統計数理(2003)
第 51 巻 第 2 号 351–372

特集「個票開示問題の統計理論」
[総合報告]

ピットマン確率分割と関連する話題

大和 元[1] ・渋谷 政昭[2]

(受付 2003 年 1 月 24 日;改訂 2003 年 5 月 12 日)

要　　旨

個票開示における 1 つの問題は，母集団一意の推定問題である．これに関連する確率分割のモデルとして，Ewens (1972)による確率分割(Ewens sampling formula)が広く知られている．このイーエンス確率分割は確率的性質として交換可能性(exchangeability)，と size-biased permutation の下での不変性をもつ．この 2 つの性質を特徴づける確率分割として，Pitman (1995, 1996c)は新しい確率分割(Pitman sampling formula)を導いた．このピットマン確率分割を簡単な壺のモデルを用いて導き，更にその性質と関連する事柄を紹介する．なお，個票開示と確率分割については本特集号の竹村(2003)，渋谷(2003)の稿を参照されたい．

キーワード：イーエンス確率分割，交換可能性，ミタッグ・レフラー分布，ピットマン確率分割，ポアソン・ディリクレ分布．

1. はじめに

自然数の確率分割として，すなわち自然数の分割全体の上の確率分布として，良く知られているのはイーエンス確率分割(Ewens sampling formula)である．自然数 n に対して，その順序のつかない分割全体を $\mathcal{M}_n$ とする．すなわち

$$\mathcal{M}_n = \left\{ (m_1, \ldots, m_n) : m_i \geq 0 \; (i = 1, \ldots, n) \text{ かつ } \sum_{j=1}^{n} j m_j = n \right\}.$$

$\mathcal{M}_n$ 上の 確率分布として，イーエンス確率分割は

$$P\big((M_1, \ldots, M_n) = (m_1, \ldots, m_n)\big) = \frac{n!\theta^k}{\theta^{[n]}} \prod_{j=1}^{n} \frac{1}{j^{m_j} m_j!} \tag{1.1}$$

で与えられる．ここで，$\theta > 0$ は母数であり，$(m_1, \ldots, m_n) \in \mathcal{M}_n$, $x^{[j]} = x(x+1)\cdots(x+j-1)$．また，$k = \sum_{j=1}^{n} m_j$ は分割の個数に等しい．

この確率分布は，Ewens (1972)により，n 個の選択的に同値な(中立な)遺伝子を異なる遺伝子型へ分割する確率として導かれた．一方，ノンパラメトリックなベイズ推定に関連して Antoniak (1974)により，独立にこの分布は導かれたが，Ewens (1972)の研究を契機として多くの集団遺伝学の研究者によりこの分布は研究されている．(参照 Johnson et al. (1997))．

このイーエンス確率分割を一般化して，Pitman (1995)は自然数 n の確率分割として，次を

[1] 鹿児島大学 理学部：〒890–0065 鹿児島市郡元 1–21–35
[2] 高千穂大学 経営学部：〒168–8508 東京都杉並区大宮 2–19–1

与えている．

$$P\big((M_1,\ldots,M_n)=(m_1,\ldots,m_n)\big)=\frac{n!\theta^{[k:\alpha]}}{\theta^{[n]}}\prod_{j=1}^{n}\left(\frac{(1-\alpha)^{[j-1]}}{j!}\right)^{m_j}\frac{1}{m_j!}\,, \tag{1.2}$$

ここで，$0\le\alpha<1$ と $\theta(>-\alpha)$ は母数で，$(m_1,\ldots,m_n)\in\mathcal{M}_n$，$x^{[j:\alpha]}=x(x+\alpha)\cdots(x+(j-1)\alpha)$，$x^{[j]}=x^{[j:1]}$, $k=\sum_{j=1}^{n}m_j$．特に，$\alpha=0$ の場合はイーエンス確率分割(1.1)に等しい．

本報告の目的は，個票開示における 1 つの問題である母集団一意の推定問題に関連する確率分割のモデルとして，ピットマン確率分割(1.2)を壺のモデルを用いて導き，この確率分割自身と関連する話題を紹介することである．

まず，2 節では，壺のモデルを用いてピットマン確率分割(1.2)を導く．続いて，3 節では関連する統計量として，分割の個数について，その分布と積率を述べる．4 節では，exchangeability と size-biased permutation について，各々の下での確率分割の特徴付けとしてピットマン確率分割を述べる．5 節では 2 母数ポアソン・ディリクレ分布を，サブオーディネーター(subordinator)の観点から述べる．6 節では，2 母数 GEM 分布からの標本の重複の同時分布として，ピットマン確率分割 を与える．7 節では，ランダム・ウオークおよびベッセル過程とベッセルブリッヂを下にピットマン確率分割を与える．8 節では，先ず分割の個数について，その極限分布を与え，ついでピットマン確率分割の極限分布を与える．最後に 9 節では，ポアソン・ディリクレ過程について述べる．

2. 壺のモデル

自然数の全体を $\mathcal{N}=\{1,2,\ldots\}$ で表し，各 $n\in\mathcal{N}$ に対して $\mathcal{N}_n=\{1,\ldots,n\}$ と置く．また，α および θ は

$$0\le\alpha<1 \;\text{ かつ }\; \theta>-\alpha \tag{2.1}$$

あるいは，

$$\alpha<0 \;\text{ かつ }\; \text{ある } m\in\mathcal{N} \text{ に対して } \theta=-m\alpha \tag{2.2}$$

を満足する母数とする．

次の様な壺のモデルを考える(Sibuya and Yamato (2001))：番号のついた玉 B_1, $\mathrm{B}_2,\ldots$ を，この順序に従い， 番号のついた壺 U_1, $\mathrm{U}_2,\ldots$ に逐次，ランダムに入れる．先ず，玉 B_1 は壺 U_1 に入れる． 次いで，$n\ (=1,2,\ldots)$ 個の玉 $\mathrm{B}_1,\ldots,\mathrm{B}_n$ を壺に入れたとし，玉は k 個の壺 $\mathrm{U}_1,\ldots,\mathrm{U}_k$ の中に入っているものとする．なお，$j=1,\ldots,k$ に対し，壺 U_j には $c_j(>0)$ 個の玉が 入っているとする，ここで $\sum_{j=1}^{k}c_j=n$.

次の玉 B_{n+1} は,

$$\text{新しい壺 } \mathrm{U}_{k+1}\text{へ，確率 }\frac{\theta+k\alpha}{\theta+n}\text{で}$$

入るか，または，既に玉が入っている壺の何れか

$$\mathrm{U}_j\text{へ確率 }\frac{c_j-\alpha}{\theta+n}\quad(j=1,\ldots,k)\text{で}$$

入る．(注：(2.2)の場合は，何れの玉 B· も壺 U_{m+1}, $\mathrm{U}_{m+2},\ldots$ には確率 1 で入らない．)

n 個の玉 $\mathrm{B}_1,\ldots,$ B_n を壺に入れたとき，これらの玉の壺への入り方を玉の添え数により捉えると，$\mathcal{N}_n$ の順序のついた分割が得られる．今，$\mathcal{A}_n^o$ を $\mathcal{N}_n$ の順序のついた分割全体とする．すなわち，

$$\mathcal{A}_n^o=\left\{(a_1,\ldots,a_k)\,:\,1\le k\le n,\ a_i\cap a_j=\emptyset\ (i\ne j)\,;\ \bigcup_{j=1}^{k}a_j=\mathcal{N}_n\right\}\,,$$

ここで，a_1 は 1 を含み，a_2 は a_1 を除いた $\mathcal{N}_n$ の中の最小数を含む，等々．n 回目の試行の後で，k 個の壺に玉が入っているものとして，$j = 1, \ldots, k$ に対して，壺 U_j に入っている玉の添え数の集合を A_j とする．このとき，$(A_1, \ldots, A_k)$ は $\mathcal{N}_n$ の順序のついたランダムな分割でその確率は

$$P((A_1, \ldots, A_k) = (a_1, \ldots, a_k)) = \frac{\theta^{[k:\alpha]}}{\theta^{[n]}} \prod_{j=1}^{k} (1-\alpha)^{[c_j - 1]}, \quad (a_1, \ldots, a_k) \in \mathcal{A}_n^o, \tag{2.3}$$

で与えられる．ここで，k は玉の入っている壺の数で，$c_j = | a_j | > 0$ は a_j に属する要素の数である $(j = 1, \ldots, k)$．この確率は，各 $j = 1, \ldots, k$ について集合 a_j にはそれに属する要素の個数 c_j にのみ依存し，また $\{c_1, \ldots, c_k\}$ については順序に依存しない．

従って，$\mathcal{N}_n$ の順序のつかない分割全体 $\mathcal{A}_n$ 上で考えても，ランダムな分割 $(A_1, \ldots, A_k)$ について同じ確率が成り立つ．

$$P((A_1, \ldots, A_k) = (a_1, \ldots, a_k)) = \frac{\theta^{[k:\alpha]}}{\theta^{[n]}} \prod_{j=1}^{k} (1-\alpha)^{[c_j - 1]}, \quad (a_1, \ldots, a_k) \in \mathcal{A}_n \,. \tag{2.4}$$

もし，$(a_1, \ldots, a_k)$ $(\in \mathcal{A}_n^o)$ のそれぞれについて属する要素の個数を考えると，自然数 n の順序のついた分割が得られる．今，$\mathcal{C}_n^o$ を n の順序のついた分割全体とする．すなわち，

$$\mathcal{C}_n^o = \{(c_1, \ldots, c_k) \,:\, 1 \le k \le n,\ c_i > 0, i = 1, \ldots, k\,;\ c_1 + \cdots + c_k = n\}\,.$$

n 回目の試行の後で，k 個の壺に玉が入っているものとして，$j = 1, \ldots, k$ に対して，壺 U_j に入っている玉の個数を C_{nj} とする．このとき，$(C_{n1}, \ldots, C_{nk})$ は $\mathcal{N}_n$ の順序のついたランダムな分割である．ある $(c_1, \ldots, c_k) (\in \mathcal{C}_n^o)$ に対して，これを与える集合 $(a_1, \ldots, a_k) (\in \mathcal{A}_n^o)$ の総数は

$$\binom{n-1}{c_1 - 1} \binom{n - c_1 - 1}{c_2 - 1} \cdots \binom{n - c_1 - \cdots - c_{k-1} - 1}{c_k - 1}$$

