数理解析研究所講究録 1437 巻 2005 年 122-127

TORIC IDEALS AND CONTINGENCY TABLES

HIDEFUMI OHSUGI (RIKKYO UNIVERSITY)
TAKAYUKI HIBI (OSAKA UNIVERSITY)

ABSTRACT. Fundamental questions on semigroup rings and toric ideals arising from contingency tables will be studied. In addition to discussing recent developments on such the topic, the algebraic background of contingency tables, including the algebraic aspects of Markov chains will be also explained.

1. ALGEBRAIC BACKGROUND OF CONTINGENCY TABLES

In commutative algebra, a Markov chain can be regarded as a system of binomial generators of the toric ideal arising from a contingency table.

An n-way contingency table is an n dimensional matrix whose entries are non-negative integers. For example, the following 2-way contingency table is given in [6, Table 2].

Example 1.1. Looking at the below 2-way contingency table T, we want to know whether "Eye color" and "Hair color" are correlated.

Eye color \ Hair color	Black	Brunette	Red	Blonde	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

In general, when an n-way contingency table is given, we are interested in n factor interaction. For the sake of simpleness, we explain the case $n = 2$ here. (See [1] for details.) We consider the following $I \times J$ contingency table T_0:

$X \setminus Y$	Y_1	Y_2	$\cdots$	Y_J	Total
X_1	n_{11}	n_{12}	$\cdots$	n_{1J}	n_{1+}
X_2	n_{21}	n_{22}	$\cdots$	n_{2J}	n_{2+}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
X_I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	n_{I+}
Total	n_{+1}	n_{+2}	$\cdots$	n_{+J}	N

Then we suppose the *null hypothesis* H_0 "there is no association between X and Y," and try to test it.

One of the popular methods which test the association of X and Y is the χ^2 test. In the χ^2 test, we compute the χ^2 statistic

$$\chi^2(T_0) = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \frac{n_{i+}n_{+j}}{N})^2}{\frac{n_{i+}n_{+j}}{N}}.$$

If the hypothesis H_0 is true, then χ^2 statistic has asymptotic χ^2 distributions with degrees of freedom $(I-1)(J-1)$. Thus we compare $\chi^2(T_0)$ with the value $\alpha_{0.05}$ of the right-hand side 5% point of χ^2 distribution. If $\chi^2(T_0) < \alpha_{0.05}$, then we conclude that H_0 is true, and X and Y are independent.

However, it is known that, if T_0 is *sparse*, that is, there are a lot of cells with $n_{i+}n_{+j}/N < 5$, then the χ^2 approximation is bad. Hence χ^2 test is not good for contingency tables such that N/IJ is small.

For such sparse contingency tables, we use *the Fisher's exact test.* Let $\mathcal{F}_{T_0}$ denote the set of tables with the same marginal distribution as T_0. For example, in Example 1.1, $\mathcal{F}_T$ is the set of all 4×4 matrix such that the sum of four rows is (108 286 71 127) and the sum of four columns is the transpose of (220 215 93 64). We now assume that H_0 is true and that $\mathcal{F}_{T_0}$ follows the multiple hypergeometric distribution. Here the multiple hypergeometric distribution is defined by

$$P(T) = \frac{\left(\prod_{i=1}^{I} n_{i+}!\right)\left(\prod_{j=1}^{J} n_{+j}!\right)}{N! \prod_{i,j} n_{ij}!}$$

for each $T \in \mathcal{F}_{T_0}$. For the Fisher's exact test, we compute $\chi^2(T)$ for *all* $T \in \mathcal{F}_{T_0}$ and P-value

$$P = \sum_{T \in \mathcal{F}_{T_0},\ \chi^2(T) \geq \chi^2(T_0)} P(T)$$

of T_0. If $P > 0.05$, then we conclude that H_0 is true, and X and Y are independent.

