マルコフ基底と実験計画, 問題集, 2012.02.22
計算の実行に必要なプログラムなどは
http://www.math.kobe-u.ac.jp/~nakayama/gb_school_stat
で入手できる.

1 マルコフ基底の計算

例題 1 2×2 分割表で行和, 列和を固定したものを考える. これについての配置行列 A とそのマルコフ基底を求めよ.

解答 配置行列は

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

である.

- **(トーリックイデアル I_A の生成系からマルコフ基底を得る方法)** トーリックイデアル I_A の生成系がマルコフ基底であったので, I_A を計算する. I_A を計算するには,

$$\langle u_1 - v_1 v_3, u_2 - v_2 v_3, u_3 - v_2 v_3, u_4 - v_2 v_4 \rangle \cap K[u_1, u_2, u_3, u_4]$$

を計算すればよい. 下では, $v_1 > v_2 > v_3 > v_4 > u_1 > u_2 > u_3 > u_4$ なる辞書式順序についてのグレブナー基底計算を行っている.

Asir での計算例

```
[1632] Id=[u1-v1*v3, u2-v1*v4, u3-v2*v3, u4-v2*v4];
[u1-v3*v1,u2-v4*v1,u3-v2*v3,-v2*v4+u4]
[1631] nd_gr(Id, [v1,v2,v3,v4,u1,u2,u3,u4],0,2);
[-u4*u1+u3*u2,v4*u3-u4*v3,v4*u1-v3*u2,-v2*v4+u4,u3-v2*v3,v2*u2-u4*v1,v2*u1-v1*u3,u2-v4*v1,u1-v3*v1]
```

よって $I_A = \langle u_1 u_4 - u_2 u_3 \rangle$. A のマルコフ基底は

1	-1
-1	1

となる.

- **(4ti2 を用いたマルコフ基底の計算)** 4ti2 を使いマルコフ基底を計算する. 行列 A を表す次のような入力ファイル cont2x2 を用意する. ここで 1 行目の 5 6 は行列の行数と列数を表し, その後に行列データを書く.

4ti2 入力ファイル cont2x2

```
4 4
1 1 0 0
0 0 1 1
1 0 1 0
0 1 0 1
```

4ti2 のマルコフ基底を求めるコマンド markov を実行する.

4ti2 コマンド markov の実行

```
$ markov cont2x2
... 略
```

マルコフ基底が出力ファイル cont2x2.mar に出力される.

4ti2 出力ファイル cont2x2.mar

```
1 4
 1 -1 -1  1
```

第 1 行目はマルコフ基底の数と変数の数を表し, それ以降の各行ベクトルがマルコフ基底である.

問題 2 次の各表について配置行列 A を求め, A についてマルコフ基底を, トーリックイデアル I_A の生成系を求める方法 (グレブナー基底を用いた消去法による方法) か, `4ti2` を用いた方法で求めよ.

1. 2×3 分割表の行和, 列和を固定したものの配置行列, すなわち
$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$
2. 5×5 分割表の行和, 列和を固定したものの配置行列
3. 3×3 表で行和, 列和, 2 つの対角和 $(x_{11}+x_{22}+x_{33}, x_{13}+x_{22}+x_{31})$ を固定したもの (魔方陣).
4. $2\times 2\times 2$ 分割表ですべての 2 次元周辺頻度 $x_{ij\cdot}, x_{i\cdot k}, x_{\cdot jk}$ を固定したもの.

問題 3 次の各配置行列 A についてマルコフ基底を `4ti2` を用いた方法で求めよ. (配置行列が大きくなるので, プログラムで配置行列を書かないと大変)

1. $3\times 3\times 3$ 分割表ですべての 2 次元周辺頻度 $x_{ij\cdot}, x_{i\cdot k}, x_{\cdot jk}$ を固定したもの.

