

1.  $2 \times 3$  分割表で行和、列和が固定されているものを考える。つまり

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

を考える。

(a) Markov 基底をトーリックイデアル  $I_A$  のグレブナ基底を計算することによりもとめよ。

$$I_A = \langle x_2x_6 - x_3x_5, x_1x_6 - x_3x_4, x_1x_5 - x_2x_4 \rangle$$

トーリックイデアルを計算する Asir のプログラム toric.rr を用いれば、

```
Asir
[1206] load("toric.rr");
[1219] toric_ideal([[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]]);
ideal :
[x5-t1*t4,x4-t1*t3,x3-t1*t2,x2-t0*t4,x1-t0*t3,
x0-t0*t2,t0*t1*t2*t3*t4*t5-1]
gb :
[x1*x5-x2*x4,x0*x5-x2*x3,x0*x4-x1*x3,t3*x5-t4*x4,
t3*x2-t4*x1,t2*x5-t4*x3,t2*x4-t3*x3,t2*x2-t4*x0,
t2*x1-t3*x0,t5*x0*x4*x5-t1,-t5*x0*x1*x5+t0,t4*t5*x0*x4-1,
t4*t5*x0*x1*x5-x2,-t3*t4*t5*x0*x3+t2]
[x1*x5-x2*x4,x0*x5-x2*x3,x0*x4-x1*x3]
[1220]
```

(b) 4ti2 を使って Markov 基底を計算せよ。

上でグレブナ基底を得られているので、それが Markov 基底でもあるが、あえて 4ti2 を使い Markov 基底を計算する。行列  $A$  を表す次のような入力ファイルを用意。

```
cont2x3.mat
5 6
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
```

4ti2 の Markov 基底を求めるコマンド markov を実行

markov の実行

```
$ markov cont2x3.mat
... 略
File cont2x3.mat.cost not found. Using default cost.
COMPUTING GENERATING SET ... PROJECT-AND-LIFT ...
Phase 1: Saturating 3 variable(s).
  Saturating (3) ... (GM) Size:      2, Time: -0.00 / 0.01 secs. Done.
  Saturated already on 2 variable(s).
Done. Size:      2, Time: -0.00 / 0.01 secs
Phase 2: Lifting 3 variable(s).
  Lifted already on 2 variable(s).
  Lifting (1) ... (N) Size:      3, Time: -0.00 / 0.01 secs. Done.
Done. Size:      3, Time: -0.00 / 0.01 secs
COMPUTING MINIMAL GENERATING SET (FAST) ...
Size:      3, Time: -0.00 / 0.01 secs. Done.
Total Time: 0.02 secs.
```

Markov 基底が次のファイルに出力される。

cont2x3.mat.mar

```
3 6
0 -1 1 0 1 -1
1 -1 0 -1 1 0
1 0 -1 -1 0 1
```

(c)  $Ax = \beta$ ,  $x \in \mathbf{N}_0^6$  を満たす  $x$  を  $\beta = (2, 3; 1, 2, 2)'$  についてすべて求めよ。

$Ax = \beta$  を満たす  $x$  を 1 つとってくる。たとえば、

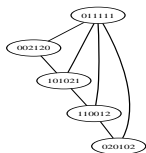
$$x = (0, 1, 1, 1, 1, 1)'$$

これに対して、上で求めた Markov 基底を足し引きして、各成分が非負であるものを列挙していけばよい。  
このような  $x$  は次の 5 つ。

$$(0, 1, 1, 1, 1, 1)', (0, 2, 0, 1, 0, 2)', (1, 1, 0, 0, 1, 2)', (1, 0, 1, 0, 2, 1)', (0, 0, 2, 1, 2, 0)'$$

(d) ファイバー  $\mathcal{F}_{(2,3;1,2,2)}$  の連結グラフを書け。

連結グラフは Markov 基底の加減で移り合うもの同士を結べばよい。



参考文献: [5, chap.3], [3, chap.5].

