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

Distribution of Derminants of Contingency Tables

Shusaku Tsumoto
Department of Medical Informatics, Faculty of Medicine,
Shimane University
89-1 Enya-cho Izumo 693-8501 Japan
Email: tsumoto@med.shimane-u.ac.jp

***Abstract*—This paper gives a empirical analysis of determinant, which empirically validates the trade-off between sample size and size of matrix. In the former studies, relations between degree of granularity and dependence of contingency tables are given from the viewpoint of determinantal divisors and sample size. The nature of determinantal divisors shows that the increase of the degree of granularity may lead to that of dependence. However, a constraint on the sample size of a contingency table is very strong, which leads to the evaluation formula where the increase of degree of granularity gives the decrease of dependency. This paper gives a further study of the nature of sample size effect on the degree of dependency in a contingency matrix. The results show that sample size will restrict the nature of matrix in a combinatorial way, which suggests that the dependency is closely related with integer programming.**

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

I. Introduction

Although independence is a very important concept, it has not been fully and formally investigated as a relation between two attributes. Tsumoto introduces linear algebra into formal analysis of a contigency table [1]–[4]. The results give the following interesting results. First, a contingency table can be viewed as comparison between two attributes with respect to information granularity. Second, 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. Especially, The degree of independence, rank plays a very important role in extracting a probabilistic model from a given contingency table.

Then, thirdly, the results of determinantal divisors show that it seems that the devisors provide information on the degree of dependencies between the matrix of the whole elements and its submatrices and the increase of the degree of granularity may lead to that of dependence [4], [5]. This gives a contradictory view from the intuition that when two attributes has many values, the dependence between these two attributes becomes low.

The key for understanding these conflicts is to consider the constraint on the sample size.

In [6] we show that a constraint on the sample size of a contingency table is very strong, which leads to the evaluation formula where the increase of degree of granularity gives the decrease of dependency.

This paper confirms this constraint by using enumerative combinatorics.

The results show that sample size will restrict the nature of matrix in a combinatorial way, which suggests that the dependency is closely related with integer programming.

II. Degree of Dependence

A. 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 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 ($n \times m$)

	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,

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

□

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

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

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.

B. Rank of Contingency Matrix ($m \times n$)

In the former paper, Tsumoto obtained the following theorem [1].

Theorem 1: Let the contingency matrix of a given contingency table be a $m \times n$ matrix. The rank of this matrix is less than $\min(m, n)$. 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 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,

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

□

C. Degree of Granularity and Dependence

Let us assume that the determinant of a give contingency matrix gives the degree of the dependence of the matrix. Then, from the results of linear algebra, we obtain the following theorem.

Theorem 2: Let A denote a $n \times n$ contingency matrix, which includes N samples. If the rank of A is equal to n, then there exists a matrix B ($n \times n$) which satisfies

$$BA \;=\; \begin{pmatrix} \rho_1 & & & \\ & \rho_2 & & O \\ & & \ddots & \\ & O & & \rho_n \end{pmatrix} = P,$$

where $\rho_1 + \rho_2 + \cdots + \rho_n = N$.

It is notable that the value of determinants of P is larger than A:

$$detA \leq detP$$

□

Thus, the following theorem is obtained [6].

Theorem 3: When a contingency matrix A holds $AB = P$, where P is a diagonal matrix, the following inequality holds:

$$detA \leq \left(\frac{N}{n}\right)^n,$$

Proof.

$$\begin{aligned} detA &= det(PB^{-1}) \\ &\leq detP \\ &= \rho_1\rho_2\cdots\rho_n \\ &\leq \left(\frac{\rho_1+\rho_2+\cdots+\rho_n}{n}\right)^n = \left(\frac{N}{n}\right)^n, \end{aligned} \qquad (1)$$

where the former equality holdes when $detB^{-1} = detB = 1$ and the latter equality holds when $\rho_1 = \rho_2 = \cdots = \rho_n = \frac{N}{n}$.

