sigma_{ijk} x^2_{\bullet\bullet\bullet} &= x_{\bullet j\bullet} (x_{ijk} \sum_{n \neq k} x_{\bullet\bullet n} - \sum_{n \neq k} x_{ijn} x_{\bullet\bullet k}) \\
&\quad + x_{\bullet\bullet k} (x_{ijk} \sum_{m \neq j} x_{\bullet m\bullet} - \sum_{m \neq j} x_{imk} x_{\bullet j\bullet}) \\
&\quad + x_{ijk} (\sum_{m \neq j} x_{\bullet m\bullet})(\sum_{n \neq k} x_{\bullet\bullet n}) \\
&\quad\quad - (\sum_{\substack{m \neq j \\ n \neq k}} x_{imn}) x_{\bullet j\bullet} x_{\bullet\bullet k} \\
&= x_{\bullet j\bullet} \Delta^{ijk}_{ijl} + x_{\bullet\bullet k} \Delta^{ijk}_{ink} + \Delta^{ijk}_{inl} \qquad (32)
\end{aligned}$$

□

As shown in this formula, an alternative sum is very important concept for statistical independence of three variables. While the first two components can be viewed as an extension of two dimensional cases, the third component seems to be a little different from the others, because it contains the odds ratio of marginal distributions. We call this part a odds ratio of second order.

In summary, residuals of three dimensional data cube can be represented as:

(marginal distribution of two dimensions)
  × (residual for an odds ratio)
\+ (a specific cell)
  × (residual for an odds ratio of marginal distribution)

D. Information Granule for $2 \times 2 \times 2 \times 2$ Data Cube

In the similar way, we can obtain a formula for the residual of four-way tables.

Proposition 2: The residual σ_{1111} of $2 \times 2 \times 2 \times 2$ -data cube is obtained as:

$$\begin{aligned}
\sigma_{1111} x^3_{\bullet\bullet\bullet\bullet} &= x_{\bullet 1\bullet\bullet} x_{\bullet\bullet 1\bullet} (x_{1111} x_{\bullet\bullet\bullet 2} - x_{1112} x_{\bullet\bullet\bullet 1}) \\
&\quad + x_{\bullet 1\bullet\bullet} x_{\bullet\bullet\bullet 1} (x_{1111} x_{\bullet\bullet 2\bullet} - x_{1121} x_{\bullet\bullet 1\bullet}) \\
&\quad + x_{\bullet\bullet 1\bullet} x_{\bullet\bullet\bullet 1} (x_{1111} x_{\bullet 2\bullet\bullet} - x_{1211} x_{\bullet 1\bullet\bullet}) \\
&\quad + x_{\bullet 1\bullet\bullet} (x_{1111} x_{\bullet\bullet 2\bullet} x_{\bullet\bullet\bullet 2} \\
&\quad\quad - x_{1122} x_{\bullet\bullet 1\bullet} x_{\bullet\bullet\bullet 1}) \\
&\quad + x_{\bullet\bullet 1\bullet} (x_{1111} x_{\bullet 2\bullet\bullet} x_{\bullet\bullet\bullet 2} \\
&\quad\quad - x_{1212} x_{\bullet 1\bullet\bullet} x_{\bullet\bullet\bullet 1}) \\
&\quad + x_{\bullet\bullet\bullet 1} (x_{1111} x_{\bullet 2\bullet\bullet} x_{\bullet\bullet 2\bullet} \\
&\quad\quad - x_{1221} x_{\bullet 1\bullet\bullet} x_{\bullet\bullet 1\bullet}) \\
&\quad + x_{1111} x_{\bullet 2\bullet\bullet} x_{\bullet\bullet 2\bullet} x_{\bullet\bullet\bullet 2} \\
&\quad\quad - x_{1222} x_{\bullet 1\bullet\bullet} x_{\bullet\bullet 1\bullet} x_{w\bullet\bullet\bullet 1}
\end{aligned}$$

If we extend the notion of residual for odds ratio (Equation(30)), the above formula can be represented as:

$$\begin{aligned}
\sigma_{1111} x^3_{\bullet\bullet\bullet\bullet} &= x_{\bullet 1\bullet\bullet} x_{\bullet\bullet 1\bullet} \Delta^{1112}_{1111} \\
&\quad + x_{\bullet 1\bullet\bullet} x_{\bullet\bullet\bullet 1} \Delta^{1121}_{1111} + x_{\bullet\bullet 1\bullet} x_{\bullet\bullet\bullet 1} \Delta^{1211}_{1111} \\
&\quad + x_{\bullet 1\bullet\bullet} \Delta^{1122}_{1111} + x_{\bullet\bullet 1\bullet} \Delta^{1212}_{1111} + x_{\bullet\bullet\bullet 1} \Delta^{1221}_{1111} \\
&\quad + \Delta^{1222}_{1111}
\end{aligned}$$

□

This can be generalized into:

Theorem 9: The residual σ_{ijkl} of $l \times m \times n \times k$-data cube is obtained as:

$$\begin{aligned}\sigma_{ijkl}x^3_{\bullet\bullet\bullet\bullet} &= x_{\bullet j\bullet\bullet}x_{\bullet\bullet k\bullet}\Delta^{ijkl}_{ijkq} \\ &+x_{\bullet j\bullet\bullet}x_{\bullet\bullet\bullet l}\Delta^{ijkl}_{ijpl} \\ &+x_{\bullet\bullet k\bullet}x_{\bullet\bullet\bullet l}\Delta^{ijkl}_{inkl} \\ &+x_{\bullet j\bullet\bullet}\Delta^{ijkl}_{ijpq} + x_{\bullet\bullet k\bullet}\Delta^{ijkl}_{inkq} + x_{\bullet\bullet\bullet l}\Delta^{ijkl}_{inpl} \\ &+\Delta^{imnp}_{ijkl} \qquad (33)\end{aligned}$$

□

In summary, the residuals of four dimensional data cube can be represented as:

(marginal distribution) × (marginal distribution)
× (residual for an odds ratio)
+ (marginal distribution)
× (residual for an odds ratio × marginal distributions)
+ (a specific cell)
× (residual for an odds ratio of marginal distributions)

As shown in Equation (33), the residual decomposition has a combinatorial nature. The first part uses the second and third indices in the two-dimensional marginal sum and the fourth index for the residual for odds ratio. The second one use the second and fourth indices in the marginal sum and the third one for the residual.

