Algebra in Combinatorics of Statistical Dependence

Shusaku Tsumoto and Shoji Hirano

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

Abstract.

This paper proposes homological analysis of statistical dependency graph. If a dependency graph model satisfy the condition of a chain complex, homological algebra can be applied. Especially, the degree of freedom can be viewed as a dual space of an original complex.

Keywords: Statistical Independence, Pearson Residual, Homology, Cohomology

1 Introduction Analysis of a contingency analysis has a long history, where χ^2-test play a central role in detecting statistical independence of two variables. The key idea of χ^2-test is a degree of freedom of χ^2-test statistics, the number of independent cells in a given table. If we assume that the marginal distributions of a column and a row are fixed, all the numbers in the cell will be determined by the values of independent cells. For example, since the degree of freedom of a 2×2 contingency table is equal to 1, if one cell is given, other three cells will be obtained under the given marginal distribution.

One interesting observation is that the formula of chi-square test statistics of a 2×2 contingency table includes the form of a determinant when a table is regarded as a 2×2 matrix. Tsumoto focuses on this observation and finds the interesting relations between linear algebra and statistical independence. from the viewpoint of granular computing.
The important result is that a degree of freedom is equal to the number of 2×2 submatrices in a contingency table, which can be viewed as a granule of statistical independence. Interestingly, the results are generalized into mulivariate contingency tables, where combinatorics of independent variables is important to determine the degree of freedom. Furthermore, symmetry of dependent variables gives classificaiton of a contingency table. This paper gives further extension of this analysis, which shows that the degree of freedom corresponds to the number of outer products of dependent variables, which shows that the degree of freedom will give a dual space of statistical dependency graph when a graphical model satisfies the condition of a chain complex.

The paper is organized as follows: Section **2** gives the results of previous studies on Pearson residuals. Section **3** gives some mathematical discussions on geometrical and combinatorial structure of the above theory. Section **4** introduces homological algera as a tool for the analysis of statistical dependence. Section **5** discusses correspondence between boundary and coboundary operators and table operations. Finally, Section **6** concludes this paper.

2 Combinartorics of Peasrson Residuals

2.1 Multiway Contingency Table

Definition 2.1 *Let $R_1, R_2, \cdots, R_n$ denote $n(\in N)$ multinominal attributes in an attribute space A which have $m_1, m_2, \cdots, m_n$ values Let $|R_j = A_{j_i}|$ denote the set of data whose jth-attribute is equal to A_{j_i} (ith-partition of j). Then, an element of a multiway contigency table, which has n attributes, is defined as:*

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

where their marginal sums are not included as elements. □

For example, in the two dimensional case, this table is arranged into the form shown in Table **??**, 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 1: 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	$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot n}$	$x_{\cdot\cdot} = \lvert U \rvert = N$

Let us denote the sum over one attribute a contingency table by "$\bullet$". Then, marginal sums over one attribute is defined as follows.

Definition 2.2 *Let a contigency table have m attributes. The marginal sum over $i_k (1 \leq k \leq m)$ is:*

$$x_{i_1 \cdots i_{k-1} \bullet i_k \cdots i_m} = \sum_{j=1}^{q_k} x_{i_1 \cdots i_{k-1} i_j i_{k+1} \cdots i_m}, \tag{1}$$

where q_k is the number of equivalence classes of i_k. □

Then, marginal sums over all the attributes is equal to the sample size:

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

2.2 Information Granule in a Contingency Table

2.2.1 Pearson Residual

Definition 2.3 *Pearson residual of the cell $i_1 \cdots i_m (m \geq 2)$, denoted by $\sigma_{i_1 \cdots i_m}$, is defined as the difference between the observed value $x_{i_1 \cdots i_m}$ and its expected value:*

$$\begin{aligned} \sigma_{i_1 \cdots i_m} =& x_{i_1 \cdots i_m} \\ & - \frac{x_{i_1 \bullet \cdots \bullet} \times x_{\bullet i_2 \cdots \bullet} \cdots \times x_{\bullet \bullet \cdots i_m}}{x_{\bullet\bullet\bullet}^{m-1}}. \end{aligned} \tag{2}$$