Unfortunately, the Fisher's exact test also has a problem. In general, it is very difficult to enumerate all elements of $\mathcal{F}_{T_0}$. For example, in Example 1.1, $\mathcal{F}_T$ consists of 1,225,914,276,768,514 elements. In such a case, it is almost impossible to compute P-value of T_0 exactly.

Here we are in the position to introduce *Markov Chain Monte Carlo method* (called MCMC method for short). For the computation of P-value, we give up above exact calculation and make use of the Markov chain to do the sampling from $\mathcal{F}_{T_0}$.

Note that, if both T and T' belong to $\mathcal{F}_{T_0}$, then $T - T'$ is an integer $I \times J$ matrix such that the sum of all entries of each rows and each columns is zero. Let $\mathcal{M}_{I\times J}$ denote the set of all integer $I \times J$ matrices which satisfies that the sum of all entries of each rows and each columns is zero. Then a *Markov basis* is a finite subset $\{T_1, \ldots, T_\ell\} \subset \mathcal{M}_{I\times J}$ satisfying that, for any $T, T' \in \mathcal{F}_{T_0}$, there exist $T_{i_1}, \ldots, T_{i_A}$ with $\varepsilon_k = \pm 1$ such that

$$T' = T + \sum_{k=1}^{A} \varepsilon_k T_{i_k},$$

$$T + \sum_{k=1}^{a} \varepsilon_k T_{i_k} \in \mathcal{F}_{T_0}$$

for all $1 \leq a \leq A$.

If a Markov basis $\{T_1, \ldots, T_\ell\}$ is given, then we can construct a Markov chain by the following algorithm:

Metropolis–Hastings algorithm

0. Choose $T \in \mathcal{F}_{T_0}$ at random and set $t = T$;
1. Repeat the following:
   1.1. Select T_i from the uniform distribution on $\{T_1, \ldots, T_\ell\}$;
   1.2. Select ε from the uniform distribution on $\{1, -1\}$ (independent of i);
   1.3. If $t + \varepsilon T_i$ is a nonnegative matrix, then set $t = t + \varepsilon T_i$ with probability

$$\min\left\{\frac{P(t+\varepsilon T_i)}{P(t)}, 1\right\}.$$

Let $\mathcal{B} = \{T_1, \ldots, T_\ell\}$ be a Markov basis and let $G_{\mathcal{B},T_0}$ denote the graph with the vertex set $\mathcal{F}_{T_0}$ where $m \in \mathcal{F}_{T_0}$ and $m' \in \mathcal{F}_{T_0}$ are joined by an edge if $m - m' \in \mathcal{B} \cup -\mathcal{B}$. The most important point is that, the graph $G_{\mathcal{B},T_0}$ must be connected. Otherwise, there exists an unreachable element of $\mathcal{F}_{T_0}$ in Metropolis–Hastings algorithm. The infinite set $\mathcal{M}_{I\times J}$ is regarded as the set of all integer solutions of some system of linear equations, and hence we can associate $\mathcal{M}_{I\times J}$ with the "toric ideal" $I_{A_{r_1\cdots r_n}}$.

A configuration in $\mathbb{R}^d$ is a finite set $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{Z}^d$ which is contained in a hyperplane in $\mathbb{R}^d$ without the origin. Let $K[\mathbf{t}, \mathbf{t}^{-1}] = K[t_1, t_1^{-1}, \ldots, t_d, t_d^{-1}]$ denote the Laurant polynomial ring in d variables over a field K. We associate a configuration $\mathcal{A} \subset \mathbb{Z}^d$ with the semigroup ring $K[\mathcal{A}] = K[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset K[\mathbf{t}, \mathbf{t}^{-1}]$, where $\mathbf{t}^{\mathbf{a}} = t_1^{a_1} \cdots t_d^{a_d}$ if $\mathbf{a} = (a_1, \ldots, a_d)$. Let $K[\mathbf{x}] = K[x_1, \ldots, x_n]$ denote the polynomial ring in n variables over K. The toric ideal $I_{\mathcal{A}}$ of $\mathcal{A}$ is the kernel of the surjective homomorphism $\pi : K[\mathbf{x}] \longrightarrow K[\mathcal{A}]$ defined by setting $\pi(x_i) = \mathbf{t}^{\mathbf{a}_i}$ for $1 \leq i \leq n$. A polynomial $f \in K[\mathbf{x}]$ of the form $u - v$, where u and v are monomials, is called a *binomial.* It is known [14] that the toric ideal $I_{\mathcal{A}}$ is generated by the binomials $u - v$ with $\pi(u) = \pi(v)$.

