Contingency Matrix Theory
– Investigation of Information Granules in Statistics –

Shusaku Tsumoto and Shoji Hirano
Department of MedicalInformatics, Shimane University
89-1 Enya-cho, Izumo 693-8501 Japan
{tsumoto,hirano}@med.shimane-u.ac.jp

Abstract

This paper focuses on how statistical independence can be observed in a contingency table when the table is viewed as a matrix. Statistical independence in a contingency table is represented as a special form of linear dependence, where all the rows or columns are described by one row or column, respectively.

Keywords: Statistical Independence, Contingency Table, Matrix Theory.

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.

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.

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1765–1772
Torremolinos (Málaga), June 22–27, 2008

2 Contingency Table from Rough Sets

2.1 Rough Sets Notations

In the subsequent sections, the following notations is adopted, which is introduced in [2]. Let U denote a nonempty, finite set called the universe and A denote a nonempty, finite set of attributes, i.e., $a : U \to V_a$ for $a \in A$, where V_a is called the domain of a, respectively. Then, a decision table is defined as an information system, $A = (U, A \cup \{\mathcal{D}\})$, where $\{\mathcal{D}\}$ is a set of given decision attributes. The atomic formulas over $B \subseteq A \cup \{\mathcal{D}\}$ and V are expressions of the form $[a = v]$, called descriptors over B, where $a \in B$ and $v \in V_a$. The set $F(B, V)$ of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation. For each $f \in F(B, V)$, f_A denote the meaning of f in A, i.e., the set of all objects in U with property f, defined inductively as follows.

1. If f is of the form $[a = v]$ then, $f_A = \{s \in U | a(s) = v\}$
2. $(f \wedge g)_A = f_A \cap g_A$; $(f \vee g)_A = f_A \vee g_A$; $(\neg f)_A = U - f_a$

By using this framework, classification accuracy and coverage, or true positive rate is defined as follows.

Definition 1
Let R and D denote a formula in $F(B, V)$ and a set of objects whose decision attribute is given as $\lceil$, respectively. Classification accuracy and coverage(true positive rate) for $R \to \mathcal{D}$ is defined as:

$$\alpha_R(D) = \frac{|R_A \cap D|}{|R_A|}(= P(D|R)), \text{ and}$$
$$\kappa_R(D) = \frac{|R_A \cap D|}{|D|}(= P(R|D)),$$

where $|A|$ denotes the cardinality of a set A, $\alpha_R(D)$ denotes a classification accuracy of R as to classification of $\mathcal{D}$, and $\kappa_R(D)$ denotes a coverage, or a true positive rate of R to $\mathcal{D}$, respectively.

2.2 Two-way Contingency Table

From the viewpoint of information systems, a contingency table summarizes the relation between two attributes with respect to frequencies. This viewpoint has already been discussed in [3, 4]. However, this study focuses on more statistical interpretation of this table.

Definition 2 *Let R_1 and R_2 denote binary attributes in an attribute space A. A contingency table is a table of a set of the meaning of the following formulas: $|[R_1 = 0]_A|$,$|[R_1 = 1]_A|$, $|[R_2 = 0]_A|$,$|[R_1 = 1]_A|$, $|[R_1 = 0 \wedge R_2 = 0]_A|$,$|[R_1 = 0 \wedge R_2 = 1]_A|$, $|[R_1 = 1 \wedge R_2 = 0]_A|$,$|[R_1 = 1 \wedge R_2 = 1]_A|$, $|[R_1 = 0 \vee R_1 = 1]_A|(= |U|)$. This table is arranged into the form shown in Table 1, where: $|[R_1 = 0]_A| = x_{11} + x_{21} = x_{\cdot 1}$, $|[R_1 = 1]_A| = x_{12} + x_{22} = x_{\cdot 2}$, $|[R_2 = 0]_A| = x_{11} + x_{12} = x_{1\cdot}$, $|[R_2 = 1]_A| = x_{21} + x_{22} = x_{2\cdot}$, $|[R_1 = 0 \wedge R_2 = 0]_A| = x_{11}$, $|[R_1 = 0 \wedge R_2 = 1]_A| = x_{21}$, $|[R_1 = 1 \wedge R_2 = 0]_A| = x_{12}$, $|[R_1 = 1 \wedge R_2 = 1]_A| = x_{22}$, $|[R_1 = 0 \vee R_1 = 1]_A| = x_{\cdot 1} + x_{\cdot 2} = x_{\cdot\cdot}(= |U|)$.*

