計算代数の direct sampler への応用

高山信毅 (神戸大学)*, 間野修平 (統数研)

参考文献等 $\boxed{\text{hgm OpenXM}}$ search.

Diaconis-Sturmfels (1998) found that any set of generators of the affine toric ideal of a matrix A gives a Markov basis for the Markov chain Monte Carlo (MCMC) simulation for the distribution associated to A. It can be constructed by a Gröbner basis computation.

http://www.math.kobe-u.ac.jp/HOME/taka/2018/rims-2018.pdf, http://www.math.kobe-u.ac.jp/HOME/taka/2018/rims-2018-en.pdf (partial) English translation.

The A-hypergeometric system is defined by the toric ideal plus first order differential operators. The partition function of the distribution is a solution of it. S.Mano found that the recurrence relation of the partition function gives an "effective" direct sampler for the distribtion. Some computer algebra algorithms can be applied to obtain the recurrence relation (contituity relation).

*平代達哉君 (神戸大, M2) の最近の結果も含む.

$\boxed{\text{Sampler とは}}$ Sampler = 与えられた分布に従う乱数 (ベクトル) を生成する

Example

$$u_1 + u_2 = \beta, u_i \in \mathbf{N}_0$$

を満たす数のベクトル $u = (u_1, u_2)$ を分布が

$$\frac{\beta!}{u_1!u_2!} p_1^{u_1} p_2^{u_2} \qquad (1)$$

となるようにランダムに生成しなさい．ここで $p_i \geq 0$, $p_1 + p_2 = 1$.

$$P(U = u) = (1) \text{式}$$

たとえば $\beta = 2$, $p_i = 1/2$ なら

$$P(U = (0,2)) = \frac{1}{4}, P(U = (1,1)) = \frac{1}{2}, P(U = (2,0)) = \frac{1}{4}$$

```
rbinom(20,size=2,prob=1/2);
[1] 1 2 1 2 1 1 0 2 0 1 1 2 2 0 2 2 1 2 1 1
```

$(1,1), (2,0), (1,1), (2,0), (1,1), (1,1), (0,2), \ldots$ に対応.

Direct sampler でこの乱数を作るには?

Input: β, p_1, p_2

Output: (u_1, u_2)

1. $(c_1, c_2) = (0, 0)$ (init count vector)
2. $e_1 = \frac{p_1}{p_1+p_2}$, $e_2 = \frac{p_2}{p_1+p_2}$.
3. $[0, 1]$ を $e_1 : e_2$ の区間 E_1, E_2 に分割.
4. $[0, 1]$ に値をもつ一様乱数 t を一つ生成.
5. **if** $t \in E_1$, **then** $c_1 + +$, $\beta - -$ **else if** $t \in E_2$, **then** $c_2 + +$, $\beta - -$.
6. **if** $\beta > 0$, **then** goto 4 **else** return $u = (c_1, c_2)$.

$\beta = 3$ でこの direct sampler の実行実験! (内職してる人はそのまま で).

Theorem (well-known)

(u_1, u_2) を得る確率は (1) 式.

Proof. Step 5 で選ばれる index の列を $i_1, i_2, \ldots, i_\beta$ とする. i_j は 1 か 2 である. $\sharp\{k \mid i_k = 1\} = c_1$, $\sharp\{k \mid i_k = 2\} = c_2$, であった. よってこの index 列を得る確率は $p_1^{c_1} p_2^{c_2}$. 同じ c_1, c_2 の時 1, 2 の並べ方は $\binom{\beta}{c_1}$ とおり. $\square$

$(3,0)$	$\frac{1}{8}$
$(2,1)$	$\frac{3}{8}$
$(1,2)$	$\frac{3}{8}$
$(0,3)$	$\frac{1}{8}$

A 分布とは?

A: $d \times n$ 行列. 整数成分 [†]. $p \in \mathbf{R}^n_{\geq 0}$. $u \in \mathbf{N}_0^n$, $\beta \in \mathbf{N}_0^d$.

$$Z_A(\beta; p) = \sum_{Au=\beta, u \in \mathbf{N}_0^n} \frac{p^u}{u!} \tag{2}$$

とおいて, $u \in \mathbf{N}_0^n$ に対して A 分布を次の式できめる.

$$P(U = u) = \frac{p^u}{u! Z_A(\beta; p)} \tag{3}$$

$u! = u_1! \cdots u_n!$.
例: $A = (1, 1)$ なら最初の例. $\beta_1! Z_A(\beta; p) = (p_1 + p_2)^\beta$.
間野の direct sampler for A-distribution: S.Mano, The A-hypergeometric System Associated with the Rational Normal Curve and Exchangeable Structures, Electronic Journal of Statistics 11 (2017), 4452–4487 [‡].

