2 元分割表の条件付き確率の差分 HGM による計算 (の実装研究)

後藤良彰, 橘義仁, 高山信毅 (神戸大学)

$r_1 \times r_2$ 分割表とは, 非負整数を成分とする $r_1 \times r_2$ 行列 u.
$p_{ij} \in \mathbf{R}_{\geq 0}^{r_1 \times r_2}$ に対して多項分布 $\frac{|u|}{u!} p^u$ を考える. 行和, 列和
$\beta \in \mathbf{N}_0^{r_1+r_2-1}$ を固定した条件付き分布を考える. その分布は

$$\frac{p^u}{u!} \frac{1}{Z(\beta; p)}, \quad Z(\beta; p) = \sum_{u \text{ の行和列和は } \beta} \frac{p^u}{u!}$$

目標:

$$\frac{p^u}{u!}\frac{1}{Z(\beta;p)}, \quad Z(\beta;p) = \sum_{u\text{ の行和列和は }\beta} \frac{p^u}{u!}$$

の厳密計算. 応用: 条件付き最尤推定等 (T-栗木-竹村: arxiv: 1510.02269). 応用問題募集.

$\beta = (37,36,12)^T$. $\begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$. 行和 (11,37), 列和 (36, 12)

$$\hat{u} = u = \begin{pmatrix} 11 & 0 \\ 25 & 12 \end{pmatrix}, \ldots, u = \begin{pmatrix} 4 & 7 \\ 32 & 5 \end{pmatrix}, \ldots, u = \begin{pmatrix} 0 & 11 \\ 36 & 1 \end{pmatrix}.$$

$Z(\beta;p) = \frac{p_1^{11}p_3^{25}p_4^{12}}{11!25!12!}{}_2F_1\left(-11,-12,26;y\right), y = \frac{p_2p_3}{p_1p_4}$(odds 比),

$\mathrm{Ker}\, A = \mathbb{Z}\bar{a}, \bar{a} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$, $u = \hat{u} + v\bar{a}$ の確率は $\frac{y^v/(\hat{u}+v\bar{a})!}{{}_2F_1(y)/\hat{u}!}$.

${}_2F_1$ は Gauss の超幾何級数.

多項式の計算といえども定義どおりの計算では時間がかかる.

$O(|\beta|^{r_1r_2-r_1-r_2+1})$. 例: $r_1 = r_2 = 5$, $O(|\beta|^{16})$

正規化定数 (分配関数) Z とその偏微分を計算するための (差分) ホロノミック勾配法 (HGM) (一般論は中山, 西山, 野呂, 小原, 清, 高山, 竹村 2011 等).

1. (一般論) $\sum$ (差分)holonomic 関数 と (その微分達) は漸化式 (差分方程式) を満たす.
2. (一般論) $\sum$ (差分)holonomic 関数 は漸化式およびその rank を導出するアルゴリズムがある.
3. (二元分割表) p_{ij} が有理数なら厳密値が計算できる.
4. (二元分割表) 漸化式およびその rank を twisted cohomology group や D-加群を用いて理論的に導出, 決定できる.
    1. $r_1 \times r_2$ 二元分割表の場合は, Aomoto-Gel'fand 超幾何系に帰着することを示して rank は $r = \binom{r_1+r_2-2}{r_1-1}$ (青本, 1977).
    2. (使いやすい形の漸化式は) Y.Goto, K.Matsumoto, Pfaffian equations and contiguity relations of the hypergeometric function of type $(k+1, k+n+2)$ and their applications, arxiv:1602.01637.

漸化式 (contiguity relation) の例: $f_n = {}_2F_1(n, b, c; x)$ とおく.

$$F_n = \begin{pmatrix} f_n \\ \theta_x f_n \end{pmatrix}, \quad M_n = \frac{1}{n-c+1} \begin{pmatrix} bx+n-c+1 & x-1 \\ -nbx & n(1-x) \end{pmatrix}$$

は $F_{n+1} = M_n^{-1} F_n$ を満たす. θ_x は Euler 作用素 $x\partial_x$ である.