Table 1: Contingency Table (2 × 2

	$R_1 = 0$	$R_1 = 1$	
$R_2 = 0$	x_{11}	x_{12}	$x_{1\cdot}$
$R_2 = 1$	x_{21}	x_{22}	$x_{2\cdot}$
	$x_{\cdot 1}$	$x_{\cdot 2}$	$x_{\cdot\cdot}$ $(= \|U\| = N)$

From this table, accuracy and coverage for $[R_1 = 0] \to [R_2 = 0]$ are defined as:

$$\alpha_{[R_1=0]}([R_2 = 0]) = \frac{|[R_1 = 0 \wedge R_2 = 0]_A|}{|[R_1 = 0]_A|} = \frac{x_{11}}{x_{\cdot 1}},$$
$$\text{and}$$
$$\kappa_{[R_1=0]}([R_2 = 0]) = \frac{|[R_1 = 0 \wedge R_2 = 0]_A|}{|[R_2 = 0]_A|} = \frac{x_{11}}{x_{1\cdot}}.$$

2.3 Contingency Table($m \times n$)

Two-way contingency table can be extended into a contingency table for multinominal attributes.

Definition 3 *Let R_1 and R_2 denote multinominal attributes in an attribute space A which have m and n values. A contingency tables is a table of a set of the meaning of the following formulas: $|[R_1 = A_j]_A|$, $|[R_2 = B_i]_A|$, $|[R_1 = A_j \wedge R_2 = B_i]_A|$, $|[R_1 = A_1 \wedge R_1 = A_2 \wedge \cdots \wedge R_1 = A_m]_A|$, $|[R_2 = B_1 \wedge R_2 = A_2 \wedge \cdots \wedge R_2 = A_n]_A|$ and $|U|$ ($i = 1, 2, 3, \cdots, n$ and $j = 1, 2, 3, \cdots, m$). This table is arranged into the form shown in Table 1, 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 2: 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}	$x_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_m	x_{m1}	x_{m2}	$\cdots$	x_{mn}	$x_{m\cdot}$
Sum	$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot n}$	$x_{\cdot\cdot} = \|U\| = N$

3 Statistical Independence in 2×2 Contingency Table

Let us consider a contingency table shown in Table 1. Statistical independence between R_1 and R_2 gives:

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

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.

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

Thus,

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

$$x_{11}x_{22} - x_{12}x_{21} = 0 \tag{1}$$

□

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.

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

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.

4 Statistical Independence in 2×3 Contingency Table

Let us consider a 2×3 contingency table shown in Table 3. Statistical independence

Table 3: 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}	$x_{2\cdot}$
	$x_{\cdot 1}$	$x_{\cdot 2}$	$x_{\cdots 3}$	$x_{\cdot\cdot}$
			$(= \|U\| = N)$	

between R_1 and R_2 gives:

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

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

$$\frac{x_{11}}{N} = \frac{x_{11} + x_{12} + x_{13}}{N} \times \frac{x_{11} + x_{21}}{N} \quad (2)$$

$$\frac{x_{12}}{N} = \frac{x_{11} + x_{12} + x_{13}}{N} \times \frac{x_{12} + x_{22}}{N} \quad (3)$$

$$\frac{x_{13}}{N} = \frac{x_{11} + x_{12} + x_{13}}{N} \times \frac{x_{13} + x_{23}}{N} \quad (4)$$

$$\frac{x_{21}}{N} = \frac{x_{21} + x_{22} + x_{23}}{N} \times \frac{x_{11} + x_{21}}{N} \quad (5)$$

$$\frac{x_{22}}{N} = \frac{x_{21} + x_{22} + x_{23}}{N} \times \frac{x_{12} + x_{22}}{N} \quad (6)$$

$$\frac{x_{23}}{N} = \frac{x_{21} + x_{22} + x_{23}}{N} \times \frac{x_{13} + x_{23}}{N} \quad (7)$$

From equation (2) and (5),

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

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

$$\frac{x_{11}}{x_{21}} = \frac{x_{12}}{x_{22}} = \frac{x_{13}}{x_{23}} = \frac{x_{11} + x_{12} + x_{13}}{x_{21} + x_{22} + x_{23}} \quad (8)$$

Thus, we obtain the following theorem:

Theorem 2 *If two attributes in a contingency table shown in Table 3 are statistical indepedent, the following equations hold:*

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

□

It is notable that this discussion can be easily extended into a $2xn$ contingency table where $n > 3$. The important equation 8 will be extended into

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

Thus,

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

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

□

It is also notable that this equation is the same as the equation on collinearity of projective geometry [1].

5 Statistical Independence in $m \times n$ Contingency Table

Let us consider a $m \times n$ contingency table shown in Table 2. Statistical independence of R_1 and R_2 gives the following formulae:

$$
\begin{aligned}
P([R_1 = A_i, R_2 = B_j]) &= P([R_1 = A_i]) \\
&\quad \times P([R_2 = B_j]) \\
&\quad (i = 1, \cdots, m, j = 1, \cdots, n).
\end{aligned}
$$

According to the definition of the table,

$$\frac{x_{ij}}{N} = \frac{\sum_{k=1}^{n} x_{ik}}{N} \times \frac{\sum_{l=1}^{m} x_{lj}}{N}. \quad (9)$$

Thus, we have obtained:

$$x_{ij} = \frac{\sum_{k=1}^{n} x_{ik} \times \sum_{l=1}^{m} x_{lj}}{N}. \quad (10)$$

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:

$$\frac{x_{i_a 1}}{x_{i_b 1}} = \frac{x_{i_a 2}}{x_{i_b 2}} \cdots = \frac{x_{i_a n}}{x_{i_b n}} = \frac{\sum_{k=1}^{n} x_{i_a k}}{\sum_{k=1}^{n} x_{i_b k}}. \quad (11)$$

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 2 are statistical indepedent, the following equations hold:*

$$\frac{x_{i_a 1}}{x_{i_b 1}} = \frac{x_{i_a 2}}{x_{i_b 2}} \cdots = \frac{x_{i_a n}}{x_{i_b n}} = const. \quad (12)$$

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

□

5.1 Three-way Table

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

$$x_{i\bullet} = \sum_{j=1}^{n} x_{ij} \quad (13)$$

$$x_{\bullet j} = \sum_{i=1}^{m} x_{ij}, \quad (14)$$

where (13) and (14) 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 (10) is reformulated as:

$$\frac{x_{ij}}{x_{\bullet\bullet}} = \frac{x_{i\bullet}}{x_{\bullet\bullet}} \times \frac{x_{\bullet j}}{x_{\bullet\bullet}} \quad (15)$$

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 "·". By use of the above relation, Equation (12) 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 (15) can be extended into multivariate cases. Let us consider a three attribute case.