A configuration $\mathcal{A}$ is called *unimodular* if the initial ideal of $I_{\mathcal{A}}$ is generated by squarefree monomials with respect to any monomial order. A configuration $\mathcal{A}$ is called *compressed* if the initial ideal of $I_{\mathcal{A}}$ is generated by squarefree monomials with respect to any reverse lexicographic order. We are interested in the following conditions:

(i) $\mathcal{A}$ is unimodular;

(ii) $\mathcal{A}$ is compressed;

(iii) there exists a monomial order $<$ such that the initial ideal of $I_{\mathcal{A}}$ with respect to $<$ is generated by squarefree monomials;

(iv) $K[\mathcal{A}]$ is normal.

Then (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) holds and each of the converse of them is false in general. We refer the reader to [8], [9] and [10] for further information.

On the other hand, a binomial f belonging to $I_{\mathcal{A}}$ is called *indispensable* ([15] and [11]) if, for an arbitrary system $\mathcal{F}$ of binomial generators of $I_{\mathcal{A}}$, either f or $-f$ appears in $\mathcal{F}$. If f is indispensable, then $-f$ is indispensable. Hence the set of indispensable binomials is of the form $F \cup -F$, where $F \cap -F = \emptyset$. In abuse of terminology, such a set F will be called *the* set of indispensable binomials of $I_{\mathcal{A}}$.

In the present paper, we study the configuration arising from a $r_1 \times r_2 \times \cdots \times r_n$ contingency table, where $r_1 \geq r_2 \geq \cdots \geq r_n \geq 2$. Let $\mathcal{A}_{r_1 r_2 \cdots r_n}$ be the set of vectors

$$\mathbf{e}^{(1)}_{i_2 i_3 \cdots i_n} \oplus \mathbf{e}^{(2)}_{i_1 i_3 \cdots i_n} \oplus \cdots \oplus \mathbf{e}^{(n)}_{i_1 i_2 \cdots i_{n-1}},$$

where each i_k belongs to $[r_k] = \{1, 2, \ldots, r_k\}$ and $\mathbf{e}^{(k)}_{j_1 j_2 \cdots j_{n-1}}$ is a unit coordinate vector of $\mathbb{Z}^{r_1 \times \cdots r_{k-1} \times r_{k+1} \cdots \times r_n}$. The toric ideal $I_{\mathcal{A}_{r_1 r_2 \cdots r_n}}$ is the kernel of the homomorphism

$$\pi : K[\{x_{i_1 i_2 \cdots i_n} \ ; \ i_k \in [r_k]\}] \longrightarrow K[\{t^{(k)}_{i_1 \cdots i_{k-1} i_{k+1} \cdots i_n} \ ; \ k \in [n], i_k \in [r_k]\}]$$

defined by $\pi(x_{i_1 i_2 \cdots i_n}) = t^{(1)}_{i_2 i_3 \cdots i_n} t^{(2)}_{i_1 i_3 \cdots i_n} \cdots t^{(n)}_{i_1 i_2 \cdots i_{n-1}}$.

Proposition 1.2. *Work with the same notation as above. Then $\mathcal{B} \subset \mathcal{M}_{r_1 \times \cdots \times r_n}$ is a Markov basis for an arbitrary T if and only if the toric ideal $I_{\mathcal{A}_{r_1 \cdots r_n}}$ is generated by the binomials $\mathbf{x}^{\beta^+} - \mathbf{x}^{\beta^-} \in K[\mathbf{x}]$ with $\beta^+ - \beta^- \in \mathcal{B}$.*

2. Recent developments

In the present section, we discuss the recent developments ([12]) on semigroup rings and toric ideals arising from contingency tables. First, using the formula in [13, p. 162], we can compute the dimension of $K[\mathcal{A}_{r_1, \ldots, r_n}]$.

