分割，超幾何系，Dirichlet過程と統計的推測

間野　修平　(統計数理研)*

1. 導入

定義 1 (Ferguson 1973 [1]) $\mathbb{R}$ 上の Borel 集合族を $\mathcal{B}(\mathbb{R})$ と記す．可測空間 $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ のランダム確率測度 F が $\mathbb{R}$ の任意の有限可測分割 $\{A_1,...,A_k\}$ に対し

$$(F(A_1),...,F(A_k)) \sim \mathrm{Dir}(\rho(A_1),...,\rho(A_k))$$

を満たすとき，F を基底測度 ρ の Dirichlet 過程という．

本稿では，確率測度 μ と $\theta>0$ について $\rho=\theta\mu$ とした Dirichlet 過程を $\mathrm{DP}(\theta;\mu)$ と記す．また，簡単のため，μ は non-atomic とする．$F\sim\mathrm{DP}(\theta;\mu)$ のとき，Dirichlet 分布の多項抽出における共役性から

$$\begin{aligned}(F_n(A_1),...,F_n(A_k)) &:= (F(A_1),...,F(A_k))|(X_1,...,X_n)\\ &\sim \mathrm{Dir}\left(\theta\mu(A_1)+\sum_{i=1}^n\delta_{X_i}(A_1),...,\theta\mu(A_k)+\sum_{i=1}^n\delta_{X_i}(A_k)\right)\end{aligned}$$

が従うので，

$$F_n\sim\mathrm{DP}\left(\theta+n,\frac{\theta\mu+\sum_{i=1}^n\delta_{X_i}}{\theta+n}\right)$$

である．$\mathbb{P}(X_1\in\cdot)=\mathbb{E}\{\mathbb{P}(X_1\in\cdot|F)\}=\mathbb{E}(F(\cdot))=\mu(\cdot)$,

$$\mathbb{P}(X_{n+1}\in\cdot|X_1,...,X_n)=\frac{\theta}{\theta+n}\mu(\cdot)+\frac{n}{\theta+n}\Lambda_n(X_1,...,X_n)(\cdot) \tag{1}$$

である．ここで，$\Lambda_n(X_1,...,X_n):=n^{-1}\sum_{i=1}^n\delta_{X_i}$ は経験分布であり，Bayes 統計の文脈では，F を事前過程，(1) を予測規則という．j 番目に初めて現れる値を X_j^*, $j\in[k]:=\{1,2,...,k\}$ とすると，$(n_1,...,n_k)$, $n_j:=\{i;X_i=X_j^*\}$ は n の正の整数による分割をなす．

$$\mathbb{P}(N_1=n_1,...,N_k=n_k)=\frac{\theta^k}{(\theta)_n}\prod_{j=1}^k(n_j-1)!$$

であるが，$(n_1,...,n_k)$ の置換について対称で，$(c_1,...,c_n)$, $c_i:=\#\{j;n_j=i\}$ の分布は

$$\mathbb{P}(C_1=c_1,...,C_n=c_n)=\frac{n!}{(\theta)_n}\prod_{i=1}^n\left(\frac{\theta}{i}\right)^{c_i}\frac{1}{c_i!} \tag{2}$$

である．ただし，$(\theta)_n:=\theta(\theta+1)\cdots(\theta+n-1)$ とした．分割の確率函数 (2) は Ewens 抽出公式とよばれる．Bayes 統計の文脈では，(2) は事前過程 $\mathrm{DP}(\theta;\mu)$ からの多項抽出における母数 θ の周辺尤度である．

2010 Mathematics Subject Classification: 60C05, 60K35, 33C45, 05A17, 62F15, 13P25
キーワード：組み合わせ確率過程, 代数統計, 超幾何系, 分割, ベイズ統計, Dirichlet 過程
* 〒190-8562 東京都立川市緑町 10-3　統計数理研究所 数理・推論研究系
e-mail: smano@ism.ac.jp