Statistical independence with three attributes is defined as:

$$\frac{x_{ijk}}{x_{\bullet\bullet\bullet}} = \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \times \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \times \frac{x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}}, \quad (16)$$

Thus,

$$x_{ijk} x_{\bullet\bullet\bullet}^2 = x_{i\bullet\bullet} x_{\bullet j\bullet} x_{\bullet\bullet k}, \quad (17)$$

which corresponds to:

$$P(A=a, B=b, C=c) = P(A=a)P(B=b)P(C=c), \quad (18)$$

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 (18), the following equations should satisfy:

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

Thus,

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

$$x_{i\bullet k} x_{\bullet\bullet\bullet} = x_{i\bullet\bullet} x_{\bullet\bullet k} \quad (20)$$

$$x_{\bullet jk} x_{\bullet\bullet\bullet} = x_{\bullet j\bullet} x_{\bullet\bullet k} \quad (21)$$

From Equation (17) and Equation (19),

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

Therefore,

$$\frac{x_{ijk}}{x_{ij\bullet}} = \frac{x_{\bullet\bullet k}}{x_{\bullet\bullet\bullet}} \quad (22)$$

In the same way, the following equations are obtained:

$$\frac{x_{ijk}}{x_{i\bullet k}} = \frac{x_{\bullet j\bullet}}{x_{\bullet\bullet\bullet}} \quad (23)$$

$$\frac{x_{ijk}}{x_{\bullet jk}} = \frac{x_{i\bullet\bullet}}{x_{\bullet\bullet\bullet}} \quad (24)$$

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 2 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.

□

5.2 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.*

$$\frac{x_{i_1i_2\cdots i_k\cdots i_n}}{x_{i_1i_2\cdots j_k\cdots i_n}} = \frac{x_{\bullet\bullet\cdots i_k\cdots\bullet}}{x_{\bullet\bullet\cdots j_k\cdots\bullet}}$$

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

$$x_{i_1i_2\cdots i_n} \times x^{n-1}_{\bullet\bullet\cdots\bullet}$$
$$= x_{i_1\bullet\cdots\bullet}x_{\bullet i_2\cdots\bullet}\times\cdots\times x_{\bullet\bullet\cdots i_k\cdots\bullet}\times\cdots\times x_{\bullet\bullet\cdots\bullet i_n}$$

□

6 Contingency Matrix

The meaning of the above discussions will become much clearer when we view a contingency table as a matrix.

Definition 4 *A corresponding matrix $C_{T_{a,b}}$ is defined as a matrix the element of which are equal to the value of the corresponding contingency table $T_{a,b}$ of two attributes a and b, except for marginal values.*

Definition 5 *The rank of a table is defined as the rank of its corresponding matrix. The maximum value of the rank is equal to the size of (square) matrix, denoted by r.*

The contingency matrix of Table $2(T(R_1, R_2))$ is defined as $C_{T_{R_1,R_2}}$ as below:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}$$

6.1 Independence of 2×2 Contingency Table

The results in Section 3 corresponds to the degree of independence in matrix theory. Let us assume that a contingency table is given as Table 1. Then the corresponding matrix ($C_{T_{R_1,R_2}}$) is given as:

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix},$$

Then,

Proposition 1 *The determinant of* $det(C_{T_{R_1,R_2}})$ *is equal to* $x_{11}x_{22} - x_{12}x_{21}$,

Proposition 2 *The rank will be:*

$$rank = \begin{cases} 2, & if\ det(C_{T_{R_1,R_2}}) \neq 0 \\ 1, & if\ det(C_{T_{R_1,R_2}}) = 0 \end{cases}$$

From Theorem 1,

Theorem 7 *If the rank of the corresponding matrix of a* $2times2$ *contingency table is 1, then two attributes in a given contingency table are statistically independent. Thus,*

$$rank = \begin{cases} 2, & dependent \\ 1, & statistical\ independent \end{cases}$$

This discussion can be extended into $2 \times n$ tables. According to Theorem 3, the following theorem is obtained.