である．これを(2.3)に乗じて，n の順序付きのランダムな分割 $(C_{n1}, \ldots, C_{nk})$ の確率を得る：

$$P(C_{n1} = c_1, \ldots, C_{nk} = c_k) = \frac{n!\theta^{[k:\alpha]}}{\theta^{[n]}} \prod_{i=1}^{k} \frac{(1-\alpha)^{[c_i - 1]}}{(\sum_{j=i}^{k} c_j)(c_i - 1)!}\,, \tag{2.5}$$

ここで，$(c_1, \ldots, c_k) \in \mathcal{C}_n^o$．

次に，n の順序の付かないランダムな分割を考える．各 $j = 1, \ldots, n$ に対して，

$$M_j = | \{i : C_{ni} = j, i = 1, \ldots, k\} |\,, \quad (C_{n1}, \ldots, C_{nk}) \in \mathcal{C}_n^o$$

とおく．すなわち，n 回の試行の後で，各 $j = 1, \ldots, n$ について，M_j は j 個の玉を含む壺の個数である．$\{M_1, \ldots, M_k\}$ は自然数 n の順序の付かないランダムな分割である．ある $(m_1, \ldots, m_n) (\in \mathcal{M}_n)$ について，これを与える集合 $(a_1, \ldots, a_k) (\in \mathcal{A}_n^o)$ の総数は

$$\frac{n!}{\prod_{j=1}^{n} (j!)^{m_j} m_j!}$$

である．これを(2.3)に乗じる事により，ピットマン確率分割(1.2)を得る．

この壺のモデルに似たものとしてピットマンによる中華料理店過程(Pitman's Chinese restaurant process)がある．ある中華料理店には，可算無限個の人が座れる可算無限個の丸テーブルがあり，この丸テーブルに番号 $1, 2, 3, \ldots$ を付けておく．最初は客のいない店に，番号 $1, 2, 3, \ldots$

の客が一人ずつ入って来て次の様に着席する：番号 1 の客はテーブル 1 に座る．番号 $1,\dots,n$ の客が順次着席し，テーブル $1,\dots,k$ に座っているものとする．各 $j=1,\dots,k$ について，テーブル j には $c_j(>0)$ 人の客が座っているとする．次の客 j は，新しい

$$\text{テーブル } k+1 \text{ に，確率 } \frac{\theta+k\alpha}{\theta+n} \text{で，}$$

座るか，又は既に他の客が座っている

$$\text{テーブル } j \text{ に，確率 } \frac{c_j-\alpha}{\theta+n} \text{で，} (j=1,\dots,k)$$

座るものとする．この時，客 $1,\dots,n$ が番号の付いたテーブルに着席した状態の確率は(2.3)で与えられる．

この中華料理店過程で客 $n+1$ は既に客の座っているテーブルを選んだとき，それぞれのテーブルでは，

既にすわっている客何れかの右隣りに，等確率で，座る

ものとする．客 $1,\dots,n$ が番号の付いたテーブルに着席した状態は，$\mathcal{N}_n$ の順列のサイクル表示を与える．テーブル j の自然数の配列を Π_j で表す． Π_1 は 1 で始まるサイクルを表す． Π_2 は，Π_1 を除いた $\mathcal{N}_n$ の中の，最小数から始まるサイクルを表す，等々． このとき，次の確率をもつ：

$$P((\Pi_1,\dots,\Pi_k)=(\pi_1,\dots,\pi_k))=\frac{\theta^{[k:\alpha]}}{\theta^{[n]}}\prod_{i=1}^{k}\frac{(1-\alpha)^{[|\pi_i|-1]}}{(|\pi_i|-1)!}, \tag{2.6}$$

ただし，$(\pi_1,\dots,\pi_k)$ は $\mathcal{N}_n$ のサイクル表示で， サイクル π_1 は 1 から始まり， サイクル π_2 は，π_1 を除く $\mathcal{N}_n$ の中の，最小数から始まる，等々．また，$|\pi_i|$ は サイクル π_i の長さを表す．

$\alpha=0$ で $\theta=1$ の場合， この確率は次のようになる，

$$P((\Pi_1,\dots,\Pi_k)=(\pi_1,\dots,\pi_k))=\frac{1}{n!}.$$

これは，$\mathcal{N}_n$ の全ての順列上の一様分布である．

3. 分割の個数

前節の壺のモデルに付随して，次のような確率変数列 $Y_1,Y_2,Y_3,\dots$ を考える；$Y_1\equiv 1$ とし，$n=1,2,\dots$ に対して

$$Y_{n+1}=1 \quad (\text{玉 } \mathrm{B}_{n+1}\text{が新しい壺に入る場合}),$$

および

$$Y_{n+1}=0 \quad (\text{玉 } \mathrm{B}_{n+1}\text{が古い壺に入る場合})$$

とする．このとき，

$$P(Y_{j+1}=1 \mid Y_1=y_1,\dots,Y_j=y_j)=\frac{\theta+(y_1+\cdots+y_j)\alpha}{\theta+j}, \tag{3.1}$$

かつ

$$P(Y_{j+1}=0 \mid Y_1=y_1,\dots,Y_j=y_j)=\frac{j-(y_1+\cdots+y_j)\alpha}{\theta+j}, \quad j=1,2,\dots \tag{3.2}$$

ただし，$y_1=1$ および $y_j=0$ 又は $=1$ $(j=2,3,\dots)$.

壺のモデルで，n 回の試行の後で，玉の入っている壺の数を K_n とすると，

$$K_n = Y_1 + \cdots + Y_n$$

と表され，上の $\{Y_n\}$ の条件付き確率から

$$P(K_{n+1} = k+1 \mid K_n = k) = \frac{\theta + k\alpha}{\theta + n}, \quad P(K_{n+1} = k \mid K_n = k) = \frac{n - k\alpha}{\theta + n}. \tag{3.3}$$

が得られる．これから，K_n の確率についての漸化式

$$P(K_{n+1} = k) = \frac{n - k\alpha}{\theta + n} P(K_n = k) + \frac{\theta + (k-1)\alpha}{\theta + n} P(K_n = k-1)$$

が得られる．これを用いて，$\alpha \neq 0$ の場合の K_n の確率が求められる．

$$P(K_n = k) = \frac{\theta^{[k:\alpha]}}{\theta^{[n]}} c(n,k,\alpha)\alpha^{-k}, \quad k = 1, 2, \ldots, n, \tag{3.4}$$

ここで，$c(n,k,\alpha) = (-1)^{n-k} C(n,k,\alpha)$ であり，$C(n,k,\alpha)$ は C-numbers 又は the generalized Stirling numbers と言われる（Pitman (1996b,1999), Yamato et al. (2001)）．C-number である $C(n,k;s)$ は次の多項式で表される恒等式で定義される．

$$(st)^{(n)} = \sum_{k=1}^{n} C(n,k;s)\, t^{(k)}, \quad t^{(k)} = t\,(t-1) \cdots (t-k+1),$$

または，同値であるが

$$(st)^{[n]} = \sum_{k=1}^{n} c(n,k;s)\, t^{[k]}.$$

$c(\cdot,\cdot;\cdot)$ は任意の s に対し，次の漸化式を満たす

$$c(n+1,k;s) = (n - k\,s)\, c(n,k;s) + s\, c(n,k-1;s),$$

ただし，$c(0,0;s) = 1$, $c(n,0;s) = 0\,(n > 0)$, and $c(0,k;s) = 0\,(k > 0)$（参照，例えば，Charalambides and Singh (1988)）．

$\alpha = 0$ の場合，$K_n = \sum_{j=1}^{n} M_j$ の分布は

$$P(K_n = k) = \begin{bmatrix} n \\ k \end{bmatrix} \frac{\theta^k}{\theta^{[n]}}, \quad k = 1, 2, \ldots, n,$$

で与えられることが，イーエンス確率分割から既に知られている（Ewens (1972)）．ここで，$\left[n \atop k \right]$ は符号無しの第 1 種スターリング数である．

一方，K_n の分布は次の様にしても求められる．各 $k = 1, \ldots, n$ について，

$$P(K_n = k) = \sum\nolimits_1 P((M_1, \ldots, M_n) = (m_1, \ldots, m_n)),$$

ここで, 和 $\sum_1$ は, 固定された k と n について, $m_1 + \cdots + m_n = k$ および $m_1 + 2m_2 + \cdots + nm_n = n$ を満たす全ての正の整数 $m_1, \ldots, m_n$ についての和である．ところで，ピットマン確率分割 (1.1) は次の様に表される．

$$P((M_1, \ldots, M_n) = (m_1, \ldots, m_n)) = \frac{(-1)^{n-k}\theta^{[k:\alpha]} n!}{\alpha^k \theta^{[n]}} \prod_{j=1}^{n} \frac{1}{m_j!} \binom{\alpha}{j}^{m_j}. \tag{3.5}$$

これへ，Charalambides and Singh (1988)， p. 2553 の C-number の表現

$$C(n,k;\alpha)=\sum\nolimits_{1}\frac{n!}{m_1!\cdots m_n!}\binom{\alpha}{1}^{m_1}\cdots\binom{\alpha}{n}^{m_n},$$

を用いて，K_n の確率分布(3.4)が得られる．

この K_n の平均は

$$EK_n^{(1)}=EK_n=\frac{\theta}{\alpha}\Big[\frac{(\theta+\alpha)^{[n]}}{\theta^{[n]}}-1\Big].$$

で与えられる(Pitman (1996b))．分散については

$$EK_n^{(2)}=\frac{\theta(\theta+\alpha)}{\alpha^2\theta^{[n]}}\{(\theta+2\alpha)^{[n]}-2(\theta+\alpha)^{[n]}+\theta^{[n]}\}$$

を用いて，

$$\mathrm{Var}(K_n)=\frac{\theta(\theta+\alpha)}{\alpha^2}\cdot\frac{(\theta+2\alpha)^{[n]}}{\theta^{[n]}}-\frac{\theta^2}{\alpha^2}\left(\frac{(\theta+\alpha)^{[n]}}{\theta^{[n]}}\right)^2-\frac{\theta}{\alpha}\cdot\frac{(\theta+\alpha)^{[n]}}{\theta^{[n]}}.$$

$\theta=0$ の場合には，K_n の上昇階乗積率が Pitman (1996b)によって求められている． $r=1,2,\ldots$ に対して，

$$E_{\alpha,0}(K_n^{[r]})=\frac{(r-1)!(r\alpha)^{[n]}}{(n-1)!\alpha}. \tag{3.6}$$