コンパクト距離空間 E 上の Borel 確率測度の集合を $\mathcal{P}(E)$，$\mu \in \mathcal{P}(E)$ と函数 f について $\langle f, \mu \rangle = \int f d\mu$ と記す．任意の自然数 m について $F \in C^2(\mathbb{R}^m)$ とし，

$$\phi(\mu) = F(\langle f_1, \mu \rangle, ..., \langle f_m, \mu \rangle)$$

とする．Fleming-Viot 過程とよばれる測度値拡散過程の生成作用素 G は $\phi(\mu)$ に

$$\begin{aligned} G\phi(\mu) =& \frac{1}{2} \sum_{i,j=1}^{m} (\langle f_i f_j, \mu \rangle - \langle f_i, \mu \rangle \langle f_j, \mu \rangle) F_{,ij}(\langle f_1, \mu \rangle, \cdots, \langle f_k, \mu \rangle) \\ &+ \sum_{i=1}^{m} \langle B f_i, \mu \rangle F_{,i}(\langle f_1, \mu \rangle, \cdots, \langle f_m, \mu \rangle) \end{aligned} \tag{3}$$

のように作用する．ただし，B は浮動係数を定める作用素である．$E = \{1/m, 2/m, ..., 1\}$,

$$Bf_i = \frac{\alpha}{2} \sum_{j=1}^{m} (f_j - f_i)$$

のとき，$\mu(t) = \sum_{i=1}^{m} x_i(t) \delta_{i/m}$，$f_j(i/m) = \delta_{i,j}$ とすれば，(3) より $(x(t); t \geq 0)$ は生成作用素

$$L = \sum_{i,j=1}^{m} \frac{x_i(\delta_{i,j} - x_j)}{2} \frac{\partial^2}{\partial x_i \partial x_j} + \frac{\alpha}{2} \sum_{i=1}^{m} \left\{ \sum_{j=1}^{m} (x_j - x_i) \right\} \frac{\partial}{\partial x_i} \tag{4}$$

に従う単体 $\{x_i; \sum_{i=1}^{m} x_i = 1\}$ の上の拡散過程であり，Wright-Fisher 拡散とよばれる．

一般に，作用素 L で生成される拡散過程の定常測度 π は

$$0 = \frac{d}{dt} \langle f, \pi \rangle = \langle Lf, \pi \rangle = \langle f, L^+ \pi \rangle, \qquad \forall f \tag{5}$$

を満たす．ただし，L^+ は L の随伴作用素である．さらに可逆測度であれば,

$$\langle Lf, g\pi \rangle = \langle Lg, f\pi \rangle, \qquad \forall f, \, g$$

を満たす．これは $L_j^+ \pi = 0, \forall j$ に等価である．ただし，

$$L_j^+ \bullet \phi = \sum_i \frac{1}{2} \frac{\partial}{\partial x_j} (a_{ij}(x) \phi) - b_j(x) \phi, \qquad L^+ = \sum_j \frac{\partial}{\partial x_j} \bullet L_j^+$$

であり，$a_{ij}(x)$, $b_j(x)$ は拡散係数，浮動係数である．分割の確率函数は多項抽出

$$(n_1, ..., n_m) \sim \mathrm{Multinom}(n; x_1, ..., x_m)$$

の定常測度 π による混合 $p(n) = \langle q_n, \pi \rangle$ である．ここで，

$$q_n(x) := \frac{n!}{n_1! \cdots n_m!} x^n, \qquad x^n := \prod_{i=1}^{m} {x_i}^{n_i}$$

とした．生成作用素 (4) の下での定常性の条件 (5) より，漸化式

$$\begin{aligned} p(n) =& \frac{n-1}{m\alpha + n - 1} \sum_{i=1}^{m} \frac{n_i - 1}{n - 1} p(n - e_i) \\ &+ \frac{\alpha}{m\alpha + n - 1} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{n_j + 1 - \delta_{ij}}{n} p(n - e_i + e_j) \end{aligned} \tag{6}$$

が従う．ただし，境界条件 $p(e_i)=1/m,\ i\in[m]$ である．漸化式 (6) の解は，π_α を母数 α の m 変量対称 Dirichlet 分布 $\mathrm{Dir}(\alpha)$ の密度として

$$\langle q_n,\pi_\alpha\rangle=\begin{pmatrix}-m\alpha\\ n\end{pmatrix}^{-1}\prod_{i=1}^{m}\begin{pmatrix}-\alpha\\ n_i\end{pmatrix} \tag{7}$$

である．π_α は Wright-Fisher 拡散 (4) の可逆測度であり，確率函数 (7) は Dirichlet 多項分布である．Fleming-Viot 過程の可逆測度は $\sum_{i=1}^m x_i\delta_{i/m}, x\sim\mathrm{Dir}(\alpha)$ である．同様に，$E=[0,1]$,

$$Bf(x)=\frac{\theta}{2}\int_0^1\{f(y)-f(x)\}dy$$

の Fleming-Viot 過程の可逆測度は Dirichlet 過程である．確率函数 (7) において $m\to\infty$, $\alpha\to 0,\ \theta\equiv m\alpha>0$ とすれば，Ewens 抽出公式 (2) が得られる．

