

計算代数統計再入門 – Pistone-Wynn 流の実験計画

竹村彰通

2008年4月7日 (Ver.2)

Abstract

ここでは、伝統的な一部実施要因計画の割り付けや別名関係の数学的な構造を、Pistone-Wynn 流の計算代数統計の観点から眺めることによって、計算代数統計への一つの入門と動機付けを与える。

1 はじめに

各要因の水準が 2 水準であるような多因子要因実験の一部実施法は、Wu and Hamada (2000) などの実験計画法の教科書で標準的に解説されている。一方、奥野・芳賀 (1969) の 8 章には以下のような記述があり、田口玄一などのわが国の貢献が強調されている。

多因子実験の計画と解析は、「直交表」を用いることによって、きわめてエレガントにおこなうことができる。これは、わが国独特の手法であって、田口玄一氏等に負うものである。諸外国では、いまだに直交表を用いず、面倒な割付け方をしているために、この有用な手法の普及がたいへん遅れている。

一部実施要因計画の別名関係の扱いには、いかにも代数的な「におい」がする。しかし、これまで Pistone-Wynn 流の解説を何回か聞いていたにもかかわらず、自分には別名関係の扱いが実は多項式環のイデアルの操作そのものであることが理解できていなかった。最近になって、Aoki and Takemura (2007) の改訂作業において、このことを明確に理解するに至った。直交表の利用が我が国で始まったとするならば、その代数的側面をこれまで見過ごして来たのは残念なことである。代数統計の観点から、実験計画法の数学への新たな貢献を目標とすべきである。

ところで，Pistone, Riccomagno and Wynn (2000) の説明も，過度に形式的であって，標準的な一部実施要因計画との関連が明確に説明されていなかったことは事実であると思う．文献としても，このあたりの理解は Pistone 等の本より後になってであり，1) 2水準系の議論は Fontana et al.(2000)，2) より一般の場合は Pistone and Rogantin (2007)，で明らかにされて来ている状況と考えられる．

2 2水準系の一部実施計画の基本 (乗法型の記法)

ある製品の製造工程において，製品の品質に影響を与える要因として，温度や圧力などいくつかの要因が考えられるとする．具体例のために，要因数を6とし，それらの要因を A, B, C, D, E, F とする．それぞれの要因を「低水準」と「高水準」のいずれかに設定して実験をおこなうこととする．「積型」の記法では，低水準を $+1$ (あるいは簡単に $+$)，高水準を -1 (あるいは $-$) と表すこととする．

水準のすべての組合せを行うとすると， $64 = 2^6$ 回の実験をおこなわなければならない．そこで，実験回数を $1/4$ の 16 回で済ませることを考える．このような実験を 2^{6-2} design などという．本来 2^6 の実験をおこなうべきところをその 2^{-2} 回で済ますという意味である．そして， $1/4$ に減らすための「定義対比」として次の二つを考える：

$$I = ABE, \quad I = ACDF \quad (1)$$

例えば， $I = ABE$ の意味は， A, B, C のそれぞれの水準 (± 1) をかけ合わせた時に，それらの積が $+1$ となるように水準の組合せが定まっていることを意味する．このルールに従えば， $A = B = +1$ の時には $E = +1$ と「割り付け」なければならない．

いま，それぞれの水準は ± 1 であるから，2乗すると常に $+1$ である．そこで

$$A^2 = B^2 = C^2 = D^2 = E^2 = F^2 = I \quad (2)$$

という操作のルールを定める．そして，(1) の第1式の両辺には E をかけ，(1) の第2式の両辺に F をかけると

$$E = AB, \quad F = ACD \quad (3)$$

と書き直すことができる．この第一式のルールは，要因 E の水準を A, B の積として定めることを示している． A, B, C, D の水準をまず自由に $2^4 = 16$ 通りの組合せでおこない， E, F の水準をそれぞれ (3) 式のように定めてやると，16回の実験の水準の組合せの設定は次の表のようになる．

