3. Gröbner bases theory in sampling problems of contingency tables

1. Conditional tests of contingency tables

- Example: Grades of Geometry and Probability

Geo\Prob	5	4	3	2	1-	Total
5	2	1	1	0	0	4
4	8	3	3	0	0	14
3	0	2	1	1	1	5
2	0	0	0	1	1	2
1-	0	0	0	0	1	1
Total	10	6	5	2	3	26

- 5×5 contingency table
- H_0 : Two grades are independent

- Tests of independence model for two-way contingency tables
  - x_{ij} : frequency of grades (i,j), $\mathbf{x} = (x_{ij})$
    $x_{i+} = \sum_j x_{ij}$, $x_{+j} = \sum_i x_{ij}$, $x_{++} = \sum_i \sum_j x_{ij}$
  - p_{ij} : probability of grades (i,j), $\sum_i \sum_j p_{ij} = 1$
  - H_0 : independence model

$$\exists \{\alpha_i\}, \{\beta_j\} \text{ s.t. } p_{ij} = \alpha_i \beta_j, \quad \forall i, j$$

  - $\mathbf{m} = (m_{ij})$: fitted value under H_0 (MLE)

$$m_{ij} = x_{++} \hat{p}_{ij} = x_{++} \frac{x_{i+}}{x_{++}} \frac{x_{+j}}{x_{++}} = \frac{x_{i+} x_{+j}}{x_{++}}$$

  - goodness-of-fit χ^2 test

$$\chi^2(\mathbf{x}) = \sum_i \sum_j \frac{(x_{ij} - m_{ij})^2}{m_{ij}} > c_\alpha \Rightarrow \mathrm{H}_0 \text{ reject}$$

Data $\mathbf{x}^o = (x^o_{ij})$

	5	4	3	2	1−	
5	2	1	1	0	0	4
4	8	3	3	0	0	14
3	0	2	1	1	1	5
2	0	0	0	1	1	2
1−	0	0	0	0	1	1
	10	6	5	2	3	26

Fitted value $\mathbf{m} = (m_{ij})$

	5	4	3	2	1−	
5	1.54	0.92	0.77	0.31	0.46	4
4	5.38	3.23	2.69	1.08	1.62	14
3	1.92	1.15	0.96	0.38	0.58	5
2	0.77	0.46	0.38	0.15	0.23	2
1−	0.38	0.23	0.19	0.08	0.12	1
	10	6	5	2	3	26

$$\chi^2(\mathbf{x}^o) = \sum_i \sum_j \frac{(x^o_{ij} - m_{ij})^2}{m_{ij}} = 25.338$$

If this value is significantly large, we can reject H_0 and conclude "There are some relations between the two grades".

- Question: Is $\chi^2(\mathbf{x}^o) = 25.338$ significantly large?

  $\Longrightarrow$ Judge whether 25.338 is large enough or not by p value.

> <u>p value</u>
>
> The probability that the value of the test statistics is larger than the observed value under H_0.
>
> $$p \text{ value} = P(\chi^2(\mathbf{x}) \geq 25.338 \mid \mathrm{H}_0)$$

- Strategies of calculating p value

  (a) Asymptotic theory

  (b) Exact calculation

  (c) Monte Carlo method

(a) Asymptotic theory

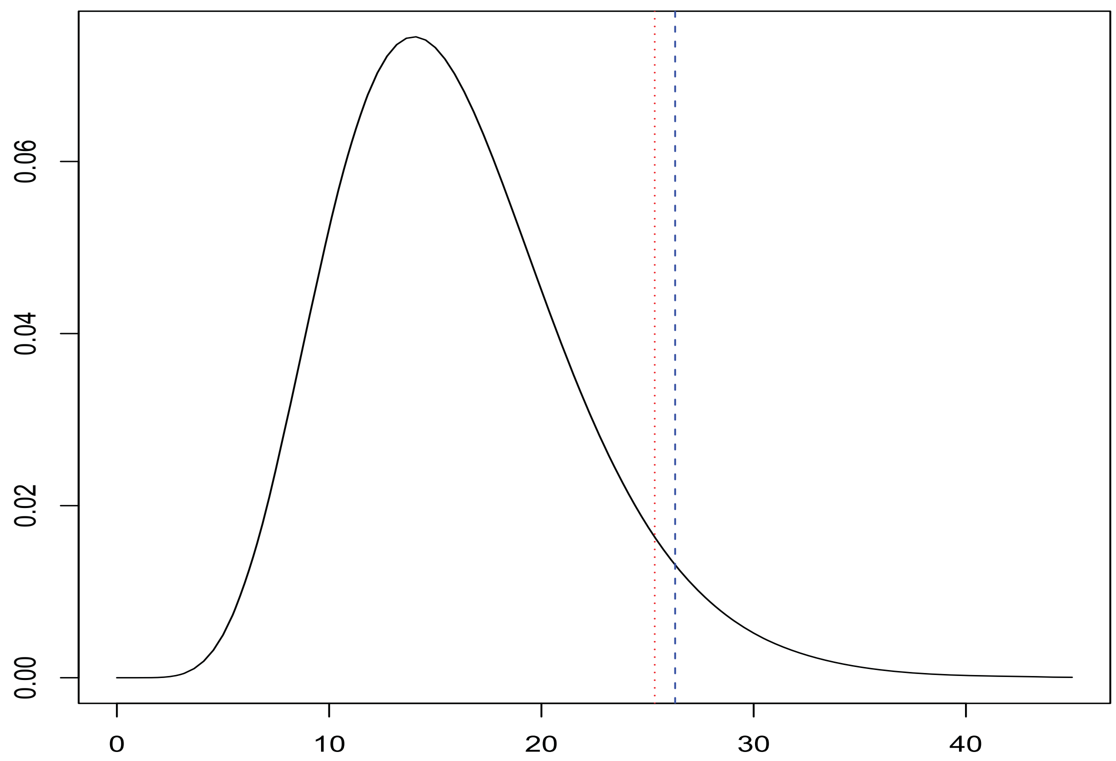
- The asymptotic distribution of $\chi^2(\mathbf{x})$ under H_0 when $n \to \infty$ is the χ^2 distribution with $(5-1)(5-1) = 16$ degree of freedom.
  (Limit is considered under some regularity condition.)

$$\Downarrow$$

- Compare the observed $\chi^2(\mathbf{x}^o) = 25.338$ with the upper probability of the asymptotic distribution χ^2_{16}.
  H_0 is not rejected by the significant level of 5% because $\chi^2_{16,0.05} = 26.30$.