2 $\mathbf{t}$ ファイバーの列挙

配置行列 A ($d\times n$ 行列) のマルコフ基底 $\mathcal{B}$ と, $\mathbf{t}\in\mathbb{Z}^d$ が与えられている時, $\mathbf{t}$ ファイバー $\mathcal{F}_{\mathbf{t}} = \{\mathbf{x}\in(\mathbb{Z}_{\geq 0})^n \mid A\mathbf{x}=\mathbf{t}\}$ の全ての元を列挙するアルゴリズムは次の通りである. $\mathbf{t}$ ファイバーの元の 1 つ $\mathbf{x}^o$ が与えられているとする.

(ファイバー $\mathcal{F}_{\mathbf{t}}$ の元を全て列挙するアルゴリズム)

入力: マルコフ基底 $\mathcal{B}$, $\mathbf{x}^o$
出力: ファイバー $\mathcal{F}_{\mathbf{t}}$ のすべての元

1. Active $\leftarrow \{\mathbf{x}^o\}$, Fiber $\leftarrow \{\mathbf{x}^o\}$
2. Active $\neq \phi$ である限り以下を続ける.
   - 2.1. $\mathbf{u}\leftarrow$ (Active の元いづれか), Active $\leftarrow$ Active $\setminus\{\mathbf{u}\}$
   - 2.2. $\mathcal{B}$ の各元 $\mathbf{v}=\mathbf{v}^+-\mathbf{v}^-$ について,
     - 2.2.1 $\mathbf{u}-\mathbf{v}^+\geq\mathbf{0}$ かつ $\mathbf{u}-\mathbf{v}\notin$ Fiber ならば,
       Active $\leftarrow$ Active $\cup\{\mathbf{u}-\mathbf{v}\}$, Fiber $\leftarrow$ Fiber $\cup\{\mathbf{u}-\mathbf{v}\}$
     - 2.2.2 $\mathbf{u}-\mathbf{v}^-\geq\mathbf{0}$ かつ $\mathbf{u}+\mathbf{v}\notin$ Fiber ならば,
       Active $\leftarrow$ Active $\cup\{\mathbf{u}+\mathbf{v}\}$, Fiber $\leftarrow$ Fiber $\cup\{\mathbf{u}+\mathbf{v}\}$
3. Fiber を出力.

例題 4 2×3 分割表で行和, 列和を固定したものを考える. $\mathbf{x}^o = \boxed{\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}}$, $\mathbf{t}=A\mathbf{x}^o=(1,1,1,1,1)'$ とする. $\mathbf{t}$ ファイバー $\mathcal{F}_{\mathbf{t}}$ を計算せよ.

解答 マルコフ基底 $\mathcal{B}$ として,
$$\boxed{\begin{array}{ccc} 0 & -1 & 1 \\ 0 & 1 & -1 \end{array}}, \boxed{\begin{array}{ccc} 1 & -1 & 0 \\ -1 & 1 & 0 \end{array}}, \boxed{\begin{array}{ccc} 1 & 0 & -1 \\ -1 & 0 & 1 \end{array}}$$
が得られる.

- **(手で計算する)** 上のアルゴリズムに従い, 地道に計算すると次の元たちが得られる.

$$\boxed{\begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{matrix}}, \boxed{\begin{matrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}}, \boxed{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{matrix}}$$

- **(プログラム enumerate_fiber) を用いて計算する)** C のプログラム enumerate_fiber.c は上のアルゴリズムに従い, 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 start.txt とマルコフ基底の元たちを書いたファイル markov.txt を用意する. 元の書き方は 4ti2 の場合と同様である. (すなわち, 1 行目に下に書く行列の行数と列数を書き, 2 行目以降にデータを書く.)

enumerate_fiber の入力ファイル start.txt

```
1 6
0 1 0 1 0 1   <- ファイバー中のいずれかの元
```

enumerate_fiber の入力ファイル markov.txt

```
3 6
 0 -1  1  0  1 -1  <- マルコフ基底の元たち
 1 -1  0 -1  1  0
 1  0 -1 -1  0  1
```

プログラム enumerate_fiber の実行は次のように行う. 入力ファイルとして start.txt と markov.txt を与える.