[†] 1 行目の成分は全部 1. rank $= d$.

[‡] `https://projecteuclid.org/euclid.ejs/1510887943`

direct sampler アルゴリズム (間野, 2018)

Input: β, p

Output: c

1. $c := (0, 0, \ldots, 0)$ (init count vector)
2. $e_i := \frac{p_i Z(\beta - a_i; p)}{\beta_1 Z(\beta; p)}$, $i = 1, \ldots, n$[§].
3. $[0, 1]$ を $e_1 : e_2 : \ldots : e_n$ に分割.
4. $[0, 1]$ に値を持つ一様乱数 t を一つ生成.
5. t が e_j の領域に入ったら c_j を 1 増やす. $\beta := \beta - a_j$.
6. $\beta \neq 0$ なら goto 2.

終了したとき c は $Ac = \beta$[¶] を満たす. なお a_i は行列 A の i 列.

[§] $\beta - a_i \notin \mathbf{N}_0^d$ なら $e_i = 0$. $|\beta| = \beta_1 + \ldots + \beta_d$

[¶] 計算途中の β でなく input の β

$\boxed{e_i \text{ の計算} \Leftarrow \text{漸化式を計算代数で}}$

1. 1980 年代後半から計算代数の問題に since D.Zeilberger. The book "$A = B$",
   `https://www.math.upenn.edu/~wilf/AeqB.html`
2. 微分差分作用素のグレブナー基底. $(I + (S-1)D_n) \cap D_{n-1}$. S は差分作用素.
3. Creative Telescoping. $(I + (S-1)R_n) \cap R_{n-1}$.

$\boxed{\text{間野 direct sampler の性質}}$ $Z_A(\beta; p)$ を定義から計算するのは β_1 や A が大きいと計算時間の点で困難が増す.

Theorem

1. A-超幾何系の *Pfaffian* はグレブナー基底で計算可能. これが漸化式も与えていて遷移確率 e_i が漸化式で計算できる[‖].
2. N 個の乱数を計算するための計算量は $O(r^2\beta_1 N)$ プラス グレブナー基底計算の計算量. ここで r は A の *normalized volume*. また有理数の四則の計算量は $O(1)$ とする.

補足:

1. MCMC[**] の計算量は $O(n'(N * T + (\text{burn-in の回数})))$ [††].
2. A 超幾何では Pfaffian = contiguity. Pfaffian の導出は計算量の多い計算であるが, 後藤, 松本の $E(k, n)$ の contiguity 関係 ⇒ 効率的な分割表の sampler [‡‡].
3. Direct sampler: 並列化可能, パラメータの調整不要 (経験不要).

[‖] 一般の実装は `tk_ds_ahg.rr`

[**] Diaconis-Sturmfels, 1998, その後は “グレブナー道場” 参照

[††] n' は I_A の生成元の最大次数. T は thinning の間隔

[‡‡] 実装は `gtt_ds.rr`

gtt_ds.rr の timing data

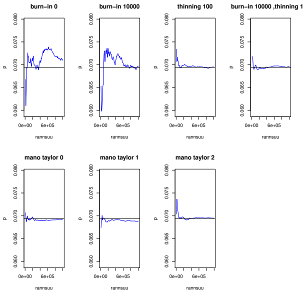
幾何/推測	A	B	C	合計
A	2	2	0	4
B	8	9	2	19
C	0	0	3	3
計	10	11	5	26

$$P = \begin{pmatrix} 1 & 9/10 & 11/10 \\ 1 & 13/10 & 99/100 \\ 1 & 1 & 1 \end{pmatrix}$$

$Au = \beta$: 行和, 列和が上の表で与えられたもので固定.
100 個の乱数の生成: 81.5s + 48.1s
3×3 で $r = 6$. $\beta_1 = 26$.
5×5 では $r = \binom{8}{4} = 70$.

bignum...

MCMC と direct sampler による p-値の計算 (平代達哉 M2 による)



図はグレブナー道場の, 幾何工学/推測工学の成績データに対する χ^2 検定統計量による p-値の計算である. p は

$$\begin{pmatrix} 1.01 & 1.05 & 0.95 & 1.06 & 0.93 \\ 0.97 & 1.05 & 0.97 & 1.07 & 0.97 \\ 0.94 & 1.03 & 0.98 & 0.99 & 1.07 \\ 0.97 & 1.01 & 0.93 & 0.99 & 1.01 \\ 1.01 & 1.01 & 0.99 & 1.03 & 1.06 \end{pmatrix}$$

当てはめ値はこの p に対する期待値を用いている. 平代君の direct sampler は間野の漸化式による方法ではなく $Z_A(\beta; p)$ の Taylor 展開を用いた近似計算によるものである. 2 次の Taylor 展開で良好な結果を得ている.