- Simple and easy. However, the fitting of the asymptotic distribution can be very poor, especially for sparse contingency tables.

- χ^2_{16} distribution
  blue: upper 5% point (26.30), red: observed $\chi^2(\mathbf{x}^o) = 25.338$



(*b*) Exact calculation (exact test)

- The distribution of $\mathbf{x}$ under H_0 (hypergeometric distribution):

$$p(\mathbf{x}) = \frac{\left(\prod_i x_{i+}!\right)\left(\prod_j x_{+j}!\right)}{x_{++}! \prod_{i,j} x_{ij}!} \quad \text{for} \quad \mathbf{x} \in \mathcal{F}$$

- The set of contingency tables with the same row sums and column sums to $\mathbf{x}^o$:

$$\mathcal{F} = \left\{ \mathbf{x} = (x_{ij}) \;\middle|\; x_{i+} = x^o_{i+},\ x_{+j} = x^o_{+j},\ x_{ij} \in \mathbb{Z}_{\geq 0} \right\}$$

- The exact p value of the goodness-of-fit χ^2 :

$$p = \Pr(\chi^2(\mathbf{x}) \geq \chi^2(\mathbf{x}^o) \mid \mathrm{H}_0) = \sum_{\mathbf{x} \in \mathcal{F}} g(\mathbf{x}) p(\mathbf{x}),$$

$$g(\mathbf{x}) = \begin{cases} 1, & \text{if } \chi^2(\mathbf{x}) \geq \chi^2(\mathbf{x}^o), \\ 0, & \text{otherwise} \end{cases}$$

- Best method, if possible. However, the cardinality of $\mathcal{F}$ becomes huge as the sample size increases, and an exact calculation becomes infeasible.

$$p(\mathbf{x}) = \frac{(4!14!5!2!1!)\,(10!6!5!2!3!)}{26!} \prod_{i=1}^{5} \prod_{j=1}^{5} \frac{1}{x_{ij}!}$$

$$\mathcal{F} = \left\{ \mathbf{x} = (x_{ij}) \;\middle|\; \begin{array}{|ccccc|c} \hline x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & 4 \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} & 14 \\ x_{31} & x_{32} & x_{33} & x_{34} & x_{35} & 5 \\ x_{41} & x_{42} & x_{43} & x_{44} & x_{45} & 2 \\ x_{51} & x_{52} & x_{53} & x_{54} & x_{55} & 1 \\ \hline 10 & 6 & 5 & 2 & 3 & 26 \end{array} \right\} \quad (\#\mathcal{F} = 229,174)$$

$$\chi^2(\mathbf{x}^o) = 25.338$$

$$p = \Pr(\chi^2(\mathbf{x}) \geq 25.338 \mid \mathrm{H}_0) = \sum_{\mathbf{x}\in\mathcal{F}} g(\mathbf{x})p(\mathbf{x}) = 0.0609007$$

$$g(\mathbf{x}) = \begin{cases} 1, & \text{if } \chi^2(\mathbf{x}) \geq 25.338, \\ 0, & \text{otherwise} \end{cases}$$

- H_0 cannot be rejected by the significant level 5%.

There is a problem when
the sample size is too large for an exact calculation, and
the fit of asymptotic distribution is poor due to sparseness.
What is an efficient strategy for such problem?

(*c*) Monte Carlo method

Estimate $p = \sum_{\mathbf{x} \in \mathcal{F}} g(\mathbf{x}) p(\mathbf{x})$ as $\hat{p} = \frac{1}{N} \sum_{t=1}^{N} g(\mathbf{x}_t)$

by samples $\mathbf{x}_1, ..., \mathbf{x}_N$ from $p(\mathbf{x})$.

- Some advantages:
  We can estimate the variance of $\hat{p}$.
  (For example, a conventional 95% CL: $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/N}$)
  We can apply this method to arbitrary test statistics.
- However, it is difficult to generate samples if $p(\mathbf{x})$ is complicated.

In this topic, we generate samples using Markov chain (Markov chain Monte Carlo method).

- Markov chain Monte Carlo method (MCMC method)

  Construct an ergodic Markov chain with the stationary distribution $p(\mathbf{x})$, conditional distribution under H_0.

  From some initial state, we move states for some large (100,000 for example) times and discards them, then use the subsequent Markov process as the samples from $p(\mathbf{x})$.

$$\mathbf{x}^{(0)} \rightarrow \mathbf{x}^{(1)} \rightarrow \mathbf{x}^{(2)} \rightarrow \cdots \rightarrow \underbrace{\mathbf{x}^{(100000)} \rightarrow \mathbf{x}^{(100001)} \rightarrow \cdots}_{\text{samples from } p(\mathbf{x})}$$

  Point: Any connected Markov chain over $\mathcal{F}$ can be modified so as to have the stationary distribution $p(\mathbf{x})$.
  (Metropolis-Hastings algorithm)

- $\mathcal{F} = \{\mathbf{x}_1, \ldots, \mathbf{x}_s\}$ (appropriately numbered)
- $\pi = (\pi_1, \ldots, \pi_s)$: distribution under H_0, $\pi_i = p(\mathbf{x}_i)$
- $\{Z_t,\ t = 0, 1, 2, \ldots\}$: Markov chain
- Transition probability matrix $Q = (q_{ij})$:

$$q_{ij} = \Pr(Z_{t+1} = \mathbf{x}_j \mid Z_t = \mathbf{x}_i)$$

- π is called a stationary distribution if it satisfies

$$\pi = \pi Q.$$

- **Theorem** (Metropolis-Hastings)
  Let π be a probability distribution on $\mathcal{F}$.
  Let $R = (r_{ij})$ be the transition probability matrix of a connected, aperiodic, and symmetric Markov chain over $\mathcal{F}$.
  Define $Q = (q_{ij})$ by

$$q_{ij} = r_{ij} \min\left(1, \frac{\pi_j}{\pi_i}\right), \; i \neq j,$$
$$q_{ii} = 1 - \sum_{j \neq i} q_{ij}.$$

  Then Q satisfies $\pi = \pi Q$.
- Then we only need to construct arbitrary connected, symmetric Markov chain over $\mathcal{F}$.
- An important advantage: it does not require the explicit evaluation of the normalizing constant of π.