プログラム enumerate_fiber の実行

```
$ ./enumerate_fiber start.txt markov.txt
n_move : 1, msize : 6
start_v :  <-- スタートのファイバーの元
 0  1  0  1  0  1
n_move : 3, msize : 6
move : 3 6 <-- move の元たち
 0 -1  1  0  1 -1
 1 -1  0 -1  1  0
 1  0 -1 -1  0  1
0 0 1 1 1 0
1 0 0 0 1 1
depth : 0, count : 3, queue : 2
depth : 1, count : 3, queue : 1
depth : 2, count : 3, queue : 0
count : 3
fiber : <-- ファイバーの元たち
0 1 0 1 0 1
0 0 1 1 1 0
1 0 0 0 1 1
```

問題 5 先の section で計算したマルコフ基底を使って, 次の $\mathbf{t}$ ファイバーを列挙せよ.

1. 2×3 分割表で行和, 列和を固定したもの.

   $\mathbf{x}^o = \boxed{\begin{matrix} 0 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}}$, $\mathbf{t} = A\mathbf{x}^o = (2,3,1,2,2)$ とおく. $\mathbf{t}$ ファイバー $\mathcal{F}_{\mathbf{t}}$. (手計算で)

2. 5×5 分割表で行和, 列和を固定したもの.

   $\mathbf{x}^o = \boxed{\begin{matrix} 2 & 1 & 1 & 0 & 0 \\ 8 & 3 & 3 & 0 & 0 \\ 0 & 2 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{matrix}}$, $\mathbf{t} = A\mathbf{x}^o = (4,14,5,2,1,10,6,5,2,3)'$ とおく. $\mathbf{t}$ ファイバー $\mathcal{F}_{\mathbf{t}}$. (計算するには 10 分程度時間がかかる)

3. 3×3 表で行和, 列和, 2 つの対角和 $(x_{11}+x_{22}+x_{33}, x_{13}+x_{22}+x_{31})$ を固定したもの (3×3 魔方陣).

   $\mathbf{x}^o = \boxed{\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}}$, $\mathbf{t} = A\mathbf{x}^o = (3,3,3,3,3,3,3,3)$ とおく. $\mathbf{t}$ ファイバー $\mathcal{F}_{\mathbf{t}}$.

4. $2\times2\times2$ 分割表ですべての 2 次元周辺頻度 $x_{ij\cdot}, x_{i\cdot k}, x_{\cdot jk}$ を固定したもの.

$\mathbf{x}^o = \boxed{\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}}, \boxed{\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}}$, $\mathbf{t} = A\mathbf{x}^o = (1,1,1,1,1,1,1,1,1,1,1,1)'$ とおく. $\mathbf{t}$ ファイバー $\mathcal{F}_{\mathbf{t}}$.

問題 6 先の section で計算したマルコフ基底を使って，次の $\mathbf{t}$ ファイバーを列挙せよ.

1. $3\times3\times3$ 分割表ですべての 2 次元周辺頻度 $x_{ij\cdot}, x_{i\cdot k}, x_{\cdot jk}$ を固定したもの.

$\mathbf{x}^o = \boxed{\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{array}}, \boxed{\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{array}}, \boxed{\begin{array}{ccc} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{array}}$,

$\mathbf{t} = A\mathbf{x}^o = (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)'$ とおく. $\mathbf{t}$ ファイバー $\mathcal{F}_{\mathbf{t}}$ (3×3 ラテン方陣と対応).

3　2 元分割表の独立性の検定

例題 7 2×2 分割表でデータ

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

が与えられているとする.

1. 「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 $p(\mathbf{x}) = \frac{1}{Z}\frac{1}{x_{11}!x_{12}!x_{21}!x_{22}!}$ を求めよ. ただし，Z は正規化定数である。

2. 「$X_{11} \geq c$ 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.

