A GRAPHICAL ASSOCIATION MEASURE FOR THE CROSS CLASSIFICATION

Wakimoto, Kazumasa
Department of Mathematics and Statistics, College of Liberal Arts & Sciences, Okayama University

https://doi.org/10.5109/13388

出版情報：Bulletin of informatics and cybernetics. 22 (3/4), pp.209-214, 1987-03. Research Association of Statistical Sciences

Bulletin of Informatics and Cybernetics Vol. 22, No. 3~4, 1987

A GRAPHICAL ASSOCIATION MEASURE FOR THE CROSS CLASSIFICATIONS

By

Kazumasa WAKIMOTO*

Abstract

An association graph is proposed to the $K \times L$ table of the cross classifications and an association measure based on the graph is also proposed to the table. The association graph makes it easy to find the detailed differences between two $K \times L$ tables. Thus graphical merit is illustrated by the example.

1. Introduction

Let us consider the finite population which contains a finite number N of units and suppose the population completely known in regard to the classifications. Then we have the $K \times L$ table for the closs classifications shown by the table 1, where

(i) Classification A divides the population into the K classes $A_1, A_2, \cdots, A_K$.

(ii) Classification B divides the population into the L classes $B_1, B_2, \cdots, B_L$.

(iii) The number of the population units that is classified as both A_i and B_j is N_{ij}. The marginal numbers are denoted by $N_{i\cdot}$ and $N_{\cdot j}$, respectively.

The traditional association measures for the table are mainly as follows. (See Reference [1]).

(i) Mean square contingency: $\phi = \sqrt{\dfrac{\chi^2}{N}}$

Table 1.

A	B				
	B_1	B_2	....	B_L	Total
A_1	N_{11}	N_{12}	....	N_{1L}	$N_{1\cdot}$
A_2	N_{21}	N_{22}	....	N_{2L}	$N_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_K	N_{K1}	N_{K2}	....	N_{KL}	$N_{K\cdot}$
Total	$N_{\cdot 1}$	$N_{\cdot 2}$	....	$N_{\cdot L}$	N

* Department of Mathematics and Statistics, College of Liberal Arts & Sciences, Okayama University, Tsushima 700, Japan.

(ii) Coefficient of contingency: $C=\sqrt{\frac{\chi^2/N}{1+\chi^2/N}}$

(iii) Tschuprow's coefficient: $T=\sqrt{\frac{\chi^2/N}{(K-1)(L-1)}}$

(iv) Cramér's coefficient: $V=\sqrt{\frac{\chi^2/N}{\mathrm{Min}(K-1,\ L-1)}}$,

where

$$\chi^2=\sum_{i=1}^{K}\sum_{j=1}^{L}\frac{\left(N_{ij}-\frac{N_{i\cdot}N_{\cdot j}}{N}\right)^2}{\frac{N_{i\cdot}N_{\cdot j}}{N}}$$

In this paper, we proposed a new association measure for the table using the linked line chart which is similar one in the previous papers. (See References [2] and [3]). In the section 2, we discuss how to make an assocation graph using the linked line chart and proposed an association measure based on the graph. In the section 3, we illustrate the comparison between the new measure and the traditional measures, and illustrate some graphical characteristics for the new measure.

2. Association Graph

Put

$$D_{ij}=\frac{\left(N_{ij}-\frac{N_{i\cdot}N_{\cdot j}}{N}\right)^2}{\frac{N_{i\cdot}N_{\cdot j}}{N}}, \qquad (i=1,2,\cdots,K,\ j=1,2,\cdots,L) \tag{2.1}$$

Then we draw an association graph which is constracted by the following steps:

Step 1. Rearrange the values $D_{11}, D_{12}, \cdots, D_{1L}, D_{21}, D_{22}, \cdots, D_{2L}, \cdots, D_{K1}, D_{K2}, \cdots, D_{KL}$ of size KL in the order of the magnitudes as follows:

$$D(1)\leqq D(2)\leqq \cdots \leqq D(KL-1)\leqq D(KL). \tag{2.2}$$

Step 2. Draw the x-axis and the y-axis intersecting at right angles at the origin O. Draw a line $\overline{OP_1}$ of the length $D(KL)$ in the direction of the x-axis starting from the origin O. Next, draw another line $\overline{P_1P_2}$ of the length $D(KL-1)$ in the direction of $\pi/2(KL-1)$ radian from the line $\overline{OP_1}$ starting from the point P_1. Continue in the same manner and draw other lines $\overline{P_2P_3}, \cdots, \overline{P_{KL-1}P_{KL}}, \cdots, \overline{P_{2KL-1}P_{2KL}}$ of the lengths $D(KL-2), \cdots, D(1), D(1), \cdots, D(KL)$ in the direction of $\pi/2(KL-1)$ radian from the before line.