2. Definition of Markov bases

- $\mathbf{x} = (x_{11}, \ldots, x_{IJ})' \in \mathbb{Z}^p_{\geq 0}$: contingency table (column vector)
- $\mathbf{t} = A\mathbf{x} \in \mathbb{Z}^d$: fixed marginal sums

  Example: independence model of 2×3 table:

$$\begin{pmatrix} x_{1+} \\ x_{2+} \\ x_{+1} \\ x_{+2} \\ x_{+3} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{21} \\ x_{22} \\ x_{23} \end{pmatrix}$$

- **Definition**

  $A \in \mathbb{Z}^{d\times p}$ is a configuration matrix if $(1,\dots,1)$ is in the row space of A.

- We assume A is a configuration matrix hereafter. This is a natural assumption for natural statistical models.

- $\mathbf{t}$-fiber: the set of tables with the same fixed marginals to $\mathbf{t}$

$$\mathcal{F}_{\mathbf{t}} = \{\mathbf{x} \in \mathbb{Z}^{p}_{\geq 0} \mid A\mathbf{x} = \mathbf{t}\}$$

  We want to construct a connected Markov chain over $\mathbf{t}$-fiber including the observation $\mathbf{x}^o$, to calculate the p value of the conditional test by MCMC method.

- Integer kernel of the configuration matrix $A \in \mathbb{Z}^{d\times p}$:

$$\mathrm{Ker}_{\mathbb{Z}}(A) = \{\mathbf{z} \in \mathbb{Z}^p \mid A\mathbf{z} = \mathbf{0}\}$$

  We call an element in $\mathrm{Ker}_{\mathbb{Z}}(A)$ a move for A.

- Definition $\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\mathbf{t}}$ is mutually accessible by $\mathcal{B} \subset \mathrm{Ker}_{\mathbb{Z}}(A)$ if there exists $N > 0,\ \mathbf{z}_j \in \mathcal{B},\ \varepsilon_j \in \{-1, 1\},\ j = 1, \ldots, N$, satisfying

$$\mathbf{y} = \mathbf{x} + \sum_{j=1}^{N} \varepsilon_j \mathbf{z}_j \text{ and } \mathbf{x} + \sum_{j=1}^{n} \varepsilon_j \mathbf{z}_j \in \mathcal{F}_{\mathbf{t}},\ n = 1, \ldots, N$$

- Mutual accessibility by $\mathcal{B}$ is an equivalence relation and each $\mathcal{F}_{\mathbf{t}}$ is partitioned into disjoint equivalence classes by $\mathcal{B}$.
  We call them $\mathcal{B}$-equivalence classes of $\mathcal{F}_{\mathbf{t}}$.

- Definition (Diaconis and Sturmfels, 1998)
  A finite set of moves $\mathcal{B} \subset \mathrm{Ker}_{\mathbb{Z}}(A)$ is called a Markov basis (for A) if $\mathcal{F}_{\mathbf{t}}$ itself is a $\mathcal{B}$-equivalence class for arbitrary $\mathbf{t}$.
- Once a Markov basis is obtained, it is easy to construct a connected and symmetric Markov chain over the $\mathbf{t}$-fiber containing any observation. Therefore we can estimate p values of the test statistics by MCMC method.
- Markov basis always exists for arbitrary A.
  (Hilbert Basis Theorem)
- There is an algorithm to calculate a Markov basis for arbitrary A.

3. Example of Markov bases: Independent model of two-way contingency tables

- $\mathbf{x} = (x_{11}, \ldots, x_{IJ})' \in \mathbb{Z}^p_{\geq 0}$: $I \times J$ contingency table $(p = IJ)$

$$\begin{array}{|ccc|} \hline x_{11} & \cdots & x_{1J} \\ \vdots & & \vdots \\ x_{I1} & \cdots & x_{IJ} \\ \hline \end{array}$$

- Fixed row and column sums : $\mathbf{t} = (x_{1+}, \ldots, x_{I+}, x_{+1}, \ldots, x_{+J})'$
- We want to construct a connected Markov chain over

$$\mathcal{F}_{A\mathbf{x}^o} = \{\mathbf{x} \in \mathbb{Z}^p_{\geq 0} \mid A\mathbf{x} = A\mathbf{x}^o\}$$

for arbitrary given $\mathbf{x}^o \in \mathbb{Z}^p_{\geq 0}$.

It is easy.

Connected Markov chain over $\mathcal{F}_{\mathbf{t}}$:

$$\text{The set of the moves} \quad \begin{matrix} + & - \\ - & + \end{matrix} \quad \text{is a Markov basis.}$$

Example of algorithms :

- $\mathbf{x} \in \mathcal{F}_{\mathbf{t}}$: Initial state
- Choose pair of rows and pair of columns randomly.
- Choose sign $\begin{matrix} + & - \\ - & + \end{matrix}$ or $\begin{matrix} - & + \\ + & - \end{matrix}$ randomly.
- Update 4 entries of $\mathbf{x} \longrightarrow \mathbf{y}$
- If $\mathbf{y} \in \mathcal{F}_{\mathbf{t}}$, then $\mathbf{y}$ is the next state.
  If $\mathbf{y} \notin \mathcal{F}_{\mathbf{t}}$ (i.e., negative entries appear), then stay at $\mathbf{x}$.

$$
\cdots \quad \Rightarrow \quad \begin{array}{|cccc|} \hline 2 & 0 & 3 & 1 \\ 1 & 1 & 0 & 3 \\ 0 & 2 & 1 & 1 \\ \hline \end{array} \begin{array}{c} \\ \leftarrow \\ \leftarrow \end{array} \quad \Rightarrow \quad \begin{array}{|cccc|} \hline 2 & 0 & 3 & 1 \\ 1 & 2 & 0 & 2 \\ 0 & 1 & 1 & 2 \\ \hline \end{array} \quad \Rightarrow \quad \cdots
$$

$$
\begin{array}{cccc} & \uparrow & & \uparrow \end{array}
$$

Markov chain constructed in this way is connected over $\mathcal{F}_{\mathbf{t}}$ for arbitrary $\mathbf{t}$.