解答　1. 実現値 $\mathbf{x}$ は次の 5 通りである.

$$\boxed{\begin{array}{cc} 4 & 0 \\ 1 & 5 \end{array}}, \boxed{\begin{array}{cc} 3 & 1 \\ 2 & 4 \end{array}}, \boxed{\begin{array}{cc} 2 & 2 \\ 3 & 3 \end{array}}, \boxed{\begin{array}{cc} 1 & 3 \\ 4 & 2 \end{array}}, \boxed{\begin{array}{cc} 0 & 4 \\ 5 & 1 \end{array}}$$

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.

2. 観測値の x_{11} は 3 より $c = 3$ として,

$$p\text{ 値} = \Pr(X_{11} \geq 3) = \frac{10}{42} + \frac{1}{42} = \frac{11}{42} = 0.261$$

p 値 > 0.05 より, H_0 は棄却されない.

例題 8 離散確率変数 x_1, x_2, x_3 , 各変数は値 $+1$ 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 は正規化定数であり,

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

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. (**参考**：[4, p.8])

解答 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である.

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

このアルゴリズムに従って作った R のプログラムが `metropolis.r` である. 実行結果は以下の通りである.

R のプログラム `metropolis.r` の実行

```
> source("metropolis.r")
metropolis(サンプル数, 初期値) で実行
e.g. metropolis(10000, c(-1,-1,-1))
> metropolis(10000, c(-1,-1,-1))
[1] -1 -1 -1  <- 得られるサンプルが順に表示される
[1] -1 -1 -1
...
[1]  1 -1 -1
[1] 2127  957  923  987  918  934  924 2230 <- 各サンプルの頻度
実験値
[1] 0.2127 0.0957 0.0923 0.0987 0.0918 0.0934 0.0924 0.2230
理論値
[1] 0.21294838 0.09568387 0.09568387 0.09568387 0.09568387 0.09568387
0.09568387 0.21294838
```

例題 9 2×3 分割表の周辺度数 (行和, 列和) $\mathbf{t} = (5, 15, 5, 5, 10)'$ とした時の $\mathbf{t}$ ファイバー

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

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, $\sum_{x \in \mathcal{F}_{\mathbf{t}}} \frac{1}{x_{11}!x_{12}!x_{13}!x_{21}!x_{22}!x_{23}!}$)

解答 2×3 分割表のマルコフ基底は

$$\mathcal{B} = \left\{ \boxed{\begin{array}{ccc} 1 & -1 & 0 \\ -1 & 1 & 0 \end{array}}, \boxed{\begin{array}{ccc} 1 & 0 & -1 \\ -1 & 0 & 1 \end{array}}, \boxed{\begin{array}{ccc} 0 & 1 & -1 \\ 0 & -1 & 1 \end{array}} \right\}$$

である.

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 $\mathcal{F}_{\mathbf{t}}$ 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.

1. ファイバー $\mathcal{F}$ の元 $\mathbf{x}$ を 1 つ取り出す (なんでもよい).
   $e \leftarrow 0$
2. $\mathcal{B} \cup (-\mathcal{B})$ からランダムに元 $\mathbf{z}$ を選ぶ.
3. $\mathbf{x} + \mathbf{z}$ 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, $r \leftarrow 0$

4. $0 \le R < 1$ の一様乱数 R をとる.

5. $r > R$ であれば, $\mathbf{x}_{\rm next} \leftarrow \mathbf{x} + \mathbf{z}$
それ以外の場合, $\mathbf{x}_{\rm next} \leftarrow \mathbf{x}$

6. $\mathbf{x}_{\rm next}$ がサンプルとして得られる.
$e \leftarrow e+$ ($\mathbf{x}_{\rm next}$ の $2x_{11} + x_{12}$)

7. $\mathbf{x} \leftarrow \mathbf{x}_{\rm next}$
step 2 に戻る.

8. $e/$ (サンプル数) が期待値の推定値.

このアルゴリズムに従い作った R のプログラムが `2x3mcmc.r` である. 次のように R 中で実行する. 正確な期待値は $15/4 = 3.75$ である.