プログラム: cont2x3.mat, ex-stat-ans.tex

2.  $3 \times 3 \times 3$  分割表で全ての 2 次元周辺頻度  $\{x_{ij+}\}, \{x_{i+k}\}, \{x_{+jk}\}$  が固定されているものを考える。

(a) この分割表に対応する行列  $A$  を求めよ。

行列の各列が

$$(x_{111}, x_{112}, x_{113}, x_{121}, x_{122}, x_{123}, x_{131}, x_{132}, x_{133}, x_{211}, x_{212}, x_{213}, x_{221}, x_{222}, x_{223}, x_{231}, x_{232}, x_{233}, \\ x_{311}, x_{312}, x_{313}, x_{321}, x_{322}, x_{323}, x_{331}, x_{332}, x_{333})'$$

に対応しているものとし、各行は

$$(x_{11+}, x_{12+}, x_{13+}, x_{21+}, x_{22+}, x_{23+}, x_{31+}, x_{32+}, x_{33+}, x_{+11}, x_{+12}, x_{+13}, x_{+21}, x_{+22}, x_{+23}, x_{+31}, x_{+32}, x_{+33}, \\ x_{+11}, x_{+12}, x_{+13}, x_{+21}, x_{+22}, x_{+23}, x_{+31}, x_{+32}, x_{+33})'$$

に対応しているものとする。

```

3x3x3cont.mat
27 27
1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 0
1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1

```

(b) 4ti2 を使って Markov 基底を計算せよ。  
81 個の元からなる Markov 基底が得られる。

```

3x3x3cont.mat.mar
81 27
-1 0 1 1 -1 0 0 1 -1 0 0 0 0 0 0 0 0 1 0 -1 -1 1 0 0 -1 1
-1 0 1 1 -1 0 0 1 -1 1 0 -1 -1 1 0 0 -1 1 0 0 0 0 0 0 0 0 0
....
1 0 -1 0 0 0 -1 0 1 0 0 0 0 0 0 0 0 0 -1 0 1 0 0 0 1 0 -1

```

(c)  $x_{ij+} = 1, x_{i+k} = 1, x_{+jk} = 1$  を満たす分割表を全て列挙せよ。

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

を並べ替えたもの 6 個と

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

を並べ替えたもの 6 個の合計 12 個。

(d) (研究課題) 与えられた Markov 基底と分割表の 1 つからファイバーのすべての元を列挙するプログラムをつくれ。またそれから連結グラフを表示するプログラムを書け。

enumerate.fiber はファイバーの元を全て列挙するようなプログラムである。ファイバーの 1 つの元を書いたファイルと move の元たちを書いたファイルを用意する。元の書き方は 4ti2 の場合と同様である。プログラムの実行は次のように行う。

プログラム enumerate\_fiber の実行

```
$ enumerate_fiber start_v4.txt move4.txt
n_move : 1, msize : 6
start_v :
 0 1 1 1 1 1
n_move : 3, msize : 6
move : 3 6
 0 -1 1 0 1 -1
 1 -1 0 -1 1 0
 1 0 -1 -1 0 1
0 0 2 1 2 0
1 0 1 0 2 1
1 1 0 0 1 2
0 2 0 1 0 2
depth : 0, count : 5, queue : 4
depth : 1, count : 5, queue : 3
depth : 2, count : 5, queue : 2
depth : 3, count : 5, queue : 1
depth : 4, count : 5, queue : 0
count : 5
fiber :
0 1 1 1 1 1
0 0 2 1 2 0
1 0 1 0 2 1
1 1 0 0 1 2
0 2 0 1 0 2
```

3. 離散確率変数  $x_1, x_2, x_3$ 、各変数は値として  $+1$  もしくは  $-1$  をとるとする。 $(x_1, x_2, x_3)$  の発生確率を次のようにとる。

$$P(x_1, x_2, x_3) = \frac{\exp(0.2(x_1x_2 + x_2x_3 + x_1x_3))}{Z}$$

ここで  $Z$  は正規化定数とよばれるもので

$$Z = \sum_{(x_1, x_2, x_3) \in \{+1, -1\}^3} \exp(0.2(x_1x_2 + x_2x_3 + x_1x_3))$$

である。この分布に従うような  $(x_1, x_2, x_3)$  の列をマルコフ連鎖モンテカルロ法を用いてサンプリングし、 $\{+1, -1\}^3$  の 8 点の出現割合がそれぞれ  $P(x_1, x_2, x_3)$  に近くなることを確かめよ。

プログラム: metropolis を用いよ。

参考文献: [4, p.8]