Step 3. Construct a polygon with $4KL$ sides as shown in Fig. 1, where P_i' is a symmetric point of P_i $(i=1, 2, \cdots, 2KL)$ with respect to the segment $\overline{OP_{2KL}}$.

We call the chart as shown in Fig. 1 which is constructed by the above steps "Association graph".

Note. From (2.1) and (2.2) we have the following equation.

$$\overline{OP_1}+\overline{P_1P_2}+\cdots+\overline{P_{KL-1}P_{KL}}=\sum_{i=1}^{KL}D(i)=\chi^2$$

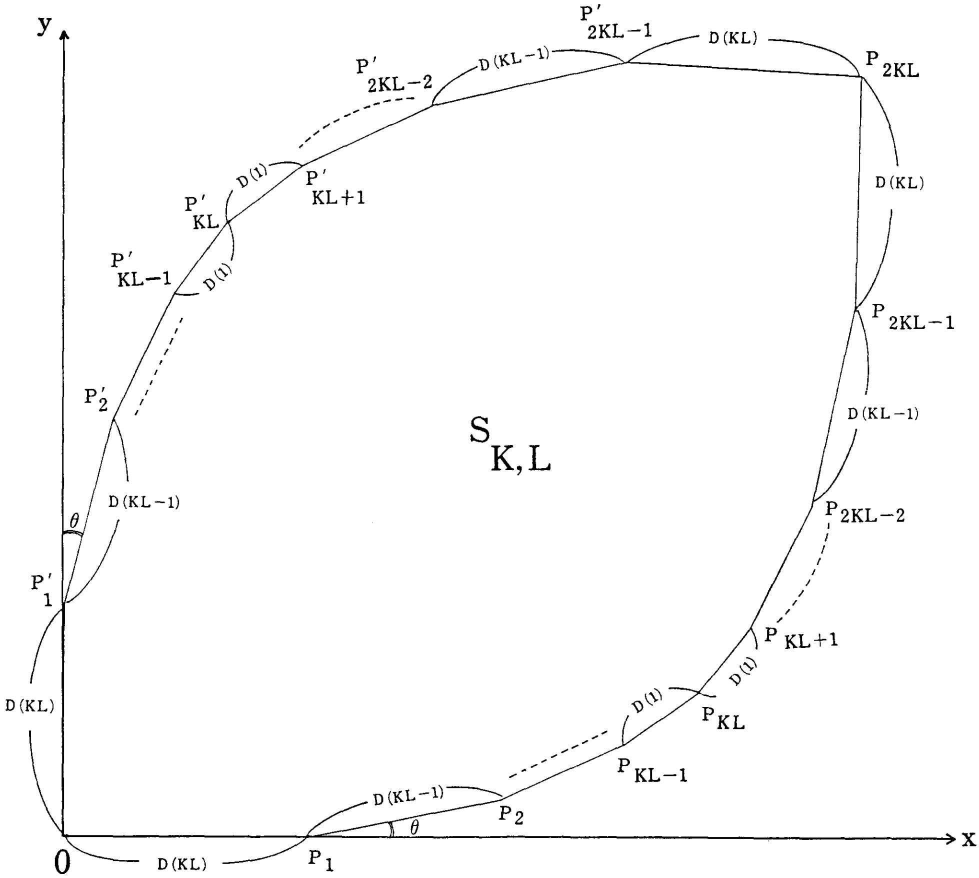

Fig. 1. Association graph, $\theta=\pi/2(KL-1)$.