R のプログラム `2x3mcmc.r` の実行結果

```
> source("2x3mcmc.r")
c2x3mcmc(n, burnin, 初期値) で実行
e.g. c2x3mcmc(10000, 10000, c(0,5,0,5,0,10))
> c2x3mcmc(10000, 10000, c(0,5,0,5,0,10))
[1] 2 1 2 3 4 8  <- 得られるサンプルが表示される
[1] 1 2 2 4 3 8
....
[1] 0 0 5 5 5 5
[1] 0 0 5 5 5 5
平均
[1] 3.7485
```

問題 10 上の例題と同じデータについて,

1. $\mathbf{t}$ ファイバーを全列挙せよ.

2. 正規化定数 Z , 各分割表の確率を計算せよ.

3. $2x_{11} + x_{12}$ の正確な期待値を計算せよ.

例題 11 5×5 の 2 元分割表

幾何/推測	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

について考察を行う. 帰無仮説を “H_0 : 推測と幾何の成績は独立” とする.

1. H_0 の下での当てはめ値を計算せよ. カイ 2 乗適合度統計量 $\chi^2(x)$ を計算せよ. 上の結果から p 値を求めよ.

2. 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.

解答
1. 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, $\hat{x}_{ij} = n\frac{x_{i+}}{n}\frac{x_{+j}}{n}$ で計算できる.
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 でこれら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(2,8,0,0,0,1,3,2,0,0,1,3,1,0,0,0,0,1,1,0,0,0,1,1,1),nrow=5,ncol=5)
t <- chisq.test(m)
> t

Pearson's Chi-squared test

data:  m
X-squared = 25.3376, df = 16, p-value = 0.06409

> t$expect
          [,1]      [,2]      [,3]       [,4]      [,5]
[1,] 1.5384615 0.9230769 0.7692308 0.30769231 0.4615385
[2,] 5.3846154 3.2307692 2.6923077 1.07692308 1.6153846
[3,] 1.9230769 1.1538462 0.9615385 0.38461538 0.5769231
[4,] 0.7692308 0.4615385 0.3846154 0.15384615 0.2307692
[5,] 0.3846154 0.2307692 0.1923077 0.07692308 0.1153846
```

この実行結果によれば, p 値は 0.06409.

2. 例題 9 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.

   1. $\mathbf{x} \leftarrow$ (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 $\chi^2 \leftarrow \chi^2(\mathbf{x})$
   2. $\mathcal{B} \cup (-\mathcal{B})$ からランダムに元 $\mathbf{z}$ を選ぶ.
      (ここで $\mathcal{B}$ は 5×5 分割表のマルコフ基底.)
   3. $\mathbf{x}+\mathbf{z}$ 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, $r \leftarrow 0$.
      (ここで $f(\mathbf{x}) = \frac{1}{Z}\frac{1}{\Pi_{i,j} x_{ij}!}$, Z は正規化定数.)
   4. $0 \leq R < 1$ の一様乱数 R をとる.
   5. $r > R$ であれば, $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x}+\mathbf{z}$
      それ以外の場合, $\mathbf{x}_{\text{next}} \leftarrow \mathbf{x}$
   6. $\mathbf{x}_{\text{next}}$ がサンプルとして得られる.
      $\chi^2(\mathbf{x}_{\text{next}}) \geq \chi^2$ ならば $c \leftarrow c+1$
   7. $\mathbf{x} \leftarrow \mathbf{x}_{\text{next}}$
      step 2 に戻る.

これを実行する R のプログラムが `5x5mcmc.r` である.

R のプログラム `5x5mcmc.r` の実行結果