□

"Partial residuals" in which one of three attributes are summarized (marginalized) are defined as follows:

$$\begin{aligned} \sigma_{\bullet i_2 \cdots i_m} &= x_{\bullet i_2 \cdots i_m} \\ &\quad - \frac{x_{\bullet i_2 \cdots \bullet} \times x_{\bullet \bullet i_3 \cdots \bullet} \cdots \times x_{\bullet \bullet \cdots i_m}}{x_{\bullet \bullet \bullet}^{m-2}}. \end{aligned} \tag{3}$$

Therefore, we obain the following theorem:

Theorem 2.1 *The Pearson residual of a m-way contingency table is reformulated as:*

$$\begin{aligned} \sigma_{i_1 \cdots i_m} &= \frac{x_{i_1 \bullet \cdots \bullet}}{x_{\bullet \bullet \cdots \bullet}} \sigma_{\bullet i_2 \cdots i_m} \\ &\quad + \frac{1}{x_{\bullet \bullet \cdots \bullet}} (x_{i_1 \cdots i_m} x_{\bullet \bullet \cdots \bullet} - x_{i_1 \bullet \cdots \bullet} x_{\bullet i_2 \cdots i_m}) \end{aligned} \tag{4}$$

□

In the subsequent sections, the second part of Equation, $x_{i_1 \cdots i_m} x_{\bullet \bullet \cdots \bullet} - x_{i_1 \bullet \cdots \bullet} x_{\bullet i_2 \cdots i_m}$ is denoted by $\sigma^{i_1}_{i_2 \cdots i_m}$. More detailed examples are shown in [ ].

2.3 Degree of Freedom

2.3.1 Formula of Degree of Freedom From Equation,

$$\begin{aligned} \sigma_{i_1 \cdots i_m} &= \frac{x_{i_1 \bullet \cdots \bullet}}{x_{\bullet \bullet \cdots \bullet}} \sigma_{\bullet i_2 \cdots i_m} \\ &\quad + \frac{1}{x_{\bullet \bullet \cdots \bullet}} (x_{i_1 \cdots i_m} x_{\bullet \bullet \cdots \bullet} - x_{i_1 \bullet \cdots \bullet} x_{\bullet i_2 \cdots i_m}) \end{aligned}$$

Although the first part includes the same number of the determinants as $\sigma_{i_2 \cdot 1_m}$ multiplied by the size of the first attribute, the weight:

$$\frac{x_{i_1 \bullet \cdots \bullet}}{x_{\bullet \bullet \cdots \bullet}} = \frac{1}{size \text{ of } the\ first\ attribute}$$

should be considered for estimation of the degree of freedom. In other words, the number of the subdeterminants can be estimated as the number of the subdeterminants of $(m-1)$-way contingency table. On the other hand, the second part is equal to:

$$\begin{aligned} \sigma^{i_1}_{i_2 \cdots i_m} &= x_{i_1 \cdots i_m} x_{\bullet \bullet \cdots \bullet} - x_{i_1 \bullet \cdots \bullet} x_{\bullet i_2 \cdots i_m} \\ &= \sum_{j_1 \neq i_1} (x_{i_1 \cdots i_m} x_{j_1 \bullet \cdots \bullet} - x_{i_1 \bullet \cdots \bullet} x_{j_1 i_2 \cdots i_m}) \end{aligned} \tag{5}$$

If the other terms $i_k (k = 2, \cdots, m)$ are not equal to j_k, the subdeterminant is not equal to 0 and the subscript of the summation of Equation is equivalent to:

$$\bigvee_{k=2}^{m} (i_k \neq j_k),$$

Therefore, the following theorem is obtained:

Theorem 2.2 *Let IND_p^m denote a set of index which is a p out of m attributes. Then, the total number of determinants of 2×2 submatrices in a residual of a m-way contingency table is given by:*

$$\zeta(1,2,\cdots,m) = \sum_{p=2}^{m} \prod_{t \in IND_p^m} (n_t - 1), \tag{6}$$

where n_t denote the number of partitions of an attribute t. □

Corollary 2.3 *The total numbers of determinants of 2×2 submatrices in m-way contingency table, denoted by $\zeta(1,2,\cdot,m)$ are equal to*

$$\zeta(1,2,\cdots,m) = n_1 n_2 \cdots n_m - \sum_{i=1}^{m} n_i + (m-1).$$

□

In this way, the degree of freedom summarizes information on combinatorial nature of Pearson residuals:

