2016年6月17日　統計数理研究所　オープンハウス

有理正規曲線に付随するA超幾何系と統計的推測

間野 修平　　数理・推論研究系 准教授

0　本ポスターの内容

計算代数の統計学への導入が進んでいます．重要なholonomic函数にGel'fand-Kapranov-Zelevinsky (1990)により導入されたA超幾何函数があります．本ポスターでは，有理正規曲線と呼ばれる代数曲線に付随するA超幾何系が交換可能な組み合わせ構造の離散確率分布に関わることを指摘し，計算代数の方法と統計的推測への応用について紹介しています．

1　偏Bell多項式と統計的推測

サイズnの標本のクラスタリングを考えます．標本点に番号を付し，クラスタによる分割の上の確率法則を考えます．交換可能であればクラスタのサイズのみに注目すれば良く，nを和とする正の整数の集合の上の確率法則を考えれば十分です．サイズiのクラスタにw_iの重みを割り当てると，クラスタ数kの条件の下で，クラスタリングの総数は

$$B_{n,k}(w) = n! \sum_{s \in \mathcal{S}_{n,k}} \frac{x^s}{s!}, \qquad x^s := \prod_{i=1}^n x_i^{s_i},\ s! := \prod_{i=1}^n s_i!,\ x_i := \frac{w_i}{i!}$$

と与えられます．s_iはサイズがiのクラスタの数で，台は

$$\mathcal{S}_{n,k} := \left\{ (s_1, ..., s_n) : \sum_{i=1}^n i s_i = n, \sum_{i=1}^n s_i = k \right\}.$$

です．多項式$B_{n,k}(w)$は偏Bell多項式，その$n!^{-1}$倍で正規化した条件付き分布は統計力学の文脈でmicrocanonical Gibbs分布と呼ばれます．自然な無条件モデルとして$q_n(s) = v_{n,k} n! x^s / s!$を考えます．

例 1 (Poisson回帰) m水準単変量の無条件モデル$S_i \sim Po(\mu_i)$, $\log \mu_i = \alpha + \beta i$を考えます．標本サイズ$k = \sum_{i=1}^m s_i$と水準の総和$n = \sum_{i=1}^m i s_i$が完備十分統計量で，条件付き分布は$B_{n,k}^{(m)-1}(1.) x^s / s!$です．$B_{n,k}^{(m)}(1.)$はLah数に関連します．$\alpha, \beta$が局外母数ならば適合を正確に相似検定できます．

例 2 (経験Bayes) Shakespeareは31,534の異なる単語を使い，14,376の単語は1度，4,343の単語を2度使いました．この様なデータからShakespeareの語彙の数mを推定できます(Efron & Thisted 1976)．語彙の使用頻度が母数$(-\alpha) > 0$のm変量対称Dirichlet分布に従うとすれば，多項抽出の周辺尤度はDirichlet多項分布

$$q_n(s) = \frac{[m]_k}{[m\alpha]_n} n! \frac{x^s}{s!}, \qquad x_i = \binom{\alpha}{i},$$

です．観察された語彙の数kはmの完備十分統計量で，αが既知のとき一様最小分散不偏推定量は$\hat{m}(k) = k + B_{n,k-1}((-\alpha).)/B_{n,k}((-\alpha).)$, $(-1)^n B_{n,k-1}((-\alpha).)$は一般化階乗係数と呼ばれます．Mosteller & Wallece (1964)はHamilton, Jefferson, Maddisonが演説の中で特定の単語を使う回数に注目しました．この様なデータでは演説の総数mは既知でαの推定に興味があります．mが局外母数ならば条件付き最尤推定できます．

2　A超幾何函数と計算代数

多項式係数の微分作用素環$R_m = \mathbb{C}(x_1, ..., x_m)\langle \partial_1, \partial_2, ..., \partial_m \rangle$を考えます．$A$超幾何多項式を$Z_A(b; x) := \sum_{\{s: As = b, s \in \mathbb{N}^m\}} x^s / s!$で定義します．$A$は整数成分の$d \times m$行列，$b \in \mathbb{C}^d$です．$A$超幾何函数を消去作用素

$$\sum_{j=1}^m a_{ij} \theta_j - b_i, \qquad \partial^{s^+} - \partial^{s^-}, \qquad i \in [d],\ s^+ - s^- \in \ker A \cap \mathbb{Z}^n, \theta_j := x_j \partial_j$$

が成す左イデアル$H_A(b)$により定義します．特に，$d = 2$の同次行列

$$A = \begin{pmatrix} 0 & 1 & 2 & \cdots & m-1 \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

は有理正規曲線を定めます．$m = 4$はtwisted cubicとして知られ，$x_1 x_3 - x_2^2, x_2 x_4 - x_3^2, x_1 x_4 - x_2 x_3$の零点です．$d = 2$の$A$超幾何系の性質は良く知られていて，有理正規曲線では$\mathrm{rank}(H_A(b)) = \mathrm{vol}(A) = m - 1$です．

命題 1 $H_A(b)$のinitial idealの標準単項式は$\{1, \theta_i : i = 3, ..., m\}$である．

補題 1 $b = (n-k, k)^\top$, $m = n - k + 1$のとき，有理正規曲線に付随するA超幾何系の多項式解は一意で，A超幾何多項式(偏Bell多項式の$n!^{-1}$倍)の定数倍である．

例を挙げた様に，統計的推測への応用では偏Bell多項式の数値計算が必要になることがあります．級数の和は避けるべきで，漸化式を用いれば良く，漸近形も議論できますが，ここではホロノミック勾配法(Nakayama et al. 2011)を紹介します．標準単項式についてPfaffian系

$$\theta_i \bullet Q = P^{(i)} Q, \qquad Q(x) = (1, \theta_3, ..., \theta_{n-k+1})^\top \bullet Z_{n,k}(x),$$