```
> source("5x5mcmc.r")
c5x5mcmc(n, burnin, 初期値) で実行
e.g. c5x5mcmc(10000, 10000, c(2,1,1,0,0,8,3,3,0,0,0,2,1,1,1,0,0,0,1,1,0,
0,0,0,1))
> c5x5mcmc(10000, 10000, c(2,1,1,0,0,8,3,3,0,0,0,2,1,1,1,0,0,0,1,1,0,0,
0,0,1))
X-squared
 25.33762
.... 略
 [1] 2 0 1 0 1 7 2 4 1 0 1 3 0 0 1 0 0 0 1 1 0 1 0 0 0
.... 略
 [1] 1 1 2 0 0 7 1 3 1 2 1 3 0 1 0 1 1 0 0 0 0 0 0 0 1
p 値の近似値
[1] 0.0639
```

p 値 > 0.05 より H_0 は棄却されない.

問題 12 5×5 の 2 元分割表

幾何/推測	5	4	3	2	1	計
5	1	0	0	0	0	1
4	0	2	1	0	0	3
3	0	1	2	0	0	3
2	0	0	0	0	1	1
1	0	0	0	1	0	1
計	1	3	3	1	1	9

に対して,

1. 上と同様の計算を行え.
2. $\mathbf{t} = (1,3,3,1,1,1,3,3,1,1)$ とおく. $\mathbf{t}$ ファイバーを全列挙して, 正確な p 値を計算せよ.

4 実験計画法

例題 13 2^{7-3} 一部実施計画 (要因 A,B,C,D,E,F,G, 別名関係 $ABDE = ACDF = BCDG = I$)

run	A	B	C	D	E	F	G	不良品数
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

について考察を行う. このデータに対して, Hamada, Nelder [1] では, H_0 : $AC/BD/E/F/G$ というモデルへの当てはまりを考察している. このモデルの当てはまりを以下の手順で検証する.

1. モデル H_0 に対する共変量行列 M を求めよ.
2. モデル H_0 の下での当てはめ値 $(m_1,\ldots,m_{16})$ を求めよ.
3. 与えられたデータ $\mathbf{y} = (y_1,\ldots,y_{16})$ について, 尤度比検定統計量
$$G(\mathbf{y}) = 2\sum_{i=1}^{16} y_i \log \frac{y_i}{m_i}$$
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.
4. M' のマルコフ基底を求めよ.
5. 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.

解答 1. 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, 具体的に行列を 4ti2 の入力ファイルの形式で書けば,

4ti2 入力ファイル covariate_mat1

```
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 M' の方)

2. 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.

R の入力ファイル 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よい.

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
 64.52677  47.25345  53.14603 151.07960  30.42595  46.79383
        7         8         9        10        11        12
115.24100  32.53337  49.42430  46.13193  70.90290 360.53502
       13        14        15        16
 35.18867  30.25510 232.14438  51.41770
```

3. $G(\mathbf{x}) = 19.09271$, $\chi^2_{0.05}(6) = 12.59159$

4. 4ti2 コマンド markov を用いてマルコフ基底を計算すると 23 個のマルコフ基底が得られる.

5. 例題 9 と同じようにマルコフ連鎖モンテカルロ法を行う. 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.

   1. $\mathbf{x} \leftarrow$ (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 $G \leftarrow G(\mathbf{x})$
   2. $\mathcal{B} \cup (-\mathcal{B})$ からランダムに元 $\mathbf{z}$ を選ぶ.
      (ここで $\mathcal{B}$ は, 行列 M' のマルコフ基底.)
   3. $\mathbf{x}+\mathbf{z}$ 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、$r \leftarrow 0$
      (ここで $f(\mathbf{x}) = \frac{1}{Z}\frac{1}{\Pi_{i,j} x_{ij}!}$, Z は正規化定数.)
   4. $0 \leq R < 1$ の一様乱数 R をとる.

5. $r > R$ であれば, $\mathbf{x}_{\rm next} \leftarrow \mathbf{x} + \mathbf{z}$
   それ以外の場合, $\mathbf{x}_{\rm next} \leftarrow \mathbf{x}$
6. $\mathbf{x}_{\rm next}$ がサンプルとして得られる.
   $G(\mathbf{x}_{\rm next}) \geq G$ ならば $c \leftarrow c + 1$
7. $\mathbf{x} \leftarrow \mathbf{x}_{\rm next}$
   step 2 に戻る.

これを実行する R のプログラムが `cov1_mcmc.r` である.

R のプログラム `cov1_mcmc.r` の実行結果