本稿では，上記の枠組みを様々に拡張したモデルの統計的推測に現れる次の 3 つの問題を議論する．

1. 分割の確率函数が与えられたとき，それに従う標本を生成する方法 (3.1 節)
2. 分割の確率函数が与えられたとき，母数を推定する方法 (3.2 節)
3. Fleming-Viot 過程が与えられ，定常測度から従う分割の確率函数は明示的に与えられていないとき，母数を推定する方法 (4 節)

ここで，母数を推定するとは，分割の確率函数を最大化する母数を求めることであり，最尤推定，もしくは，分割の確率函数が周辺尤度であることから，経験 Bayes とよばれる．3 節は対数アフィンモデルからの多項抽出の条件付き分布に関する一般論であり，1,2 の問題は例として扱う．その議論のために，2 節で Gel'fand らの A 超幾何系に関する事項を準備する．3 節は [2]，4 節は [3] の紹介であり，詳細，背景と周辺の話題は [4] に解説されている．本稿の引用は網羅的ではないので，参考文献については [4] を参照されたい．

2. A 超幾何系とホロノミック勾配法

定義 2 (Gel'fand et al. 1990 [5]) ランク d の非負整数値 $d\times m$ 行列 A とベクトル $b\in\mathbb{C}^d$ について，次の消去作用素たちが定める線形偏微分方程式系を A 超幾何系 $H_A(b)$ とよぶ．

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

$H_A(b)$ は Weyl 代数の左イデアルで，A 超幾何イデアルとよばれる．

定義 3 A 超幾何系 $H_A(b)$ の原点周りの級数解

$$Z_A(b;x):=\sum_{\{c;Ac=b,c\in\mathbb{N}^m\}}\frac{x^c}{c!},\qquad c!:=\prod_{i=1}^m c_i! \tag{8}$$

を A 超幾何級数とよぶ．ただし，$b\notin A\mathbb{N}_0^m$, $\mathbb{N}_0:=0\cup\mathbb{N}$, $\mathbb{N}:=\{1,2,...\}$ のときは，$Z_A(b;x)=0$ と規約する．

本稿では，Aを同次行列，つまりAの行空間が$(1,...,1)$を含むとする．2行の同次行列Aは一般性を失うことなく

$$A = \begin{pmatrix} 0 & i_1 & i_2 & \cdots & i_{m-1} \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \tag{9}$$

と表せる．ここで，$0 < i_1 < i_2 < \cdots < i_{m-1}$は互いに素な整数である．$i_{m-1} = m-1$, $b_1, b_2 \in \mathbb{N}$のとき，条件$Ac = b$は$b_1 + b_2$のb_2個の正の整数による分割を定める．ここで，b_2を分割の長さという．

A超幾何多項式(8)の数値評価を考える．定義どおりに足し上げることは，条件$Ac = b$を満たすcが膨大にあるので現実的ではない．そのような足し上げを避けてホロノミック関数を評価する方法に差分ホロノミック勾配法(Holonomic Gradient Method; HGM)[6]がある．一般に，HGMはホロノミックイデアルIの偏微分方程式系

$$\theta_i \bullet Q = P_i Q, \qquad i \in [\mathrm{rank}(I)], \tag{10}$$

を用いる．ここで，Qは標準単項式のベクトルである．このような偏微分方程式系をPfaffian系とよぶ．Pfaffian系は，原理的にはGröbner基底に対する標準形として求められるが，そのような計算は階数が増えるに従い大きなコストを要するので現実的ではない．そこで，陽に求めるか，効率的な計算方法を工夫する必要がある．

A超幾何イデアル$H_A(b)$に差分HGMを適用することを考える．A超幾何函数は隣接関係

$$\theta_i Z_A(b; x) = x_i Z_A(b - a_i; x), \qquad i \in [m] \tag{11}$$

を満たす．ここで，a_iは行列Aの第i列ベクトルである．2行のA超幾何系については

$$Q(b; x) = (1, \theta_3, ..., \theta_{n-k+1})^\top \bullet Z_A(b; x)$$

であり，多項式解が満たすPfaffian系を陽に得ることができる．Pfaffian系(10)と隣接関係(11)を合わせることで，Qの漸化式

$$x_i Q(b - a_i; x) = P_i(b; x) Q(b; x), \qquad i = 1, 2$$

が得られる．これを用いることで，A超幾何多項式を数値評価するための効率的なアルゴリズムを構成することができる([2]のAlgorithm 6.3).