この場合のマルコフ連鎖モンテカルロ法のアルゴリズムは次の通り。

0.  $\mathbf{x} = (x_1, x_2, x_3)$  に適当な初期状態を設定。  
例えば、 $\mathbf{x} \leftarrow (-1, -1, -1)$
1.  $x_1, x_2, x_3$  のいずれかをランダムに選択。  
選んだ変数を  $x_i$  とすれば、 $x_i$  の値を  $-x_i$  に置き換えたものを  $\mathbf{x}'$  とする。
2.  $r \leftarrow \frac{P(\mathbf{x}')}{P(\mathbf{x})}$
3.  $0 \leq R < 1$  の一様乱数  $R$  をとる
4.  $r > R$  であれば、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x}'$   
それ以外の場合、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x}$
5.  $\mathbf{x}_{\text{next}}$  がサンプルとして得られる
6.  $\mathbf{x} \leftarrow \mathbf{x}_{\text{next}}$   
step 1 に戻る

プログラム metropolis の実行

```

$ ./metropolis
-1 -1 -1
-1 1 -1
 1 1 -1
...
 1 1 1
 1 1 1
 1 1 -1
 1 -1 -1
-1 -1 -1
index 0 : 2198
index 1 : 1001
index 2 : 890
index 3 : 926
index 4 : 995
index 5 : 991
index 6 : 932
index 7 : 2067
total : 10000
z : 8.556622
index 0 : 0.219800 0.212948
index 1 : 0.100100 0.095684
index 2 : 0.089000 0.095684
index 3 : 0.092600 0.095684
index 4 : 0.099500 0.095684
index 5 : 0.099100 0.095684
index 6 : 0.093200 0.095684
index 7 : 0.206700 0.212948

```

4.  $2 \times 3$  分割表 周辺頻度  $(5, 15; 5, 5, 10)$  のファイバー  $\mathcal{F}$ 、すなわち

$$\mathcal{F} = \{(x_{ij}) \mid x_{ij} \in \mathbf{N}, x_{1+} = 5, x_{2+} = 15, x_{+1} = 5, x_{+2} = 5, x_{+3} = 10\}$$

から、分布  $\frac{1}{Z} \frac{1}{x_{11}!x_{12}!x_{13}!x_{21}!x_{22}!x_{23}!}$  に従って元をマルコフ連鎖モンテカルロ法を用いてサンプリングし、 $2x_{11} + x_{12}$  の期待値を推定せよ。(Z は正規化定数で  $\sum_{x \in \mathcal{F}} \frac{1}{x_{11}!x_{12}!x_{13}!x_{21}!x_{22}!x_{23}!}$ )

$2 \times 3$  分割表のマルコフ基底は

$$\mathcal{B} = \left\{ \begin{array}{|c|c|c|} \hline 1 & -1 & \\ \hline -1 & 1 & \\ \hline & & \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline 1 & & -1 \\ \hline -1 & & 1 \\ \hline & & \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline & 1 & -1 \\ \hline & -1 & 1 \\ \hline & & \\ \hline \end{array} \right\}$$

与えられた確率分布  $f(\mathbf{x}) = \frac{1}{Z} \prod_{i,j} \frac{1}{x_{ij}!}$  に従って、ファイバー  $\mathcal{F}$  から元をサンプリングする、マルコフ連鎖モンテカルロ法は以下の手順である。

0. ファイバー  $\mathcal{F}$  の元  $\mathbf{x}$  を 1 つ取り出す (なんでもよい)  
 $e \leftarrow 0$
1.  $\mathcal{B} \cup (-\mathcal{B})$  からランダムに元  $\mathbf{z}$  を選ぶ
2.  $\mathbf{x} + \mathbf{z}$  の各要素が全て非負ならば、 $r \leftarrow \frac{f(\mathbf{x} + \mathbf{z})}{f(\mathbf{x})}$   
 それ以外の場合、 $r \leftarrow 0$
3.  $0 \leq R < 1$  の一様乱数  $R$  をとる
4.  $r > R$  であれば、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x} + \mathbf{z}$   
 それ以外の場合、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x}$
5.  $\mathbf{x}_{\text{next}}$  がサンプルとして得られる  
 $e \leftarrow e + (\mathbf{x}_{\text{next}} \text{ の } 2x_{11} + x_{12})$

6.  $x \leftarrow x_{\text{next}}$   
 step 1 に戻る

7.  $e/(\text{サンプル数})$  が期待値の推定値

これを実行するプログラムが 2x3mcmc である。  
 2x3mcmc (乱数 seed) (サンプル数) のように実行する。  
 正確な期待値は  $15/4 = 3.75$  である。