Corollary 2.4 *Let IND_p^m denote a set of index which is a p out of m attributes The degree of freedom of a m-way contingency table is given by:*

$$\zeta(1,2,\cdots,m) = \sum_{p=2}^{m} \prod_{t \in IND_p^m} (n_t - 1),$$

where n_t is the number of partitions in an attribute t, if all the variables are assumed to be independent. □

When some of attributes are dependent, the corresponding term will be eliminated. For example, m_2 and m_3 of three-way attribute is dependent, but m_1 is conditionally independent of this pair:

$$\begin{aligned}
\zeta(1,2,3) &= (n_1-1)(n_2-1)(n_3-1) \\
&\quad +(n_3-1)(n_1-1) \\
&\quad +(n_1-1)(n_2-1) \\
&= n_1 n_2 n_3 - n_2 n_3 - n_1 + 2,
\end{aligned} \tag{7}$$

which is the same formula given in [ ].

3 Symmetry in Pearson Residuals

3.1 Determinants as Pencil of Lines

Equation shows that 2×2-subdeterminants are information granules of statistical independence. Since a 2×2-subdeterminant gives a line in a projective plane in classic projective geometry, a set of the subdetermiants can be viewed a *pencil of lines* in a space of projective plane whose coordinates are given as a cell in a given contingency table: dependence and independence can be captured as geometrical structure of a pencil in a projective space. Thus, dependent or independent relations of a multivariate table gives complex geometrical structure, which suggests that a tool of algebra can be applied to analysis

3.2 Symmetry Let $a_1, a_2, \cdots, a_m$ denote m attributes in a m-way contingency table and e_i be equal to $n_i - 1$ where n_i is a number of partition in attribute a_i. Then, $e_j e_k$ gives the number of 2×2subdeterminants when a_j and a_k are dependent. In general, $e_{j_1} e_{j_2} \cdots e_{j_l}$ gives the number of $l - way$ subdeterminants when these l attributes are dependent.

Let $E^l_{12\cdots m}$ denote a set of $e_{j_1} e_{j_2} \cdots e_{j_l}$, each element of which is a product of l selected from m attributes. For each l dimension, a polynomial symmetric over S_m can be derived as:

$$s^l_m = \sum_{E_{12\cdots m}} e_{j_1} e_{j_2} \cdots e_{j_l},$$

Then, a polynomial symmetric over S_m is represented as:

$$\bigoplus_{k=2}^{l} s^k_m = \bigoplus_{k=2}^{l} \sum_{E_{12\cdots m}} e_{j_1} e_{j_2} \cdots e_{j_l} \tag{8}$$

Then, relations between symmetric group and geometrical structure can be discussed. For example, since conditional dependence is defined as statistical dependence of a set of variables, denoted by V_1 under the assumptions where the values of the set of other variables (V_2) are fixed, the symmetry of V_2 will be lost, that is , "breakdown of symmetry".

Theorem 3.1 *Let $a_1, a_2, \cdots, a_m$ denote m attributes in a m-way contingency table and e_i be equal to $n_i - 1$ where n_i is a number of partition in attribute a_i. A formula $e_{j_1} e_{j_2} \cdots e_{j_l}$ is equal to the number of $l - way$ subdeterminants when these l attributes are dependent. Let $E^l_{12\cdots m}$ denote a set of $e_{j_1} e_{j_2} \cdots e_{j_l}$, each element of which is a product of l selected from m attributes. Then, a polynomial symmetric over S_m is given as:*

$$\bigoplus_{k=2}^{l} s^k_m = \bigoplus_{k=2}^{l} \sum_{E_{12\cdots m}} e_{j_1} e_{j_2} \cdots e_{j_l}. \tag{9}$$

□

Then,

Corollary 3.2 *A model with partial independence can be derived by removal of attributes whose values a fixed. For example,*

$$\bigoplus_{k=2}^{l} s^k_m - \sum_{k=2}^{m} \sum_{T^m_k} e_{j_1} e_{j_2} \cdot e_{j_k} \tag{10}$$

gives the number of subdeterminants where T^m_k gives a set of k-pair statistical independent attributes out of m attributes. □

Tsumoto shows that structure of symmetric group will give some global information on statistical dependency model in a multivariate contingency table. However, this tool focuses on interchangeability of dependence relations among variables. In order to investigate other properties of geometrical structure, other tool is needed as shown in the next section.