4. Example of Markov bases: No three-factor interaction model of three-way tables

$$\mathbf{x} = (x_{ijk}) \in \mathbb{Z}_{\geq 0}^{IJK}, \quad 1 \leq i \leq I,\ 1 \leq j \leq J,\ 1 \leq k \leq K$$

$$\boxed{\begin{array}{ccc} x_{111} & \cdots & x_{11K} \\ \vdots & & \vdots \\ x_{1J1} & \cdots & x_{1JK} \end{array}} \quad \boxed{\begin{array}{ccc} x_{211} & \cdots & x_{21K} \\ \vdots & & \vdots \\ x_{2J1} & \cdots & x_{2JK} \end{array}} \quad \cdots \quad \boxed{\begin{array}{ccc} x_{I11} & \cdots & x_{I1K} \\ \vdots & & \vdots \\ x_{IJ1} & \cdots & x_{IJK} \end{array}}$$

- $\mathbf{t} = A\mathbf{x}$: two-dimensional marginals $\{x_{ij+}\}, \{x_{i+k}\}, \{x_{+jk}\}$
- We want to construct a connected Markov chain over

$$\mathcal{F}_{A\mathbf{x}^o} = \{\mathbf{x} \in \mathbb{Z}_{\geq 0}^{IJK} \mid A\mathbf{x} = A\mathbf{x}^o\}$$

for arbitrary given $\mathbf{x}^o \in \mathbb{Z}_{\geq 0}^{IJK}$.

surprisingly difficult !

Remark : The corresponding model is important in applied statistics.

- Log-linear model

  $\log p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$

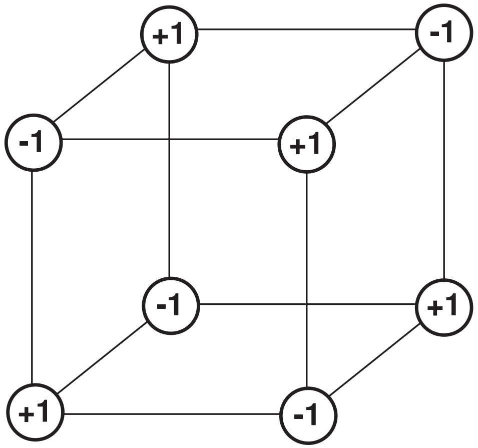
- Null hypothesis: "There is no three-factor interaction"

$$\mathrm{H}_0 : (\alpha\beta\gamma)_{ijk} = 0, \quad \text{for all } i, j, k$$

- Fixed marginals for the test H_0: $\{x_{ij+}\}$, $\{x_{i+k}\}$, $\{x_{+jk}\}$
- Conditional probability function:
  $p(x_{ijk} \mid x_{ij+}, x_{i+k}, x_{+jk}) \propto \left(\prod_{i,j,k} x_{ijk}!\right)^{-1}$

A direct sampling from $p(x_{ijk} \mid x_{ij+}, x_{i+k}, x_{+jk})$ is difficult, because H_0 is not a decomposable model. We are happy if we can use MCMC.

Simple extension of 2×2 move ($2 \times 2 \times 2$ basic move)



$i = i_1$

$j \backslash k$	k_1	k_2
j_1	$+1$	-1
j_2	-1	$+1$

$i = i_2$

$j \backslash k$	k_1	k_2
j_1	-1	$+1$
j_2	$+1$	-1

This is a natural extension of the basic move of two-way contingency tables, $\begin{matrix} + & - \\ - & + \end{matrix}$, to the three-way contingency tables.

$\longrightarrow$ Markov chain constructed from the basic move is not connected in general.
(The set of the basic move is not a Markov basis.)

Example: $3 \times 3 \times 3$ contingency tables ($\#\mathcal{F}_{\mathbf{t}} = 18$)

$$\mathbf{t} \;:\; \begin{array}{|ccc|c} \hline & & & 2 \\ & & & 1 \\ & & & 1 \\ \hline 2 & 1 & 1 & 4 \end{array} \quad \begin{array}{|ccc|c} \hline & & & 1 \\ & & & 2 \\ & & & 1 \\ \hline 1 & 2 & 1 & 4 \end{array} \quad \begin{array}{|ccc|c} \hline & & & 1 \\ & & & 1 \\ & & & 2 \\ \hline 1 & 1 & 2 & 4 \end{array} \quad \begin{array}{|ccc|c} \hline 2 & 1 & 1 & 4 \\ 1 & 2 & 1 & 4 \\ 1 & 1 & 2 & 4 \\ \hline 4 & 4 & 4 & 12 \end{array}$$

$1: \boxed{\begin{matrix} 2&0&0\\0&1&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\1&1&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\0&0&1\\1&1&0 \end{matrix}}$

$2: \boxed{\begin{matrix} 2&0&0\\0&1&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\1&0&1\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\0&1&0\\1&0&1 \end{matrix}}$

$3: \boxed{\begin{matrix} 2&0&0\\0&1&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\0&1&1\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\1&0&0\\0&1&1 \end{matrix}}$

$4: \boxed{\begin{matrix} 2&0&0\\0&1&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\1&1&0\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\0&0&1\\1&0&1 \end{matrix}}$

$5: \boxed{\begin{matrix} 2&0&0\\0&0&1\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\1&1&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\0&1&0\\1&0&1 \end{matrix}}$

$6: \boxed{\begin{matrix} 2&0&0\\0&0&1\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\0&2&0\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\1&0&0\\0&0&2 \end{matrix}}$

$7: \boxed{\begin{matrix} 1&1&0\\1&0&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&2&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\0&0&1\\1&1&0 \end{matrix}}$

$8: \boxed{\begin{matrix} 1&1&0\\1&0&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&1&1\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\0&1&0\\1&0&1 \end{matrix}}$

$9: \boxed{\begin{matrix} 1&1&0\\1&0&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\0&2&0\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&0&1\\0&1&1 \end{matrix}}$

$10: \boxed{\begin{matrix} 1&1&0\\0&0&1\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&2&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\1&0&0\\0&1&1 \end{matrix}}$

$11: \boxed{\begin{matrix} 1&1&0\\0&0&1\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 0&0&1\\1&1&0\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&1&0\\0&0&2 \end{matrix}}$

$12: \boxed{\begin{matrix} 1&0&1\\1&0&0\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&2&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\0&0&1\\1&0&1 \end{matrix}}$