3. A超幾何分布と統計的推測

Takayamaら[7]により定義されたA超幾何分布は，対数アフィンモデルからの交換可能な標本の条件付き分布であり，計数データの統計モデルとして現れる．大きさnの標本を考える．セルたちを$[m]$で表し，$t_i \in [m]$, $i \in [n]$をi番目の標本点が属するセルとする．標本は計数ベクトル$(c_1, ..., c_m)$, $c_i := \#\{j; t_j = i\}$で表せるが，

$$(c_1, ..., c_m) \sim \mathrm{Multinom}(n; p_1(\xi), ..., p_m(\xi))$$

とし，対数アフィンモデル$\log p_i(\xi) = \sum_{j=1}^{d+g} a_{ji}\xi_j - \phi_i(\xi)$ を仮定すれば，

$$\mathbb{P}(C = c) = \frac{n!}{c!} \exp\left\{ \sum_{j=1}^{d+g} \xi_j b_j - \sum_{i=1}^{m} \phi_i(\xi) \right\}, \qquad c! := \prod_{i=1}^{m} c_i!$$

となる．ここで，$(b_1, ..., b_{d+g})$ は $b_j = \sum_{i=1}^m a_{ji}c_i$, $j \in [d+g]$ で定まる十分統計量である．ここで，$j \in [d]$ の条件は $d \times m$ 行列 $A = (a_{ji})$ と数ベクトル $b \in \mathbb{N}_0^d$ により $Ac = b$ と表せるが，その条件付き分布は

$$\mathbb{P}(C = c | AC = b) \propto \frac{1}{c!} \exp\left\{ \sum_{j=d+1}^{d+g} \xi_j b_j \right\} = \frac{x^c}{c!}, \qquad \log x_i := \sum_{j=d+1}^{d+g} a_{ji}\xi_j \tag{12}$$

である．正規化定数が A 超幾何多項式になることから，$\{c; Ac = b\}$ を台とし，確率函数 (12) が定める分布を A 超幾何分布とよぶ．行列 A の同次性から標本の大きさの条件

$$\sum_{i=1}^m c_i = n \tag{13}$$

が従う．

例 1 (2 元分割表) 2×2 の分割表

$$\begin{array}{cc|c} n_{11} & n_{12} & n_{1\cdot} \\ n_{21} & n_{22} & n_{2\cdot} \\ \hline n_{\cdot 1} & n_{\cdot 2} & n_{\cdot\cdot} \end{array}$$

の標準的なモデルは $(N_{11}, N_{12}, N_{21}, N_{22}) \sim \mathrm{Multinom}(p_{11}, p_{12}, p_{21}, p_{22})$ である．

$$\psi_1 := \log \frac{p_{12}}{p_{22}}, \qquad \psi_2 := \log \frac{p_{21}}{p_{22}}, \qquad \lambda := \log \frac{p_{11}p_{22}}{p_{12}p_{21}}$$

とすれば，

$$\begin{aligned} \mathbb{P}(N = n) =& \frac{n_{\cdot\cdot}!}{n_{11}!n_{12}!n_{21}!n_{22}!} \\ & \times \exp\left\{ n_{11}\lambda + n_{1\cdot}\psi_1 + n_{\cdot 1}\psi_2 - n_{\cdot\cdot} \log\left(1 + e^{\psi_1} + e^{\psi_2} + e^{\lambda+\psi_1+\psi_2}\right) \right\} \end{aligned}$$

が得られる．通常はオッズ比 $y := e^\lambda$ に興味があるので，ψ_1 と ψ_2 を局外母数とする．周辺和 $n_{1\cdot}$ と $n_{\cdot 1}$ は局外母数の十分統計量だから，条件付き分布は

$$\mathbb{P}(N_{11} = n_{11} | N_{\cdot\cdot} = n_{\cdot\cdot}, N_{1\cdot} = n_{1\cdot}, N_{\cdot 1} = n_{\cdot 1}) = \frac{1}{Z_A(b; x)} \frac{y^{n_{11}}}{n_{11}!n_{12}!n_{21}!n_{22}!} \tag{14}$$