一般の積率について，Pitman (1996b)は次のような漸近的な評価を与えている．

$$E_{\alpha,\theta}(K_n^r)\sim n^{\alpha r}\frac{\Gamma(\frac{\theta}{\alpha}+r+1)\Gamma(\theta+1)}{\Gamma(\theta+r\alpha+1)\Gamma(\frac{\theta}{\alpha}+1)}.$$

Yamato and Sibuya (2000)は積率の厳密な評価を $r=1,2,\ldots,n$ に対して与えている．先ず，K_n の下降階乗積率について，(3.3)から次の漸化式が示される．$r=1,2,\ldots,n$ に対して，

$$EK_{n+1}^{(r)}=\left(1+\frac{\alpha r}{n+\theta}\right)EK_n^{(r)}+\frac{r[\theta+(r-1)\alpha]}{n+\theta}EK_n^{(r-1)}. \tag{3.7}$$

これを用いて K_n の下降階乗積率が帰納法で示される．$r=1,2,\ldots,n$ に対して

$$E_{\alpha,\theta}(K_n^{(r)})=\frac{\theta^{[r:\alpha]}}{\alpha^r\theta^{[n]}}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}(\theta+j\alpha)^{[n]}.$$

第 2 種スターリング数 $\left\{ {r \atop j} \right\}$ の満たす関係 $x^r=\sum_{j=1}^{r}\left\{ {r \atop j} \right\}x^{(j)}$ を用いて，K_n の通常の積率が求められる．

$$E_{\alpha,\theta}(K_n^r)=\sum_{j=0}^{r}(-1)^{r-j}\left(1+\frac{\theta}{\alpha}\right)^{[j]}R\left(r,j,\frac{\theta}{\alpha}\right)\frac{(\theta+j\alpha+1)^{[n-1]}}{(\theta+1)^{[n-1]}}, \tag{3.8}$$

ここで， $R(r,j,\lambda)$ は任意の y と λ および $r=1,2,\ldots$ に対して，

$$\sum_{j=0}^{r}y^{(j)}R(r,j,\lambda)=(y+\lambda)^r$$

を満たす関数である． 関数 $R(r,j,\lambda)$ については，Carlitz (1980)を参照のこと．更に，符号無し第 1 種スターリング数の満たす関係 $x^{[r]}=\sum_{j=1}^{r}\left[{r \atop j} \right]x^j$ を用いて，これから上昇階乗積率が求められる．

$$E_{\alpha,\theta}(K_n^{[r]})=\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}\left(\frac{\theta}{\alpha}\right)^{[j]}\left(\frac{\theta}{\alpha}\right)^{(r-j)}\frac{(\theta+j\alpha)^{[n]}}{\theta^{[n]}}$$

(参照 Yamato and Sibuya (2000)).

4. Exchangeability と size-biased permutation

本節では，ピットマン確率分割を exchangeability と size-biased permutation の二つの観点から捉えた特徴付けを述べる．2 節の壺のモデルで，$a_1 = (1, 2, \ldots, c_1)$ に対し，最初の壺に入る確率

$$P(A_1 = (1, 2, \ldots, c_1)) = \frac{(1-\alpha)^{[c_1-1]}(\theta+\alpha)^{[n-c_1]}}{(1+\theta)^{[n-1]}}, \quad c_1 = 1, \ldots, n,$$

を得る．$|a_1| = c_1$ を満たす各 a_1 に対し，この確率は，$P(A_1 = a_1)$ に等しい．何故なら確率 $P((A_1, \ldots, A_k) = (a_1, \ldots, a_k))$ は各部分集合 a_j の要素自身には依存していないからである．(2.5)で与えられる n の順序のついたランダムな分割$(C_{n1}, \ldots, C_{nk})$の確率に対して，次の周辺確率と条件付き確率を得る．

命題 4.1.

$$P(C_{n1} - 1 = x) = \binom{n-1}{x} \frac{(1-\alpha)^{[x]}(\theta+\alpha)^{[n-1-x]}}{(1+\theta)^{[n-1]}}, \quad x = 0, 1, \ldots, n-1.$$

すなわち， $C_{n1} - 1$ は負の超幾何分布 $NgHg(n-1, 1-\alpha, \theta+\alpha)$ に従う．

$c_1 + \cdots + c_{l-1} < n$ を満たす正整数 $l (\leq n)$ と $c_1, \ldots, c_{l-1}$ に対して，

$$P(C_{nl} = c_l \mid C_{n1} = c_1, \ldots, C_{n,l-1} = c_{l-1}) = \binom{n - c_1 - \cdots - c_{l-1} - 1}{c_l - 1}$$
$$\times \frac{(1-\alpha)^{[c_l-1]}(\theta+l\alpha)^{[n-c_1-\cdots-c_l]}}{(1+\theta+(l-1)\alpha)^{[n-c_1-\cdots-c_{l-1}-1]}}, \quad c_l = 1, 2, \ldots, n - c_1 - \cdots - c_{l-1}$$

すなわち，条件 $C_{n1} = c_1, \ldots, C_{n,l-1} = c_{l-1}$ の下で，$C_{nl} - 1$ は負の超幾何分布 $NgHg(n - c_1 - \cdots - c_{l-1} - 1, 1-\alpha, \theta + m\alpha)$ に従う．

ここで，負の超幾何分布はその母数がベータ分布に従う 2 項分布で表される事に注意しよう．このことから，相対的比率 $(C_{n1}/n, C_{n2}/n, \ldots)$ は次で与えられるランダムな比率 $(P_1, P_2, \ldots)$ に確率 1 で収束する．

$$P_1 = W_1, \quad P_j = (1-W_1)\cdots(1-W_{j-1})W_j, \quad j = 2, 3, \ldots, \tag{4.1}$$

ここで， $W_1, W_2, \ldots$ は独立で，各 W_j $(j = 1, 2, \ldots)$ はベータ分布 beta$(1-\alpha,\, \theta + j\alpha)$ に従う確率変数である．ただし，母数が(2.2)の場合には $W_m = 1$ で，$j > m$ に対しては W_j は定義されないものとする．

母数 $0 \leq \alpha < 1$ および $\theta > -\alpha$ について，(4.1)で与えられるランダムな比率は 2 母数 GEM 分布 (α, θ) として，知られている．$\alpha = 0$ かつ $\theta > 0$ の場合は GEM 分布として知られている (参照，例えば， Johnson et al. (1997), Chapter 41)．

Pitman (1995)は exchangeability の見地から次の命題 4.2 を示している．

ここでは，$\mathcal{N}_n = \{1, 2, \ldots, n\}$ 及び $\mathcal{N} = \{1, 2, \ldots\}$ の順序の付かない分割全体とその上の分布(確率分割)を考える．$\mathcal{N}_n$ の確率分割 $\Pi_n = (A_i)$ が exchangeable とは，$\mathcal{N}_n$ の全ての置換の下で Π_n の分布が不変である事である．$\mathcal{N}$ の確率分割とは，$\mathcal{N}_n$ の確率分割の列 $\{\Pi_n\}$, $n = 1, 2, \ldots$ を言う．ただし，$m < n$ に対して，Π_n の $\mathcal{N}_m$ への制限が Π_m ($\mathcal{N}_m$ の確率分割)であるものとする．全ての n に対して，Π_n の分布が不変であるとき，$\Pi = \{\Pi_n\}$ は exchangeable と言う．

命題 4.2. (Pitman (1995))(2.1) 或いは (2.2)で与えられる母数 α 及び θ に対して，$\mathcal{N}$ の exchangeable な確率分割の列 $\Pi_n = (A_1, \ldots, A_n)$, $n = 1, 2, \ldots$ は 2 節の壺のモデルで記述される．その確率は (2.4)で与えられる．これに付随する n の順序の付かない確率分割 $(M_1, \ldots, M_n)$ の確率は(1.2)で与えられる．P_i を A_i の相対頻度の概収束の極限とする．すなわち，

$$P_i = \lim_{n\to\infty} \frac{|\, A_i \cap \mathcal{N}_n \,|}{n} \quad \text{a.s.} \quad (\text{確率 } 1 \text{ で}) \quad i = 1, 2, \ldots$$

このとき，P_i $(i = 1, 2, \ldots)$ は(4.1)で表される．ただし，母数が (2.2)の場合には $W_m = 1$ で，$j > m$ に対しては W_j は定義されないものとする．

次に，size-biased permutation を考える．$\mathbf{P} = (P_1, P_2, \ldots)$ を $\mathcal{N}$ 上のランダムな離散分布とする．すなわち，$\mathbf{P}$ は次で与えられる Δ のランダムな要素である．

$$\Delta = \left\{ \mathbf{x} = (x_1, x_2, \ldots) \;:\; x_j \geq 0,\; j \in \mathcal{N},\; \sum_{j=1}^{\infty} x_j = 1 \right\}.$$

N_1, N_2, $\ldots$ は $\mathcal{N} \cup \{\infty\}$ の値をとる確率変数列で次の様に定める；

$$P\big(N_1 = i \mid \mathbf{P} = (p_1, p_2, \ldots)\big) = p_i\,, \quad i = 1, 2, \ldots$$

$\forall k \in \mathcal{N}$ および異なる $\forall\, j_1, \ldots, j_k \in \mathcal{N}$ に対して，

(1) $(p_1, p_2, \ldots)$ が少なくとも k 個の正の要素を持つときには，

$$P(N_k = j_k \mid \mathbf{P} = (p_1, p_2, \ldots), N_1 = j_1, \ldots, N_{k-1} = j_{k-1}) = \frac{p_{j_k}}{1 - p_{j_1} - \cdots - p_{j_{k-1}}}\,.$$

(2) $(p_1, p_2, \ldots)$ が $k-1$ 個以下の正の要素を持つときには，

$$P(N_k = \infty \mid \mathbf{P} = \mathbf{p}) = 1\,.$$

$\mathbf{P} = (P_1, P_2, \cdots)$ の size-biased permutation は

$$\widetilde{\mathbf{P}} = (P_{N_1}, P_{N_2}, \ldots)\,,$$

で与えられる．ただし，$P_\infty = 0$ とする．(参照，例えば，Pitman (1996a, 1996c)および Gnedin (1998)．)

$\mathbf{P}$ の size-biased permutation $\widetilde{\mathbf{P}}$ が元の $\mathbf{P}$ と 同じ分布をもつとき，$\mathbf{P}$ は size-biased permutation の下で不変(invariant)と言う．