4 Degree of Freedom and Homological Calculus Equation is a little complicated and it is difficult to see the meaning except for the action of symmetric group. The most import problem is that we have to eliminate independence variables explicitly, which makes the representation power weak. The other point is that we may have a situation when three variables are statistical dependent (independent) although all the combinations of two of three variables are statistical independent (dependent), which analysis based on symmetric group cannot capture. Thus, by investigating the nature of statistical dependence much further, new representation should be explored.

4.1 Main Ideas The main idea is that geometrical structure of a statistical dependence model corresponds to its degree of freedom of a contingency table as shown in Figure 1. Let us assume three variables provided in a contingency table. In the case of one dimension, the degree of freedom of one attribute will be the number of its partition minus 1 if the number of examples is fixed. If we add one more dependent attribute, the degree of freedom is equal to $(n_1-1)(n_2-1)$, where n_1 and n_2 denote the number of partitions of the first and second attribute. In the same way, dependency graph of three attribute has $(n_1-1)(n_2-1)(n_3-1)$ as its degree of freedom. If we consider dependency of three attributes with full dependency, the degree of freedom is equal to:

$$\begin{aligned}(n_1-1)(n_2-1)(n_3-1) \quad &+(n_1-1)(n_2-1)\\ &+(n_2-1)(n_3-1)\\ &+(n_3-1)(n_1-1).\end{aligned}$$

The main idea is that decomposition of a dependency graph gives a homological sequence shown in Figure 1. Formal definition of mappings will be given in subsequent subsections.

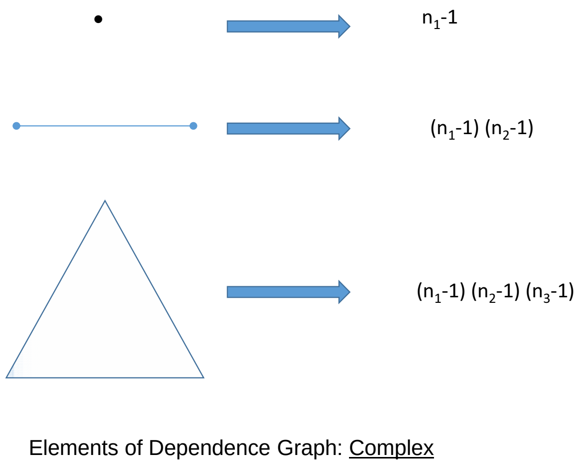

n_1-1
(n_1-1) (n_2-1)
(n_1-1) (n_2-1) (n_3-1)
Elements of Dependence Graph: Complex

Figure 1: Correspondence between Dependency Graph and its Degree of Freedom

4.2 Basic Definition The key components of the above section are the degree of freedom of each attribute. When two attributes are dependent, the degree of freedom is obtained by their product of the degree of freedom. Furthermore, the product is invariant over permutation.

Definition 4.1 *Let $A = \{a_1, a_2, \cdots, a_m\}$ denote a set of m attributes in a m-way contingency table and e_i be equal to $n_i - 1$ where n_i is a number of partition in attribute a_i. Then, a linear sum of m attributes gives a vector space spanned by A:*

$$vec(A) = \sum_{i=1}^{m} k_i a_i,$$

where $k_i \in \mathbb{Z}$. Thus, $vec(A)$ can be viewed as $\mathbb{Z}$-module. $\square$

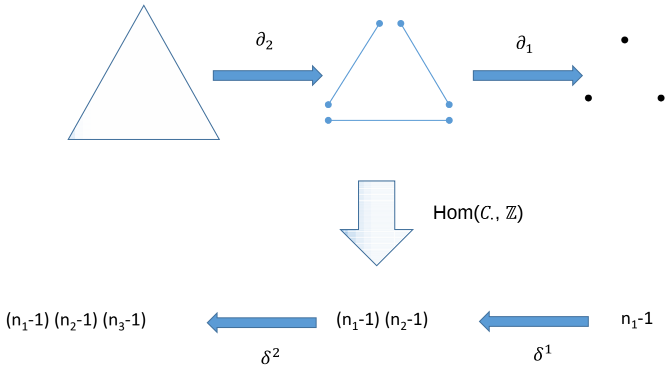


Figure 2: Homological Sequence of Three Attributes

As a different type of operation, we define $a_1 \wedge a_2$ as a matrix generated by a_1 and a_2:

$$a_1 \wedge a_2 = Mat(a_1, a_2).$$

It is notable that a_1 and a_2 are reversibly obtained by $Mat(a_1, a_2)$ as marginal sums. For example, when a matrix is given as:

$$\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix},$$

corresponding two vectors are: $a_1 = (2+2, 2+3) = (4,4)$ and $a_2 = (2+1, 2+3) = (3,5)$.