```
> source("cov1_mcmc.r")
cov1_mcmc(n, burnin) で実行
e.g. cov1_mcmc(10000, 10000)
> cov1_mcmc(10000, 10000)
likelihood ratio stat :
       1
19.09271
 [1]  69  32  54 149  45  43 118  31  44  45  71 379  37  35 213  52
 ...     <- 得られたサンプルたちが表示される (略)
 [1]  64  39  51 154  34  52 117  30  44  45  79 367  41  31 221  48
n        <- サンプル数
[1] 10000
count    <- 条件に合うサンプル数
[1] 31
p 値の近似値
[1] 0.0031
```

問題 14 例題 13 の同じデータについて, モデル $AB/AC/BD/E/F/G$ の当てはまりを検証せよ.

例題 15 例題 13 の 2^{7-3} 一部実施計画について, 次の考え方をもとにグレブナー基底を使って別名関係を全て列挙せよ.
(考え方)
因子 A, B, C, D, E, F, G が値 $\{+1, -1\}$ の値をとることと, 定義関係 $ABDE = ACDF = BCDG = 1$ を満たすことから, その関係式たちを生成元とするイデアル

$$I = \langle A^2-1, B^2-1, C^2-1, D^2-1, E^2-1, F^2-1, G^2-1, \\ ABDE-1, ACDF-1, BCDG-1\rangle$$

を考える. 例えば, この一部実施計画において, 「因子 A, B の 2 因子交互作用」と「因子 F, G の 2 因子交互作用」は別名関係にあるが, イデアル I を使った見方をすれば, $AB - FG \in I$ (AB と FG が I 中の関係式で移り合う) ということである. これを判定するには, I の適当な項順序 $<$ についてのグレブナー基底 G を求め, $AB - FG$ を G で割り, 余りが 0 かどうかを判定すればよい (グレブナー基底によるイデアル所属問題). 上と同じことであるが, AB を G で割った余りと FG を G で割った余りが同じ単項式であることをいってもよい.

解答 Asir においてグレブナー基底計算で, AB, FG が別名関係にあるかどうかと, AC, BD が別名関係にあるかどうかを判定すると次のようになる.

Asir での別名関係の判定

```
[1361] Id=[a^2-1,b^2-1,c^2-1,d^2-1,e^2-1,f^2-1,g^2-1,a*b*d*e-1,
a*c*d*f-1,b*c*d*g-1];
[1362] VL=[a,b,c,d,e,f,g];
[1364] GB=nd_gr(Id,VL,0,0);        <- イデアル Id のグレブナー基底 GB
[a^2-1,-b*a+g*f,b^2-1,-c*a+g*e,-c*b+g*d,c^2-1,-d*a+f*c,d*b-g*c,-g*b+d*c,
d^2-1,-e*a+g*c,-e*b+f*c,-g*a+e*c,e*d-g*f,e^2-1,-f*a+g*b,-g*a+f*b,f*d-g*e,
g*d-f*e,f^2-1,g^2-1]
[1365] p_nf(a*b-f*g, GB, VL, 0);   <- a*b-f*g を GB で割った余り
0                                  <- 0 だから AB, FG は別名関係
[1367] p_nf(a*c-b*d, GB, VL, 0);   <- a*c-b*d を GB で割った余り
g*c-g*e                            <- g*c-g*e だから AC, BD は別名関係にない
```

別名関係の列挙は, 単項式 $A^{i_1}B^{i_2}C^{i_3}D^{i_4}E^{i_5}F^{i_6}G^{i_7}$ $(i_1, \ldots, i_7 \in \{0, 1\})$ について, I のグレブナー基底 G で割り, その余りとして出てきた単項式で分類すればよい.

Asir での別名関係の列挙 (`alias-2.rr`)