さて (2) 式のルールのもとで (1) の二つの式を辺々かけあわせると

$$I = ABE = ACDF = BCDEF \quad (4)$$

Table 1: 2^{6-2} 計画の例

A	B	C	D	E	F
+	+	+	+	+	+
+	+	+	-	+	-
+	+	-	+	+	-
+	+	-	-	+	+
+	-	+	+	-	+
+	-	+	-	-	-
+	-	-	+	-	-
+	-	-	-	-	+
-	+	+	+	-	-
-	+	+	-	-	+
-	+	-	+	-	+
-	+	-	-	-	-
-	-	+	+	+	-
-	-	+	-	+	+
-	-	-	+	+	+
-	-	-	-	+	-

という関係を得る。 I を含むようなこのような関係を contrast subgroup と言う。 I を含まない等号関係としては、(4) 式に例えば A をかけることによって

$$A = BE = CDF = ABCDEF$$

などを得る。このような関係を「別名関係」という。統計的には、これらの4個の主効果および交互作用が一部実施のために完全に「交絡」しており、別々には推定できないことを意味している。

3 加法型への書き換えと線形符号との同値性

ここでは前節の例を加法型に書き換えることにより2水準系の一部実施計画と線形符号が数学には同じものである事を具体的に説明する。

前節では水準を ± 1 とし、水準の組合せの演算を AB などの積の形で表した。しかしながら数学的には2を法とする加法の演算を考えたほうが、議論はある意味では明確となる。これは $\text{mod } 2$ での演算で考えれば、加法と乗法の両方

を組み合わせる事ができるからである¹．特に 2 を法とする加法と乗法で $F_2 = \{0, 1\}$ は「体」をなしており、 F_2 の要素からなるベクトルを通常の線形代数の手法で扱うことができる．いま「低水準」と「高水準」を 0 および 1 で符号化する．つまり

$$+ \rightarrow 0, \quad - \rightarrow 1$$

と符号化を変えてやる．そうすると、前節の積の演算が mod 2 での “exclusive or” の加法の演算と同じであることがわかる．

前節の 2^{6-2} design の例で、6 個の factor の水準を 6 個の「ビット」と考え x_1, \dots, x_6 と表す．まず最初の 4 ビットについては full factorial (完全実施計画) の形に書く．このことは、最初の 4 ビットはそのまま「送信する」ことを意味している．

x_1	x_2	x_3	x_4	
0	0	0	0	
0	0	0	1	
0	0	1	0	
0	0	1	1	
0	1	0	0	
0	1	0	1	
0	1	1	0	
0	1	1	1	(5)
1	0	0	0	
1	0	0	1	
1	0	1	0	
1	0	1	1	
1	1	0	0	
1	1	0	1	
1	1	1	0	
1	1	1	1	

5 ビット目と 6 ビット目は、誤り訂正のために一定のルールで値を定めて付加する．(3) 式を加法的に表せば

$$\begin{aligned} x_5 &\equiv x_1 + x_2 \pmod{2} \\ x_6 &\equiv x_1 + x_3 + x_4 \pmod{2} \end{aligned}$$

¹実は、この節の内容は数年前の自分のメモの一部再利用である．以前は、加法型にすることによって、体の構造を用いることができることを利点としていたが、このことは本当ではないかも知れない．積は、和の繰り返しと見ることができるから、積型の記法でも実現できるはずである．

となる。つまり最初の4ビットの1の数の偶奇を求め、それに応じて5ビット目と6ビット目を定めることとなる。これを実行すれば以下となる。

$$\begin{array}{cccccc}
 x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 1 \\
 0 & 0 & 1 & 0 & 0 & 1 \\
 0 & 0 & 1 & 1 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 \\
 0 & 1 & 0 & 1 & 1 & 1 \\
 0 & 1 & 1 & 0 & 1 & 1 \\
 0 & 1 & 1 & 1 & 1 & 0 \\
 1 & 0 & 0 & 0 & 1 & 1 \\
 1 & 0 & 0 & 1 & 1 & 0 \\
 1 & 0 & 1 & 0 & 1 & 0 \\
 1 & 0 & 1 & 1 & 1 & 1 \\
 1 & 1 & 0 & 0 & 0 & 1 \\
 1 & 1 & 0 & 1 & 0 & 0 \\
 1 & 1 & 1 & 0 & 0 & 0 \\
 1 & 1 & 1 & 1 & 0 & 1
 \end{array} \tag{6}$$