幾何/推測	5	4	3	2	1	合計
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

分割表, Timing データ

99 万のサンプル.

MCMC	CPU 時間
burn-in:0 thinnng:無し	362,809
burn-in:10000 thinnng:無し	378,440
burn-in:0 thinnng:100	17,063,450
burn-in:10000 thinnng:100	17,064,158
MANO	
Taylor 0 次	27,174,019
Taylor 一次	289,105,633
Taylor 二次	14,849,937,181

CPU 時間 * は 1,000,000 で 1 秒.

*`clock()` で計測. Xeon E5-4650 CPU, 2.7 GHz; 256 GB of memory.

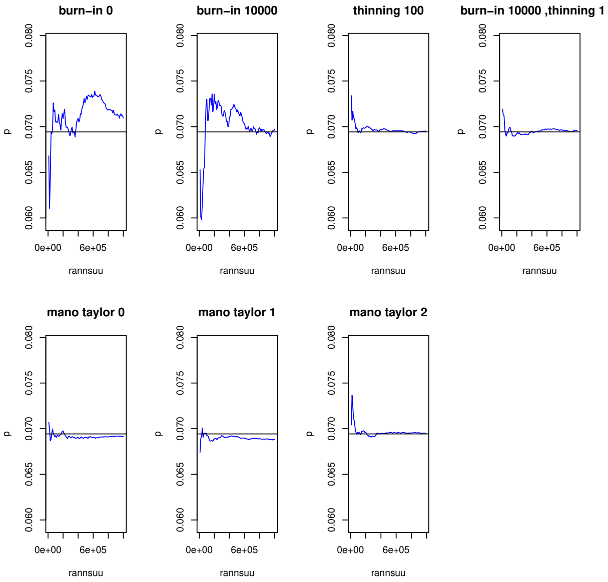


Figure: MCMC と direct sampler による p-値の計算 (平代達哉 M2 による)[拡大図] 右端は thining 100(が切れてる)

A 超幾何系 (とばす)

A : $d \times n$ 行列. 整数成分. A の列ベクトルは a_i. a_i は $\mathbf{Z}^d$ を生成.
$\beta = (\beta_1, \ldots, \beta_d) \in \mathbf{C}^d$ (parameters).

$$\mathbf{C}\langle x_1, \ldots, x_n, \partial_1, \ldots, \partial_n \rangle, \; x_j x_j = x_j x_i, \partial_i \partial_j = \partial_j \partial_i, \partial_i x_j = x_j \partial_i + \delta_{ij}$$

を D または D_n と書く.

Definition

A-hypergeometric system または GKZ hypergeometric system (GKZ, 1989), $H_A(\beta)$, $M_A(\beta) = D_n / H_A(\beta)$:

$$\begin{aligned}
(E_i - \beta_i) \bullet f &= 0, \; E_i - \beta_i = \sum_{j=1}^{n} a_{ij} x_j \partial_j - \beta_i, \; (i = 1, \ldots, d) \\
\Box_u \bullet f &= 0, \; \Box_u = \prod_{\{i \,|\, 1 \le i \le n, u_i > 0\}} \partial_i^{u_i} - \prod_{\{j \,|\, 1 \le j \le n, u_j < 0\}} \partial_j^{-u_j}
\end{aligned}$$

with $u \in \mathbf{Z}^n$ running over all u such that $Au = 0, u \neq 0$.

I_A は $\Box_u$ 達が $\mathbf{C}[\partial_1, \ldots, \partial_n]$ で生成するイデアル.

$\boxed{\text{例 (とばす)}}$

$$A(F_C,2)=\begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$\mathrm{degree}(I_A)=\mathrm{vol}(A)$.

Example

Macaulay2 commands to evaluate the volume (the degree) of $A(0134)$. Here, o5 is I_A.

```
loadPackage "FourTiTwo"
M=matrix "1,1,1,1; 0,1,3,4"
R=QQ[a..d]
I=toricGroebner(M,R)
   o5 = ideal (b^3  - a^2*c, b*c - a*d, - a*c^2  + b^2*d, c^3  - b*d^2
degree(I)
   o6 = 4
```

contigyity と 例 (とばす)