$13: \boxed{\begin{matrix} 1&0&1\\1&0&0\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\0&1&1\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&1&0\\0&0&2 \end{matrix}}$

$14: \boxed{\begin{matrix} 1&0&1\\0&1&0\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&1&1\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\1&0&0\\0&0&2 \end{matrix}}$

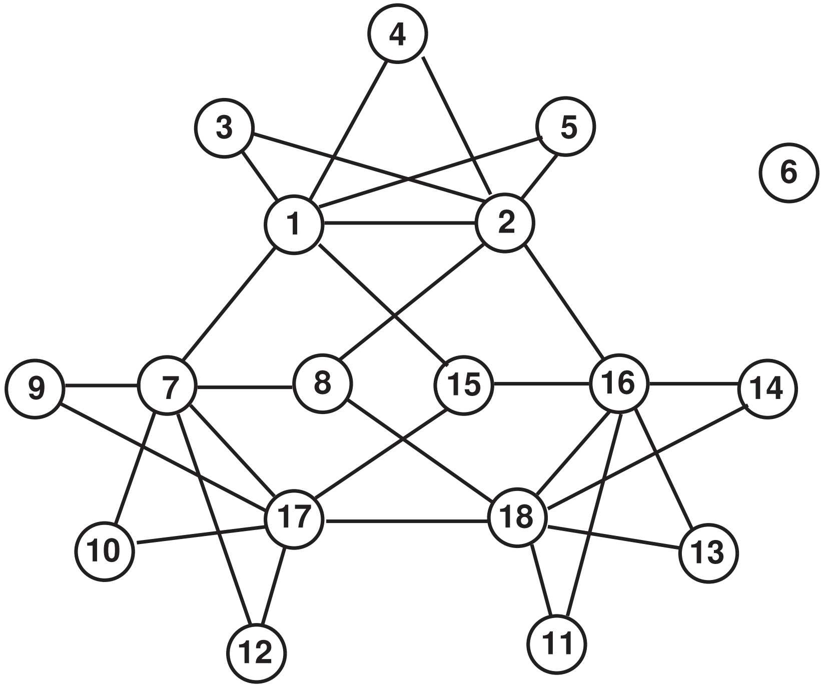
$15: \boxed{\begin{matrix} 1&0&1\\0&1&0\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\1&1&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&0&1\\0&1&1 \end{matrix}}$

$16: \boxed{\begin{matrix} 1&0&1\\0&1&0\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 0&1&0\\1&0&1\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&1&0\\0&0&2 \end{matrix}}$

$17: \boxed{\begin{matrix} 0&1&1\\1&0&0\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&2&0\\0&0&1 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&0&1\\0&1&1 \end{matrix}}$

$18: \boxed{\begin{matrix} 0&1&1\\1&0&0\\1&0&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&1&1\\0&1&0 \end{matrix}} \boxed{\begin{matrix} 1&0&0\\0&1&0\\0&0&2 \end{matrix}}$

Transition graph obtained from the set of basic moves



State 6: $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

is isolated, and there are two equivalence classes in $\mathcal{F}_{\mathbf{t}}$.

Adding a move

$i = i_1$

$j \backslash k$	k_1	k_2
j_1	$+1$	-1
j_2	-1	$+1$

$i = i_2$

$j \backslash k$	k_1	k_2
j_1	-1	$+1$
j_2	$+1$	-1

for any pairs of i, j, k forces some negative entries.

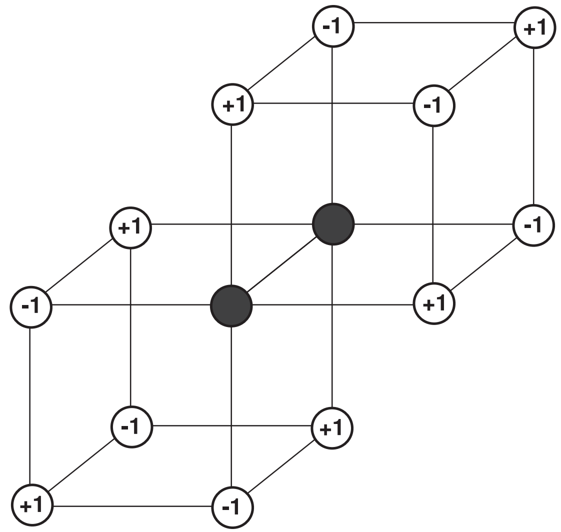
In $3 \times 3 \times 3$ case, moves of degree 6:

$i = i_1$

$j \backslash k$	k_1	k_2	k_3
j_1	0	$+1$	-1
j_2	-1	0	$+1$
j_3	$+1$	-1	0

$i = i_2$

$j \backslash k$	k_1	k_2	k_3
j_1	0	-1	$+1$
j_2	$+1$	0	-1
j_3	-1	$+1$	0



are needed. In fact, {basic move, moves of degree 6} is a Markov basis for this case.

5. Markov bases and ideals

- Contingency table: $\mathbf{x} = (x_1, \ldots, x_p)'$,
  Fixed marginal: $\mathbf{t} = (t_1, \ldots, t_d)'$
  $\mathbf{t} = A\mathbf{x}$, $A = (a_{ij}) \in \mathbb{Z}^{d \times p}$: configuration matrix

$$\mathrm{Ker}_{\mathbb{Z}}(A) = \{\mathbf{z} \in \mathbb{Z}^p \mid A\mathbf{z} = \mathbf{0}\}$$

- $K[\mathbf{u}] = K[u_1, \ldots, u_p]$: polynomial ring over a field K,
  $\mathbf{u} = \{u_1, \ldots, u_p\}$: variables

- For each $\mathbf{z} = (z_1, \ldots, z_p)' \in \mathrm{Ker}_{\mathbb{Z}}(A)$, define a binomial $f_{\mathbf{z}} \in K[\mathbf{u}]$ by

$$f_{\mathbf{z}} = \prod_{z_i > 0} u_i^{z_i} - \prod_{z_i < 0} u_i^{-z_i}.$$

  $f_{\mathbf{z}}$ is homogeneous if A is a configuration.

- Example: $\mathbf{u} = \{u_{11}, \ldots, u_{33}\}$

$$\mathbf{z} : \boxed{\begin{array}{ccc} 2 & -2 & 0 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{array}} \quad \Leftrightarrow \quad f_{\mathbf{z}} = u_{11}^2 u_{23} u_{32}^2 - u_{12}^2 u_{21} u_{31} u_{33}$$

- **Definition**

  Let $A \in \mathbb{Z}^{d \times p}$ be a configuration matrix.