これが表1と全く同じ形であることは明らかである。

加法型に書くことのメリットは、有限体に関してではあるが、通常の線形代数と同様に理解ができる点にある。2元体 $F_2 = \{0, 1\}$ の元を要素とする p 次元のベクトルについて、要素ごとに mod 2 で演算をおこなうこととする。さらに2本のベクトル

$$(x_1, \dots, x_p), \quad (y_1, \dots, y_p)$$

の「内積」² を

$$x_1 y_1 + \dots + x_p y_p \pmod{2}$$

と定義すれば、 $\{0, 1\}^p$ は通常の R^p のような線形空間と考えてよい。この観点からは contrast subgroup $I = ABE = ACDF = BCDEF$ は

$$L = \{(0, 0, 0, 0, 0, 0), (1, 1, 0, 0, 1, 0), (1, 0, 1, 1, 0, 1), (0, 1, 1, 1, 1, 1)\}$$

の4点からなる2次元の線形部分空間 L であり、その基底として $I = ABE = ACDF$ の2本のベクトル

$$(1, 1, 0, 0, 1, 0), \quad (1, 0, 1, 1, 0, 1)$$

²本来は双対空間であるが、内積をいれて1次形式ともとのベクトルを同一視して考えてもよい。

をとる事ができる. そして (6) の行からなる 16 点は L の直交補空間である.

符号理論では contrast subgroup $I = ABE = ACDF = BCDEF$ は「パリティ検査行列」に対応する. (6) の各行をビット列として考えよう. いまノイズのある通信路を使って $x_1 \dots x_6$ の 6 ビットを送るとする. 2 ビット分は誤り訂正のため「パリティ」である. L の 2 個の基底を並べた行列

$$H' = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

をパリティ検査行列とすると, 6 ビットからなる符号 $x = (x_1, \dots, x_6)$ が送られて来た時に

$$xH \equiv 0 \pmod{2}$$

ならば誤りがなかったと判断する事ができる.

4 積型記法のメリットとは

前節のように説明されると, いかにも加法型の記法のほうがカッコいい. 有限体上の線形代数で見通しよく理解できる感じがする. それに対して, 2 節の積型の記法は「単なる説明」であって, 数学的にはそれほど根拠がないのではないかという印象がある. 自分は長いことそう考えてきた. しかし実は積型の記法のほうがより「代数的」であることを最近になって理解した.

そもそも Diaconis and Sturmfels 流のマルコフ基底に関する計算代数統計を勉強しはじめた時に, 分割表を monomial に対応させること自体が不思議な感じがした. move を考える時に, 負の要素をなぜ負のべきに対応せずに, monomial の差として表すのか, がわからなかった. monomial の記法は「積型」である. 積型に書いて, はじめて代数の手法が使えるわけである. そんなものかと思っていた. しかし, Pistone-Wynn 流の実験計画の代数統計でも同じことが起きていることに, 気がつかなかった.

2 節の積型の記法の代数的な説明は次のように (言われてみれば) 自明なものである. いま A, B, C, D, E, F の 6 個の文字を「不定元」と考えて, これらの不定元からなる多項式環 $k(A, B, C, D, E, F)$ を考える. k は適当な体でよい. $k(A, B, C, D, E, F)$ の中に, 次の 8 個の多項式で生成されるイデアルを考える.

$$I = \langle A^2 - 1, B^2 - 1, C^2 - 1, D^2 - 1, E^2 - 1, F^2 - 1, ABE - 1, ACDF - 1 \rangle \quad (7)$$

Pistone and Wynn の言うところの “design ideal” である. そうすると, 2 節の形式的な操作は, 商環 $k(A, B, C, D, E, F)/I$ の操作と全く同じである. 2 個の

monomial が別名関係にあるための必要十分条件は、それらの差が design ideal I に属することである。例えば

$$BE - ABCDEF \in I$$

である。この観点に立てば、例えば別名関係の判定はイデアル帰属問題であるから、 I のグレブナー基底を求めておけば別名関係が判定できることになる。いずれにしても、ここで強調すべき点は 2 節のような積型の記法での一部実施計画の説明が「説明の便宜」のためではなく、代数的に由緒正しいものであるということである。このことは、積型の記法がなぜ加法型の記法によって駆逐されなかったのか、ということの理由にもなっていると思われる。