3. An Association Measure Based on the Association Graph

In Fig. 1, let us denote the area of the polygone which is enclosed by the $4KL$ sidelines $\overline{OP_1}$, $\overline{P_1P_2}$, $\cdots$, $\overline{P_{2KL-1}P_{2KL}}$, $\overline{P_{2KL}P'_{2KL-1}}$, $\cdots$, $\overline{P'_2P'_1}$, $\overline{P'_1O}$ by S^2_{KL}. Then we have

$$S^2_{KL}=2\sum_{i=1}^{KL}\sum_{j=1}^{i}D(i)D(j)\sin\frac{i-j}{2(2KL-1)}\pi$$

$$+\sum_{i=1}^{KL}\sum_{j=1}^{KL}D(i)D(j)\sin\frac{i+j-1}{2(2KL-1)}\pi \tag{3.1}$$

Put

$$W_{KL}=\frac{M}{N}S_{KL}, \tag{3.2}$$

where $M=\max(K, L)$. We call the measure given by (3.2) "Area Coefficient". We disscuss about the upper bound of W_{KL} to lie it between 0 and 1.

Let us suppose that $K\geqq L$. Let Φ be the set of all vector $\boldsymbol{N}_0=(N_{11}, \cdots, N_{LL}, N_{L,L+1}, \cdots, N_{LK})$ satisfying the following equation:

$$\sum_{i=1}^{L}N_{ii}+\sum_{i=L+1}^{K}N_{iL}=N. \tag{3.3}$$

Then it is well known that the traditional association measures based on the value χ^2 are attained equally at any N_0 in Φ. However, the measure W_{KL} is not equal to all vector N_0 in Φ. Therefore it may be interesting to find the upper bound. The upper bound of W_{KL} may be attained at a value N_0^* in Φ. Though we do not obtain the exact upper bound, we have the following inequality in the case of $K=L=E$.

$$
\begin{aligned}
0 \leqq W_{KL}^2 \leqq 2\Big\{ & \frac{(E-1)^4}{N^2}\sum_{i=1}^{E(E-1)}\sum_{j=1}^{i}\sin(i-j)\theta \\
& + \frac{(E-1)^2}{N^2}\sum_{i=E(E-1)+1}^{E^2}\sum_{j=1}^{E(E-1)}\sin(i-j)\theta \\
& + \frac{(E-1)^4}{N^2}\sum_{i=E(E-1)+1}^{E^2}\sum_{j=E(E-1)+1}^{i}\sin(i-j)\theta\Big\} \\
& + \frac{1}{N^2}\sum_{i=1}^{E(E-1)}\sum_{j=1}^{E(E-1)}\sin(i+j-1)\theta \\
& + \frac{2(E-1)^2}{N^2}\sum_{i=1}^{E(E-1)}\sum_{j=E(E-1)+1}^{E^2}\sin(i+j-1)\theta \\
& + \frac{(E-1)^4}{N^2}\sum_{i=E(E-1)+1}^{E^2}\sum_{j=E(E-1)+1}^{E^2}\sin(i+j-1)\theta \\
& + \varepsilon,
\end{aligned}
\tag{3.4}
$$

where $\theta=\dfrac{\pi}{2(Mm-1)}$ and the part except ε in the right hand is the value of W_{KL}^2.

In the inequality (3.4), ε may be very small compared with the value of W_{KL}^2. For example, in the case of $K=L=3$ and $N=9$, we have the Table 2.

Table 2.

$(N_{11},\ N_{22},\ N_{33})$	χ^2	$W_{3\cdot3}$
(1, 1, 7)	18	6.271
(1, 2, 6)	18	6.302
(1, 3, 5)	18	6.313
(1, 4, 4)	18	6.312
(2, 2, 5)	18	6.301
(2, 3, 4)	18	6.296
(3, 3, 3)	18	6.251

4. Example to Show the Graphical Merit

Assume that two 4×3 tables are given by the table 3 and table 4. Then we can draw the association graph as shown by Fig. 2 for the both tables.

In the table 3 and the table 4, we obtain the association measures including the area coefficient as shown by the table 5. Then we cannot recognize the difference for the strength of association between the table 3 and the table 4 by looking two measures in the table 5, because two measures for both tables are very closed. However, we will recognize some difference between both tables by looking two patterns of the

Table 3.