Theorem 8 *If the rank of the corresponding matrix of a* $2 \times n$ *contigency table is 1, then two attributes in a given contingency table are statistically independent. Thus,*

$$rank = \begin{cases} 2, & dependent \\ 1, & statistical\ independent \end{cases}$$

6.2 Independence of 3×3 Contingency Table

When the number of rows and columns are larger than 3, then the situation is a little changed. It is easy to see that the rank for statistical independence of a $m \times n$ contingency table is equal 1.0 as shown in Theorem 4. Also, when the rank is equal to $\min(m, n)$, two attributes are dependent.

Then, what kind of structure will a contingency matrix have when the rank is larger than 1,0 and smaller than $\min(m, n) - 1$? For illustration, let us consider the following $3times3$ contingecy table.

Example 1 *Let us consider the following corresponding matrix:*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

The determinant of A is:

$$\begin{aligned} det(A) &= 1 \times (-1)^{1+1} det \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} \\ &\quad +2 \times (-1)^{1+2} det \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} \\ &\quad +3 \times (-1)^{1+3} det \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\ &= 1 \times (-3) + 2 \times 6 + 3 \times (-3) = 0 \end{aligned}$$

Thus, the rank of A is smaller than 2. On the other hand, since $(123) \neq k(456)$ *and* $(123) \neq k(789)$*, the rank of A is not equal to 1.0 Thus, the rank of A is equal to 2.0. Actually, one of three rows can be represented by the other two rows. For example,*

$$(4\,5\,6) = \frac{1}{2}\{(1\,2\,3) + (7\,8\,9)\}.$$

Therefore, in this case, we can say that two of three pairs of one attribute are dependent to the other attribute, but one pair is statistically independent of the other attribute with respect to the linear combination of two pairs. It is easy to see that this case includes the cases when two pairs are statistically independent of the other attribute, but the table becomes statistically dependent with the other attribute.

In other words, the corresponding matrix is a mixure of statistical dependence and independence. We call this case *contextual independent.* From this illustration, the following theorem is obtained:

Theorem 9 *If the rank of the corresponding matrix of a* 3×3 *contigency table is 1, then two attributes in a given contingency table are statistically independent. Thus,*

$$rank = \begin{cases} 3, & dependent \\ 2, & contextual\ independent \\ 1, & statistical\ independent \end{cases}$$

It is easy to see that this discussion can be extended into $3 \times n$ contingency tables.

6.3 Independence of $m \times n$ Contingency Table

Finally, the relation between rank and independence in a multi-way contingency table is obtained from Theorem 4.

Theorem 10 *Let the corresponding matrix of a given contingency table be a $m \times n$ matrix. If the rank of the corresponding matrix is 1, then two attributes in a given contingency table are statistically independent. If the rank of the corresponding matrix is* $\min(m, n)$ *, then two attributes in a given contingency table are dependent. Otherwise, two attributes are contextual dependent, which means that several conditional probabilities can be represented by a linear combination of conditional probabilities. Thus,*

$$rank = \begin{cases} \min(m,n) & dependent \\ 2, \cdots, & \\ \quad \min(m,n)-1 & contextual\ independent \\ 1 & statistical\ independent \end{cases}$$

References

[1] H.S.M. Coxeter, editor. *Projective Geometry*. Springer Verlag, New York, 2nd edition, 1987.

[2] A. Skowron and J. Grzymala-Busse. From rough set theory to evidence theory. In R. Yager, M. Fedrizzi, and J. Kacprzyk, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 193–236. John Wiley & Sons, New York, 1994.

[3] Y.Y. Yao and S.K.M. Wong. A decision theoretic framework for approximating concepts. *International Journal of Man-machine Studies*, 37:793–809, 1992.

[4] Y.Y. Yao and N. Zhong. An analysis of quantitative measures associated with rules. In N. Zhong and L. Zhou, editors, *Methodologies for Knowledge Discovery and Data Mining, Proceedings of the Third Pacific-Asia Conference on Knowledge Discovery and Data Mining LNAI* **1574**, pages 479–488, Berlin, 1999. Springer.