An application of computer algebra to direct samplers

Nobuki Takayama(神戸大学)*, Shuhei Mano(統数研)

References hgm OpenXM search.

http://www.math.kobe-u.ac.jp/HOME/taka/2018/rims-2018.pdf

*Tatsuya Hiradai(神戸大, M2) の最近の結果も含む.

What is a sampler? Sampler = generate random vectors with a given distribution.

Example

Generate random vectors $u = (u_1, u_2)$ satisfying

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

with the distribution

$$\frac{\beta!}{u_1!u_2!} p_1^{u_1} p_2^{u_2} \tag{1}$$

where $p_i \geq 0$, $p_1 + p_2 = 1$.

$$P(U = u) = (1)$$

When $\beta = 2$, $p_i = 1/2$, then

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

```
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
```

stands for random vectors $(1,1), (2,0), (1,1), (2,0), (1,1), (1,1), (0,2), \ldots$

How to generate these random vectors by a direct sampler?

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. Divide $[0, 1]$ by $e_1 : e_2$. It is divided into E_1, E_2.
4. Generate a random number in $[0, 1]$ with the uniform distribution.
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 $u = (c_1, c_2)$.

Theorem (well-known)

The change of getting (u_1, u_2) *is* (1).

Proof. Let $i_1, i_2, \ldots, i_\beta$ be the sequence obtained in the step 5. i_j is 1 or 2. Note $\sharp\{k \mid i_k = 1\} = c_1$, $\sharp\{k \mid i_k = 2\} = c_2$. Then the chance of getting this index sequence is $p_1^{c_1} p_2^{c_2}$. When we have the count vector (c_1, c_2), the number of positions of $1, 2$ is $\binom{\beta}{c_1}$. □

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

What is the A distribution?

A: $d \times n$ matrix. Integer entries[†]. $p \in \mathbf{R}^n_{\geq 0}$. $u \in \mathbf{N}^n_0$, $\beta \in \mathbf{N}^d_0$. Put

$$Z_A(\beta; p) = \sum_{Au=\beta, u \in \mathbf{N}^n_0} \frac{p^u}{u!} \tag{2}$$

The A distribuion of $u \in \mathbf{N}^n_0$ is

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

$u! = u_1! \cdots u_n!$.
Example: When $A = (1, 1)$, it is the distribution of the previous slide. $\beta_1! Z_A(\beta; p) = (p_1 + p_2)^\beta$.

Mano's 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 [‡].

[†] The first row consists of 1's. rank $= d$.

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

Direct sampler algorithm(S.Mano, 2018)

Input: β, p

Output: c

1. $c := (0, 0, \ldots, 0)$ (init count vector)
2. $e_i := \frac{p_i Z(\beta - a_i; p)}{\beta_1 Z(\beta; p)}$, $i = 1, \ldots, n^{\S}$.
3. Divide $[0, 1]$ with $e_1 : e_2 : \ldots : e_n$.
4. Generate a random number t in $[0, 1]$ with the uniform distribution.
5. If t is in the segment of e_j, then increase c_j by 1. $\beta := \beta - a_j$.
6. If $\beta \neq 0$, then goto 2.

The output c satisfies $Ac = \beta^{\P}$. a_i is the i-th column of the matrix A.

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

¶ This β is not an intermedfiate beta, and is the input β

Evaluation of $e_i \Leftarrow$ Recurrence relation by computer algebra

1. Since D.Zeilberger from the late 1980's. The book "$A = B$", `https://www.math.upenn.edu/~wilf/AeqB.html`
2. Gröbner basis in the ring of differential difference operators $(I + (S-1)D_n) \cap D_{n-1}$. S is a difference operator.
3. Creative Telescoping. $(I + (S-1)R_n) \cap R_{n-1}$.

Properties of Mano's direct sampler $Z_A(\beta; p)$ 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. *Obtain a Pfaffian system of A-hypergeometric system (by Gröbner basis). This gives a recurrence relation for $Z_A(\beta; p)$ and the transition probability e_i can be evaluated by the recurrence relation.* [‖].
2. *The complexity of getting N random vectors is $O(r^2\beta_1 N)$ plus the complexity of computing the Gröbner basis. r is the normalized volume of A. Here, we assume the complexity of the arithmetics of rational numbers is $O(1)$.*

Note:

1. The complexity of MCMC[**] is $O(n'(N * T + \text{(the number of burn-in)}))$ [††].
2. Goto-Matsumoto gave recurrence relations of $E(k, n)$ by the twisted cohomology groups ⇒ efficient sampler. [‡‡].
3. Direct sample: parallelizable, need no tuning of parameters.