Patil and Taillie (1977)は次の問題を提起した：独立であるが，おそらく同一分布には従わない確率変数列 $(W_1, W_2, \ldots)$ に対し，関係(4.1)で表されるランダムな離散分布 $(P_1, P_2, \ldots)$ を考える．これが size-biased permutation の下で不変となるのは，確率変数列 $(W_1, W_2, \ldots)$ が如何なる場合か？ この問題は Pitman (1996c)により，次のように解決された．

命題 4.3. (Pitman(1996c))ランダムな比率 $\mathbf{P} = (P_1, P_2, \ldots)$ は，$P_n \geq 0$，$\sum_n P_n = 1$，$P_1 = W_1 < 1$，かつ独立な確率変数列 W_i により $P_n = (1 - W_1) \cdots (1 - W_{n-1}) W_n$ と表されるものとする．このとき，$\mathbf{P}$ が size-biased permutation の下で不変であるのは，次の 4 つの条件 (i), (ii)a, (ii)b, (ii)c の内何れか 1 つが成り立つときである．

(i) 全ての n に対して $P_n > 0$ a.s. (確率 1 で)；この場合，ある母数 $0 \leq \alpha < 1$, $\theta > -\alpha$ があって，W_n は，各 $n = 1, 2, \ldots$ について，ベータ分布

$$\text{beta}(1 - \alpha, \theta + n\alpha)$$

に従う．

または

（ii） ある正の整数 m があって，$\{n : P_n > 0\} = \{1, \ldots, m\}$ a.s.；この場合，次の(a), (b), (c)の何れかが成り立つ．

（a） ある母数 $\beta > 0$ に対し，W_n，$n = 1, \ldots, m$ は beta $(1+\beta, m\beta - n\beta)$ に従う，あるいは

（b） $n = 1, \ldots, m$ に対して，$W_n = 1/(m-n+1)$ a.s.，すなわち，$P_n = 1/m$ a.s.，または

（c） $m = 2$，かつ $F(dw) = (1-w)\Pr(W_1 \in dw)/E(1-W_1)$ で定義される $(0,1)$ 上の分布 F は $1/2$ に関し対称である．

5. 2 母数ポアソン・ディリクレ分布とサブオーディネーター

母数 $\alpha = 0$ および $\theta > 0$ により，(4.1)で与えられるランダムな比率 $(P_1, P_2, \ldots)$ は GEM 分布と言われ，これを減少順に並べた順序統計量($P_{(1)} \geq P_{(2)} \geq \cdots$)はポアソン・ディリクレ分布(Poisson-Dirichlet distribution)と言われる．ポアソン・ディリクレ分布の size-biased permutation は GEM 分布である (参照，例えば Johnson et al. (1997), Chapter 41)．母数 $0 \leq \alpha < 1$ および $\theta > -\alpha$ により，(4.1)で与えられるランダムな比率 $(P_1, P_2, \ldots)$ は 2 母数 GEM 分布，GEM(α, θ) である．これを減少順に並べた順序統計量($P_{(1)} \geq P_{(2)} \geq \cdots$)は 2 母数ポアソン・ディリクレ分布，PD$(\alpha, \theta)$，と言われる．2 母数ポアソン・ディリクレ分布の size-biased permutation は 2 母数 GEM 分布である(参照，Pitman and Yor (1997))．特に，GEM$(0, \theta)$ および PD$(0, \theta)$ はそれぞれ，母数 $\theta (> 0)$ の(1 母数の)GEM 分布 GEM(θ) および PD 分布 PD(θ) である．2 母数 GEM 分布は次の様な特徴付けができる．

命題 5.1. (Yamato et al. (2001)) Δ 上のランダムな比率 $\boldsymbol{V} = (V_1, V_2, \ldots)$ が 2 母数 GEM(α, θ) 分布をもつのは，次の場合にかぎる，

$$(5.1)\qquad E\Big[V_1^{c_1-1}(1-V_1)V_2^{c_2-1}(1-V_1-V_2)V_3^{c_3-1}\cdots(1-V_1-\cdots-V_{k-1})V_k^{c_k-1}\Big] = \frac{\theta^{[k:\alpha]}}{\theta^{[c_1+\cdots+c_k]}}\prod_{i=1}^{k}(1-\alpha)^{[c_i-1]}, \quad \forall k, c_1, \ldots, c_k \in \mathcal{N}.$$

2 母数ポアソン・ディリクレ分布で，特に母数 $\theta = 0$ の場合 PD$(\alpha, 0)$ については，次の様な特徴付けが示されている．

命題 5.2. (Pitman and Yor (1997)) $0 < \alpha < 1$ とし，$(P_{(1)} \geq P_{(2)} \geq \cdots)$ は 2 母数ポアソン・ディリクレ分布 PD$(\alpha, 0)$ に従うとする．このとき，

$$R_n = \frac{P_{(n+1)}}{P_{(n)}}, \quad n = 1, 2, \ldots$$

は互いに独立で，かつベータ分布 beta$(n\alpha, 1)$ に従う．すなわち，$P(R_n \leq x) = x^{n\alpha}$ $(0 \leq x \leq 1)$.

逆に，$R_n, n = 1, 2, \ldots$ は互いに独立で，かつベータ分布 beta$(n\alpha, 1)$ に従うものとする．このとき，関係

$$P_{(1)} = \frac{1}{1 + R_1 + R_1R_2 + R_1R_2R_3 + \cdots}, \quad P_{(n+1)} = P_{(1)}R_1R_2\cdots R_n \ \ (n = 1, 2, \ldots)$$

により，$R_n, n = 1, 2, \ldots$ から一意に定まる $P_{(n)}, n = 1, 2, \ldots$ は 2 母数ポアソン・ディリクレ分布 PD$(\alpha, 0)$ に従う．

母数が $0 < \alpha < 1$ 及び $\theta > 0$ の場合に，2 母数ポアソン・ディリクレ分布に従う確率変数列の，subordinator に基づく具体的な表現が Pitman and Yor (1997) に与えられている．次の定理の表現は Carlton (1999) に沿ったものである．

定理 5.1. (Pitman and Yor (1997), Carlton (1999)) $0 < \alpha < 1$, $\theta > 0$ および $C > 0$ とする．$\{Z_t, t \geq 0\}$ はレヴィ測度 (Lévy measure) $d\Lambda(x) = \alpha C x^{-\alpha-1} e^{-x} dx$ をもつ subordinator とする．確率変数 T は，$\{Z_t, t \geq 0\}$ に独立で，ガンマ分布 $\mathrm{Ga}(\theta/\alpha, \beta)$ に従うものとする．ここで，$\beta^{-1} = C\Gamma(1-\alpha)$．また，$V_1(Z_T) \geq V_2(Z_T) \geq V_3(Z_T) \geq \cdots \geq 0$ は区間 $[0, T]$ における Z_t のジャンプを大きさの順序に並べたものである．この時，

$$\left(\frac{V_1(Z_T)}{Z_T}, \frac{V_2(Z_T)}{Z_T}, \frac{V_3(Z_T)}{Z_T}, \ldots \right) \tag{5.2}$$

は 2 母数ポアソン・ディリクレ分布 $\mathrm{PD}(\alpha, \theta)$ に従う．さらに，確率変数列(5.2)は Z_T に独立で，Z_T はガンマ分布 $\mathrm{Ga}(\theta, 1)$ に従う．

上の定理に現れたサブオーディネーター(subordinator)について，以下に補足説明をする(参照，例えば，Kingman (1975)， Carlton (1999)，Pitman (2002))．

確率過程 $\{Z_t, t \geq 0\}$ は次の性質をもつとき，subordinator と言われる：

(i) 全ての $t \geq 0$ に対して，$0 \leq Z_t < \infty$ かつ $Z_0 = 0$.

(ii) Z_t は独立増分を持つ；全ての $0 \leq t_0 < t_1 < \cdots < t_n$ に対して，増分 $Z_{t_1} - Z_{t_0}, \ldots, Z_{t_n} - Z_{t_{n-1}}$ は独立である．

(iii) Z_t は定常増分を持つ；全ての $0 \leq s < t$ に対して，増分 $Z_t - Z_s$ は $t - s$ にのみ依存する．

(iv) Z_t は確率的に非減少である；全ての $0 \leq s < t$ に対して，$Z_t - Z_s \geq 0$ a.s.

レヴィ過程(Lévy process)の特別な場合であり，Sato (1999)では increasing Lévy process と言われている．Z_t の分布のラプラス変換

$$E[e^{-uZ_t}] = e^{-t\psi(u)}$$

については，レヴィ・ヒンチン公式(Lévy-Khintchine formula)の特別な場合として，ラプラス指数(Laplace exponent)$\psi(u)$ は次のように表される．

$$\psi(u) = cu + \int_0^\infty (1 - e^{-ux}) \Lambda(dx),$$

ここで，Λ は $(0, \infty)$ 上の測度で，Z_1 または $\{Z_t, t \geq 0\}$ のレヴィ測度と言われる．c はドリフト(drift)係数と言われ，以下では $c = 0$ の場合を考える．

$\{Z_t, t \geq 0\}$ は次の形で表わされる：

$$Z_t = \sum_{0 < s \leq t} \Delta_s \quad (t \geq 0),$$

ここで，$\Delta_s = Z_s - Z_{s-}$ である．$\{\Delta_s > 0, s > 0\}$ は intensity measure $\Lambda(dx)$ をもつ $(0, \infty)$ 上のポアソン点過程である．すなわち，$(0, \infty)$ のボレル集合 B に対し $\mid \Delta \in B \mid$ (B に入る Δ の個数) は平均 $\int_B \Lambda(dx)$ のポアソン分布に従う．上の定理中の $V_1(Z_T) \geq V_2(Z_T) \geq V_3(Z_T) \geq \cdots \geq 0$ は，区間 $[0, T]$ 上での Δ_s を大きさの順に並べたものである．

一例として，標準ガンマ過程 $\{\gamma_s, s \geq 0\}$ を考えよう．ここで，γ_s の密度は標準ガンマ分布 $\mathrm{Ga}(s,1)$ で与えられる．これに対するラプラス指数 $\psi(u)$ は

$$\psi(u) = \log(1+u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \cdots$$

で，レヴィ測度は $\Lambda(dx) = x^{-1}e^{-x}dx$ である．今，$\theta > 0$ に対して，区間 $[0,\theta]$ におけるガンマ過程 $\{\gamma_s, s \geq 0\}$ のジャンプを大きさの順序に並べたものを $V_1(\gamma_\theta) \geq V_2(\gamma_\theta) \geq V_3(\gamma_\theta) \geq \cdots \geq 0$ とする．このとき，次が知られている．