性質: f が $H_A(\beta)$ の解なら, $\partial_i f$ は $H_A(\beta - a_i)$ の解となる.
f および f の偏微分を basis vector F とした Pfaffian

$$\partial_i F = P_i F$$

を作ると, P_i は contiguity

$$P_i(\beta)F(\beta; x) = F(\beta - a_i; x)$$

を与える. $\Rightarrow$ $\boxed{\text{期待値の比 } e_i \text{ の計算が漸化式で可能}}$

例: `A=[[1,1,1],[0,1,2]]`. Pfaffian は

$$\partial_2 - \begin{pmatrix} \frac{\beta_2}{x_1} & -\frac{2x_3}{x_2} \\ \frac{2\beta_2(\beta_2-1)x_1}{4x_2(x_1x_3-x_2^2)} & \frac{-4(\beta_2-1)x_1x_3+(\beta_2-2\beta_1)x_2^2}{4x_2(x_1x_3-x_2^2)} \end{pmatrix}$$

```
load("tk_ds_ahg.rr")$ C=tk_ds_ahg.build_contiguity_0([[1,1,1],[0,1,2]])
```

例: Z の直接計算は時間がかかる 1

Contiguity relation/Recurrence relation

$$\partial_i \bullet Z_A(\beta; x) = Z_A(\beta - a_i; x)$$

(the contiguity relation)
Numerical evaluation of hypergeometric polynomial becomes hard problem when $\dim \mathrm{Ker}\, A$ and the rank of $H_A(\beta)$ increase and β becomes larger.

Example:

$$F_C(a,b,c;y) = \sum_{k \in \mathbf{N}_0^n} \frac{(a)_{|k|}(b)_{|k|}}{\prod k_i! \prod (c_i)_{k_i}} y^k, \quad A = \begin{pmatrix} \mathbf{1} & \mathbf{1} \\ E_{n+1} & -E_{n+1} \end{pmatrix}$$

where $(a)_m = a(a+1)\cdots(a+m-1)$ and $|k| = k_1 + \cdots + k_n$.
$n = 4,\ a = -179 - N,\ b = -139 - N,\ c = (37, 23, 13, 31)$,
$y = (31/64, 357/800, 51/320, 87/160)$

N	Evaluating series	method of Macaulay type matrix
0	6822s (1.89 hour)	61399s (about 17 hours)
100	138640s (1 day and about 14.5 h)	73126s(about 20.3 hours)
200	More than 2 days	84562s (about 23.5 hours)

例: Z の直接計算は時間がかかる 2

```
N=200
A=[[1,0,0,1,0,1,0,1,0,1],[0,1,0,1,0,1,0,1,0,1],[0,0,1,-1,0,0,0,0,0,0],[0,0,0,0,1,-1,0,0,0,0],[0,0,0,0,0,0,
Beta=[452,412,-37,-23,-13,31]
at ([x1,x2,x3,x4,x5,x6,x7,x8,x9,x10]=[140/411,40/137,25/822,31/411,14/411,17/274,17/822,5/137,10/137,29/82
oohg_native=0, oohg_curl=1
EV([x3])=[484018240471728953822203320553380653219481012643866487201043272204554116427335942534923953734369
863656998391689243859475296234352137555517730222159221047221525046528456147511166276222765024345097422807774
305750092193523229313167685161576286201466399466487213469381535663734384193880974741829514261324096233334 5
344275350822035203131054916726819435165178778325389866000027699548897905993488167196392728277735383730885 1
/194422284984251555304384242912588859511600655333063789436840056072076800834495255696040312940357668265848
206368590575510231394395404443601780545808586417609373178438189812637405870280353563181965119049387640350 1
941772514489533194749781746840208705674606008876031734288671532476200701856516011956451597268538379935874 1
320906272014298259515698562808086396098869061102204255115706387649155785914644280004302208683409377394435 4
9573932056327206030262721912023810463723569352286063413912998077871191506911]
Time=84562.4
```

N	Evaluating of series	method of Macaulay type matrix
0	6822s (1.89 hour)	61399s (about 17 hours)
100	138640s (1 day and about 14.5 h)	73126s(about 20.3 hours)
200	More than 2 days	84562s (about 23.5 hours)

Intel Xeon E5-4650 (2.7GHz) with 256G memory, the computer algebra system Risa/Asir (20140528).