2x3mcmc の実行結果

```
./2x3mcmc 0 100000
2 0 3 3 5 7
2 0 3 3 5 7
2 0 3 3 5 7
2 0 3 3 5 7
1 1 3 4 4 7
0 1 4 5 4 6
0 2 3 5 3 7
1 2 2 4 3 8
2 1 2 3 4 8
1 2 2 4 3 8
...
2 1 2 3 4 8
3 0 2 2 5 8
2 0 3 3 5 7
2 1 2 3 4 8
ave : 3.746200
```

5.  $2 \times 2$  分割表でデータ

疾病/喫煙	あり	なし	
あり	3	1	4
なし	2	4	6
	5	5	10

が与えられているとする。

(a) ファイバー  $\mathcal{F}_{(4,6;5,5)}$  の要素を全列挙し、

$H_0$  : 喫煙と疾病は無関係

が真と仮定した時のそれぞれの出現確率  $p(x) = \frac{1}{Z} \frac{1}{x_{11}!x_{12}!x_{21}!x_{22}!}$  を求めよ。ただし、 $Z$  は正規化定数である。

ファイバー  $\mathcal{F}_{(4,6;5,5)}$  の元は次の 5 つである。

<table border="1"><tr><td>4</td><td>0</td></tr><tr><td>1</td><td>5</td></tr></table>	4	0	1	5	,	<table border="1"><tr><td>3</td><td>1</td></tr><tr><td>2</td><td>4</td></tr></table>	3	1	2	4	,	<table border="1"><tr><td>2</td><td>2</td></tr><tr><td>3</td><td>3</td></tr></table>	2	2	3	3	,	<table border="1"><tr><td>1</td><td>3</td></tr><tr><td>4</td><td>2</td></tr></table>	1	3	4	2	,	<table border="1"><tr><td>0</td><td>4</td></tr><tr><td>5</td><td>1</td></tr></table>	0	4	5	1
4	0																											
1	5																											
3	1																											
2	4																											
2	2																											
3	3																											
1	3																											
4	2																											
0	4																											
5	1																											

それぞれの確率は

$$\frac{1}{42}, \frac{10}{42}, \frac{20}{42}, \frac{10}{42}, \frac{1}{42}$$

となる。

(b) 「 $X_{11} > c$  ならば  $H_0$  を棄却」という検定方式を考える。与えられたデータの  $p$  値を計算せよ。

観測値の  $x_{11}$  は 3 より  $c = 3$  として、

$$p \text{ 値} = \Pr(X_{11} \geq 3) = \frac{10}{42} + \frac{1}{42} = \frac{11}{42}$$

$p \text{ 値} > 0.05$  より、 $H_0$  は棄却されない。

6.  $5 \times 5$  2元分割表でデータ

幾何/代数	5	4	3	2	1	計
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
計	10	6	5	2	3	26

が与えられているとする。([5], [6] より引用)

帰無仮説 ( $H_0$ ) : 代数と幾何の成績は独立。

とする。

(a)  $H_0$  の下での当てはめ値を計算せよ。(R or 手計算)

$H_0$  の下での当てはめ値 (理論度数) の計算。セル  $(i, j)$  の当てはめ値は

$$\hat{x}_{ij} = n \frac{x_{i+} x_{+j}}{n^2}$$

で計算できる。

(b) カイ 2 乗適合度統計量  $\chi^2(x)$  を計算せよ。(R or 表)

(c) 上の結果から漸近的  $p$  値を求めよ。

カイ 2 乗検定統計量の式

$$\chi^2(\mathbf{x}) = \sum_{i,j} \frac{(x_{ij} - \hat{x}_{ij})^2}{\hat{x}_{ij}}$$

にしたがって計算を行えば、 $\chi^2(\mathbf{x}) = 25.3376$  という結果が得られる。

R でこれを実行するには次のように行う。

```
R でのカイ 2 乗検定の実行
m <- matrix(c(2,8,0,0,1,3,2,0,0,1,3,1,0,0,0,0,
              1,1,0,0,0,1,1,1),nrow=5,ncol=5)
chisq.test(m)
```