である．この確率函数は A 超幾何分布で，

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \qquad c = \begin{pmatrix} n_{11} \\ n_{12} \\ n_{21} \\ n_{22} \end{pmatrix}, \qquad b = \begin{pmatrix} n_{1\cdot} \\ n_{2\cdot} \\ n_{\cdot 1} \\ n_{\cdot 2} \end{pmatrix}, \qquad x = \begin{pmatrix} y \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

である．ここで，A 超幾何多項式 $Z_A(b; x)$ は Gauss の超幾何多項式

$${}_2F_1(-n_{1\cdot}, -n_{\cdot 1}, n - n_{1\cdot} - n_{\cdot 1} + 1; y) = \sum_{i \geq 0} \frac{(-n_{1\cdot})_i (-n_{\cdot 1})_i}{(n - n_{1\cdot} - n_{\cdot 1} + 1)_i} \frac{y^i}{i!}$$

に比例する．A 超幾何分布 (14) は一般化超幾何分布とよばれ，行と列の独立性の帰無仮説 $y = 1$ の下で超幾何分布に帰着する．以上に基づく独立性の相似検定は Fisher の正確検定とよばれる．

例 2 (Gibbs 分割) Gibbs 分割は

$$\mathbb{P}(C_1 = c_1, ..., C_n = c_n) = \frac{n!}{B_n(v,w)} v_{n,l(c)} \prod_{i=1}^{n} \left(\frac{w_i}{i!}\right)^{c_i} \frac{1}{c_i!}, \qquad l(c) := c_1 + \cdots + c_n \quad (15)$$

で与えられる．$w_i = (1-\alpha)_{i-1}$, $\alpha < 1$, $v_{n,k} = (\theta)(\theta+\alpha)\cdots(\theta+(k-1)\alpha)$ のときはPitman 分割とよばれ，対応する事前過程は2母数Poisson-Dirichlet (Pitman-Yor) 過程とよばれる．$\alpha = 0$ でEwens 抽出公式 (2) に帰着する．ここで，$B_n(v,w) = \sum_{k=1}^{n} v_{n,k} B_{n,k}(w)$, $B_{n,k}(w)$ は偏Bell 多項式である．分割の長さ $l(c) = k$ は $(v_{n,k})$ の十分統計量で，条件付き分布は

$$\mathbb{P}(C_1 = c_1, ..., C_n = c_n | Ac = b) = \frac{n!}{B_{n,k}(w)} \prod_{i=1}^{n} \left(\frac{w_i}{i!}\right)^{c_i} \frac{1}{c_i!} \qquad (16)$$

となる．この確率函数は A 超幾何分布で，A は (9) において $i_{m-1} = m-1$，$m = n-k+1$ としたもの，$b^\top = (n-k, k)$, $x_i = w_i/i!$, $Z_A(b;x) = B_{n,k}(w)/n!$ である．

3.1. A 超幾何分布に独立に従う計数ベクトルの直接抽出

統計的推測において，例1の相似検定のように，与えられた分布に従う標本の抽出が必要になることがある．分布の正規化定数の計算が簡単ではないとき，正規化定数を要しないMarkov 連鎖Monte Carlo 法 (MCMC) が良く用いられるが，MCMC は定常分布からの抽出を要し，標本に自己相関があるので，与えられた分布に独立に従う標本を直接抽出できるのであれば，その方が望ましいと思われる．A 超幾何分布 (12) については，2節で紹介したように，規格化定数である A 超幾何多項式 (8) を効率的に数値評価できることを期待できる．その利用を前提として，以下に紹介するアルゴリズム1により A 超幾何分布に独立に従う計数ベクトルを直接抽出できる．

同次性の条件 (13) は消去作用素 $\sum_{i=1}^{m} \theta_i - n$ に等しく，これを A 超幾何多項式に作用させると

$$\sum_{j=1}^{m} x_j Z_A(b - a_j; x) = n Z_A(b; x).$$

が得られる．両辺を右辺で除すと

$$\sum_{j=1}^{m} p_A(b; j) = 1 \qquad (17)$$

となる．ただし，$p_A(b;j)$ は A 超幾何分布における j 番目のセルの割合の期待値

$$p_A(b;j) := \frac{\mathbb{E}(C_j)}{n} = \frac{Z_A(b - a_j; x)}{Z_A(b;x)} \frac{x_j}{n}$$

である．(17) より，A 超幾何多項式を状態空間とし，ステップごとに次数が1下がるMarkov 連鎖において，$p_A(b;j)$ を $Z_A(b;x)$ から $Z_A(b-a_j;x)$ への推移確率とみなせることが分かる．以上の観察から，次の逐次的な直接抽出のアルゴリズムを構成できる．

アルゴリズム 1 ([2]) A 超幾何多項式 (8) を規格化定数とする A 超幾何分布 (12) からの直接抽出.

