Bell多項式とGibbs分割

間野 修平　　**数理・推論研究系 准教授**

0　本ポスターの内容

確率分割の研究においては，様々な良い性質をもつがゆえにPitman分割が主な対象で，一般化としてのGibbs分割が提案されたのは最近です．従来のPitman分割へのアプローチはその性質に依存していて，一般のGibbs分割には適用できないものが多いです．本ポスターは，順序統計量について，任意のGibbs分割に適用できるアプローチを紹介しています．

1　Bell多項式

自然数の集合$[n] := \{1, ..., n\}$をkの部分集合への分割を$\mathcal{P}^k_{[n]}$とし，$\mathcal{P}_{[n]} := \cup^n_{k=1}\mathcal{P}^k_{[n]}$とします．2つの非負の数列$v. := (v_1, v_2, ...)$と$w. := (w_1, w_2, ...)$により，分割の数に$v.$，部分集合の大きさに$w.$で定まる数を割り当てると，そのような構造物の数は

$$B_n(v., w.) := \sum_{k=1}^{n} v_k B_{n,k}(w.), \qquad B_{n,k}(w.) := \sum_{\{A_1,...,A_k\}\in\mathcal{P}^k_{[n]}} \prod_{i=1}^{k} w_{|A_i|}.$$

で与えられます．多重度$m_i = \#\{j : |A_j| = i\}$を用いると，

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

と表せます．これは$w.$の多項式ですが，Bell多項式(1927)とよばれます．

	k				
n	1	2	3	4	5
1	w_1				
2	w_2	w_1^2			
3	w_3	$3w_1w_2$	w_1^3		
4	w_4	$4w_1w_3 + 3w_2^2$	$6w_1^2w_2$	w_1^4	
5	w_5	$5w_1w_4 + 10w_2w_3$	$10w_1^2w_3 + 15w_1w_2^2$	$10w_1^3w_2$	w_1^5

表1．Bell多項式

2　Gibbs分割

分割の形で与えられるデータを統計解析に供するためには分割の確率モデルが必要ですが，そのようなモデルを確率分割とよびます．交換可能性は妥当な仮定ですから，

$$\mathbb{P}(\Pi_n = \{A_1, ..., A_k\}) = p_n(|A_1|, ..., |A_k|)$$

を満たす対称な函数を導入し，exchangeable partition probability function (EPPF)とよびます．Gibbs分割は広いクラスの確率分割で，EPPFは

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

もしくは，多重度で表すと

$$p_n(m_1, ..., m_n) = n! v_{n,k} \prod_{i=1}^{n} \left(\frac{w_i}{i!}\right)^{m_i} \frac{1}{m_i!}$$

です．次の自然な要請をすることがあります．

(1) 1つの標本点を除いた標本が同じEPPFに従う(Gnedin & Pitman 2005)

$$w_i = (1-\alpha)(2-\alpha)\cdots(i-1-\alpha), \qquad i \geq 2, \qquad w_1 = 1.$$

(2) $v_{n,k}$のn, kへの依存が分離(Pitman 2006)，すなわち

$$v_{n,k} = \frac{v_k}{B_n(v., w.)}.$$

例1 要請(1)と(2)を課すと良く知られているPitman分割(1995，Ewens 1972)になります．$\mathbb{R}$の確率測度μと任意の有限分割$\{B_1, ..., B_k\}$に対して

$$(D(B_1), ..., D(B_k)) \sim Dirichlet(\theta\mu(B_1), ..., \theta\mu(B_k))$$

となるランダム測度Dを母数θのDirichlet過程(Ferguson 1973)とよび，ノンパラメトリック・ベイズにおける共役事前分布として使われます．Dirichlet過程からの標本は$\alpha = 0, v_k = \theta^k$のPitman分割に従い，その構成は中華料理店過程として知られています．また，順序統計量$n_{(\cdot)}n^{-1}$の極限分布はPoisson-Dirichlet分布として知られています．

3　関連づけられたBell多項式

組み合わせ論において良く知られているFaà di Brunoの公式を用いることで，(1)の和を$n_{(i)} \leq r$のみについてとるBell多項式$B^{(r),(i)}_{n,k}(v., w.)$の指数母函数

$$\sum_{n=k}^{\infty} B^{(r)(i)}_{n,k}(w.)\frac{u^n}{n!} = \sum_{j=0}^{i-1} \frac{(\breve{w}_{(r+1)}(u))^j}{j!}\frac{(\breve{w}^{(r)}(u))^{k-j}}{(k-j)!}$$

を得ることができます．ここで，$\breve{w}_{(r+1)}(u), \breve{w}^{(r)}(u)$ はそれぞれ$w.$の指数母函数において$w_1 = \cdots = w_r = 0, w_{r+1} = w_{r+2} = \cdots = 0$としたものです．このようにして得られるBell多項式は，組み合わせ論においてassociated Stirling number of the first kind のように“associated ...”とよばれる拡張に対応するので，関連づけられたBell多項式とよんでおきます．

4　順序統計量

分割の数K_nの分布は$v_{n,k}B_{n,k}(w.)$で与えられますので，条件つき分布は

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

で与えられ，K_nは$v.$の十分統計量です．これは統計力学における$w.$を微視的状態の数とするミクロカノニカルGibbs分布で，順序統計量は条件をつけた状態の数え上げから得られます．例えば，順序統計量の周辺分布は次のように表せることが直ちに分かります．

$$\mathbb{P}(|\Pi_n|_{(1)} \leq r) = \sum_{k=\lceil n/r \rceil}^{n} v_{n,k} B^{(r)(1)}_{n,k}(w.),$$

$$\mathbb{P}(|\Pi_n|_{(i)} \leq r) = \sum_{k=1}^{i-1} v_{n,k}B_{n,k}(w.) + \sum_{k=i}^{n} v_{n,k}B^{(r)(i)}_{n,k}(w.), \qquad 2 \leq i \leq n,$$

ここで，$|\Pi_n|_{(i>K_n)} = 0$としました．このように，順序統計量の議論は関連づけられたBell多項式とその混合の解析に帰着します．例えば，

$$\mathbb{P}(|\Pi_n|_{(1)} \leq r) = \frac{n!}{2\pi\sqrt{-1}B_n(v., w.)} \oint \frac{\breve{v}_n(\breve{w}^{(r)}(u))}{u^{n+1}} du, \qquad n \to \infty,\ r = o(n)$$

のような漸近評価をすることができます．このような結果は形式的ですが，任意のGibbs分割にあてはまりますし，Pitman分割について知られている性質を容易に導くこともできます．ただし，解析の難しさはモデルの性質に依存しますし，モデルに固有の性質を用いるアプローチの方がモデルのより深い理解につながることもあるでしょう．