本稿では、2 水準の場合を述べてきた。2 水準の場合だと $(+, -) \rightarrow (0, 1)$ の対応がいかに単純で、どちらで考えても同じではないかと思われるかも知れない。しかしながら、例えば 3 水準の実験を考えるような場合にはより本質的な差が見えてくるのである ([1])。例えば Pistone and Wynn の本の 4.6 節には

$$\begin{aligned} f(x) &= \frac{1}{2} + \frac{1}{4}x_1x_2x_4 - \frac{1}{4}x_1x_2x_3 + \frac{1}{4}x_1x_2x_4x_5 + \frac{1}{4}x_1x_3x_4x_5 \\ &= \frac{1}{2} + \frac{x_1x_2(x_4(x_3 + 1) + x_3(x_5 - 1))}{4} \end{aligned}$$

の零点の集合として次のような奇妙な例が与えられている。

1	1	1	1	1
1	1	-1	1	1
1	1	-1	1	-1
1	1	-1	-1	-1
1	-1	1	1	-1
1	-1	1	-1	1
1	-1	1	-1	-1
1	-1	-1	-1	1
-1	1	1	1	-1
-1	1	1	-1	1
-1	1	1	-1	-1
-1	1	-1	-1	1
-1	-1	1	1	1
-1	-1	-1	1	1
-1	-1	-1	1	-1
-1	-1	-1	-1	-1

すでに述べたように、マルコフ基底に関する議論の中では、積型の記法によって大きな展開が見られた。Pistone and Wynn も実験計画に関してすでに同様の

貢献をしていたわけであるが、そのことの意味が必ずしも理解されていなかったように思われる。

References

- [1] Satoshi Aoki and Akimichi Takemura. (2007). Markov basis for design of experiments with three-level factors [arXiv:0709.4323v2](#)
- [2] Roberto Fontana, Giovanni Pistone and Maria-Piera Rogantin. (2000). Classification of two-level factorial fractions. *J. Statist. Plann. Inference*, **87**, 149–172.
- [3] 奥野忠一・芳賀敏郎 (1969). 実験計画法 . 培風館 .
- [4] Pistone, G., Riccomagno, E. and Wynn, H. P. (2000). *Algebraic Statistics, Computational Commutative Algebra in Statistics*. Chapman & Hall.
- [5] Giovanni Pistone and Maria-Piera Rogantin. (2007). Indicator function and complex coding for mixed fractional factorial designs. [arXiv:math/0703365v1](#)
- [6] Wu, C. F. J. and Hamada, M. *Experiments: Planning, analysis, and parameter design optimization*. (2000) Wiley Series in Probability and Statistics: Texts and References Section. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.

壺のモデルからみたスワッピングとマルコフ基底

竹村彰通（東京大学・大学院情報理工学系研究科）

「官庁統計データの公開における諸問題の研究」集会
2007年11月17日

概要

ここでは2元表という一番簡単な分割表を例にとりて、壺のモデルの観点から、グレブナー基底, sorting, 対称群の作用などがどのように理解できるかをクイズ形式で示す。さらに一般の分解可能モデルへの拡張を示唆する。なお、文末にクイズ解答用の草稿用ページがあるので利用していただきたい。

1 2元表の場合の分割表と北西隅ルール

ここでは例題として簡単な 3×3 分割表を考える。

問1 周辺が以下の分割表の中身をうまくうめてください。

表 1: 周辺表の例

*	*	*	6
*	*	*	5
*	*	*	5
7	5	4	16

解答例: 行優先で右下からうめることを考える。まず右下を $\min(4, 5) = 4$ とする。

*	*	*	6
*	*	*	5
*	*	4	1
7	5	0	12

この段階で、実は3列目が $(0\ 0\ 4)'$ となることは見えているが、行優先の時にはそこはまだ確定しないことにする。そこで次に $(3,2)$ 要素を $\min(1, 5) = 1$ でうめると $(3,1)$ 要素も 0 となり

$$\begin{array}{ccc|c} * & * & * & 6 \\ * & * & * & 5 \\ 0 & 1 & 4 & 0 \\ \hline 7 & 4 & 0 & 11 \end{array}$$