1. $t_1 = j$ を確率 $p_A(b;j)$ で抽出.

2. $i=2,...,n$ について，$t_i=j$ を確率 $p_A(b-(a_{t_1}+\cdots+a_{t_{i-1}});j)$ で抽出.

例 3 (Gibbs 分割の直接抽出) 予測規則 (1) は逐次的な直接抽出を与えるが，このような抽出は確率分割に偏交換可能性があるときのみ可能である [8]．偏交換可能性があり対称な確率分割は交換可能で，例 2 に紹介した Pitman 分割はその例である．Gibbs 分割 (15) は対称であるが，交換可能なのは $w_i=(1-\alpha)_{i-1}$, $\alpha<1$ の場合に限られる [9]．交換可能でない Gibbs 分割からの抽出には MCMC を用いることもできるが，アルゴリズム 1 を利用することで次のように直接抽出することができる.

1. 分割の長さを確率 $\mathbb{P}(l(c)=k)=B_{n,k}(w)/B_n(w)$ で抽出.
2. 分割 $(n_1,...,n_k)$ を A 超幾何分布 (16) からアルゴリズム 1 により抽出.

3.2. 2 行の行列 A に付随する A 超幾何分布の最尤推定

A 超幾何分布は指数型分布族であり，対数尤度は定数項を除いて

$$\sum_{i=1}^m \xi_i c_i-\psi(\xi), \qquad \psi(\xi):=\log Z_A(b;e^{\xi}), \qquad \xi_i:=\log x_i\in\mathbb{R}$$

で与えられる．計数ベクトル $(c_1,...,c_n)$ は自然母数 $(\xi_1,...,\xi_m)$ の十分統計量である．本節では

$$A=\begin{pmatrix} 0 & 1 & 2 & \cdots & n-k \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}, \qquad b=\begin{pmatrix} n-k \\ k \end{pmatrix} \tag{18}$$

が定める A 超幾何分布 (12) の最尤推定を議論する.

同次多項式の指数ベクトルの凸包を Newton ポリトープといい，(18) が定める A 超幾何多項式については分割の集合 $\mathcal{P}_{n,k}:=\{\lambda;\lambda\vdash n, l(\lambda)=k\}$ に対応する計数ベクトル $(c_1,...,c_{n-k+1})\in\mathbb{N}_0^{n-k+1}$ の凸包である．n の分割の集合 $\mathcal{P}_n=\cup_{k=1}^n\mathcal{P}_{n,k}$ の凸包は分割ポリトープとよばれるが，それを P_n と記す．分割ポリトープにはピラミッド性とよばれる性質がある．すなわち，P_n の $c_n>0$ を満たす頂点は e_n のみであり，他の頂点は $c_n=0$ の超平面上にある．その底面は P_{n-1} を写像

$$(c_1,c_2,...,c_{n-1})\in\mathbb{N}_0^{n-1}\mapsto(c_1+1,c_2,...,c_{n-1},0)\in\mathbb{N}_0^n$$

で表される平行移動により埋め込んだものである．このような埋め込みにより分割ポリトープたちは鎖 $P_1\subset P_2\subset\cdots$ をなす．Newton ポリトープから分割ポリトープへのアフィン写像を構成し，分割ポリトープのピラミッド性のために十分統計量の写像が分割ポリトープの内部にないことを用いて，次の定理を示すことができる.

定理 1 ([2]) (18) が定める A 超幾何分布 (12) の最尤推定量は確率 1 で存在しない.