Proposition 2.1. *The dimension of $K[\mathcal{A}_{r_1, \ldots, r_n}]$ is equal to*

$$(-1)^{n-1} + \sum_{k=1}^{n-1} (-1)^{n-k-1} \sum_{i_1 < \cdots < i_k} r_{i_1} \cdots r_{i_k}$$

Indispensable binomials have been completely determined for the following three classes of $I_{\mathcal{A}_{r_1 r_2 \cdots r_n}}$.

(1) $n = 2$ (unimodular, Segre product of polynomial rings),
(2) $n \geq 3$ and $r_1 \times r_2 \times 2 \times \cdots \times 2$ (Lawrence lifting)
(3) $r_1 \times 3 \times 3$, $r_1 \times 4 \times 3$, $4 \times 4 \times 4$ (computed by Aoki–Takemura [2], [3]).

In particular, for all of (1) – (3), a minimal set of binomial generators is unique.

Conjecture 2.2. *The toric ideal of the configuration $\mathcal{A}_{r_1 r_2 \cdots r_n}$ is generated by indispensable binomials.*

On the other hand, Boffi–Rossi [5] computed a lexicographic Gröbner basis of the toric ideal $I_{\mathcal{A}_{r_1 33}}$ whose initial ideal is generated by squarefree monomials. We now discuss the following two properties of the toric ideal of the configuration $\mathcal{A}_{333}$.

Theorem 2.3. *No reduced Gröbner basis of $I_{\mathcal{A}_{333}}$ coincides with the set of indispensable binomials (=* the *minimal set of binomial generators) of $I_{\mathcal{A}_{333}}$.*

Theorem 2.4. *The configuration $\mathcal{A}_{333}$ is compressed.*

We characterize the configurations $\mathcal{A}_{r_1r_2\cdots r_n}$ for which there exists a monomial order $<$ such that the reduced Gröbner basis of $I_{\mathcal{A}_{r_1r_2\cdots r_n}}$ with respect to $<$ is the set of indispensable binomials of $I_{\mathcal{A}_{r_1r_2\cdots r_n}}$.

Theorem 2.5. *Let $n \leq 3$. Then the following conditions are equivalent for $\mathcal{A}_{r_1r_2\cdots r_n}$:*
(i) *either $n = 2$ or $r_3 = 2$;*
(ii) *$\mathcal{A}_{r_1r_2\cdots r_n}$ is unimodular ;*
(iii) *there exists a monomial order $<$ such that the reduced Gröbner basis of $I_{\mathcal{A}_{r_1r_2\cdots r_n}}$ with respect to $<$ is the set of indispensable binomials of $I_{\mathcal{A}_{r_1r_2\cdots r_n}}$;*
(iv) *there exists a monomial order $<$ such that the reduced Gröbner basis of $I_{\mathcal{A}_{r_1r_2\cdots r_n}}$ with respect to $<$ is a minimal set of binomial generators of $I_{\mathcal{A}_{r_1r_2\cdots r_n}}$.*

We study normality of semigroup rings arising from contingency tables. We classify all normal semigroup rings $K[\mathcal{A}_{r_1r_2\cdots r_n}]$ except for $K[\mathcal{A}_{553}]$, $K[\mathcal{A}_{543}]$ and $K[\mathcal{A}_{443}]$.

