Gibbs 確率分割における統計的推測への代数的接近

統計数理研究所　　間野 修平

$n(\in \mathbb{N})$ 個の要素を k 個のクラスタに割り当てる組み合わせ構造を考える．非負の整数列 $\{v_{n,k}\}$, $\{w_i\}$, $i,k \in [1,2,...,n]$ を，それぞれ，大きさ i のクラスタがとる組み合わせ構造の数，n 個の要素を k 個のクラスタに割り当てる方法の数とすると，可能な組み合わせ構造の総数は $B_n(v_{n\cdot}, w.) := \sum_{k=1}^n v_{n,k} B_{n,k}(w.)$ で与えられる．ただし，

$$B_{n,k}(w.) = n! \sum_{s.:\sum s_i = k, \sum i s_i = n} \prod_{i=1}^n \left(\frac{w_i}{i!}\right)^{s_i} \frac{1}{s_i!}$$

は偏 Bell 多項式，s_i は大きさ i のクラスタの数である．非負の数列 v, w について，自然数 n の自然数による分割 $(n_1, ..., n_k)$, $\sum_{i=1}^k n_i = n$ が

$$p_n(n_1, ..., n_k) = v_{n,k} \prod_{i=1}^k w_{n_i}$$

のような確率法則に従う確率分割 Π_n を，v, w を母数とする Gibbs 分割という [1]．2013 年の連合大会では，Faà di Bruno の公式を通じてクラスタの大きさの極値の漸近的性質を母函数の解析に帰着できることを指摘し，Gibbs 分割の重要な例である Pitman 分割に関する結果を報告した [2]．本講演では，Gibbs 分割における統計的推測について報告する．

母数 w に興味があるとし，v を局外母数とすることは自然である．例えば，Pitman 分割では $w_i = (1-\alpha)_{i-1}$, $-\infty < \alpha < 1$ であるが，$\alpha = 0$ は Dirichlet 過程からの標本であり，Ewens 分割として知られる特別な位置を占めていて，α の符号により性質が本質的に異なる．クラスタの総数は局外母数 v の十分統計量であり，条件付分布は

$$\mathbb{P}(S_i = s_i; i = 1, ..., n \| \Pi_n| = k) = \frac{n!}{B_{n,k}(w.)} \prod_{i=1}^n \left(\frac{w_i}{i!}\right)^{s_i} \frac{1}{s_i!}$$

で与えられる．そこで，w に関する相似検定，条件付き最尤法を考える．条件付き最尤法については

$$\hat{w} = \mathrm{argmax}_w \left(\sum_{i=1}^n s_i \log w_i - \log B_{n,k}(w.) \right).$$

に帰着するが，組み合わせに因る大きな数を扱うので，厳密な計算は簡単ではない．ところで，A-超幾何函数は

$$Z_A(b;p) = n! \sum_{s.:As=b} \prod_{i=1}^n \frac{p^{s_i}}{s_i!}.$$

で定義されるが，偏 Bell 多項式は A-超幾何函数 $Z_A(b; w./\cdot!)$ である．ここで，

$$A = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} n \\ k \end{pmatrix}$$

である．配置行列 A の toric ideal に着目することで，偏 Bell 多項式を微分方程式系の解として与えることができ，数値計算が可能になる．

参考文献

[1] Pitman, J. (2006) “Combinatorial Stochastic Processes”, Springer.

[2] Mano, S. Extreme sizes in Gibbs-type exchangeable random partitions. Ann. Inst. Stat. Math. to appear. arXiv: 1306.2056.