最尤推定量を得るためには 2 つの方法がある．1 つは，標本が独立に A 超幾何分布に従う複数の計数ベクトル $\{c^{(1)},...,c^{(N)}\}$ からなるとすることで，$N\geq 2$ では十分統計量が Newton ポリトープの内部になり，最尤推定量が存在することがある．もう 1 つは，曲指数族を用いることである．曲指数族により母数空間を制限すれば，最尤推定量が

存在することがある．曲指数族の最尤推定は情報幾何の枠組みで議論することができる．双対座標は$\eta := \mathbb{E}(C_i)$，Fisher 計量は

$$g_{ij} = \frac{\partial^2 \psi(\xi)}{\partial \xi_i \partial \xi_j} = \frac{Z_A(b-a_i-a_j;x)}{Z_A(b;x)} x_i x_j 1_{\{b-a_i-a_j \geq 0\}} - \eta_i \eta_j + \eta_i \delta_{i,j}$$

である．ここに現れるA超幾何多項式は差分 HGM を用いて効率的に数値評価できるので，自然勾配法を用いて最尤推定量を数値的に求めることができる．

例 4 (交換可能な Gibbs 分割の母数の推定) 例 3 で紹介した交換可能な Gibbs 分割の長さによる条件付き分布として現れるA超幾何分布 (16) においては，母数(w_i)は 1 つの母数αによりパラメトライズされる．交換可能な Gibbs 分割において，母数$(v_{n,k})$とαの同時推定は難しいことが知られているが，$(v_{n,k})$を局外母数とみなせば，条件付き分布 (16) の最尤法によりαを推定することができる．

4. Fleming-Viot 過程の定常測度における母数の推定

Kingman のコアレセント (合併) とよばれる分割値連続時間 Markov 過程がある．その推移規則は次の通りである．状態が$\{A_1,...,A_l\}$にあるとき，可能な推移は，レート 1 で 2 つの部分を選び，合併して 1 つの部分とすることである．分割の長さ$(L_t; t \geq 0)$，$L_0 = n$は$l \to l-1$の推移がレート$l(l-1)/2$で起こる死滅過程で，1 に吸収される．この過程の実現は葉たち$[n]$から根$\{1,...,n\}$に向かって二分木を生成することである．木の枝の上にレート$\theta/2$の Poisson 過程に従い印をつける．葉iとjを結ぶ木の上の唯一の経路に印がないとき，同値類$i \sim j$を入れる．この手続きにより，Ewens 抽出公式に従う分割が得られる．

双対過程の概念は無限粒子系の確率解析によく使われる．

定義 4 状態空間E_xとE_yに値を取る Markov 過程$(X_t; t \geq 0)$, $X_0 = x$と$(Y_t; t \geq 0)$, $Y_0 = y$が核$k(x,y)$について

$$\mathbb{E}_x(k(X_t,y)) = \mathbb{E}_y(k(x,Y_t)), \qquad \forall x \in E_x,\, y \in E_y \tag{19}$$

を満たすとき，X_tとY_tは$k(x,y)$について双対という．

X_tの生成作用素をG_x，Y_tの生成作用素をG_yとする．$\forall x \in E_x, y \in E_y$について$G_x k(x,y) = G_y k(x,y)$が満たされれば (19) が成り立つので，$G_x$と$k$から$G_y$を同定することができる．Wright-Fisher 拡散 (4) については，生成作用素Lと核$k(x,n) = x^n/\mathbb{E}_{\pi_\alpha}(x^n)$について

$$Lk(x,n) = -\frac{n(n-1+m\alpha)}{2}k(x,n) + \frac{n(n-1+m\alpha)}{2}\sum_{i=1}^m \frac{n_i}{n}k(x,n-e_i) = G_n k(x,n),$$

が得られ，双対過程$(N_t; t \geq 0)$, $N_0 = n$を読み取ることができる．$m \to \infty$, $\alpha \to 0$, $\theta \equiv m\alpha > 0$とすれば，$N_t$は$l \to l - e_i$の推移がレート$l(l-1+\theta)/2 \times l_i/l$で起こる死滅過程であることがわかる．$l(l-1)/2$は合併，$l\theta/2$は印のレートであるから，$N_t$はコアレセントを与える．

Fleming-Viot 過程が与えられているけれども，定常測度から従う分割の確率函数が明示的に得られないときに，分割の確率函数を最大化する母数を求めたいことがある．

与えられた母数における分割の確率函数の値を求めるには，1節において述べたことから，(6) に類似の漸化式を $p(n)$ について解けば良い．非常に多変数の漸化式なので代数的に解くことは現実的ではない．そこで，確率シミュレーションにより $p(n)$ を数値的に評価することを考える．