この実行結果によれば、漸近的  $p$  値は 0.06409

(d) マルコフ連鎖モンテカルロ法を用いてサンプリングを行い、 $p$  値を近似的に求めよ。

Q5 と同じようにマルコフ連鎖モンテカルロ法を行う。今回は  $p$  値を計算しなければならないので、得られたサンプル  $\mathbf{x}$  について、 $\chi^2(\mathbf{x}) \geq 25.3376$  を判定して真であれば、カウント  $c$  を増やす。サンプリング終了時に、 $\frac{c}{(\text{サンプル数})}$  が  $p$  値の近似になる。アルゴリズムを改めて書けば次のようになる。

0.  $\mathbf{x} \leftarrow$  (与えられているデータ)  
 カウント  $c \leftarrow 0$   
 基準のカイ 2 乗統計量  $\chi^2 \leftarrow \chi^2(\mathbf{x})$
1.  $B \cup (-B)$  からランダムに元  $\mathbf{z}$  を選ぶ
2.  $\mathbf{x} + \mathbf{z}$  の各要素が全て非負ならば、 $r \leftarrow \frac{f(\mathbf{x}+\mathbf{z})}{f(\mathbf{x})}$   
 それ以外の場合、 $r \leftarrow 0$
3.  $0 \leq R < 1$  の一様乱数  $R$  をとる
4.  $r > R$  であれば、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x} + \mathbf{z}$   
 それ以外の場合、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x}$
5.  $\mathbf{x}_{\text{next}}$  がサンプルとして得られる  
 $\chi^2(\mathbf{x}_{\text{next}}) \geq \chi^2$  ならば  $c \leftarrow c + 1$

6.  $\mathbf{x} \leftarrow \mathbf{x}_{\text{next}}$   
 step 1 に戻る

これを実行するプログラムが 5x5mcmc である。  
 5x5mcmc (乱数 seed) (サンプル数) のように実行する。

```

5x5mcmc の実行結果
./5x5mcmc 0 100000
chisq : 25.337619
0 2 1 1 0 7 4 1 1 1 1 0 3 0 1 1 0 0 0
1 1 0 0 0 0
0 2 1 1 0 7 4 1 1 1 1 0 3 0 1 1 0 0 0
1 1 0 0 0 0
0 2 1 1 0 7 4 1 1 1 1 0 3 0 1 1 0 0 0
1 1 0 0 0 0
....
2 2 0 0 0 7 2 2 0 3 1 0 2 2 0 0 1 1 0
0 0 1 0 0 0
2 2 0 0 0 7 2 2 0 3 1 0 2 2 0 0 1 1 0
0 0 1 0 0 0
2 2 0 0 0 7 2 2 0 3 1 0 2 2 0 0 1 1 0
0 0 1 0 0 0
tchi_count : 6302
approximate p-value : 0.063020
  
```

$p$  値  $> 0.05$  より  $H_0$  は棄却されない。

7.  $2^{7-3}$  一部実施計画 (要因  $A, B, C, D, E, F, G$ , 別名関係  $ABDE = ACDF = BCDG = I$ ) のデータとして次のようなものが与えられているとする。([1] より引用)

run	A	B	C	D	E	F	G	y
1	1	1	1	1	1	1	1	69
2	1	1	1	-1	-1	-1	-1	31
3	1	1	-1	1	1	-1	-1	55
4	1	1	-1	-1	-1	1	1	149
5	1	-1	1	1	-1	1	-1	46
6	1	-1	1	-1	1	-1	1	43
7	1	-1	-1	1	-1	-1	1	118
8	1	-1	-1	-1	1	1	-1	30
9	-1	1	1	1	-1	-1	1	43
10	-1	1	1	-1	1	1	-1	45
11	-1	1	-1	1	-1	1	-1	71
12	-1	1	-1	-1	1	-1	1	380
13	-1	-1	1	1	1	-1	-1	37
14	-1	-1	1	-1	-1	1	1	36
15	-1	-1	-1	1	1	1	1	212
16	-1	-1	-1	-1	-1	-1	-1	52

階層モデル  $H_0: AC/BD/E/F/G$  のデータへの当てはまりを、以下の手順で検証せよ。

- (a) モデル  $H_0$  に対する共変量行列  $M$  を求めよ。

階層モデル  $AC/BD/E/F/G$  だから、主効果  $A, B, C, D, E, F, G$  と 2 因子交互作用  $AC, BD$  であるから、共変量行列は

$$M = (1, d_A, d_B, d_C, d_D, d_E, d_F, d_G, d_{AC}, d_{BD})$$

で具体的には、



```

covariate_mat1.mat
10 16
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
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

