有限母集団分割の推測

慶應義塾大学　理工学部　渋谷　政昭

目的　数の確率分割データの解析では, 伝統的に無限母集団分割モデルから出発している. ここでは有限母集団分割からのランダム・サンプルである標本分割から母集団を推測する.

数の分割からのランダム・サンプル　数 ν の κ 個の正整数への分割 $\tau = (\tau_1, \tau_2, \ldots,)$ からの, 大きさ n の標本を考える. $S_{k,\ell}$ により τ_k から得られた大きさ ℓ の部品 (クラスター) の数を表し ($1 \le k \le \kappa$; $0 \le \ell \le \min(n, k\tau_k)$), さらに $S_{+,\ell} := \sum_{k=1}^{\kappa} S_{k,\ell}$ とする. $S = (S_{+,1}, \ldots, S_{+,n})$ の観測値 s (標本分割) から τ を推定する:

$$\hat{\tau}(s) := \mathrm{argmax}_{\tau \in \mathcal{P}_\nu} \mathbb{P}\{S = s; \tau\}$$

$$\mathbb{P}\{(S_{k,\ell}) = (s_{k,\ell}); \tau\} = \sum_{k=1}^{\kappa} \tau_k! \prod_{\ell=0}^{\min(k,n)} \frac{1}{s_{k,\ell}!} \binom{k}{\ell}^{s}_{k,\ell} \Big/ \binom{\nu}{n}, \tau \in \mathcal{P}_{\nu,\kappa}.$$

推定精度は $\hat{\tau}(S)|\tau_o$ の中心 (モード, 期待値) からの距離で測る. 問題は計算法である.

表 1: Sample random partitions from $\mathcal{P}_8$ to $\mathcal{P}_5$, a part, probability($\times 56$)

$\mathcal{P}_5 \backslash \mathcal{P}_8$	62	61^2	53	521	51^3	4^2	431	42^2	421^2	41^4	3^22	3^21^2	32^21	321^3	31^5	2^4	2^31^2	2^21^4
5	6	6	1	1	1	0	0	0	4	0	0	0	0	0	0	0	0	0
41	30	30	15	15	15	8	4	4	4	4	0	0	0	0	0	0	0	0
32	20	0	40	10	0	48	18	8	20	24	8	2	6	1	0	0	0	0
31^2	0	20	0	20	30	0	16	16	12	0	12	8	14	9	10	0	0	0
2^21	0	0	0	10	0	0	18	28	16	24	36	22	18	9	0	24	12	4
21^3	0	0	0	0	10	0	0	0	0	4	0	24	18	31	30	32	36	32
1^5	0	0	0	0	0	0	0	0	0	0	0	0	0	6	16	0	8	20

表 2: Estimates $\hat{\tau}(s) : s \in \mathcal{P}_{12,5}, \hat{\tau} \in \mathcal{P}_{16}, \mathcal{P}_{20}$

$s \in \mathcal{P}_{12,5}$	81^4	721^3	631^3	62^21^2	541^3	5321^2	52^31	4^221^2	43^21^1	432^21	42^4	3^321	3^22^3
$\hat{\tau} \in \mathcal{P}_{16}$	$11\ 1^5$	92^21^3	841^4	82^4	6^21^4	7431^2	82^4	652^21	5^241^2	4^32^2	63^22^2	43^4	43^4
$\hat{\tau} \in \mathcal{P}_{20}$	$14\ 1^6$	$12\ 2^31^2$	$10\ 51^5$	$10\ 2^5$	871^5	8532^2	83^4	7^22^3	6^251^3	5^24^22	4^5	4^5	4^5

パラメトリック・モデル　ノンパラメトリック手法で漸近論を考えるためには分割 τ の漸近的性質を仮定しなければならないが, その可能範囲が広いために一般論の展開が難しい. 手始めにピットマン確率分割 (Ewense Piman Sampling Formula, EPSF (θ, α)) を取り上げる. もし τ が EPSF (θ, α)) に従うならば, その標本確率分割も, 同じパラメータの EPSF に従う (EPSF の分割構造). つまり EPSF が任意の標本分割の共役事前分布に相当する.

EPSF が確率分割の中では最も扱いやすいものの. その性質は数値的に議論することになる. まずノンパラメトリック手法との比較が課題である.

Sibuya, M. (2014) Prediction in Ewens-Pitman Sampling Formula and random samples from number-partitions, *Ann. Inst. Statist. Math.*,**66**, 833–864.