  The binomial ideal in $K[\mathbf{u}]$,

$$I_A = \langle \{ f_{\mathbf{z}} \mid \mathbf{z} \in \mathrm{Ker}_{\mathbb{Z}}(A) \} \rangle,$$

  is called a toric ideal of the configuration A.

- **Theorem** (Diaconis and Sturmfels, 1998)

$$\{\mathbf{z}_1, \dots, \mathbf{z}_L\}, \mathbf{z}_i \in \mathrm{Ker}_{\mathbb{Z}}(A) \text{ is a Markov basis for } A.$$

$$\iff I_A = \langle f_{\mathbf{z}_1}, \dots, f_{\mathbf{z}_L} \rangle$$

- Example: Independence model of two-way tables.

$$I_A = \langle u_{ij}u_{i'j'} - u_{ij'}u_{i'j},\ 1 \le i < i' \le I, 1 \le j < j' \le J \rangle$$

Therefore the set of the moves

$$u_{ij}u_{i'j'} - u_{ij'}u_{i'j} \Leftrightarrow \begin{array}{c|cc|} & j & j' \\ \hline i & 1 & -1 \\ i' & -1 & 1 \\ \hline \end{array}$$

is a Markov basis.

- Computation of Markov basis.
  - $A = (a_{ij}) \in \mathbb{Z}^{d \times p}$: configuration matrix.
  - $K[\mathbf{v}] = K[v_1, \ldots, v_d]$: polynomial ring
  - Homomorphism

$$\begin{array}{rcccl} \psi_A & : & K[\mathbf{u}] & \to & K[\mathbf{v}] \\ & & u_j & \mapsto & v_1^{a_{1j}} v_2^{a_{2j}} \cdots v_d^{a_{dj}} \end{array}$$

  - Toric ideal I_A is also expressed as $I_A = \mathrm{Ker}(\psi_A)$.

◦ Example: Independence model of two-way tables.

u_{11}	$\cdots$	u_{1J}	r_1
$\vdots$		$\vdots$	$\vdots$
u_{I1}	$\cdots$	u_{IJ}	r_I
c_1	$\cdots$	c_J	

The homomorphism is

$$\psi_A(u_{ij}) = r_j c_j \quad \text{for } i = 1, \ldots, I, j = 1, \ldots, J.$$

For a move $f_{\mathbf{z}} = u_{ij}u_{i'j'} - u_{ij'}u_{i'j}$,

$$\begin{aligned} \psi_A(u_{ij}u_{i'j'} - u_{ij'}u_{i'j}) &= \psi_A(u_{ij})\psi_A(u_{i'j'}) - \psi_A(u_{ij'})\psi_A(u_{i'j}) \\ &= (r_i c_j)(r_{i'}c_{j'}) - (r_i c_{j'})(r_{i'}c_j) = 0, \end{aligned}$$

i.e., $f_{\mathbf{z}} \in \mathrm{Ker}(\psi_A)$.

- **Corollary** Define an ideal of $K[\mathbf{u}, \mathbf{v}]$ as

$$I_A^* = \langle -\psi_A(u_j) + u_j,\ j = 1, \ldots, p \rangle \subset K[\mathbf{u}, \mathbf{v}].$$

  Then we have $I_A = I_A^* \cap K[\mathbf{u}]$.

- Therefore for the reduced Gröbner basis G^* of I_A^* for any monomial order satisfying $\{v_1, \ldots, v_d\} > \{u_1, \ldots, u_p\}$ (i.e., elimination order), $G^* \cap K[\mathbf{u}]$ is a reduced Gröbner basis of I_A. (Elimination theory.)

- Example: Gröbner basis for no three-factor interaction model of $3 \times 3 \times 3$ tables
  - 27 basic moves of degree 4 (all) $\quad (27 = 3 \times 3 \times 3)$

+1	−1	0
−1	+1	0
0	0	0

−1	+1	0
+1	−1	0
0	0	0

0	0	0
0	0	0
0	0	0

  - 54 moves of degree 6 (all) $\quad (54 = 3!3!/2 \times 3)$

+1	−1	0
−1	0	+1
0	+1	−1

−1	+1	0
+1	0	−1
0	−1	+1

0	0	0
0	0	0
0	0	0

- (only) 28 moves of degree 7

$$\begin{array}{|ccc|}\hline 0 & 0 & 0 \\ 0 & -1 & +1 \\ 0 & +1 & -1 \\ \hline\end{array} \quad \begin{array}{|ccc|}\hline +1 & 0 & -1 \\ -1 & +1 & 0 \\ 0 & -1 & +1 \\ \hline\end{array} \quad \begin{array}{|ccc|}\hline -1 & 0 & +1 \\ +1 & 0 & -1 \\ 0 & 0 & 0 \\ \hline\end{array}$$

- (only) 1 move of degree 9

$$\begin{array}{|ccc|}\hline -2 & +1 & +1 \\ +1 & 0 & -1 \\ +1 & -1 & 0 \\ \hline\end{array} \quad \begin{array}{|ccc|}\hline +1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & +1 \\ \hline\end{array} \quad \begin{array}{|ccc|}\hline +1 & -1 & 0 \\ -1 & 0 & +1 \\ 0 & +1 & -1 \\ \hline\end{array}$$

- In general, a Markov basis obtained as a Gröbner basis lacks minimality and symmetry.

6. Structure of minimal Markov basis

- In the research on the structure of Markov basis, important keywords are minimality and invariance.
- Markov bases obtained as Gröbner bases lack minimality and symmetry. (Computation depends on the monomial order.)
- Today we show the structure of minimal Markov basis and examples.
- References:

  Aoki and Takemura (2003). *Aust. N. Z. J. Statist.*, **45**, 229–249.

  Takemura and Aoki (2003). *Ann. Inst. Statist. Math.*, **56**, 1–17.