Theorem 2.6. *Work with the same notation as above. Then we have*

$r_1 \times r_2$ *or* $r_1 \times r_2 \times 2 \times \cdots \times 2$	*unimodular*
$r_1 \times 3 \times 3$	*normal*
$5 \times 5 \times 3$ *or* $5 \times 4 \times 3$ *or* $4 \times 4 \times 3$	*UNKNOWN if normal or not*
otherwise, i.e., $n \geq 4$ *and* $r_3 \geq 3$ $n = 3$ *and* $r_3 \geq 4$ $n = 3$, $r_3 = 3$, $r_1 \geq 6$ *and* $r_2 \geq 4$	*not normal*

Even though a unique minimal set of generators of $I_{\mathcal{A}_{r_143}}$ is given in [3], it seems to be difficult to know if $K[\mathcal{A}_{543}]$ and $K[\mathcal{A}_{443}]$ are normal. On the other hand, we do not know if $\mathcal{A}_{r_133}$ with $r_1 \geq 4$ are compressed.

Question 2.7. *Are $K[\mathcal{A}_{553}]$, $K[\mathcal{A}_{543}]$ and $K[\mathcal{A}_{443}]$ normal?*

Question 2.8. *Is $\mathcal{A}_{r_133}$ compressed for $r_1 \geq 4$?*

References

[1] A. Agresti, "Categorical Data Analysis," John Wiley & Sons, 2002.
[2] S. Aoki and A. Takemura, Minimal basis for connected Markov chain over $3\times3\times K$ contingency tables with fixed two-dimensional marginals, *Australian and New Zealand Journal of Statistics*, **45** (2003), 229 – 249.
[3] S. Aoki and A. Takemura, The list of indispensable moves of the unique minimal Markov basis for $3 \times 4 \times K$ and $4 \times 4 \times 4$ contingency tables with fixed two-dimensional marginals, *METR Technical Report*, 03-38.
[4] G. Boffi and F. Rossi, Gröbner bases related to 3-dimensional transportation problems, *Quaderni matematici Univ. Trieste*, II, **482** (2000).

[5] G. Boffi and F. Rossi, A stability property of certain Gröbner bases, preprint, Feb. 2003.
[6] P. Diaconis and B. Sturmfels, Algebraic algorithms for sampling from conditional distributions, *Annals of Statistics* **26** (1998), 363 – 397.
[7] H. Ohsugi, J. Herzog and T. Hibi, Combinatorial pure subrings, *Osaka J. Math.*, **37** (2000), 745 – 757.
[8] H. Ohsugi and T. Hibi, A normal (0,1)-polytope none of whose regular triangulations is unimodular, *Discrete Compt. Geom.* **21** (1999), 201 – 204.
[9] H. Ohsugi and T. Hibi, Compressed polytopes, initial ideals and complete multipartite graphs, *Illinois J. Math.*, **44** (2000), 391 – 406.
[10] H. Ohsugi and T. Hibi, Convex polytopes all of whose reverse lexicographic initial ideals are squarefree, *Proc. Amer. Math. Soc.*, **129** (2001), Number 9, 2541 – 2546.
[11] H. Ohsugi and T. Hibi, Indispensable binomials of finite graphs, *J. Algebra and its applications*, to appear.
[12] H. Ohsugi and T. Hibi, Toric ideals arising from contingency tables, preprint, 2004.
[13] F. Santos and B. Sturmfels, Higher Lawrence configuraions, *J. Combinatorial Theory, Ser. A*, **103** (2003), 151 – 164.
[14] B. Sturemfels, "Gröbner bases and convex polytopes," Amer. Math. Soc., Providence, RI, 1995.
[15] A. Takemura and S. Aoki, Some characterizations of minimal Markov basis for sampling from discrete conditional distributions, *Annals of the Institute of Statistical Mathematics*, (2003), to appear.

Hidefumi Ohsugi
Department of Mathematics
Rikkyo University
Toshima, Tokyo 171-8501, Japan
E-mail:ohsugi@rkmath.rikkyo.ac.jp

Takayuki Hibi
Department of Pure and Applied Mathematics
Graduate School of Information Science and Technology
Osaka University
Toyonaka, Osaka 560-0043, Japan
E-mail:hibi@math.sci.osaka-u.ac.jp