となる。さらに $(2,3)$ 要素を $\min(0, 5) = 0$ とし、続いて他の2行目をうめると

$$\begin{array}{ccc|c} * & * & * & 6 \\ 1 & 4 & 0 & 0 \\ 0 & 1 & 4 & 0 \\ \hline 6 & 0 & 0 & 6 \end{array}$$

となり、結果は

$$\begin{array}{ccc|c} 6 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 1 & 4 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}$$

となる。

一方最後の列から、列優先でうめる方法を考える。第3列が下から上へ

$$\begin{array}{ccc|c} * & * & 0 & 6 \\ * & * & 0 & 5 \\ * & * & 4 & 1 \\ \hline 7 & 5 & 0 & 12 \end{array}$$

次に第2列が、下から上へ

$$\begin{array}{ccc|c} * & 0 & 0 & 6 \\ * & 4 & 0 & 1 \\ * & 1 & 4 & 0 \\ \hline 7 & 0 & 0 & 7 \end{array}$$

と定まり、結局同じ表に到達する。

以上の例が示唆していることは次のことである。右下の要素からできるだけ大きい値を入れて行く操作は、グレブナー基底の観点からは revlex 式の term order を考えていることにあたる。そして、2元表においては、列優先の revlex でも行優先

の revlex でも、結果が同じになる．結果表は「余り」(standard monomial) にあたる．つまり 2 元表においては、二つの異なる term order に基づいて実際に割算を行うと、常に同じ standard monomial に到達する．すなわち、最適化の観点からは、異なる term order が同じ最適解を与える例となっている．

最適化された解は、対角付近に頻度が集まっており、順位相関を最大化していると見ることもできる．このような解は、離散最適化の分野で「北西隅ルール」とよばれている．北西隅ルールは列の周辺分布を行の周辺分布に変換する「輸送問題」の許容解であり、もし輸送コストが“Monge 性”と呼ばれる性質を持つ時は、コストを最小化した解となっている事が知られている．これらに点については松井知己氏よりご教示いただいた．“Northwest corner rule”などの用語で google で検索すれば多くの文献が出てくる．

2 Sorting で考える

次に“sorting”を検討する (Sturmfels の本にあり、さらに大杉・日比の研究で発展されている方法．[1, 2] を参照)．同時頻度をうめた次の表をもとに考える．

表 2: 中身をうめた表の例

3	2	1	6
2	1	2	5
2	2	1	5
7	5	4	16

これを単項式に対応させる:

$$x_{11}^3 x_{12}^2 x_{13} x_{21}^2 x_{22} x_{23}^2 x_{31}^2 x_{32}^2 x_{33}$$

行の添字のみ集めてくると、周辺頻度より

1 が 6 個, 2 が 5 個, 3 が 5 個

である．列の添字は同様に

1 が 7 個, 2 が 5 個, 3 が 4 個

である．ここで、列の添字のほうが、グループとして、行の添字より後と考える．例えば、列の添字を +3 して x_{ij} を $x_{i,j+3}$ と表す．この場合、例えば x_{11} を x_{14} と表すことになる．その上で、行の添字と列の添字を区別せずに sort すると

1 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4 4 4 5 5 5 5 5 6 6 6 6

となる．ここで， x_{ij} の 1 桁目の i の部分に上の数字を左から順にうめていく：

$$x_{1?}x_{1?}x_{1?}x_{1?}x_{1?}x_{1?}x_{2?}x_{2?}x_{2?}x_{2?}x_{2?}x_{2?}x_{3?}x_{3?}x_{3?}x_{3?}x_{3?}$$

ここで一巡したので，つぎに後半を 2 桁目にうめていく

$$x_{14}x_{14}x_{14}x_{14}x_{14}x_{14}x_{24}x_{25}x_{25}x_{25}x_{25}x_{25}x_{35}x_{36}x_{36}x_{36}x_{36}$$