Thus, the maximum value of the determinant of A is at most $\left(\frac{N}{n}\right)^n$. Since N is constant for the given matrix A, the degree of dependence will decrease very rapidly when n becomes very large. That is,

$$detA \sim n^{-n}.$$

Thus,

Corollary 1: The determinant of A will converge into 0 when n increases into infinity.

$$\lim_{n\to\infty} detA = 0.$$

□

This results suggest that when the degree of granularity becomes higher, the degree of dependence will become lower, due to the constraints on the sample size.

However, it is notable that N/n is very important. If N is very large, the rapid decrease will be observed N is close to n.

III. Distribution of Determinant

The next interest is how is the statistical nature of the derminant for $M(m, n, N)$.

First, since a 2×2 matrix is a basic one, let us examine the nature of $\det M(2, 2, N)$.

A. Total number of $M(2, 2, N)$

Let the four elements of $M(2, 2, N)$ be denoted as a,b,c,d. That is, $x_{11} = a$, $x_{12} = b$, $x_{21} = c$, and $x_{22} = d$. Then, $a + b + c + d = N$.

Let us assume that $a = 0$. Then, $b+c+d = N$. Recursively, we can assume that $b = 0$. Then, for this pair $(a, b) = 0$, we have $(N+1)$ pairs which satisfies $c+d = N$. In this way, the total number of $M(2, 2, N)$ is obtained as:

$$\sum_{i=0}^{N} \frac{(N+1-i)\times(N+2-i)}{2}.$$

Simple calculation shows that the above formula is equal to:

$$\frac{1}{6}(N+1)(N+2)(N+3).$$

That is,

Theorem 4: The total number of a contingency matrix M(2,2,N) is equal to:

$$\frac{1}{6}(N+1)(N+2)(N+3).$$

$(Proof\ Sketch)$

The total combination of $M(2,2,N)$ is given as:

$$\begin{aligned}\sum_{i=0}^{N}\left(\sum_{k=1}^{(N-i)+1} k\right) &= \sum_{i=0}^{N} \frac{(N+1-i)\times(N+2-i)}{2} \\ &= \sum_{i=0}^{N}\left\{\frac{1}{2}(N+1)(N+2)\right. \\ &\quad \left. -\frac{1}{2}(2N+3)i+\frac{1}{2}i^2\right\} \\ &= \frac{1}{6}(N+1)(N+2)(N+3)\end{aligned} \qquad (2)$$

□

Intuitively, this formula can be interpreted as follows. We have four parameters, a,b,c,d, which will take a value between 0 and N. Thus, the original freedom is 4, and the order of total number can be N^4. However, since a constraint $a+b+c+d=N$ is given, we have only three free parameters, thus the order of total number of $M(2,2,N)$ is approximately of N^3:

$$\#\ of\ M(2,2,N) \approx \mathcal{O}(N^3).$$

B. Total number of $det = 0$

Enumeration of total number of $det = 0$ is very difficult. However, upper bound can be calculated as follows. When a and d is fixed, we have obtained two constraints:

$$\begin{aligned} b+c &= N-(a+d) \\ bc &= ad \end{aligned}$$

Thus, (b,c) can be obtained as a solution for quadratic equations. If the pair (b,c) is integer, we will have obtained two solutions $(ad-bc=0)$ for each pair: (b,c) and (c,b).

Therefore, the upper bound of the number of solutions is equal to:

$$\sum_{i=0}^{N}\left(\sum_{k=1}^{(N-i)+1} 2\right) = (N+1)(N+2)$$

Theorem 5: The upper bound of total number of a contingency matrix M(2,2,N) with determinant being 0 is equal to:

$$(N+1)(N+2)$$

Thus, the probability that the determinant of a matrix $M(2,2,N)$ is equal to 0 is at most:

$$\frac{(N+1)(N+2)}{\frac{1}{6}(N+1)(N+2)(N+3)} = \frac{6}{N+3}.$$

□

Then, how is the lower bound ? This is the case when (b,c) does not have any integer solution for a given quadratic equations except for trivial solutions. The simple trivial solutions are: $a=0$ or $d=0$ with $b=0$ or $c=0$. Then, for $a=0$, $b=0$, we may have a solution for $c+d=N$, N pairs ($c \neq 0, d \neq 0$). Totally, 4N pairs. If we consider the cases when three values are equal to 0, such as $a=b=c=0$, we have 4 pairs. Thus, totally. we have 4(N+1) pairs.