```
[1351] load("alias-2.rr");
[a^2-1,b^2-1,c^2-1,d^2-1,e^2-1,f^2-1,g^2-1,e*d*b*a-1,f*d*c*a-1,
g*d*c*b-1]
.... 略
[[g*f*c,g*e*b,g*d*a,f*e*a,f*d*b,e*d*c,c*b*a,g*f*e*d*c*b*a],
 [f*c,e*b,d*a,g*f*e*a,g*f*d*b,g*e*d*c,g*c*b*a,f*e*d*c*b*a],
 [g*a,f*b,e*c,g*f*d*c,g*e*d*b,f*e*d*a,d*c*b*a,g*f*e*c*b*a],
 [g*b,f*a,d*c,g*f*e*c,g*e*d*a,f*e*d*b,e*c*b*a,g*f*d*c*b*a],
 [g*c,e*a,d*b,g*f*e*b,g*f*d*a,f*e*d*c,f*c*b*a,g*e*d*c*b*a],
 [g*d,f*e,c*b,g*f*c*a,g*e*b*a,f*d*b*a,e*d*c*a,g*f*e*d*c*b],
 [g*e,f*d,c*a,g*f*c*b,g*d*b*a,f*e*b*a,e*d*c*b,g*f*e*d*c*a],
 [g*f,e*d,b*a,g*e*c*b,g*d*c*a,f*e*c*a,f*d*c*b,g*f*e*d*b*a],
 [a,g*f*b,g*e*c,f*d*c,e*d*b,g*f*e*d*a,g*d*c*b*a,f*e*c*b*a],
 [b,g*f*a,g*d*c,f*e*c,e*d*a,g*f*e*d*b,g*e*c*b*a,f*d*c*b*a],
 [c,g*e*a,g*d*b,f*e*b,f*d*a,g*f*e*d*c,g*f*c*b*a,e*d*c*b*a],
 [d,g*f*e,g*c*b,f*c*a,e*b*a,g*f*d*b*a,g*e*d*c*a,f*e*d*c*b],
 [e,g*f*d,g*c*a,f*c*b,d*b*a,g*f*e*b*a,g*e*d*c*b,f*e*d*c*a],
 [f,g*e*d,g*b*a,e*c*b,d*c*a,g*f*e*c*a,g*f*d*c*b,f*e*d*b*a],
 [g,f*e*d,f*b*a,e*c*a,d*c*b,g*f*e*c*b,g*f*d*c*a,g*e*d*b*a],
 [1,g*f*e*d,g*f*b*a,g*e*c*a,g*d*c*b,f*e*c*b,f*d*c*a,e*d*b*a]]
```

この方針で自動的に計算する Asir のプログラムが `alias-2.rr` である. 最後の行に表示されているリストの各要素が別名関係を表している. 例えば, リストの 9 番目の要素

`[a,g*f*b,g*e*c,f*d*c,e*d*b,g*f*e*d*a,g*d*c*b*a,f*e*c*b*a]`

は, 別名関係

$$A = BFG = CEG = CDF = BDE = ADEFG = ABCDG = ABCEF$$

を表している.

問題 16 2 水準 $\{-1, 1\}$ の 4 因子 A, B, C, D が別名関係 $ACD = 1$ を持つ 2^{4-1} 一部実施計画を考える. グレブナー基底を使った方法で, 別名関係を全て列挙せよ.

参考文献

[1] M. Hamada, J.A. Nelder, Generalized linear models for quality-improvement experiments, Journal of Quality and Technology **29** (1997), 92–304.

[2] R. Hemmecke, P.N. Malkin, Computing generating sets of lattice ideals and Markov bases of lattices, Journal of Symbolic Computation **44** (2009), 1463–1476.

[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] 4ti2 team, 4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces, `http://www.4ti2.de/`

[6] J.A. De Loera, et al., LattE, `http://www.math.ucdavis.edu/~latte/`

[7] R Development Core Team, R: A Language and Environment for Statistical Computing, `http://www.r-project.org/`

[8] M. Noro, et al., Risa/Asir, `http://www.math.kobe-u.ac.jp/Asir/asir-ja.html`