Thus, let I denote an index set ($I = \{1, 2, 3, 4\}$) and i_I denote an attribute for a contingency table. Then, the equation (33) can be represented as:

$$\begin{aligned}\sigma_{i_1i_2i_3i_4}x^3_{\bullet\bullet\bullet\bullet} &= \sum_{\{m,n\}\in I-\{1\}} s(m)s(n)\Delta(I-\{1,m,n\}) \\ &+ \sum_{\{m\}\in I-\{1\}} s(m)\Delta(I-\{1,m\}) \\ &+\Delta(I-\{1\}), \qquad (34)\end{aligned}$$

where $s(m)$ denotes the one-dimensional marginal sum with fixed i_m and $\Delta(I - \{i_1, i_m, i_n\})$ denotes the residual of odds ratio in which i_1, i_m and i_n are fixed.

Or, the above formula can be rearranged as:

$$\begin{aligned}\sigma_{i_1i_2i_3i_4}x^3_{\bullet\bullet\bullet\bullet} &= \sum_{\{m\}\in I-\{1\}} \Delta(m) \prod_{n\in I-\{1,m\}} s(n) \\ &+ \sum_{\{m,n\}\in I-\{1\}} \Delta(m,n) \prod_{p\in I-\{1,m,n\}} s(p) \\ &+\Delta(I-\{1\}), \qquad (35)\end{aligned}$$

E. Information Granule for n-dimensional Data Cube

The above discussions gives an expectation that the residuals for n-dimensional data table ($n > 2$) should be represented as:

$$\sum_{k=1}^{n-2}(\text{marginal distribution})^k \times(\text{residual for an odd ratio})$$

Also, the relations of the number of " $\bullet$ " should be:

$$(\text{left side}) : n^{n-1} \rightarrow (\text{right side}) : (n-1)^{n-1}$$

It is notable that each component of n-dimensional residuals include $(n-1)^{n-1}$ $\bullet$s, which can be viewed as an assignment of $\bullet$s. Thus, the following theorem will be obtained:

Theorem 10: Let I denota an index set of n-dimensional data cube. The residual σ_I is obtained as:

$$\begin{aligned}\sigma_I x^{n-1}_{\bullet\bullet\bullet\bullet} &= \sum_{\{m\}\in I-\{1\}} \Delta(m) \prod_{n\in I-\{1,m\}} s(n) \\ &+ \sum_{\{m,n\}\in I-\{1\}} \Delta(m,n) \prod_{p\in I-\{1,m,n\}} s(p) \\ &+\cdots \\ &+ \sum_{\{m,n\}\in I-\{1\}} \Delta(I-\{1,m,n\})s(m)s(n) \\ &+ \sum_{\{m\}\in I-\{1\}} \Delta(I-\{1,m\})s(m) \\ &+\Delta(I-\{1\}). \qquad (36)\end{aligned}$$

□

Thus, the residual of n-dimensional data cube can be expanded into a series of the product of marginal sum and the residuals of odds ratio. Intuitively, high-dimensional residual will give a strong constraints for being zero, which indicates that high-dimensional statistical independence will be difficult to be achived.

It will be our future work to investigate the formal nature of this expansion.

VII. Conclusion

This paper focuses on statistical independence of three variables from the viewpoint of linear algebra. While information granules of statistical independence of two variables can be viewed as determinants of 2×2- submatrices, those of three variables consist of several combination s of subderminants of matrices generated from a data cube, although the formula is rather complicated. Thus, even in the case of three attributes, 2×2 submatrices play an important role in measuring the degree of statistical independence. However, compared with two dimension cases, the contribution of the concept of the derminants of submatrices are weaker.

It will be our future work to search for the corresponding determinants for $2 \times 2 \times 2$-data cube.

References

[1] S. Tsumoto and S. Hirano, "Statistical independence and determinants in a contingency table - interpretation of pearson residuals based on linear algebra -," *Fundam. Inform.*, vol. 90, no. 3, pp. 251–267, 2009.

[2] ——, "Contingency matrix theory ii: Degree of dependence as granularity," *Fundam. Inform.*, vol. 90, no. 4, pp. 427–442, 2009.

[3] S. Tsumoto, "Contingency matrix theory: Statistical dependence in a contingency table," *Inf. Sci.*, vol. 179, no. 11, pp. 1615–1627, 2009.

[4] H. Coxeter, Ed., *Projective Geometry*, 2nd ed. New York: Springer Verlag, 1987.

[5] S. Tsumoto and S. Hirano, "Meaning of pearson residuals - linear algebra view -," in *Proceedings of IEEE GrC 2007*. IEEE press, 2007.

[6] ——, "Statistical independence and contingency matrix," in *ICDM Workshops*. IEEE Computer Society, 2008, pp. 643–648.