ここで， $4, 5, 6 \rightarrow 1, 2, 3$ と添字を戻すと

$$x_{11}^6 x_{21} x_{22}^4 x_{32} x_{33}^4$$

となり，sorting を一回施すだけで最適化の解を得る．

以上の sorting の操作は，グレブナー基底の理論で現れるものであるが，他の分野ではあまり見かけない操作のように思われる．そもそも sorting と「行あるいは列優先で，右下から可能な限り大きな値でうめる」ということが同じ結果を導くこと自体が，自明ではない．

3 2 元表を壺で考える

前節の sorting は実は壺のモデルで考えるとわかりやすい．壺のモデルは使い古した素朴な考え方だが，素朴なりに強力である．問 1 を次のように書き換える．

問 2 いま，壺の中に，1 から 16 まで番号のふった 16 枚のカードがあったとします．各カードには 2 桁の数字を書く欄があります．ここにこれから数字を書いていきます．それぞれの欄とも 1, 2, 3 のいずれかの数字を書きます．そして，第 1 桁には

1 を書くカードは 6 枚

2 を書くカードは 5 枚

3 を書くカードは 5 枚

としてください．また第 2 桁には

1 を書くカードは 7 枚

2 を書くカードは 5 枚

3 を書くカードは 4 枚

としてください．このような数字の埋め方で，最も規則的で簡単な方法を考えてください．

解答例: 第 1 桁には，カードの番号順に，1 を 6 枚に書き，2 を 5 枚に書き，3 を残りの 5 枚に書く．第 2 桁も同様に書く．

問 3 以上のような 16 枚への数字のうめ方と異なるうめ方をした場合，必ず 2 枚のカードを選ぶことができ，第 2 桁の数字をとりかえることによって，規則的な埋め

方に近づけることができることを示してください．表 2 に対応する例が原稿の最後の草稿用ページにあるので，それで確認してみてください．

第 2 桁の数字をとりかえることは，秘匿措置における swapping に対応する ([3]) . swapping は，次数が 2 の move に密接に対応している．それは分割表の頻度で考えると $\begin{matrix} +1 & -1 \\ -1 & +1 \end{matrix}$ の形となるからである．standard monomial に近づけるように swapping するのは，グレブナー基底の要素での割算に対応している．

4 対称群の作用で考える

独立モデルのもとで周辺を与えた $I \times J$ の 2 元分割表の条件つき分布は超幾何分布であり，確率関数は

$$\frac{\prod_{i=1}^I x_{i+}! \prod_{j=1}^J x_{+j}!}{n! \prod_{ij} x_{ij}!}$$

の形で与えられる．この分布からの直接の sampling は，壺のモデルを考えると以下のように簡明である．この簡単な事実を，対称群 S_n の作用として考えることができることを，最近になって北大数学の吉田知行先生が何回かの講演で指摘されている．これはコロンブスの卵的な指摘であるが，実は非常に重要と思われる．吉田先生の講演資料は吉田先生に連絡すれば入手することができる．

いま表 1 のように周辺が所与として，超幾何分布に従って 2 元表を発生させたいとする．そのために，草稿用ページの間 2 のように，1 から 16 までの連番と 2 桁の空白があるような 16 枚のカードを壺に入れる．そして，壺の中をよく混ぜてから，順番にカードを取り出して，まず第 1 桁目には

- 6 枚のカードに 1
- 5 枚のカードに 2
- 5 枚のカードに 3

と順にうめていく．次に，カードを壺にもどして再度よく混ぜてから第 2 桁目には

- 7 枚のカードに 1
- 5 枚のカードに 2
- 4 枚のカードに 3

と順にうめる．これで分割表がランダムに生成できて，その分布は超幾何分布である．

さて，以上では壺の中を 2 回まぜているが，これは対称群 S_n , $n = 16$, をそれぞれの桁に独立に作用させていると理解することができる．すなわち $S_n \times S_n$ がカードの集合に作用している．ただし，2 元表の場合は，実は 1 回目の混合操作は不要なことは明らかである．すなわちこの場合には S_n を 1 回だけ第 2 桁に作用させればよい．