Theorem 6: The lower bound of total number of a contingency matrix M(2,2,N) with determinant being 0 is equal to:

$$4(N+1)$$

Thus, the probability that the determinant of a matrix $M(2,2,N)$ is equal to 0 is at least:

$$\frac{4(N+1)}{\frac{1}{6}(N+1)(N+2)(N+3)} = \frac{24}{(N+2)(N+3)}.$$

□

Thus, it is expected that the number of matrices with 0 determinant vibrates between $4(N+1)$ and $(N+1)(N+2)$. The variance will beccome larger when N grows. In other words, the probability of $det = 0$ will vibrate between $\mathcal{O}(1/N^2)$ and $\mathcal{O}(1/N)$. The variance will become larger when N grows.

It is notable that the above discussion can be applied to a general case, such as $ad-bc=k$, or other constraint. For example, if we have a constraint such as $a/(a+b)$ or $a/(a+c)$, then we can analyze a constraint for accuracy or coverage. It will be our future work to investigate such cases.

IV. Empirical Validations

For empirical validations, we calculate the whole combination of a 2 × 2 matrix with fixed sample size ($0 \leq N \leq 100$) $M(2,2,N)$.

A. Total Number of $M(2,2,N)$

Figure 1 plots the relation between sample size N and the total number of $M(2,2,N)$. This figure clearly shows that the relation is polynomial.

On the other hand, Figure 2, which plots the relation between sample size and the total number of matrices with zero determinant, gives an interesting feature. As discussed in Section III, the total number vibrates and the amplitude of the vibration becomes larger when N grows. Furthermore, the lower bound of the total number can be approximately equal to a linear function, whereas the upper bound is to a quadratic function.

Finally, the ratio of the number of matrices with zero determinant to the total number of $M(2,2,N)$ is plotted as Figure 3. This figure also confirms the results obtained in Section III.

B. Statistics of Determinant

Figure 4 and 5 show the distributions of the determinant of $M(2,2,10)$ and $M(2,2,50)$. The distribution are symmetric, and the median and average are exactly equal to 0. Furthermore, the number of matrices with 0 determinant is very high, compared with other values.

Figure 6 plots the distribution of $|\det M(2,2,50)|$, which suggests that the distribution is like $1/N$. However, it is notable that the vibration is observed for a given determinant value.

It is also notable that since the ratio of $det = 0$ rapidly decreases as N grows, the number of matrices with 0 determinant becomes smaller.

Table II and III shows the statistics of those matrices.