A	B_1	B_2	B_3	Total
A_1	8	24	35	67
A_2	66	10	18	94
A_3	12	28	41	81
A_4	44	5	9	58
Total	130	67	103	300

Table 4.

A	B_1	B_2	B_3	Total
A_1	47	16	4	67
A_2	46	33	15	94
A_3	30	15	36	81
A_4	7	3	48	58
Total	130	67	103	300

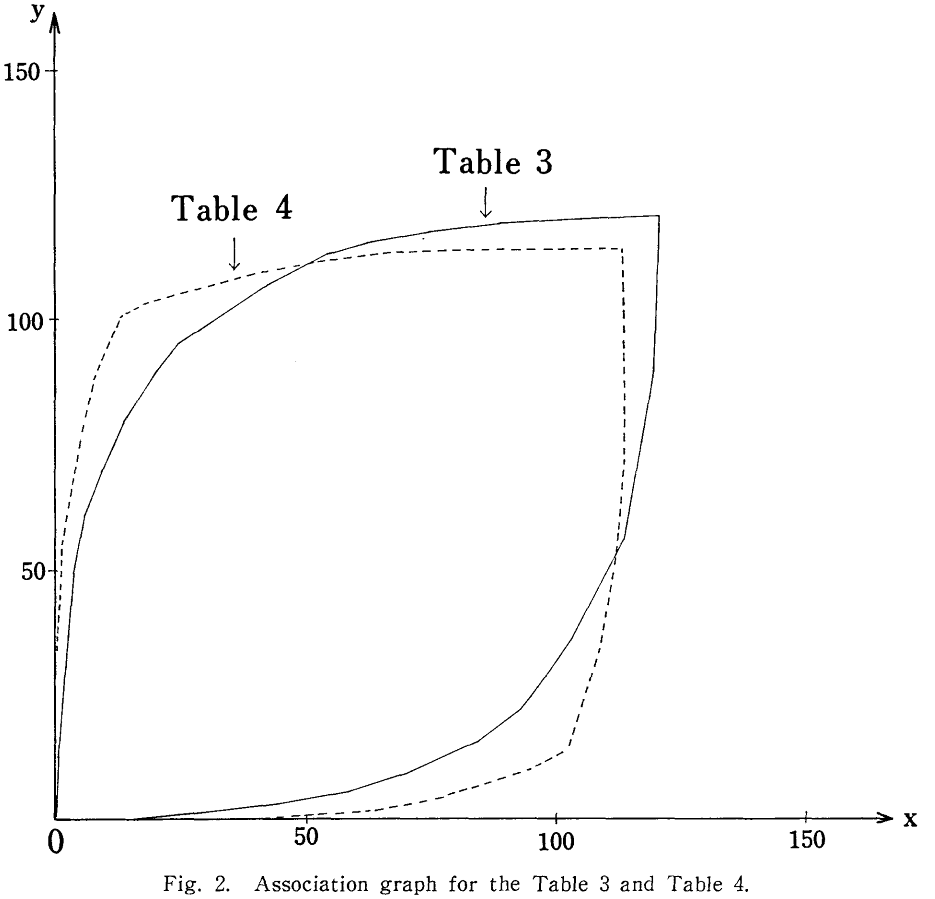

Fig. 2. Association graph for the Table 3 and Table 4.

Table 5. The association measures for the table 3 and the table 4.

	ϕ	C	T	V	$W_{3\times 4}$
Table 3	0.594	0.511	0.243	0.420	1.46
Table 4	0.597	0.512	0.244	0.422	1.51

dotted line and the real line in Fig. 2. For example, we will recognize that $D(12)$ and $D(11)$ in the table 4 are larger extremely than these in the table 3.

References

[1] GOODMAN, L. A. and KRUSKAL, W. H.: Measures of Association for Cross Classifications, Springer-Verlag, (1979).
[2] SHIRAHATA, S., WAKIMOTO, K. and TARUMI, T.: *A Goodness of Fit Test Based on the Linked Line Chart*, Australian J. Statist., **27** (1985), 163-171.
[3] WAKIMOTO, K. and SHIRAHATA, S.: *A Coefficient of Concordance Based on the Chart of Linked Lines*, J. Japan Statist. Soc., **14** (1984), 189-197.

Communicated by Ch. Asano
Received December 27, 1985