```

のようになる。(ただし、これは転置行列の方  $M'$ )

(b) モデル  $H_0$  の下での当てはめ値  $\{m_1, \dots, m_{16}\}$  を求めよ。

R を用いて当てはめ値を計算するには次のようにする。まず次のようなファイルを用意しておく。

```

2_7-3.dat
A, B, C, D, E, F, G, x
1, 1, 1, 1, 1, 1, 1, 69
1, 1, 1, -1, -1, -1, -1, 31
1, 1, -1, 1, 1, -1, -1, 55
1, 1, -1, -1, -1, 1, 1, 149
1, -1, 1, 1, -1, 1, -1, 46
1, -1, 1, -1, 1, -1, 1, 43
1, -1, -1, 1, -1, -1, 1, 118
1, -1, -1, -1, 1, 1, -1, 30
-1, 1, 1, 1, -1, -1, 1, 43
-1, 1, 1, -1, 1, 1, -1, 45
-1, 1, -1, 1, -1, 1, -1, 71
-1, 1, -1, -1, 1, -1, 1, 380
-1, -1, 1, 1, 1, -1, -1, 37
-1, -1, 1, -1, -1, 1, 1, 36
-1, -1, -1, 1, 1, 1, 1, 212
-1, -1, -1, -1, -1, -1, -1, 52

```

R で次のような計算を行えばいい。

```

R での当てはめ値の計算
> dat<-read.table(file="2_7-3.dat", header=T, sep=",")
> dat.glm<-glm(x~A+B+C+D+E+F+G+A*C+B*D, dat, family=poisson)
> fitted(dat.glm)
      1      2      3      4      5      6      7      8
64.52677 47.25345 53.14603 151.07960 30.42595 46.79383 115.24100 32.53337
      9     10     11     12     13     14     15     16
49.42430 46.13193 70.90290 360.53502 35.18867 30.25510 232.14438 51.41770

```

(c) 与えられたデータ  $\{y_1, \dots, y_{16}\}$  について、尤度比検定統計量

$$G(y) = 2 \sum_{i=1}^{16} y_i \log \frac{y_i}{m_i}$$

を計算せよ。また、この値を、自由度 6 のカイ 2 乗分布の上側 5% 点と比較せよ。

$$G(\mathbf{x}) = 19.09271$$

$$\chi_{0.05}^2(6) = 18.5476$$

(d)  $M'$  の Markov 基底を求めよ。

4ti2 を用いて Markov 基底を計算すると 23 個の Markov 基底が得られる。

(e) データに対するモデル  $H_0$  の当てはまりを、尤度比検定統計量にもとづき検証する。マルコフ連鎖モンテカルロ法により、このデータの  $p$  値を計算せよ。

Q5 と同じようにマルコフ連鎖モンテカルロ法を行う。今回は  $p$  値を計算しなければならないので、得られたサンプル  $\mathbf{x}$  について、 $G(\mathbf{x}) \geq 19.09271$  を判定して真であれば、カウント  $c$  を増やす。サンプリング終了時に、 $\frac{c}{(\text{サンプル数})}$  が  $p$  値の近似になる。アルゴリズムを改めて書けば次のようになる。

0.  $\mathbf{x} \leftarrow$  (与えられているデータ)  
    カウント  $c \leftarrow 0$   
    基準の尤度比検定統計量  $G \leftarrow G(\mathbf{x})$
1.  $B \cup (-B)$  からランダムに元  $\mathbf{z}$  を選ぶ
2.  $\mathbf{x} + \mathbf{z}$  の各要素が全て非負ならば、 $r \leftarrow \frac{f(\mathbf{x} + \mathbf{z})}{f(\mathbf{x})}$   
    それ以外の場合、 $r \leftarrow 0$
3.  $0 \leq R < 1$  の一様乱数  $R$  をとる
4.  $r > R$  であれば、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x} + \mathbf{z}$   
    それ以外の場合、 $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x}$
5.  $\mathbf{x}_{\text{next}}$  がサンプルとして得られる  
     $G(\mathbf{x}_{\text{next}}) \geq G$  ならば  $c \leftarrow c + 1$
6.  $\mathbf{x} \leftarrow \mathbf{x}_{\text{next}}$   
    step 1 に戻る