It is notable that this transformation is linear and can be represented as a matrix. In the above example, the transformation is given by:

$$\begin{pmatrix} 4 & 3 \\ 4 & 5 \end{pmatrix} = X \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix},$$

In this case, $\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$ has a reverse, so X is obtained as:

$$X = \frac{1}{4} \begin{pmatrix} 9 & -2 \\ 7 & 2 \end{pmatrix}$$

However, if the determinant of original table is equal to 0, the situation is much more complex: we should use the generalized inverse for calculation. But, since this case is corresponding to statistical independence, let say $Mat(\bullet \wedge \bullet) = 0$. That is, for two attributes a_i, a_j $(i, j = 1, \cdots, m)$,

$$a_i \wedge a_j = \begin{cases} Mat(a_i, a_j) & det(a_i, a_j) \neq 0 \\ 0 & det(a_i, a_j) = 0, \end{cases} \tag{11}$$

where $det(a_i, a_j)$ gives a determinant of a matrix (a_i, a_j) and Mat is a corresponding matrix operation. Thus, although there are many ways to make a matrix from a_1 and a_2, for a given table, a matrix corresponds to one function which partitions a_1 and a_2 as shown the above. Thus, a matrix can be viewed as a partition function of $\mathbb{Z} \times \mathbb{Z}$.

Since a vector and matrix can be viewed as a specific form of tensor, the above discussion can be discussed in the context of tensor calculus: it is easy to see that this framework satisfies the axiom of tensor space. Moreover, a space have elements of different grades, since a_1 and $a_1 \wedge a_2$ are the first and second grade, respectively.

4.3 Degree of Freedom as a Mapping Since the degree of freedom of a_i is the number of elementary partition of a_i minus 1, it can be viewed as a function of a_i:

$$df(a_i) = e_i.$$

It can be rewritten as:

$$df = Hom(a_i, \mathbb{Z}),$$

which is a homomorphism under addition. In the above example of matrix, $df(a_1 \wedge a_2) = (2-1)*(2-1) = 1$ and $df(a_1) = df(a_2) = 1-1 = 0$ because the second element of a_1 and a_2 can be described as a linear sum of the first element: $4 = 1 \times 4$, $5 = \frac{5}{3} \times 3$.

It is notable that df will give a dual space of a_i. Let us denote $df(a_i)$ by a_i^*. Then, it is easy to show that $A^* = df(A) = \{df(a_1), df(a_2), \cdots, df(a_m)\}$ gives a dual (tensor) space of $A = a_1, a_2, \cdots, a_m$.

Thus, dependence and independence can be easily described as an outer product, alternating tensor product of $df(a_i)$ $(i = 1, 2, \cdots, m)$. Since the calculation of the degree of freedom starts from two attributes as a matrix calculus, let us select two attributes first. If we take two dependent attributes a_i and a_j, then $a_i \wedge a_j$ gives:

$$a_i \wedge a_j = det(M(a_i, a_j))e_i \wedge e_j,$$

where a rectangular matrix $M(a_i, a_j)$ is generated by $\{a_i, a_j\}$ and e_i and e_j denote the orthonormal basis generated by a_i and a_j. Here, the determinant is given by Cullin's determinant, which is an extension of ordinary matrix. Then, when two attributes a_1 and a_2 are independent, since the rank of matrix is equal to 1, $M(a_i, a_j)$ is represented as (v_i, kv_i).

Thus,

$$df(a_1 \wedge a_2) \quad = df(a_1 \wedge ka_1) \quad = 0,$$

where $k \in Z$. On the other hand, if both are dependent:

$$df(a_1 \wedge a_2) = df(a_1)df(a_2) = (n_1 - 1) * (n_2 - 1),$$

where n_1- and $n_2 - 1$ denote the degree of freedom of a_1 and a_2.