[‖] Implementation tk_ds_ahg.rr

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

gtt_ds.rr の timing data

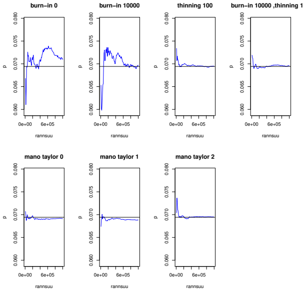
geom/stat	A	B	C	sum
A	2	2	0	4
B	8	9	2	19
C	0	0	3	3
sum	10	11	5	26

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

$Au = \beta$: the row sums and the column sums are fixed with the values above.
100 random vectors: 81.5s + 48.1s
$r = 6$, $\beta_1 = 26$.
When 5×5, $r = \binom{8}{4} = 70$.

bignum...

Evaluation of a p-value by MCMC and direct sampler(Tatsuya Hiradai (N



The figure is evaluations of p-values by the χ^2 test statistics. Here, p is

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

The direct sampler by Hiradai does not use the recurrence and use the approximation of $Z_A(\beta; p)$ by the Taylor expansion at $p = \mathbf{1}$. It works well by the Taylor expansion upto the degree 2.

Contingency table, timing data

geom/stat	5	4	3	2	1	sum
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
sum	10	6	5	2	3	26

990,000 samples.

MCMC		CPU time
	burn-in:0, no thinnng	362,809
	burn-in:10000, no thinnng	378,440
	burn-in:0 thinnng:100	17,063,450
	burn-in:10000 thinnng:100	17,064,158
MANO		
	Taylor 0th	27,174,019
	Taylor 1th	289,105,633
	Taylor 2th	14,849,937,181

CPU time* 1,000,000=1 second.

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

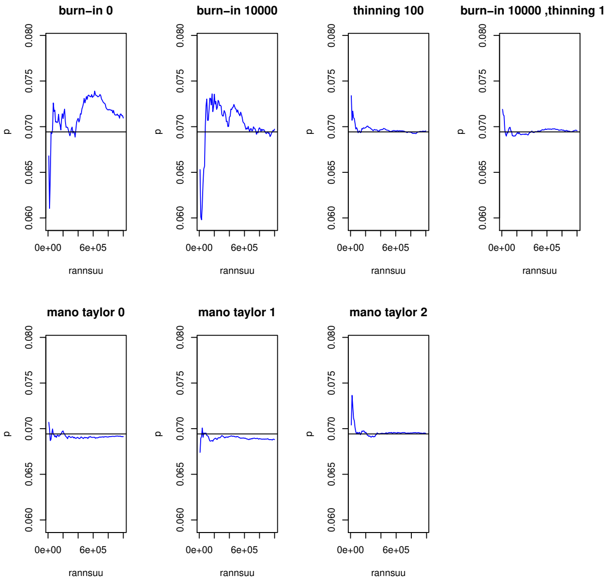


Figure: Evaluation of a p-value by MCMC and direct sampler(Tatuya Hiradai (M2))[拡大図] 右端は thining 100(が切れてる)

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), $H_A(\beta)$, $M_A(\beta) = D_n / H_A(\beta)$:

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

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.

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. $\Rightarrow$ **期待値の比 e_i 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]])
```

Example: Naive evaluation of Z is time consuming 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)

Example: Naive evaluation of Z is time consuming 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
8636569983916892438594752962343521375555177302221592210472215250465284561475111662762276502434509742280774
3057500921935232293131676851615762862014663994664872134693815356637343841938809747418295142613240962333345
3442753508220352031310549167268194351651787783253898660000276995488979059934881671963927282777353837308851
/1944222849842515553043842429125888595116006553330637894368400560720768008344952556960403129403576682658484
2063685905755102313943954044436017805458085864176093731784381898126374058702803535631819651190493876403501
9417725144895331947497817468402087056746060088760317342886715324762007018565160119564515972685383799358741
3209062720142982595156985628080863960988690611022042551157063876491557859146442800043022086834093773944354
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).

Software 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 ]
```

Creative telescoping is useful.

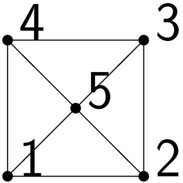


Figure: Graph for A

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

When the vertex i and the vertex j is connected, set 1 on the i-th column and the j-th column (Figure 2).

Creative telescoping is useful.

Want recurrences of $Z_A(b;\mathbf{1})$ w.r.t b

Answer by `HolonomicFunctions.m` (Christopher Kouchan), `https://risc.jku.at/m/christoph-koutschan/`

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

Here, $S_i f(b_i) = f(b_i+1)$ (difference operator w.r.t. b_i).

Creative telescoping is useful 2

Input to Mathematica

```
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]}]
```

Heuristics (by C.Kouchan) to find a smaller denominator polynomial is a point.

Summary and analogy

Interesting A ⇒ Want a direct sampler ⇒

Recurrence relations by computer algebra

Introductory book on computer algebra and recurrence relations: The book "$A = B$", https://www.math.upenn.edu/~wilf/AeqB.html

Gröbner basis of I_A	⇒	MCMC [†]
Recurrence of A-hypergeometric fn[‡]	⇒	Mano's direct sampler [§].

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.

[†] see, e.g., the book Gröbner Bases: statistics and software systems

[‡] = contiguity

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