が成立します．Pfaffian $P^{(\cdot)}$の計算は一般にGröbner基底の計算を伴い時間がかかりますが，漸化式に注目すると陽な形が得られます．Pfaffian系は1階微分方程式系ですので，閉じた簡単な表示を持つ偏Bell多項式を初期値とし，数値積分により望みのxでの偏Bell多項式が得られます．

3　統計的推測と情報幾何

特に条件付き最尤法について紹介します．尤度は不定元の対数を自然母数とする指数型分布族です．A超幾何多項式が定める格子点はすべて凸包(Newton polytope)の内点ではないことから次の定理が得られます．

定理 1 有理正規曲線に付随するA超幾何分布において，最尤推定量は確率1で存在しない．

この主張は標本をサイズ$N = 1$の多変量計数データとみなせば不思議ではなく，サイズ$N \geq 2$の標本，曲指数型分布族を扱うことは必然です．Newton polytopeを統計多様体とする情報幾何が有効です．部分多様体への直交射影の存在が1次漸近有効な最尤推定量の存在の必要十分条件です．推定多様体はm測地線で，双対座標系ηの直線です．最尤法の自然勾配法による実装では，自然勾配$\partial_j \eta_i$はFisher計量で，PfaffianによりA超幾何多項式を経由することなく計算できます．

$$g_{ij} = \sum_{l=1}^{n-k-i} (\tilde{P}^{(i)})^{-1}_{j-2,l} \eta_{l+2} I_{\{n-k+2 \geq i+j\}} - \eta_i \eta_j + \eta_i \delta_{ij}, \qquad 3 \leq i, j \leq n-k+1.$$

例 3 (Nonparametric Bayes) 事前分布として知られる2母数Dirichlet過程はDirichlet分布の無限次元版です．無条件モデルは$v_{n,k} = v_k / \sum_k v_k B_{n,k}(w)$, $v_k = (\theta)(\theta+\alpha)\cdots(\theta+(k-1)\alpha)$, $w_i = (1-\alpha)_{i-1}$です．母数空間は$\{\alpha < 0, \theta = -m\alpha\}$,$\{0 \leq \alpha < 1, \theta > -\alpha\}$で，Dirichlet多項分布を含みます．クラスタ数kは母数θの完備十分統計量です．

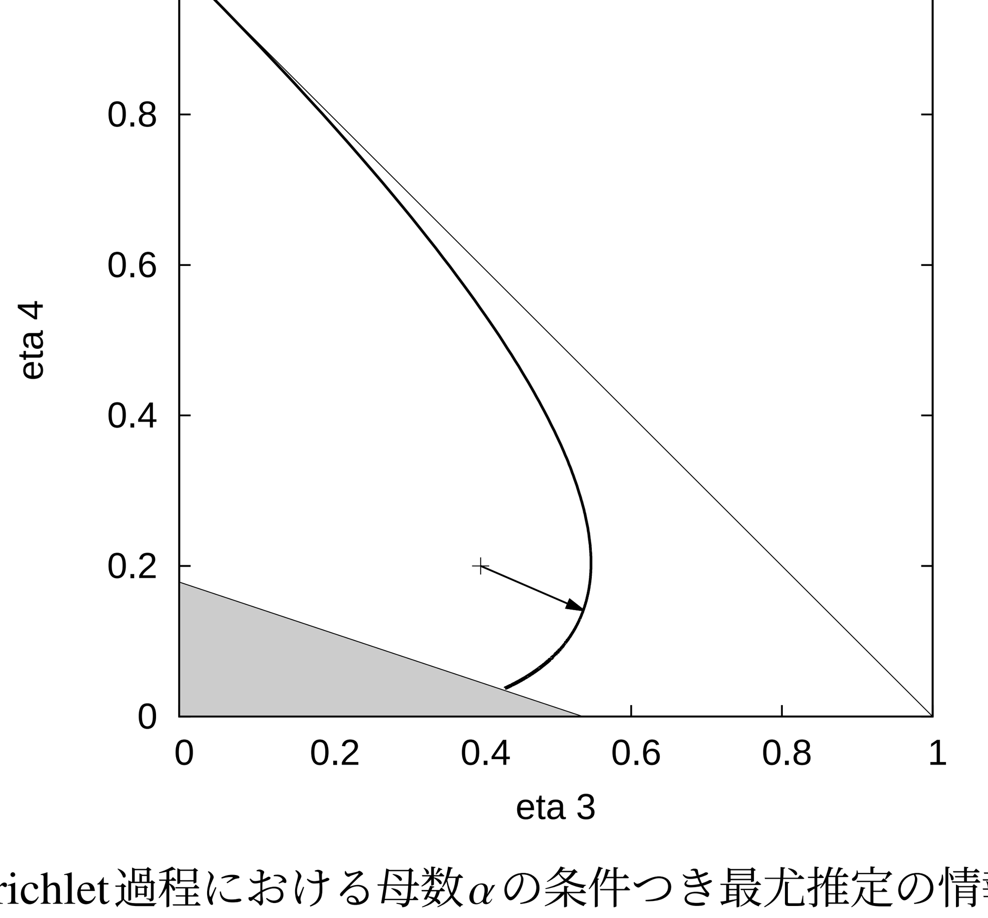

図1 2母数Dirichlet過程における母数αの条件つき最尤推定の情報幾何．標本サイズ10，クラスタ数7．対角線下側の3角形がNewton polytopeのη_3-η_4平面への射影で，曲線が部分多様体．その境界($\alpha = 1, -\infty$)は特異．最尤推定(直交射影)の推定多様体の例を矢印で示す．影は$\alpha = -\infty$の法扇で，最尤推定量が存在しない領域．