When we take 3 attributes, we can append this attribute as $a_1 \wedge a_2 \wedge a_3$. Then, full dependence can be described as:

$$a_1 \wedge a_2 \wedge a_3 + a_1 \wedge a_2 + a_2 \wedge a_3 + a_3 \wedge a_1.$$

The formula shown in Equation is given by:

$$\begin{aligned} df(a_1 \wedge a_2 \wedge a_3) &= df(a_1 \wedge a_2 \wedge a_3 + a_1 \wedge a_2 \\ &\quad +a_2 \wedge a_3 + a_3 \wedge a_1) \\ &= df(a_1 \wedge a_2 \wedge a_3) + df(a_1 \wedge a_2) \\ &\quad +df(a_2 \wedge a_3) + df(a_3 \wedge a_1) \\ &= (n_1 - 1)(n_2 - 1)(n_3 - 1) \\ &\quad +(n_1 - 1)(n_2 - 1) \\ &\quad +(n_2 - 1)(n_3 - 1) \\ &\quad +(n_3 - 1)(n_1 - 1) \end{aligned}$$

If a_1 and a_2 is independent, since $df(a_1 \wedge a_2) = 0$, we should remove this term from the above equation.

Since the $df(a_i)$ can ve viewed as a dual vector, we can rewrite the above equation as:

$$\begin{aligned}
a_1^* \wedge a_2^* \wedge a_3^* &= a_1^* \wedge a_2^* \wedge a_3^* + a_1^* \wedge a_2^* + a_2^* \wedge a_3^* \\
&\quad + a_3 \wedge a_1^* \\
&= a_1^* \wedge a_2^* \wedge a_3^* + a_1^* \wedge a_2^* + a_2^* \wedge a_3^* \\
&\quad + a_3^* \wedge a_1^* \\
&= (n_1 - 1)(n_2 - 1)(n_3 - 1) \\
&\quad + (n_1 - 1)(n_2 - 1) \\
&\quad + (n_2 - 1)(n_3 - 1) \\
&\quad + (n_3 - 1)(n_1 - 1)
\end{aligned}$$

By using these ideas, Theorem can be reformulated as follows.

Theorem 4.1 *Let $A = \{a_1, a_2, \cdots, a_m\}$ denote a set of m attributes in a m-way contingency table and e_i be equal to $n_i - 1$ where n_i is a number of partition in attribute a_i. Let DEP^m denote a set of dependent variables of A. Then, since the correspondence between a_i and e_i is given as a function: $f(a_i) = e_i$, dual tensor space can be defined by $df(A)$, which can be denoted by $A^* = \{a_1^*, a_2^*, \cdots, a_m^*\}$ Then, a polynomial symmetric over S_m is represented as:*

$$\bigoplus_{k=2}^{l} s_m^k = \bigoplus_{k=2}^{l} \sum_{DEP^k} (-1)^{\sigma(j)} a_{j_1}^* \wedge a_{j_2}^* \wedge \cdots \wedge a_{j_k}^*, \tag{12}$$

where $\sigma(j)$ denotes the the number of substitutions over $j_1, j_2, \cdots j_k$. □

4.4 Chain Complex Let A_n^* denote a space spanned by a set of dual of outer product of n attributes. Since this space is an Abelian group, we can consider a sequence $A_n, A_{n-1}, \cdots, A_2, A_1$. Let us define the boundary map from A_n^* to A_{n-1}^* as:

$$\partial^n : A_n^* \to A_{n-1}^* : a_1^* \wedge a_2^* \cdots \wedge a_n^* \mapsto \sum_{i=1}^{n} (-1)^{i-1} \sigma([a_1^* \wedge a_2^* \wedge \ldots \hat{a}_i^* \ldots \wedge a_n^*]), \tag{13}$$

where the hat denotes the omission of an attribute. Then, the following theorem is obtained.