ソフトウエア gtt_ds.rr, tk_ds_ahg.rr.

```
[1822] load("gtt_ds.rr");
[2720] gtt_ds.direct_sampler([[4,14,3],[10,6,5]],
                  [[1,9/10,11/10],[1,13/10,99/100],[1,1,1]]);
[ 0 1 3 ]
[ 8 5 1 ]
[ 2 0 1 ]
[2721] gtt_ds.direct_sampler([[4,14,3],[10,6,5]],[[1,9/10,11/10],[1,13/
[ 3 1 0 ]
[ 6 4 4 ]
[ 1 1 1 ]
[2722] gtt_ds.direct_sampler([[4,14,3],[10,6,5]],[[1,9/10,11/10],[1,13/
[ 2 1 1 ]
[ 6 4 4 ]
[ 2 1 0 ]
```

計算代数が有効に使えた例

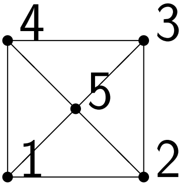


Figure: Graph for A

$$A^T = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{4}$$

図 2 の頂点 i, j が繋がってる時 i 列目と j 列目に 1 を書く.

計算代数 (creative telescoping) が有効に使えた例 続き

$Z_A(b;\mathbf{1})$ の b についての漸化式を求めたい

`HolonomicFunctions.m` (Christopher Kouchan),
`https://risc.jku.at/m/christoph-koutschan/` による答え.

$$
\begin{aligned}
& ((1+b_1)(1+2b_1)(1+b_1+b_2+b_3+b_4-b_5)(1+b_1-b_2+b_3-b_4+b_5))S_1 \\
+ \quad & (1+b_1+b_3)(1+2b_1+2b_3)(b_1-b_2+b_3-b_4-b_5), \\
& (1+b_2)(1+2b_2)(-1+b_1-b_2+b_3-b_4-b_5)(1+b_1+b_2+b_3+b_4-b_5)S_2 \\
+ \quad & (1+b_2+b_4)(1+2b_2+2b_4)(b_1-b_2+b_3-b_4+b_5), \\
& (1+b_3)(1+2b_3)(1+b_1+b_2+b_3+b_4-b_5)(1+b_1-b_2+b_3-b_4+b_5)S_3 \\
+ \quad & (1+b_1+b_3)(1+2b_1+2b_3)(b_1-b_2+b_3-b_4-b_5), \\
& (1+b_4)(1+2b_4)(-1+b_1-b_2+b_3-b_4-b_5)(1+b_1+b_2+b_3+b_4-b_5)S_4 \\
+ \quad & (1+b_2+b_4)(1+2b_2+2b_4)(b_1-b_2+b_3-b_4+b_5), \\
& (-1+b_1-b_2+b_3-b_4-b_5)(1+b_1-b_2+b_3-b_4+b_5)S_5 \\
- \quad & (-b_1-b_2-b_3-b_4+b_5)
\end{aligned}
$$

ここで $S_i f(b_i) = f(b_i+1)$ (b_i についての差分作用素).

計算代数が有効に使えた例 続き 2

前のページの出力を得るための Mathematica への入力.

```
ann4 = Annihilator[(1/Factorial[u1])*(1/Factorial[u2])*(1/
     Factorial[u3])*(1/
     Factorial[-b1 - b2 - b3 + u1 + u2 + b4 + b5])*(1/
     Factorial[2*b3 - u2 - u3])*(1/Factorial[2*b2 - u1 - u2])*(1/
     Factorial[b1 - b2 - b3 + u2 + u3 - b4 + b5])*(1/
     Factorial[b1 + b2 + b3 - u1 - u2 - u3 + b4 - b5])*1, {S[b1],
   S[b2], S[b3], S[b4], S[b5], S[u1], S[u2], S[u3]}]
FindCreativeTelescoping[ann4, {S[u1] - 1, S[u2] - 1,
  S[u3] - 1}, {S[b1], S[b2], S[b3], S[b4], S[b5]}]
```

なるべく小さい分母多項式を見つける Heuristics (by C.Kouchan) が実装されていて効率的.

まとめとアナロジー

$\boxed{\text{統計で興味のある } A} \Rightarrow \boxed{\text{Direct sampler を作りたい}} \Rightarrow \boxed{\text{漸化式を求める計算代数の問題}}$

計算代数で漸化式を求める入門書: The book "$A = B$",
`https://www.math.upenn.edu/~wilf/AeqB.html`

I_A のグレブナー基底	$\Rightarrow$	MCMC が作れる †
A-超幾何 の漸化式 ‡	$\Rightarrow$	間野の direct sampler が作れる §.

1. 素手で (理論的考察で) 漸化式を作れれば, random vector を生成する高速アルゴリズムが作れる.
2. 計算代数の手法で $Z_A(\beta; p)$ の β についての漸化式が作れれば高速な direct sampler が作れる.

† JST CREST 日比チーム編, グレブナー道場, 2011, 共立出版

‡ contiguity とも呼ぶ

§ S.Mano, Partitions, Hypergeometric Systems, and Dirichlet Processes in Statistics, JSS Research Series in Statistics (2018), Springer