命題 5.3. （Kingman (1975)）

$$\left(\frac{V_1(\gamma_\theta)}{\gamma_\theta}, \frac{V_2(\gamma_\theta)}{\gamma_\theta}, \frac{V_3(\gamma_\theta)}{\gamma_\theta}, \ldots\right)$$

は，1 母数ポアソン・ディリクレ分布 PD(θ)(=PD$(0,\theta)$) に従う．

次に，安定(stable) な subordinator $\{\xi_t, t \geq 0\}$ を考えよう．$0 < \forall\alpha < 1$ および $a > 1$ とし，レヴィ測度は $\Lambda(dx) = ax^{-\alpha-1}dx$ で与えられるとする．その積分は無限大に発散するが，ψ は収束し，

$$\psi(u) = bu^\alpha, \quad b = a\alpha^{-1}\Gamma(1-\alpha)\,.$$

このとき，

$$E[e^{-u\xi_t}] = e^{-tbu^\alpha}\,. \tag{5.3}$$

この安定な subordinator $\{\xi_t, t \geq 0\}$ について，次が知られている．

命題 5.4. （Pitman and Yor (1997)）(1) 任意の $s > 0$ に対して，区間 $[0,s]$ における安定な subordinator $\{\xi_t, t \geq 0\}$ のジャンプを大きさの順序に並べたものを $V_1(\xi_s) \geq V_2(\xi_s) \geq V_3(\xi_s) \geq \cdots \geq 0$ とする．このとき，

$$\left(\frac{V_1(\xi_s)}{\xi_s}, \frac{V_2(\xi_s)}{\xi_s}, \frac{V_3(\xi_s)}{\xi_s}, \ldots\right)$$

は，2 母数ポアソン・ディリクレ分布の特別な場合である PD$(\alpha, 0)$ に従う．

(2) 固定された $t > 0$ に対して，区間 $[0,t]$ における安定な subordinator $\{\xi_t, t \geq 0\}$ のジャンプを大きさの順序に並べたものを $V_1(t) \geq V_2(t) \geq V_3(t) \geq \cdots \geq 0$ とする．このとき，

$$\left(\frac{V_1(t)}{t}, \frac{V_2(t)}{t}, \frac{V_3(t)}{t}, \ldots\right)$$

は，2 母数ポアソン・ディリクレ分布の特別な場合である PD$(\alpha, 0)$ に従う．

先の安定な subordinator $\{\xi_t, t \geq 0\}$ に戻って，ξ_t の分布のラプラス変換(5.3)を更に次の様に表そう．

$$E[e^{-u\xi_t}] = e^{-b(ut^{1/\alpha})^\alpha} = E[e^{-ut^{1/\alpha}\xi_1}].$$

この subordinator $\{\xi_t, t \geq 0\}$ は指数(exponent)α で安定であり，

$$\xi_t \stackrel{d}{=} t^{1/\alpha}\xi_1$$

が成り立つ．ただし，$\stackrel{d}{=}$ は分布の意味で同値である事を表わす．

今，$a=\alpha/\Gamma(1-\alpha)$ ととる．従って，$b=1$ である．このとき，

$$E\left[e^{-u\xi_1}\right]=e^{-u^\alpha}=\int_0^\infty e^{-ux}f_\alpha(x)dx\,.$$

ここで，$f_\alpha(x)$ は ξ_1 の(安定)密度で

$$f_\alpha(x)=\frac{1}{\pi}\sum_{j=0}^{\infty}\frac{(-1)^{j+1}}{j!}\sin(\pi\alpha j)\frac{\Gamma(\alpha j+1)}{x^{\alpha j+1}}\,. \tag{5.4}$$

さて，$S_1=\xi_1^{-\alpha}$ とおく．式(5.3)で，変数変換 $u=x^{-\alpha}$ を行うことにより，S_1 の密度が得られる．

$$g_\alpha(u)=\frac{1}{\pi\alpha}\sum_{j=0}^{\infty}\frac{(-1)^{j+1}}{j!}\sin(\pi\alpha j)\Gamma(\alpha j+1)u^{j-1}\,,$$

これは，後の 8.1 節で現われる母数 α のミタッグ・レフラー分布(Mittag-Leffler distribution)の密度である．すなわち，母数 α の付随する安定(stable)な subordinator $\{\xi_t, t\geq 0\}$ について，$\xi_1^{-\alpha}$ は母数 α のミタッグ・レフラー分布に従う．

6. 2 母数 GEM 分布からの標本

式(3.1)で与えられる 2 母数 GEM(θ,α) 分布に従う, $\mathbf{W}=(W_1,W_2,\dots)$ からの，大きさ n の標本を $\mathbf{X}=(X_1,\dots,X_n)$ とする．すなわち，$(w_1,w_2,\dots)\in\Delta$ に対し，

$$P(\mathbf{X}=(x_1,\dots,x_n)\mid \mathbf{W}=(w_1,w_2,\dots))=\prod_{j=1}^{n}w_{x_j}\,,\quad x_j\in\mathcal{N}\,,\quad j=1,\dots,n\,. \tag{6.1}$$

この無作為標本に対して，次のような統計量を考える．$(Y_1,\dots,Y_n)$ を次のように定義する：

$$Y_1=1 \quad \text{かつ} \quad Y_j=\prod_{i=1}^{j-1}I[X_j\neq X_i]\,,\quad j=2,\dots,n\,, \tag{6.2}$$

ここで，$I[\cdot]$ は，カッコ $[\cdot]$ 内の事象が起こったとき値 1 を取り，起こらないときは値 0 を取る．すなわち，X_j が新しい観測値のとき $Y_j=1$ で，これまでの観測 $X_1,\dots,X_{j-1}$ の何れかに一致するとき $Y_j=0$． この系列 $(Y_1,\dots,Y_n)$ は 3 節の系列 $(Y_1,\dots,Y_n)$ に確率的に同値である．

$$K_n=\sum_{j=1}^{n}Y_j\,, \tag{6.3}$$

は，n 個の観測値 $X_1,\dots,X_n$ の中の異なる観測値の個数である．また，$\nu_1=1$ とおき，

$$\nu_j=\min\left\{\ell\,;\,\sum_{i=1}^{\ell}Y_i=j\right\},\quad j=2,\dots,k\,, \tag{6.4}$$

とおく．さらに，

$$\widetilde{X_j}=X_{\nu_j}\,,\quad j=1,\dots,k\,, \tag{6.5}$$

とおく．特に，$\widetilde{X_1}=X_1$ である．$\widetilde{X_2}$ は $X_1,X_2,\dots,X_n$ の中で，この添え数の順序で，X_1 と異なる 2 番目に新しい観測値である．一般に，$j=3,4,\dots$ に対し，$\widetilde{X_j}$ は $X_1,X_2,\dots,X_n$ の中で，この添え数の順序で，j 番目に新しい観測値である．$\widetilde{X_j}$ に等しい観測値の個数を

$$C_{nj}=\sum_{i=1}^{n}I[X_i=\widetilde{X_j}]\,,\quad j=1,\dots,k\,, \tag{6.6}$$

で表す．$\mathbf{V} = (V_1, V_2, \dots)$ を $\mathbf{W}$ の size-biased permutation とすると，任意の k $(1 \le k \le n)$と $c_1, \dots, c_k \in \mathcal{N}$ に対して，

$$P\big(X_1 = \cdots = X_{c_1}, X_{c_1+1} = \cdots = X_{c_1+c_2}, \dots, X_{c_1+\cdots+c_{k-1}+1} = \cdots = X_{c_1+\cdots+c_k}\big) \\ = E[V_1^{c_1-1}(1-V_1)V_2^{c_2-1}(1-V_1-V_2)V_3^{c_3-1}\cdots(1-V_1-\cdots-V_{k-1})V_k^{c_k-1}]\,, \tag{6.7}$$

（Lemma 3 of Sibuya and Yamato（1995））．今， $\mathbf{W} = (W_1, W_2, \dots)$ は 2 母数 GEM(θ, α) 分布を持つとする．2 母数 GEM(θ, α) 分布は size-biased permutation の下で不変である (Pitman（1996a））から，$\mathbf{W}$ の分布の仮定より $\mathbf{V}$ は同じ 2 母数 GEM(θ, α) 分布をもつ．従って,（5.1）により次を得る．

命題 6.1. （Yamato et al.（2001））$\mathbf{W} = (W_1, W_2, \dots)$ は 2 母数 GEM(θ, α) 分布に従うものとする．任意の整数 k $(1 \le k \le n)$ と整数 $c_1, \dots, c_k \in \mathcal{N}$ $(c_1 + \cdots + c_k = n)$に対して，

$$P(X_1 = \cdots = X_{c_1}, X_{c_1+1} = \cdots = X_{c_1+c_2}, \dots, X_{c_1+\cdots+c_{k-1}+1} = \cdots = X_{c_1+\cdots+c_k}) \\ = \frac{\theta^{[k:\alpha]}}{\theta^{[n]}} \prod_{i=1}^{k} (1-\alpha)^{[c_i-1]}\,. \tag{6.8}$$

この確率は(2.3)に等しい． 従って，2 節と同様の方法で，2 母数 GEM $(\theta,\, \alpha)$分布からの標本について$(C_{n1}, \dots, C_{nk})$の確率が(2.4)で与えられる事が分かる．

7. ランダム・ウオークとベッセル過程

7.1 ランダム・ウオーク

確率変数列 $Z_1, Z_2, \dots$ は独立で同一分布に従い，半軸上には集中していないものとする．ここでは,

$$W_0 = 0\,, \quad W_n = Z_1 + \cdots + Z_n \ \ (n = 1, 2, \dots)$$

で表されるランダム・ウオークを考える．第 1 はしご点(first ladder epoch)N_1 とは

$$\{N_1 = i\} = \{W_1 \le 0, \dots, W_{i-1} \le 0, W_i > 0\}\,.$$

で定義され，第 1 象限の中に初めて入る時点の事である．第 1 はしご点 N_1 以降のランダム・ウオークの動きは，確率的にランダム・ウオーク全体に同値である．従って，第 2 はしご点 N_2 については，$N_1 = i$ の条件の下で,

$$\{N_2 = j\} = \{Z_{i+1} \le 0, \dots, Z_{i+1} + \cdots + Z_{i+j-1} \le 0, Z_{i+1} + \cdots + Z_{i+j} > 0\}\,.$$