Theorem 4.2

$$\partial^n \partial^{n-1} A_n^* = 0$$

Thus, $(A., \partial.)$ is a chain complex. Proof

$$\begin{aligned}
\partial_n \partial_{n-1} A_n^* &= \partial\partial a_1^* \wedge a_2^* \cdots \wedge a_n^* \\
&= \partial \left(\sum_{i=1}^{n} (-1)^{i-1} a_1^* \ldots \wedge \hat{a}_i^* \ldots \wedge a_n^* \right) \\
&= \sum_{j=1}^{n} (-1)^{j-1} \left(\sum_{i=1}^{n} (-1)^{i-1} a_1^* \ldots \wedge \hat{a}_i^* \right. \\
&\quad \left. \wedge \ldots \wedge \hat{a}_j^* \ldots \wedge a_n^* \right) \\
&= \sum_{i,j=1}^{n} \left(a_1^* \ldots \wedge \hat{a}_i^* \ldots \wedge \hat{a}_j^* \ldots \wedge a_n^* \right. \\
&\quad \left. - a_1^* \ldots \wedge \hat{a}_i^* \ldots \wedge \hat{a}_j^* \ldots \wedge a_n^* \right) \\
&= 0
\end{aligned}$$

□

Example 4.1

$$\begin{aligned}
\partial^3 \partial^2 (a_1^* \wedge a_2^* \wedge a_3^*) &= \partial\partial (n_1 - 1)(n_2 - 1)(n_3 - 1) \\
&= \partial \left((n_2 - 1)(n_3 - 1) \right. \\
&\quad -(n_1 - 1)(n_3 - 1) \\
&\quad \left. +(n_1 - 1)(n_2 - 1) \right) \\
&= (n_2 - 1) - (n_3 - 1) - (n_1 - 1) \\
&\quad +(n_3 - 1) + (n_1 - 1) - (n_2 - 1) \\
&\quad = 0
\end{aligned}$$

Since it is easy to see that the outer space of a_i gives a chain complex, the dual space a_i^* is its dual space. Thus, homology of a sequence $H_n(A)$can be considered:

$$H_n(A) = \ker \partial_n / \mathrm{im} \partial_{n+1}.$$

$$H^n(A) = \ker \partial^{n+1} / \mathrm{im} \partial^n$$

Example 4.2 *In the above example,*

$$\begin{aligned}
C_2 &= \{(n_1 - 1)(n_2 - 1)(n_3 - 1)\} \\
C_1 &= \{(n_1 - 1)(n_2 - 1), (n_2 - 1)(n_3 - 1), \\
&\quad (n_1 - 1)(n_3 - 1)\} \\
C_0 &= \{(n_1 - 1), (n_2 - 1), (n_3 - 1)\}
\end{aligned}$$

$$\begin{aligned}
im\partial^2(A) = &\quad (n_2 - 1)(n_3 - 1) \\
&-(n_1 - 1)(n_3 - 1) + (n_1 - 1)(n_2 - 1)
\end{aligned}$$

Then, $\partial^2(A)$, $\partial^1(A)$, and $\partial^0(A)$ can be represented in a matrix form shown as below.

$$\begin{aligned}
&\partial_2(n_1-1)(n_2-1)(n_3-1)\\
&\quad=\begin{pmatrix} 1 & -1\,1 \end{pmatrix}\begin{pmatrix}(n_1-1)(n_2-1)\\(n_2-1)(n_3-1)\\(n_1-1)(n_3-1)\end{pmatrix}\\
&\partial_1\begin{pmatrix}(n_1-1)(n_2-1)\\(n_2-1)(n_3-1)\\(n_1-1)(n_3-1)\end{pmatrix}\\
&\quad=\begin{pmatrix}-1 & 1 & 0\\ 0 & -1 & 1\\ -1 & 0 & 1\end{pmatrix}\begin{pmatrix}(n_1-1)\\(n_2-1)\\(n_3-1)\end{pmatrix}\\
&\partial_0\begin{pmatrix}(n_1-1)\\(n_2-1)\\(n_3-1)\end{pmatrix}\\
&\quad=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}
\end{aligned}\tag{14}$$

Since the rank of each matrix gives the dimension of ∂, the difference between matrix size and its rank is equivalent of the dimension of $\ker\partial$ *From these equations, we obtained the ranks of im and* $\ker$ *as follows.*

	Im	*Ker*
∂^3	*0*	*0*
∂^2	*1*	*0*
∂^1	*2*	*1*
∂^0	*0*	*3*

Thus, H^0, H^1 and H^2 is obtained as follows.

$$\begin{aligned}
H^2 &= \frac{0}{1}=0\\
H^1 &= \frac{0}{\mathbb{Z}\oplus\mathbb{Z}}=0\\
H^0 &= \frac{\mathbb{Z}}{0}=\mathbb{Z}
\end{aligned}$$

Thus, the cohomological sequence is obtained as follow.