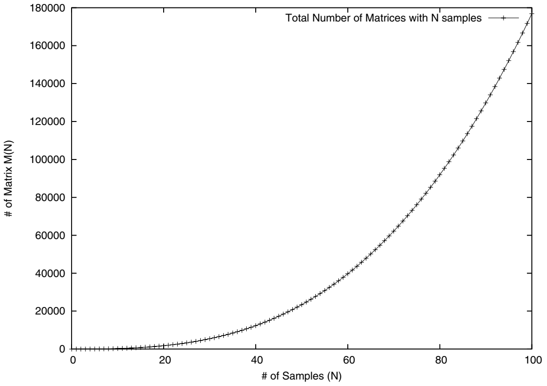

Fig. 1. Total Number of $M(2, 2, N)$

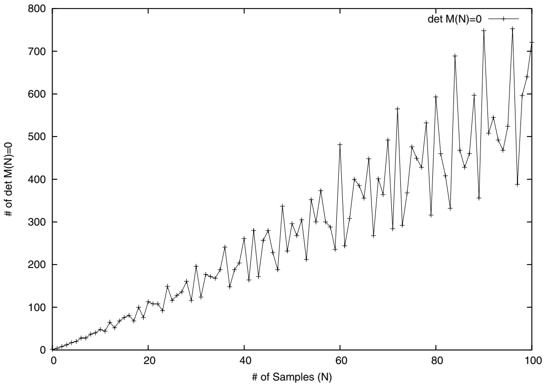

Fig. 2. Number of Matrices with [Det=0] in $M(2, 2, N)$

V. Conclusion

In this paper, the nature of the dependence of a contingency matrix and the statistical nature of the determinant are examined.

Especially, the constraint on the sample size N of a contingency table will determine the number of 2×2 matrices. As N grows, the ratio of matrices with zero determinant rapidly decreases, which shows that the number of matrix with statistical dependence will increase. However, due to the nature of the determinant, the average of absolute value of the determinant also increase with the order of N^2, whereas the increase in the size of total number of matrix is of N^3.

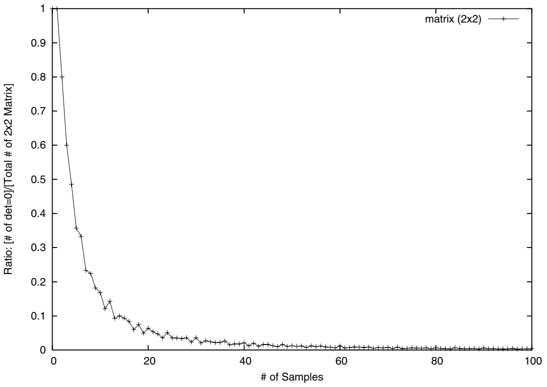

Fig. 3. Ratio of [Det=0] in $M(2, 2, N)$

TABLE II
Statistics of $M(2, 2, 10)$

	det	|det|
min	-25	0
25%	-6	2
50%	0	6
75%	6	11.75
max	25	25

TABLE III
Statistics of $M(2, 2, 50)$

	det	|det|
min	-625	0
25%	-140	60
50%	0	140
75%	140	256
max	625	625

This is a preliminary work on the statistical nature of the determinant, and it will be our future work to investigate the nature of 3×3 or higher dimensional contingency matrices.

References

[1] S. Tsumoto, "Statistical independence as linear independence," in *Electronic Notes in Theoretical Computer Science*, A. Skowron and M. Szczuka, Eds., vol. 82. Elsevier, 2003.

[2] S. Tsumoto and S. Hirano, "Statistical independence and contingency matrix," in *ICDM Workshops*. IEEE Computer Society, 2008, pp. 643–648.

[3] ——, "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.

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

[5] ——, "Determinantal divisors for the degree of independence of a contingency matrix," in *Proceedings of NAFIPS 2004*. IEEE press, 2004.

[6] ——, "Degree of dependence as granularity in contingency table," in *Proceedings of IEEE GrC 2005*, T. Hu and T. Lin, Eds. IEEE press, 2005.