以下同様にして，はしご点の列 $0 < N_1 < N_2 < N_3 < \cdots$ が考えられる．N_1 の母関数 $E[x^{N_1}] = \tau(x)$ は

$$\log \frac{1}{1-\tau(x)} = \sum_{j=1}^{\infty} \frac{x^j}{j} P(W_n > 0)\,. \tag{7.1}$$

を満たす(Feller（1971）, Chapter XII, 7)．

命題 7.1. （Pitman（1999））ある $0 < \alpha < 1$ があって，全ての n に対して，$P(W_n > 0) = \alpha$ と仮定する．K_n を時点 n 迄のはしご点の総数とし，A_n を n がはしご点である事象とする．

このとき，$1 \leq k \leq n$ に対して，

$$P(K_n = k,\ A_n) = \frac{k!}{n!} c(n, k; \alpha)$$

$$P(A_n) = \frac{\alpha^{[n]}}{n!} \tag{7.2}$$

従って，これより

$$P(K_n = k \mid A_n) = \frac{k!}{\alpha^{[n]}} c(n, k; \alpha)\,. \tag{7.3}$$

この確率(7.3)は 確率(3.4)，ただし $\theta = \alpha$，に等しい．ある $0 < \alpha < 1$ があって全ての n に対して $P(W_n > 0) = \alpha$ と言う条件は，原点の周りで対称な確率変数 Z_i により $\alpha = 1/2$ で満足される．また， Z_i がコーシー分布 $C(\mu, 1)$ に従うならば，μ の選び方により条件は $0 < \alpha = P(Z > -\mu) < 1$ で満たされる．

このランダム・ウオークで，はしご点列の間の区間を考える事にし，

$$\widetilde{N}_1 = N_1, \widetilde{N}_{i+1} = N_{i+1} - N_i\ (i = 1, \ldots, k-1)$$

とおく．このとき,(前半で述べた理由により)，$\widetilde{N}_1, \widetilde{N}_2, \ldots, \widetilde{N}_k$ は独立で，N_1 と同じ分布をもつ．その母関数は,(7.1)から $E(x^{N_1}) = 1 - (1-x)^\alpha$ となる (Pitman (1999))． したがって，その確率関数は $P(N_1 = j) = (-1)^{j-1}\binom{\alpha}{j}$ となり，次を得る．

$$P(\widetilde{N}_1 = j_1, \widetilde{N}_2 = j_2, \ldots, \widetilde{N}_k = j_k) = (-1)^{n-k} \prod_{j=1}^{k} \binom{\alpha}{j_i}, \quad j_1 + \cdots + j_k = n\,. \tag{7.4}$$

(7.2)と(7.4)から,

$$P(\widetilde{N}_1 = j_1, \widetilde{N}_2 = j_2, \ldots, \widetilde{N}_k = j_k \mid A_n) = \frac{n!}{\alpha^{[n]}} (-1)^{n-k} \prod_{j=1}^{k} \binom{\alpha}{j_i}, \quad j_1 + \cdots + j_k = n\,. \tag{7.5}$$

ここで，次のようにおく．

$$M_i = \mid \{l : \widetilde{N}_l = i, l = 1, \ldots, k\} \mid, \quad i = 1, \ldots, n\,.$$

すなわち，M_i ははしご間の長さが i のものの個数である．全部で k 個のはしご点があるとき，はしご間の長さについて，$(M_1, \ldots, M_n) = (m_1, \ldots, m_n)$， $(m_1, \ldots, m_n) \in \mathcal{M}_n$ を与えるはしご間の総数は $k!/[m_1! \cdots m_n!]$ である． したがって，この数を (7.5)に乗じて $(m_1, \ldots, m_n) \in \mathcal{M}_n$ についての確率を得る．

命題 7.2. ランダム・ウオークのはしご間の長さについて

$$P\big((M_1, \ldots, M_n) = (m_1, \ldots, m_n) \mid A_n\big) = \frac{(-1)^{n-k} k! n!}{\alpha^{[n]}} \prod_{j=1}^{n} \frac{1}{m_j!} \binom{\alpha}{j}^{m_j},$$

が成り立つ．これは $\theta = \alpha$ の場合のピットマン確率分割 (1.2) に等しい．

7.2 ベッセル過程とベッセルブリッヂ

Pitman (1997)は，ベッセル過程(Bessel process)とベッセルブリッヂ(Bessel bridge)を用いて，母数が $\theta = 0$ と $\theta = \alpha$ の場合に，ピットマン確率分割を導いている．これを紹介する．

平面上の連続な道を持つ確率過程が横軸と交わる，あるいは接する，点全体に対して，その補集合の各区間を excursion 区間と言う．ここでは，区間 $[0,1]$ 上の確率過程 $(B_t, 0 \le t \le 1)$ を考え，$B_0 = 0$ とする．

命題 7.3. (Pitman (1997)) $U_1, U_2, \dots, U_n$ は連続な道を持つ確率過程 B に独立で，区間 $[0,1]$ 上の一様分布に従うものとする．各 $1 \le \forall j \le n$ について，n 個の点 $U_1, U_2, \dots, U_n$ の中の丁度 j 個を含む B の *excursion* 区間の個数を M_j で表す．$(m_1, \dots, m_n) \in \mathcal{M}_n$ とする．このとき，次元 $2-2\alpha(0<\alpha<1)$ のベッセル過程 B に対して，

$$P\big((M_1, \dots, M_n) = (m_1, \dots, m_n)\big) = n(k-1)!\alpha^{k-1} \prod_{j=1}^{n} \left(\frac{(1-\alpha)^{[j-1]}}{j!} \right)^{m_j} \frac{1}{m_j!},$$

すなわち，(1.2)で $\theta = 0$ の場合に当たる．また，次元 $2-2\alpha$ のベッセルブリッヂ B に対しては，

$$P\big((M_1, \dots, M_n) = (m_1, \dots, m_n)\big) = \frac{n!\alpha^k k!}{\alpha^{[n]}} \prod_{j=1}^{n} \left(\frac{(1-\alpha)^{[j-1]}}{j!} \right)^{m_j} \frac{1}{m_j!},$$

すなわち，(1.2)で $\theta = \alpha$ の場合に当たる．

今，$(\beta_t, t \ge 0)$ をブラウン運動とする．任意の $\delta \ge 0$ と $x \ge 0$ に対して確率微分方程式

$$Z_t = x + 2\int_0^t \sqrt{Z_s}\, d\beta_s + \delta t$$

の一意な解 $(Z_t, t \ge 0)$ は，x に始まる δ-次元のベッセル過程の 2 乗である．これを $\mathrm{BESQ}^\delta(x)$ で表すと，$\mathrm{BESQ}^\delta(x)$ は $\delta \le 2$ のとき再帰的で，$\delta \le 1$ のとき確率 1 で 0 に到達する．$0 < \delta < 2$ のとき，0 への滞在時間のルベーグ測度は 0 である．$\sqrt{\mathrm{BESQ}^\delta(x^2)}$ は x に始まる次元 δ のベッセル過程と言われる．整数の次元 N については，N 次元ブラウン運動 $(\beta_t, t \ge 0)$ に対して，その原点からの距離である $|\,\beta_t\,|$ が N 次のベッセル過程である．

任意の $\delta > 0$ に対して確率微分方程式

$$X_t = 2\int_0^t \sqrt{X_s}\, d\beta_s + \int_0^t \left(\delta - \frac{2X_s}{1-s} \right) ds$$

の一意な解 $(X_t, t \ge 0)$ は，ベッセルブリッヂの平方である.（参照，舟木(1996)， Revuz and Yor (1999).）

8. 極限分布

8.1 分割の個数

式(1.1)で与えられるイーエンス確率分割については，$M_j\ (j = 1, 2, \dots)$ は母数 θ/j のポアソン分布に収束する（例えば，Arratia and Tavaré (1992), Sibuya (1993)）．また，$K_n/\log n$ は θ へ概収束する（Korwar and Hollander (1973)）．ここでは，ピットマン確率分割についての漸近的性質を紹介する．

式(3.8)で与えられる K_n の r 次の積率を次の様に表す，

$$E_{\alpha,\theta}(K_n^r) = \sum_{j=0}^{r} (-1)^{r-j} \left(1 + \frac{\theta}{\alpha}\right)^{[j]} R\left(r, j, \frac{\theta}{\alpha}\right) \frac{\Gamma(\theta + j\alpha + n)\Gamma(\theta+1)}{\Gamma(\theta+n)\Gamma(\theta+j\alpha+1)}.$$

Γ-関数の性質 $\Gamma(\theta + j\alpha + n)/\Gamma(\theta+n) \approx n^{j\alpha}$ と $R(r, r, \theta/\alpha) = 1$（(3.12) of Carlitz (1980)）を用いる事により，次を得る．

補題 8.1. （Yamato and Sibuya (2000)）

$$\mu'_r = \lim_{n\to\infty} E\left[\left(\frac{K_n}{n^\alpha}\right)^r\right] = \left(1+\frac{\theta}{\alpha}\right)^{[r]}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1+r\alpha)} = \left(\frac{\theta}{\alpha}\right)^{[r]}\frac{\Gamma(\theta)}{\Gamma(\theta+r\alpha)}\,.$$

特別な場合として，

$$\lim_{n\to\infty} E\left[\frac{K_n}{n^\alpha}\right] = \frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta+\alpha)}\,, \tag{8.1}$$

$$\lim_{n\to\infty} \mathrm{Var}\left[\frac{K_n}{n^\alpha}\right] = \frac{\Gamma(\theta+1)}{\alpha^2}\left\{\frac{\theta+\alpha}{\Gamma(\theta+2\alpha)} - \frac{\Gamma(\theta+1)}{[\Gamma(\theta+\alpha)]^2}\right\}.$$

この μ'_r は密度関数が次式で与えられる分布の r 次の積率である；

$$\frac{\Gamma(\theta+1)}{\Gamma(\frac{\theta}{\alpha}+1)} x^{\frac{\theta}{\alpha}} g_\alpha(x)\,, \tag{8.2}$$

ここで， g_α は母数 α のミタッグ・レフラー分布(Mittag-Leffler distribution)の密度関数である．密度関数 g_α は，任意の実数 $p > -1$ に対して，

$$\int_0^\infty x^p g_\alpha(x)dx = \frac{\Gamma(p+1)}{\Gamma(p\alpha+1)}\,, \tag{8.3}$$

を満たす一意な関数である(Pitman (1999), pp. 20–21)．この一意性より，密度関数(8.2)で与えられる分布も，その積率により一意に決定される．今，L を密度関数(8.2)で与えられる分布に従う確率変数とする．特に，$\theta = 0$ の場合には，L の密度関数は g_α で，ミタッグ・レフラー分布である．このとき，補題 8.1 よりモーメント法により次が得られる．