$$0\rightarrow\mathbb{Z}\rightarrow 0\rightarrow 0$$

In a dual way, homological sequence of the outer product of dependency relations can be obtained as follows: since ∂_n is given as the transpose of ∂^n in a matrix representation, the *table of rank of im and* ker *is equivalent. Thus,* H_0*,* H_1 *and* H_2 *is obtained as follows.*

$$\begin{aligned} H_2 &= \frac{0}{0} = 0 \\ H_1 &= \frac{\mathbb{Z}}{\mathbb{Z}} = 0 \\ H_0 &= \frac{\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}}{\mathbb{Z} \oplus \mathbb{Z}} = \mathbb{Z} \end{aligned}$$

Thus, the homological sequence is obtained as follow.

$$0 \to 0 \to \mathbb{Z} \to 0$$

□

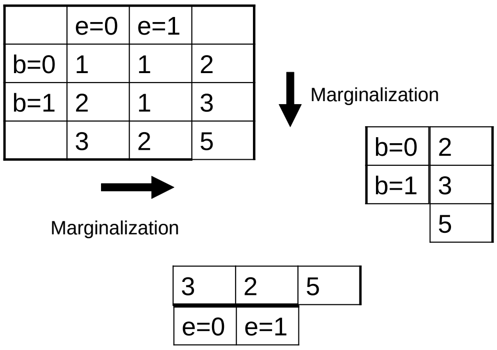
In the same way, both the homological and cohomology sequences for the model with only two variables dependent are:

$$0 \to \mathbb{Z} \to \mathbb{Z} \to 0$$

which corresponds to the homology of a corresponding dependency graph.

5 Discussion Figure 3 illustrates how boundary and coboundary operators are used in the context of contingency table analysis. Boundary operators will reduce the degree of freedom, which corresponds to marginalization as shown in Figure 3. On the other hand, coboundary operators corresponds to partition as shown in Figure 4.

Although boundary and coboundary operators are dual to each other, corresponding operations show that although boundary is one choice, but coboundary may give many possible ways. In other words, partition or coboundary suffers from combinatorial problems. Thus, cohomological analysis may give insights to formal discussions on partition.



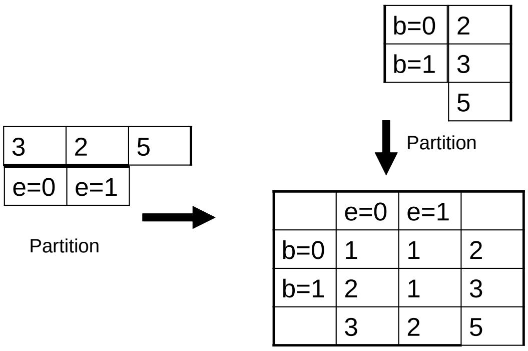
	e=0	e=1	
b=0	1	1	2
b=1	2	1	3
	3	2	5

b=0	2
b=1	3
	5

3	2	5
e=0	e=1	

Figure 3: Marginalization as Boundary Operator

6 Conclusion This paper focuses on the formula of degree of freedom and investigate its nature. First, if we assume that a dependency graph satisfies the condition of a complex, a boundary operator ∂ for the formula of degree of freedom can be defined and the duplicated operation will be 0: $\partial\partial = 0$, which leads to the basic step to homological algebra. Second, the formula can be viewed as a homomorphism from structure to integer, denoted by $Hom(Structure, \mathbb{Z})$, thus the hierarchy of the formula of degree of freedom,



b=0	2
b=1	3
	5

3	2	5
e=0	e=1	

	e=0	e=1	
b=0	1	1	2
b=1	2	1	3
	3	2	5

Figure 4: Partition as Coboundary Operator

which corresponds to the hierarchy of the dependency graph, generates a cocomplex. Thus, cohomology of the formula can be considered. By using this framework, the complex nature of dependency graph is translated into the algebraic structure of (co-)homological sequence, and (co)homology groups characterize the dependency graph. Thus, several tools in homological algebra can be applied to analysis of statistical independence

This study is a preliminary step of the analysis of statistical (in)dependence based on homological algebra. It will be our future work to investigate further the property a of contingency table from the viewpoint of algebra.

Acknowledgment This research is supported by Grant-in-Aid for Scientific Research (B) 24300058 from Japan Society for the Promotion of Science(JSPS).