- Structure of minimal Markov basis
  - $A \in \mathbb{Z}^{d\times p}$: configuration matrix
  - $\mathrm{Ker}_{\mathbb{Z}}(A) = \{\mathbf{z} \in \mathbb{Z}^p \mid A\mathbf{z} = \mathbf{0}\}$: integer kernel (set of moves)
  - $|\mathbf{t}| = |\mathbf{x}| = \sum_i x_i$ for $\mathbf{t} = A\mathbf{x}$
  - $\deg(\mathbf{z}) = |\mathbf{z}^+| = |\mathbf{z}^-|$ for $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^- \in \mathrm{Ker}_{\mathbb{Z}}(A)$
  - $\mathcal{M}_n = \{\mathbf{z} \in \mathrm{Ker}_{\mathbb{Z}}(A) \mid \deg(\mathbf{z}) \le n\}$: moves with the degree less than or equal to n
  - $\mathcal{B}_{\mathbf{t}} = \{\mathbf{z} \in \mathrm{Ker}_{\mathbb{Z}}(A) \mid \mathbf{z}^+, \mathbf{z}^- \in \mathcal{F}_{\mathbf{t}}\}$: moves belonging to $\mathcal{F}_{\mathbf{t}}$
  - $\mathcal{F}_{\mathbf{t}} = \mathcal{F}_{\mathbf{t},1} \cup \cdots \cup \mathcal{F}_{\mathbf{t},K_{\mathbf{t}}}$: $\mathcal{M}_{|\mathbf{t}|-1}$-equivalence class of $\mathcal{F}_{\mathbf{t}}$

- **Theorem** (Takemura and Aoki, 2004)
  Let $\mathcal{B} \subset \mathrm{Ker}_{\mathbb{Z}}(A)$ is a minimal Markov basis for A. Then for each $\mathbf{t}$, $\mathcal{B} \cap \mathcal{B}_{\mathbf{t}}$ is $K_{\mathbf{t}} - 1$ moves connecting different $\mathcal{M}_{|\mathbf{t}|-1}$-equivalence classes of $\mathcal{F}_{\mathbf{t}}$ into a tree.

  Conversely, for each $\mathbf{t}$ with $K_{\mathbf{t}} \geq 2$, let $\{\mathbf{z}_{\mathbf{t},1}, \ldots, \mathbf{z}_{\mathbf{t},K_{\mathbf{t}}-1}\}$ is a set of $K_{\mathbf{t}} - 1$ moves connecting different $\mathcal{M}_{|\mathbf{t}|-1}$-equivalence classes of $\mathcal{F}_{\mathbf{t}}$ into a tree. Then

  $$\bigcup_{\mathbf{t}:\ K_{\mathbf{t}} \geq 2} \{\mathbf{z}_{\mathbf{t},1}, \ldots, \mathbf{z}_{\mathbf{t},K_{\mathbf{t}}-1}\}$$

  is a minimal Markov basis.

- The uniqueness of minimal Markov basis depends on the uniqueness of the ways connecting different $\mathcal{M}_{|\mathbf{t}|-1}$-equivalence classes of $\mathcal{F}_{\mathbf{t}}$ into a tree.

- In particular, for $\mathbf{t}$-fiber with two elements,

$$\mathcal{F}_{\mathbf{t}} = \{\mathbf{x}, \mathbf{y}\},$$

$\mathbf{x} - \mathbf{y} \in \mathrm{Ker}_{\mathbb{Z}}(A)$ belongs to all Markov basis.

- Definition $\mathbf{z} = \mathbf{x} - \mathbf{y}$ is called an indispensable move when $\mathcal{F}_{A\mathbf{x}} = \{\mathbf{x}, \mathbf{y}\}$ is a two-element set.

- Corollary Unique minimal Markov basis exists if and only if the set of indispensable moves is a Markov basis. In this case, the set of indispensable moves is a unique minimal Markov basis.

Example

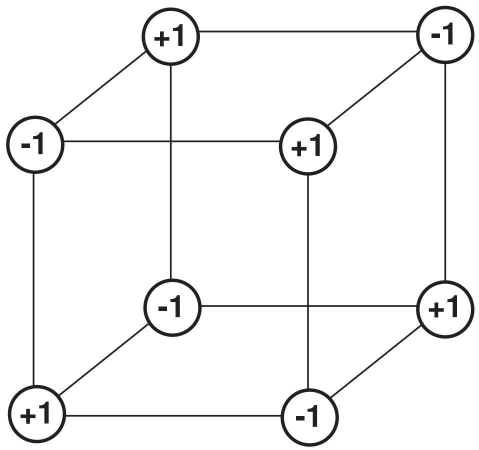
- Independence model of two-way contingency table:

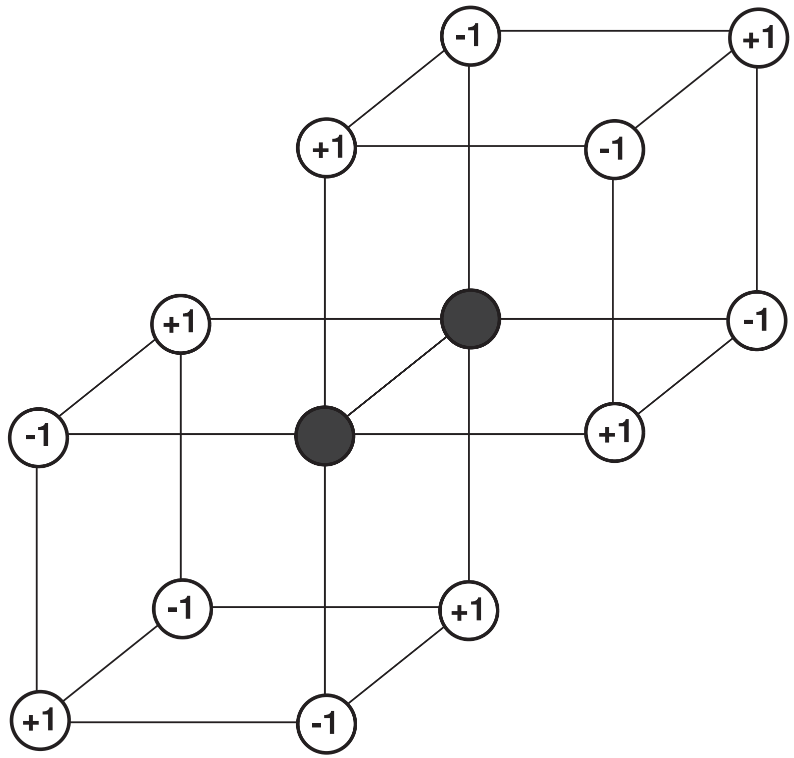
  degree 2 basic move $\begin{array}{|rr|} \hline 1 & -1 \\ -1 & 1 \\ \hline \end{array}$ is an indispensable, because it is a difference of a two-element set

$$\mathcal{F} = \left\{ \begin{array}{|cc|} \hline 1 & 0 \\ 0 & 1 \\ \hline \end{array}, \begin{array}{|cc|} \hline 0 & 1 \\ 1 & 0 \\ \hline \end{array} \right\}.$$

  The set of indispensable moves becomes a Markov basis.
  Therefore this is a unique minimal Markov basis.