命題

$n \times n$ 行列の掛け算に要する計算量を $O(n^3)$ とするとき後藤, 松本 (2016) の方法で contiguity relation を得るための計算量の上限は以下である.

1. r_1 を固定し, $r_2 \to \infty$ を考えるとき $O(r_2^{3r_1})$.
2. r_2 を固定し, $r_1 \to \infty$ を考えるとき $O(r_1^{3r_2})$.
3. $r_1 = r_2$ の条件下で, $r_1 \to \infty$ を考えるとき $O(2^{6r_1})$.

r_i を一つ固定した状況では多項式計算時間で計算できる.
漸化式の導出をすれば, あとは単にベクトルを行列で変換することを繰り返すのみ. しかし, 有理数の計算に時間がかかる!

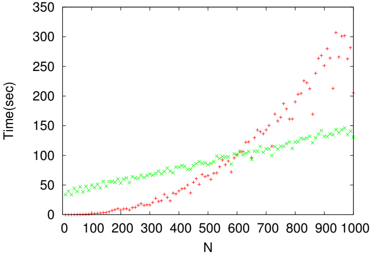
```
[3035] test_p_value2();
U=[[2,1,1,0,0],[8,3,3,0,0],[0,2,1,1,1],
   [0,0,0,1,1],[0,0,0,0,1]]
P=[[10013/5000,13/2000,49/10000,7/1000,73/10000],
   [1/250,15029/5000,1/1000,11/2000,31/5000],
   [1/5000,11/1250,1,11/5000,9/10000],
   [17/10000,1/125,7/1000,10007/10000,37/5000],
   [21/10000,93/10000,23/5000,49/5000,503/500]]
Approx prob=1.3284e-25
Exact value=
855846712181344038598886695213030440763392000000000000000/
6442688285694587112804522551090808114541208382354533987
51313623557989754139368967
```

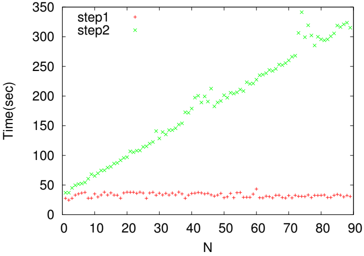
計算代数分野で標準的な手法である, 中国式剰余定理および分散計算を用いる計算法 (J.Faugerre (1999), 野呂, 横山 (1999), 佐々木, 竹島, (1989) など) を適用してアルゴリズムを作る. このアルゴリズムについて以下の計算量評価が得られる.

命題

n 桁の数の四則演算が $O(n^2)$ で計算できると仮定する. C をプロセス数, N_p を中国式剰余定理に用いる d_p 桁の素数の数と仮定すると $F(\beta; p)$ は $\max\left(O(\frac{|\beta| r^2 N_p d_p^2}{C}), O(\frac{r N_p^4 d_p^4}{C})\right)$ を計算量の上限として計算できる.

上記の計算量を実現するアルゴリズムを実装: asir-contrib の `gtt_ekn.rr` (約 2900 行, 約 59000 バイト).[1] 左の図: 漸化式による計算を素朴に行なった場合と命題 2 の方法を用いた場合の比較 (2×2). 右は 5×5 分割表での漸化式の生成の時間 (赤) と Z の計算時間 (きみどり).





$$\beta = N(4, 4, 4, 4, 4; 2, 3, 5, 5, 5).$$

まとめ

1. 定義どおりの計算方法では一次変換の回数 n に対して計算量の上限は $O(n^2)$ となるが, モジュラーメソッドを利用し, d_p, N_p を固定するという条件の下で正確な値が計算できる範囲の n において, $O(n)$ となる.
2. 実装研究として Package を作成し性能を評価した. そのパッケージでの計算実験でモジュラーメソッドが差分ホロノミック勾配法に有効であることが分かった.

---

[1] asir_contrib_update(|update=1);