命題 8.1. （Pitman (1996b)） K_n/n^α は $n\to\infty$ のとき，密度関数(8.3)を持つ分布へ法則収束する．すなわち，

$$K_n/n^\alpha \xrightarrow{d} L$$

が成り立つ．

Pitman (1996)は，更に，$n\to\infty$ のとき K_n/n^α が L へ確率 1 で，ならびに任意の $p>0$ に対し p 次の平均の意味で，収束する事を示している．

$\theta = 0$ の場合には，K_n の裾確率について，Bingham et al. (1987), p. 337 により近似的に次が成り立つ．

$$P(K_n > x) \approx \exp\left[-(1-\alpha)\alpha^{\alpha/(1-\alpha)}x^{1/(1-\alpha)}\right], \quad 0<\alpha<1, \quad (x\to\infty)\,.$$

$\theta = 0$ の場合の K_n の上昇階乗積率を(3.6)式で与えたが，Feng and Hoppe (1998)はこれを用いて先ず，

$$\lim_{n\to\infty}\frac{1}{n}\log E[\exp(\lambda K_n^{\alpha,0})] = \Lambda_\alpha(\lambda)$$

を導き，さらに

$$\lim_{n\to\infty}\frac{1}{n}\log E[\exp(\lambda K_n^{\alpha,\theta})] = \Lambda_\alpha(\lambda)$$

を導いている．ここで，

$$\Lambda_\alpha(\lambda) = -\log\left[1-(1-e^{-\lambda})^{1/\alpha}\right], \quad \lambda>0$$

$\lambda \le 0$ に対しては，$\Lambda_\alpha(\lambda) = 0$ である．また，

$$I^\alpha(x) = \sup_\lambda \{\lambda x - \Lambda_\alpha(\lambda)\}$$

とおく．更に，区間 $[0,1]$ 上の右連続で，左極限をもつ実数値関数の全体に位相として一様収束を付随させた空間を $\mathbf{D}[0,1]$ とする．

命題 8.2. （Feng and Hoppe (1998)）母数 $0 < \alpha < 1$ および $\theta > -\alpha$ について，K_n/n の確率分布を ν_n とすると，次の大偏差原理(large deviation principle)が成り立つ：$\mathbf{D}[0,1]$ のボレル集合 A に対して，

$$-\inf_{f\in A^o} I_\alpha(f) \le \liminf_{n\to\infty} \frac{1}{\log n}\nu_n(A) \le \limsup_{n\to\infty} \frac{1}{\log n}\nu_n(A) \le -\inf_{f\in A^-} I_\alpha(f)$$

ここで，A^o と A^- はそれぞれ A の内部と閉包である．

8.2 ピットマン確率分割

ピットマン確率分割の確率(1.2)式について，$\sum_{j=1}^n jm_j = n$ を満たす全ての非負整数 $m_1, m_2, \ldots, m_n$ 上での(1.2)の右辺の総和は 1 に等しい．この性質を用いて，非負整数 $r_1, \ldots, r_n$ に対し次が示される．

$$E\left[\prod_{j=1}^n M_j^{(r_j)}\right] = \frac{\theta^{[r:\alpha]}(\theta + r\alpha)^{[n-s]} n^{(s)}}{\theta^{[n]}} \prod_{j=1}^n \left(\frac{(1-\alpha)^{[j-1]}}{j!}\right)^{r_j}$$
$$= \frac{(\theta+\alpha)^{[r-1:\alpha]}\Gamma(\theta+1)}{\Gamma(\theta+r\alpha)} \prod_{j=1}^n \left(\frac{(1-\alpha)^{[j-1]}}{j!}\right)^{r_j} \frac{\Gamma(\theta+r\alpha+n-s)n^{(s)}}{\Gamma(\theta+n)}. \tag{8.4}$$

ここに，$r = r_1 + \cdots + r_n$，$s = \sum_{j=1}^n jr_j \le n$．この (8.4)から，任意の $j(=1,2,\ldots)$ と $r(=0,1,2,\ldots)$ について，

$$E[M_j^{(r)}] = \frac{\theta^{[r:\alpha]}\Gamma(\theta)}{\Gamma(\theta+r\alpha)} \left(\frac{(1-\alpha)^{[j-1]}}{j!}\right)^r \frac{\Gamma(\theta+r\alpha+n-jr)n^{(jr)}}{\Gamma(\theta+n)}.$$

上式から，次の極限の評価が得られる．

$$\lim_{n\to\infty} E\left[\frac{j!M_j}{\alpha(1-\alpha)^{[j-1]}n^\alpha}\right]^r = \left(1+\frac{\theta}{\alpha}\right)^{[r]} \frac{\Gamma(\theta+1)}{\Gamma(\theta+1+r\alpha)}.$$

特に $r = 1, 2$ の場合には，$j = 1, 2, \ldots$ について，

$$\lim_{n\to\infty} E\left[\frac{M_j}{n^\alpha}\right] = \frac{(1-\alpha)^{[j-1]}\Gamma(\theta+1)}{j!\Gamma(\theta+\alpha)},$$

$$\lim_{n\to\infty} \mathrm{Var}\left[\frac{M_j}{n^\alpha}\right] = \Gamma(\theta+1)\left\{\frac{\theta+\alpha}{\Gamma(\theta+2\alpha)} - \frac{\Gamma(\theta+1)}{[\Gamma(\theta+\alpha)]^2}\right\}\left(\frac{(1-\alpha)^{[j-1]}}{j!}\right)^2.$$

特に，M_1 については，

$$\lim_{n\to\infty} E\left[\frac{M_1}{n^\alpha}\right] = \frac{\Gamma(\theta+1)}{\Gamma(\theta+\alpha)}, \tag{8.5}$$

$$\lim_{n\to\infty} \mathrm{Var}\left[\frac{M_1}{n^\alpha}\right] = \Gamma(\theta+1)\left\{\frac{\theta+\alpha}{\Gamma(\theta+2\alpha)} - \frac{\Gamma(\theta+1)}{[\Gamma(\theta+\alpha)]^2}\right\}.$$

式(8.4)から，正の整数 l と非負の整数 $r_1,\dots,r_l$ に対して，

$$\lim_{n\to\infty} E\left[\prod_{j=1}^{l}\left(\frac{j!M_j}{\alpha(1-\alpha)^{[j-1]}n^{\alpha}}\right)^{r_j}\right] = \mu'_{r_1+\cdots+r_l}\,, \tag{8.6}$$

ここで， μ'_r は補題 8.1 で与えられている．この関係を $j, l(=1,2,\dots,j\neq l)$ に対して， $r_j=1$ および $r_l=1$ として用いて次を得る．

$$\lim_{n\to\infty}\mathrm{Cov}\left(\frac{M_j}{n^{\alpha}},\frac{M_l}{n^{\alpha}}\right) = \Gamma(\theta+1)\left\{\frac{\theta+\alpha}{\Gamma(\theta+2\alpha)} - \frac{\Gamma(\theta+1)}{[\Gamma(\theta+\alpha)]^2}\right\}\frac{(1-\alpha)^{[j-1]}(1-\alpha)^{[l-1]}}{j!l!}\,.$$

μ'_r は L の r 次の積率であるから, (8.6)式から次が得られる．

命題 8.3. （Yamato and Sibuya (2000)）正の整数 l について，

$$\left(\frac{M_1}{\alpha n^{\alpha}},\frac{2!M_2}{\alpha(1-\alpha)^{[1]}n^{\alpha}},\dots,\frac{l!M_l}{\alpha(1-\alpha)^{[l-1]}n^{\alpha}}\right) \xrightarrow{d} (L,L,\dots,L)$$

あるいは

$$\left(\frac{M_1}{n^{\alpha}},\frac{M_2}{n^{\alpha}},\dots,\frac{M_l}{n^{\alpha}}\right) \xrightarrow{d} L\left(\alpha,\frac{\alpha(1-\alpha)^{[1]}}{2!},\dots,\frac{\alpha(1-\alpha)^{[l-1]}}{l!}\right). \tag{8.7}$$

特に，確率分割の中で一意なものの個数 M_1 については,

$$\frac{M_1}{n^{\alpha}} \xrightarrow{d} \alpha L\,.$$

式 (8.7)の右辺に現れる項については， 系列

$$\left(\alpha,\frac{\alpha(1-\alpha)^{[1]}}{2!},\dots,\frac{\alpha(1-\alpha)^{[l-1]}}{l!},\dots\right)$$

は $\mathcal{N}$ 上の渋谷分布(Devroye (1993))である．すなわち,

命題 8.4. $(M_1/n^{\alpha}, M_2/n^{\alpha},\dots)/L$ の分布は，漸近的に渋谷分布である．

最後に，ピットマン確率分割を Khmaladze (1987)による LNRE (large number of rare events) の観点から見てみよう． M_1 と K_n へ Khmaladze (1987)による LNRE の定義を適用すると，(8.1)と(8.5)とにより次が得られる．

$$\lim_{n\to\infty}\frac{EM_1}{EK_n} = \alpha > 0\,,\quad \lim_{n\to\infty} EK_n = \infty\,.$$

M_j については,

$$\lim_{n\to\infty}\frac{EM_j}{EK_n} = \frac{\alpha(1-\alpha)^{[j-1]}}{j!}(>0)\,.$$

従って,

命題 8.5. 母数 $0<\alpha<1$ と $\theta>-\alpha$ を持つピットマン確率分割は，Khmaladze (1987)の Definition 2 の意味で， LNRE 系列である．

一方，$\alpha = 0$ の場合であるイーエンス確率分割は LNRE 系列ではない．この場合には，$EM_1 = \alpha n/(\alpha + n - 1)$ および $EK_n \sim 1 + \alpha \log n$ であるからである．

9. ポアソン・ディリクレ過程

$(\mathcal{X}, \mathcal{B})$ を可測空間とし，$\mathbf{B}$ をその上のランダムな確率測度とする．

定義 9.1. （Pitman (1996a), Carlton (1999)）μ は $(\mathcal{X}, \mathcal{B})$ 上の確率測度とし，$\mathcal{X}$ 上の値を取る確率変数列 $Y_1, Y_2, \ldots$ は独立で同一の確率測度 μ に従うものとする．$0 \le \alpha < 1$ および $\theta > -\alpha$ とし，ランダムな比率 $(P_1, P_2, \ldots)$ は(4.1)で与えられるものとする． このとき，ランダムな確率測度

$$\mathbf{B} = \sum_{j=1}^{\infty} P_j \delta_{Y_j} \tag{9.1}$$

を母数 μ, α, θ をもつ，$(\mathcal{X}, \mathcal{B})$ 上のポアソン・ディリクレ過程という．略して，$\mathbf{B} \in \mathcal{PD}(\mu; \alpha, \theta)$ と表す．ここで，測度 δ_x は 1 点 x にのみ重み 1 を与える．