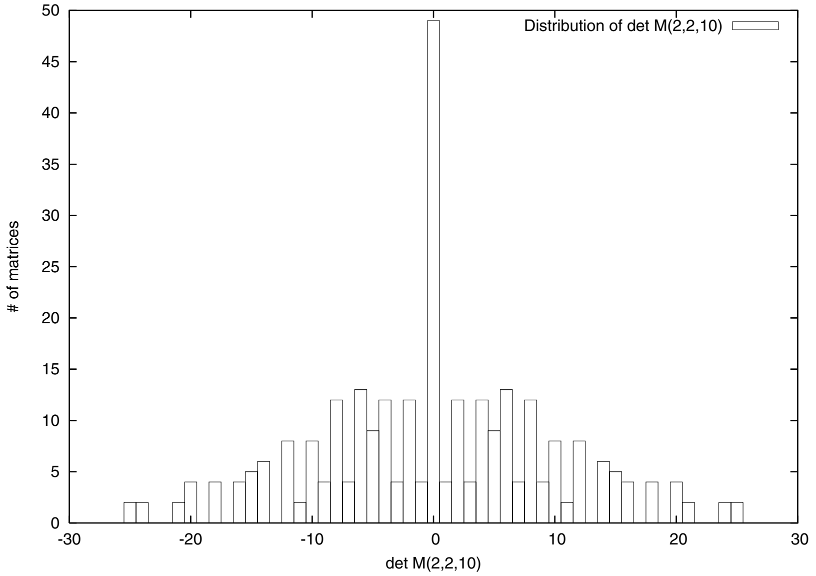

Fig. 4. Distribution of $\det M(2,2,10)$

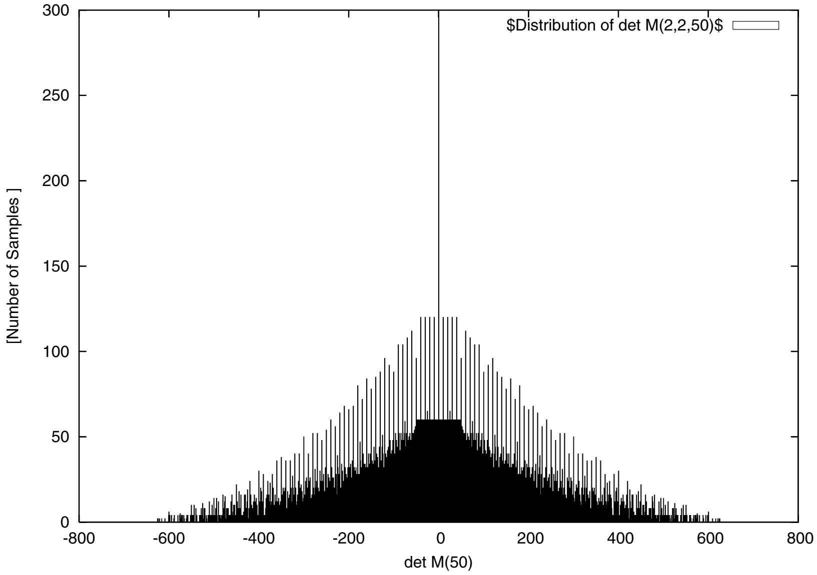

Fig. 5. Distribution of $\det M(2,2,50)$

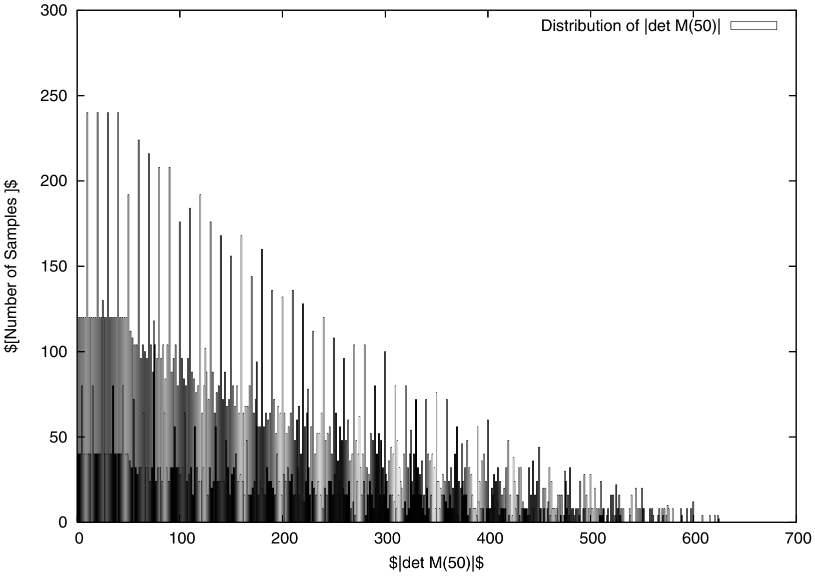

Fig. 6. Distribution of $|\det M(2,2,50)|$