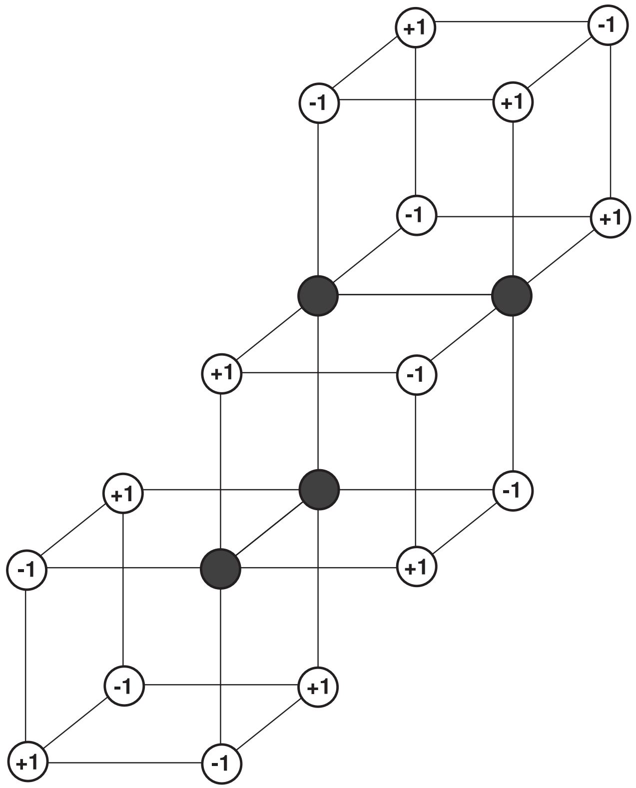
- No three-factor interaction model of $3 \times 3 \times 3$ table:
  The set of basic moves of degree 4 and moves of degree 6 is a unique minimal Markov basis.



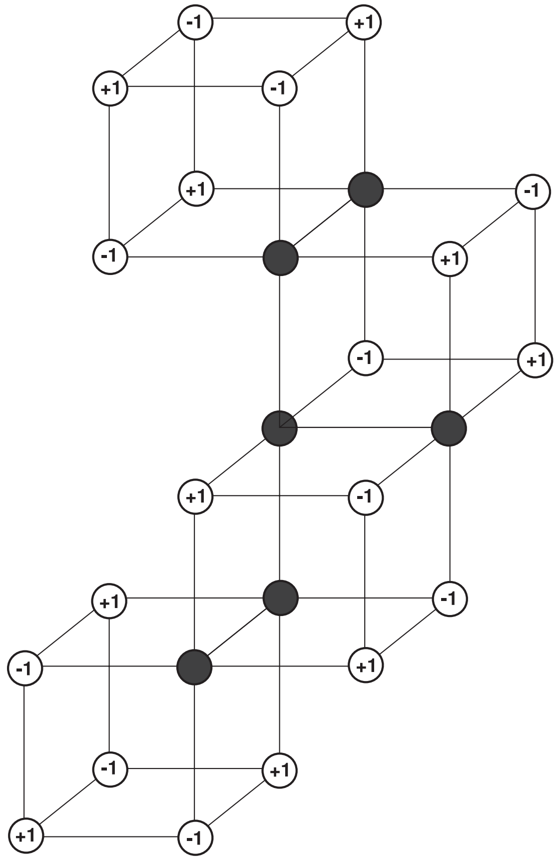


- $3 \times 3 \times 4$ case

  The set of moves of degree $4, 6, 8$ is a unique minimal Markov basis.



- $3 \times 3 \times K$ cases ($K \geq 5$)
  The set of moves of degree $4, 6, 8, 10$ is a unique minimal Markov basis.



- In $3 \times 3 \times K$ cases, no new types of moves are needed for $K > 5$.
  - This is an interesting result, because such a result cannot be obtained by algebraic algorithms.
    If we can calculate Gröbner basis for $3 \times 3 \times 6$ case, we have to calculate Gröbner basis from the first for $3 \times 3 \times 7$ case.
  - Our method is very simple. We check the patterns "by hand" by considering symmetry between the variables.
    First our method was called "by hand method".
    We name it in our book as "method of distance reduction".

- Complete independence model of $2 \times 2 \times 2$ table:
  (fixed marginals : $\{x_{i++}\},\ \{x_{+j+}\},\ \{x_{++k}\}$)

  3 indispensable moves such as $\begin{array}{|rr|}\hline 1 & -1 \\ -1 & 1 \\ \hline\end{array}\ \begin{array}{|cc|}\hline 0 & 0 \\ 0 & 0 \\ \hline\end{array}$ and 3 moves connecting 4 elements fiber

$$\left\{ \begin{array}{|cc|}\hline 1 & 0 \\ 0 & 0 \\ \hline\end{array}\ \begin{array}{|cc|}\hline 0 & 0 \\ 0 & 1 \\ \hline\end{array},\ \begin{array}{|cc|}\hline 0 & 1 \\ 0 & 0 \\ \hline\end{array}\ \begin{array}{|cc|}\hline 0 & 0 \\ 1 & 0 \\ \hline\end{array},\ \begin{array}{|cc|}\hline 0 & 0 \\ 1 & 0 \\ \hline\end{array}\ \begin{array}{|cc|}\hline 0 & 1 \\ 0 & 0 \\ \hline\end{array},\ \begin{array}{|cc|}\hline 0 & 0 \\ 0 & 1 \\ \hline\end{array}\ \begin{array}{|cc|}\hline 1 & 0 \\ 0 & 0 \\ \hline\end{array} \right\}$$

  into a tree constitute a minimal Markov basis.
  Unique minimal Markov basis does not exist.
  There are ${}_6C_3 - 4 = 16$ minimal Markov bases.