第2回目の混合操作を省略して、規則的にうめれば、前節で述べた sorting となり、北西隅型の standard monomial が得られている。

以上のように、2元表の超幾何分布の場合には、対称群上の一様分布に対応させることにより、超幾何分布からの直接の sampling が可能である。しかしながら、一般には MCMC のほうが適用範囲が広いので $\begin{matrix} +1 & -1 \\ -1 & +1 \end{matrix}$ の形の move を用いた MCMC を考察してみる。そうすると、吉田知行先生によって指摘されているように、この move を元の壺に引き戻して考えれば、実は S_n の互換を「ランダムに」作用させていることがわかる。このように S_n に引き戻して考えると、分割表のファイバー上の MCMC が、 S_n の互換によって生成されるランダムウォークに対応していることがわかる。そして、 S_n 上のランダムウォークの収束速度の評価は、Diaconis の仕事にあるように、対称群の表現の指標を用いて解析することができる。

5 分解可能モデルへの一般化

ここまでの2元表の議論のほとんどは一般の分解可能モデルに素直に一般化される。ただし、対称群の表現論を用いる MCMC の解析はまだ全く未開拓である。以下では $2 \times 2 \times 2 \times 2$ の一例を示すことにより、一般化の雰囲気を示すこととする。いま固定する周辺を $\{1, 2\}, \{2, 3\}, \{3, 4\}$ とする。

問4 いま、壺の中に、1から16まで番号のふった16枚のカードがあったとします。各カードには4桁の数字を書く欄があります。ここにこれから数字を書いていきます。それぞれの欄とも1,2のいずれかの数字を書きます。そして

第(1,2)桁には

(1,1)の組合せが6枚、(1,2)が3枚、(2,1)が2枚、(2,2)が5枚

第(2,3)桁には

いずれの組合せも4枚ずつ

第(3,4)桁には

(1,1)の組合せが3枚、(1,2)が5枚、(2,1)が3枚、(2,2)が5枚

になるようにしてください。このような数字の埋め方で、最も規則的で簡単な方法を考えてください。

解答例: (1,2)桁は単に順にうめて行く。(2,3)桁は、2桁目の数字があうカードを順にひろいながら、3桁目を順に埋めて行く。(3,4)桁は、3桁目の数字があうカードを順にひろいながら、4桁目を順に埋めて行く。

問5(難) それでは、以上のような数字の埋め方と異なる埋め方をした場合、必ず2枚のカードを選ぶことができ、書いてある数字を入れ換えることによって、必ず以上の埋め方に近づくことができることを示してください。

問 6 問 4 と同じ周辺頻度を持つ表を，超幾何分布から直接サンプリングするにはどうすればよいかを示してください．

解答例: (1,2) 桁は単に順にうめて行く．(2,3) 桁は，2 桁目の数字によって壺を部分壺に層別して，部分壺ごとに混ぜればよい．(3,4) 桁は，3 桁目の数字によって壺を部分壺に層別して，部分壺ごとに混ぜればよい．

問 5 の解答は困難であるが，分解可能モデルについては `sorting` の操作が可能であり，2 次の `move` からなるグレブナー基底が存在することに対応している．問 6 を MCMC 版に焼直せば，MCMC の収束速度の解析も可能になるとと思われる．

参考文献

- [1] Aoki, S., Hibi, T., Ohsugi, H. and Takemura, A. (2007a). Markov basis and Groebner basis of Segre-Veronese configuration for testing independence in group-wise selections. [arXiv:0704.1074v1](https://arxiv.org/abs/0704.1074v1)
- [2] Aoki, S., Hibi, T., Ohsugi, H. and Takemura, A. (2007b). A generalizations of Segre-Veronese configurations. In preparation.
- [3] Takemura, A. and Hara, H. (2007). Conditions for swappability of records in a microdata set when some marginals are fixed. *Computational Statistics*, **22**, 173–185.

草稿用ページ

問1

$$\begin{array}{r|l} & 6 \\ & 5 \\ & 5 \\ \hline 7 & 5 & 4 & 16 \end{array}$$

問2

1			5			9			13		
2			6			10			14		
3			7			11			15		
4			8			12			16		

問3

1	1	2	5	2	1	9	3	3	13	3	1
2	2	3	6	1	1	10	3	1	14	2	1
3	3	2	7	2	2	11	1	2	15	1	1
4	1	1	8	1	3	12	3	2	16	2	3

問4

1				5				9				13			
2				6				10				14			
3				7				11				15			
4				8				12				16			