これを実行するプログラムが cov1.mcmc である。

cov1.mcmc (乱数 seed) (サンプル数) のように実行する。

cov1.mcmc の実行結果

```
./cov1_mcmc 0 100000
likelihood ratio statistic : 19.092715
69 48 45 157 23 49 116 34 51 47 73 353 41 22 233 56
68 48 45 158 23 50 115 34 52 47 73 352 41 21 234 56
68 48 45 158 23 50 115 34 52 47 73 352 41 21 234 56
69 48 44 157 23 49 116 35 52 47 73 353 41 21 233 56
.....
64 48 56 161 25 52 110 25 47 46 71 350 40 28 238 56
65 49 55 160 24 51 111 26 47 46 71 350 40 28 238 56
64 49 55 160 24 52 111 26 47 46 72 350 40 28 238 55
64 49 55 160 24 52 111 26 47 46 72 350 40 28 238 55
tlratio_count : 4025
approximate p-value : 0.004025
```

モデル  $AB/AC/BD/E/F/G$  での同様の計算

(a) 共変量行列  $M$

```
covariate_mat2.mat
11 16
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 -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 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 1 1 1 1
```

(b) R で同じ計算を行えば、

```
R での当てはめ値の計算
> dat<-read.table(file="2_7-3.dat", header=T, sep=",")
> dat.glm<-glm(x~A+B+C+D+E+F+G+A*B+A*C+B*D, dat, family=poisson)
> fitted(dat.glm)
      1      2      3      4      5      6      7      8
63.12502 44.15620 49.96722 146.75155 33.08098 48.63780 120.01126 35.26996
      9     10     11     12     13     14     15     16
49.99003 48.73134 74.91772 365.36090 33.03148 29.24715 226.87628 47.84509
>
```

(c)

$$G(y) = 16.8381, \chi_{0.05}^2(5) = 11.0705$$

(d)  $M'$  の Markov 基底を 4ti2 を用いて計算すれば 41 個の Markov 基底が得られる。

(e) cov2.mcmc を実行した結果は次の通り。

```
cov2.mcmc の実行結果
./cov2_mcmc 1 100000
likelihood ratio statistic : 16.838141
p : 0.496995 0.135229
9.003644 62 38 49 155 36 53 115 33 45 41 82 371 44 31 218 44
p : 1.240357 0.939720
8.553470 62 39 50 153 36 52 116 33 44 41 82 372 43 33 218 43
p : 2.046756 0.645705
7.119319 62 40 50 152 35 52 117 33 44 42 82 371 42 33 219 43
p : 1.052914 0.914755
7.024061 64 40 48 152 34 51 118 34 44 42 82 371 43 32 218 44
p : 1.267671 0.494924
...
p : 2.105334 0.048496
4.524540 63 46 49 146 37 43 115 42 55 42 71 371 34 30 227 46
p : 0.991831 0.143881
4.548710 64 45 48 147 37 43 115 42 55 42 71 371 35 29 226 47
p : 0.502882 0.824895
4.548710 64 45 48 147 37 43 115 42 55 42 71 371 35 29 226 47
p : 2.006670 0.186261
3.159083 64 45 48 147 36 44 116 41 54 43 72 370 35 29 226 47
tlratio_count : 401
approximate p-value : 0.004010
```

## 参考文献

- [1] S.Aoki, A.Takemura, Markov chain Monte Carlo tests for designed experiments, arXiv:math.ST0611463v1
- [2] S.Aoki, A.Takemura, Markov basis for design of experiments with three-level factors, arXiv:0709.4323v2
- [3] B.Sturmfels, Gröbner Bases and Convex Polytopes, 1995, American Mathematical Society, University lecture series 8.
- [4] 伊庭 幸人, 種村 正美, 大森 裕浩, 和合 肇, 佐藤 整尚, 高橋 昭彦, “統計科学のフロンティア 12, 計算統計 II マルコフ連鎖モンテカルロ法とその周辺”, 岩波書店, 2005
- [5] 日比孝之編, グレブナー基底の現在, 数学書房, 2006
- [6] 青木敏, マルコフ連鎖、モンテカルロ法による分割表の解析, 応用統計学会第 27 回シンポジウム発表資料, 2005