上の定義では，2 母数 GEM 分布に従うランダムな比率 $(P_1, P_2, \ldots)$ を用いているが，2 母数ポアソン・ディリクレ分布を用いても出来る．$\alpha = 0$ の場合， Ferguson (1973)によるディリクレ過程(Dirichlet process)になる．non-atomic な事前分布に対する事後分布が Pitman (1996a) に述べられており，Carlton (1999)は証明と共に述べている． 確率測度 μ が non-atomic とは，全ての $x \in \mathcal{X}$ に対して $\mu(x) = 0$ が成り立つ事である．

命題 9.1. （Carlton (1999)）μ は non-atomic とし，$\mathbf{B} \in \mathcal{PD}(\mu; \alpha, \theta)$ とする．$X_1, \ldots, X_n$ を $\mathbf{B}$ からの大きさ n の標本とする．K_n は 異なる X_i の個数で，$\widetilde{X_j}$ は X_i の中で j 番目の異なる値とし c_j は X_i の中で $\widetilde{X_j}$ に等しいものの数とする． このとき，

$$\mathbf{B} \mid X_1, \ldots, X_n \stackrel{d}{=} \sum_{j=1}^{K_n} W_j \delta_{\widetilde{X_j}} + W_{K_n+1} \mathbf{B}_{K_n}\,. \tag{9.2}$$

ただし，$\mathbf{B}_{K_n} \in \mathcal{PD}(\mu; \alpha, \theta + K_n \alpha)$ で，$(W_1, \ldots, W_{K_n+1})$ はディリクレ分布 $D(c_1 - \alpha, \ldots, c_k - \alpha, \theta + K_n \alpha)$ に従い，これらは独立である．

従って，ポアソン・ディリクレ過程は，ディリクレ過程とは異なり，ノンパラメトリックな事前分布として，conjugate ではない．

さて，$\mathcal{X} = \mathcal{R}$ とし，$\mathbf{B} \in \mathcal{PD}(\mu; 0, \theta)$，すなわち，$\mathbf{B}$ は $\mathcal{R}$ 上のディリクレ過程に従うものとする．ディリクレ過程に関する汎関数の確率的性質について，紹介する．もし μ がコーシー分布ならば，$\mathbf{B}$ の平均 $\int_{\mathcal{X}} x \mathbf{B}(dx)$ も同じコーシー分布をもつ(Yamato (1984))．（これは，$\mathcal{R}$ 上のポアソン・ディリクレ過程についても言える．）Diaconis and Kemperman (1996)は，$\theta = 1$ の場合に，$X = \int_{\mathcal{X}} f(x) \mathbf{B}(dx)$ について，

$$E\left\{\frac{1}{1 - tX}\right\} = e^{-\int_{\mathcal{X}} \log(1 - t f(x)) \mu(dx)}$$

なる関係を示している．

今， $\mathcal{X}=[0,1]$ とし，平均 $\int_0^1 x\mathbf{B}(dx)$ の確率測度を ν とする． このとき，上の関係は

$$\int_0^1 \frac{1}{z-u}\nu(du)=\exp\int_0^1 \log\frac{1}{z-u}\mu(dx),\quad z\in\mathbb{C}\setminus\mathbb{R}$$

へと帰着される．ここで, $\mathbb{C}$, $\mathbb{R}$ はそれぞれ複素数全体, 実数全体である．$\alpha,\theta\neq 0$ とし, $\mathcal{X}=[0,1]$ 上のポアソン・ディリクレ過程 $\mathbf{B}\in\mathcal{PD}(\mu;\alpha,\theta)$ に対して，Tsilevich (1999)は関係

$$\left(\int_0^1 (z-u)^{-\theta}\nu(du)\right)^{-\frac{1}{\theta}}=\left(\exp\int_0^1 (z-u)^{\alpha}\mu(dx)\right)^{\frac{1}{\alpha}},\quad z\in\mathbb{C}\setminus\mathbb{R}$$

を示している．

謝 辞

本稿は，日本学術振興会科学研究費・基盤研究(A)(課題番号 14208023，研究代表者：竹村彰通，『個票データの秘匿措置と開示データの利用に関する研究』)の研究の一環として，関連する事項を纏めたものである．査読の方々には適切なご指摘を頂き，感謝致しております．

参 考 文 献

Antoniak, C. A. (1974). Mixtures of Dirichlet processes with applications to Bayesian non-parametric problems, *Ann. Statist.*, **2**, 1152–1174.

Arratia, R. and Tavaré, S. (1992). The cycle structure of random permutations, *Ann. Probab.*, **20**, 1567–1591.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Encyclopedia of Mathematics and Its Applications*, Regular Variation, Cambridge University Press, Cambridge.

Carlitz, L. (1980). Weighted Stirling numbers of the first and second kind–I, *Fibonacci Quart.*, **18**, 147–162.

Carlton, M. C. (1999). Applications of the two-parameter Poisson-Dirichlet distribution, A dissertation of Ph. D., Department of Statistics, University of California, Los Angeles.

Charalambides, Ch. A. and Singh, J. (1988). A review of the Stirling numbers, their generalizations and statistical applications, *Comm. Statist. Theory Methods*, **17**, 2533–2595.

Diaconis, P. and Kemperman, J. (1996). Some new tools for Dirichlet priors, *Bayesian Statistics* 5 (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), 97–106, Clarendon Press, Oxford.

Devroye, L. (1993). A triptych of discrete distributions related to the stable law, *Statist. Probab. Lett.*, **18**, 349–351.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles, *Theoret. Popn. Biol.*, **3**, 87–112.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed., Wiley, New York.

Feng, S. and Hoppe, F. (1998). Large deviation principles for some random combinatorial structures in population genetics and Brownian motion, *Ann. Appl. Probab.*, **8**, 975–994.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *Ann. Statist.*, **1**, 209–230.

舟木直久(1996).『確率微分方程式』，岩波講座(現代数学の基礎)，岩波書店，東京.

Gnedin, A. V. (1998). On convergence and extensions of size-biased permutation, *J. Appl. Probab.*, **35**, 642–650.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, Wiley, New York.

Khmaladze, E. V. (1987). The statistical analysis of large number of rare events, Tech. Report, MS-R8804, Centre for Mathematics and Computer Science, CWI, Amsterdam.

Kingman, J. F. C. (1975). Random discrete distributions, *J. Roy. Statist. Soc. Ser. B*, **37**, 1–15.

Korwar, R. M. and Hollander, M. (1973). Contribution to the theory of Dirichlet process, *Ann. Probab.*, **1**, 705–711.

Patil, G. P. and Taillie, C. (1977). Diversity as a concept and its implications for random communities, *Bull. Internat. Statist. Inst.*, **47**, 497–515.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions, *Probab. Theory Related Fields*, **12**, 145–158.

Pitman, J. (1996a). Some developments of the Blackwell-MacQueen urn scheme, *Statistics, Probability and Game Theory*; *Papers in honor of David Blackwell* (eds. T. S. Ferguson, L. S. Shapley and J. B. MacQueen), 245–267, Institute of Mathematical Statistics, Hayward, California.

Pitman, J. (1996b). Notes on the two parameter generalization of Ewens' random partition structure (unpublished manuscript).

Pitman, J. (1996c). Random discrete distributions invariant under size-biased permutation, *Adv. Appl. Prob.*, **28**, 525–539.

Pitman, J. (1997). Partition structures derived from Brownian motion and stable subordinators, *Bernoulli*, **3**, 79–96.

Pitman, J. (1999). Brownian motion, bridge, excursion and meander characterized by sampling at independent uniform times, *Electronic J. Probability*, **4**, Paper 11, 1–33.

Pitman, J. (2002). Combinatorial stochastic processes, Tech. Report, No. 261, Department of Statistics, University of California, Berkeley.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator, *Ann. Probab.*, **25**, 855–900.

Revuz, J. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed., Wiley, New York.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.

Sibuya, M. (1993). A random clustering process, *Ann. Inst. Statist. Math.*, **45**, 459–465.

渋谷政昭(2003). 孤立個体数の推測，統計数理，**51**(2), 261–295.

Sibuya, M. and Yamato, H. (1995). Ordered and unordered random partitions of an integer and the GEM distribution, *Statist. Probab. Lett.*, **25**, 177–183.

Sibuya, M. and Yamato, H. (2001). Pitman's model of random partitions, 数理解析研究所講究録, **1240**, 64–73.

竹村彰通(2003). 個票開示問題の研究の現状と課題，統計数理，**51**(2)，241-260.

Tsilevich, N. V. (1999). Distribution of the mean value for certain random measures, *J. Math. Sci.*, **96**, 3616–3623

Yamato, H. (1984). Characteristic functions of means of distributions chosen from a Dirichlet process, *Ann. Probab.*, **12**, 262–267.

Yamato, H. and Sibuya, M. (2000). Moments of some statistics of Pitman sampling formula, *Bull. Inform. Cybernet.*, **32**, 1–10.

Yamato, H., Sibuya, M. and Nomachi, T. (2001). Ordered sample from two-parameter GEM distribution, *Statist. Probab. Lett.*, **55**, 19–27.

372 Proceedings of the Institute of Statistical Mathematics Vol. 51, No. 2, 351–372 (2003)

Pitman Sampling Formula and Related Topics

Hajime Yamato
(Faculty of Science, Kagoshima University)

Masaaki Sibuya
(Faculty of Management, Takachiho University)

Associated with the disclosure of a microdata set, there is a problem of estimating the number of unique individuals in a population. As a stochastic model of random partition of a finite population related to this problem, Ewens sampling formula (Ewens (1972)) is well-known. This formula is exchangeable and invariant under size-biased permutation. As a random partition having these two properties, Pitman (1995, 1996c) derived the Pitman sampling formula. We introduce this formula by using a simple urn model, and review its properties and related topics. For the disclosure of a microdata set and random partition, see Takemura (2003) and Sibuya (2003) of this special issue.

Key words: Ewens sampling formula, exchangeability, Mittag-Leffler distribution, Pitman sampling formula, Poisson-Dirichlet distribution.