標本の分割を x_0 とする．木を時間を遡って葉 x_0 から根 $x_{-T}, T < \infty$ の向きに生成することを考える．x_t を時刻 t の分割，もしくは木の断面とする．漸化式は前向き方程式

$$p(x_t) = \sum_{\{x_{t-1}\}} p(x_t|x_{t-1})p(x_{t-1}).$$

であり，問題は $p(x_0)$ を数値評価することである．前向き推移確率 $p(x_t|x_{t-1})$ を知っているので時間について前向きに生成するのは簡単だが，x_0 が得られる確率は極めて低いので効率が悪い．そこで，時間について後向きに生成することを考える．困難は，後向き推移確率 $p(x_{t-1}|x_t)$ を知らないことである．Bayes の法則から $p(x_{t-1}|x_t) = p(x_t|x_{t-1})p(x_{t-1})/p(x_t)$ であるが，$p(x_{t-1})/p(x_t)$ は不明である．そこで，後向き確率を近似し，次の逐次的インポータンスサンプリングにより $p(x_0)$ を推定することを考える．

$$p(x_0) = \mathbb{E}_{\hat{p}}\left[\frac{p(x_0|x_{-1})}{\hat{p}(x_{-1}|x_0)} \cdots \frac{p(x_{-T+1}|x_{-T})}{\hat{p}(x_{-T}|x_{-T+1})}p(x_{-T})\right] \approx \frac{1}{N}\sum_{i=1}^{N}\left\{\prod_{j=0}^{-T+1}\frac{p(x_j^{(i)}|x_{j-1}^{(i)})}{\hat{p}(x_{j-1}^{(i)}|x_j^{(i)})}p(x_{-T})\right\}.$$

ここで，$\mathbb{E}_{\hat{p}}$ は後向き推移確率の近似 $\hat{p}(x_{t-1}|x_t)$ により生成される後向き経路 $\{x_{-1}, ..., x_{-T}\}$ について期待値をとることを表し，$x^{(1)}, ..., x^{(N)}$ は経路の独立な実現である．もし近似が完璧，つまり $\hat{p}(x_{t-1}|x_t) = p(x_{t-1}|x_t)$ であれば，インポータンスウェイトは $p(x_0)$ であり，推定量の分散が 0 なので推定の効率は最善である．

後向き推移確率の近似を求める手続きを Dirichlet 多項分布を解とする漸化式 (6) に沿って説明する．分割の確率函数 $p(n) = \langle q_n, \pi\rangle$ の近似 $\hat{p}(n)$ に交換可能性を要求すれば

$$\rho(i|n-e_i)\hat{p}(n-e_i) = \frac{n_i}{n}\hat{p}(n) \tag{20}$$

となる．ここで，$\rho(i|n-e_i)$ は標本が $n-e_i$ のときに次の標本点として i が選ばれる確率の近似である．De Iorio と Griffiths [10] は，ρ と $\hat{p}$ を定めるのに，拡散の可逆測度 π が従う条件 $\langle L_j\partial q_n/\partial x_j, \pi\rangle = 0, \forall j$ を用いることを提案した．これを Wright-Fisher 拡散 (4) に課すと

$$n_j(n-1+m\alpha)\hat{p}(n) = n(n_j-1)\hat{p}(n-e_j) + \alpha\sum_{i=1}^{m}(n_i+1-\delta_{ij})\hat{p}(n+e_i-e_j), \qquad j \in [m]$$

が得られ，さらに (20) より

$$\rho(i|n) = \frac{\alpha+n_i}{m\alpha+n}, \qquad i \in [m]$$

が得られる．これは Dirchlet 分布の予測規則に他ならないので，$\hat{p}(n) = p(n)$ である．近似が完璧なのは交換可能性と可逆性が満たされているからである．以上の手続きは，交換可能性，可逆性が満たされていない場合に近似として有効である．

例 5 (偏りのある投票者モデルの母数の推定) [3] で考察された偏りのある投票者モデルを紹介する．グラフの集合 $(I_i), i \in [m]$ があり，グラフ I_i はサイトたち $[N]$ からなり，個々のサイトは意見 1 か 0 を持つとする．$\eta_t(x), x \in I_i$ で時刻 t におけるサイト x の意見を表す．次の規則で推移する連続時間 Markov 過程 $\{\eta_t; t \geq 0\}$ を考える．

1. サイト $x \in I_i$ の意見が I_i の他のサイトにレート λ_N で伝わる．

2. サイト $x \in I_i$ の意見が $I_{j\neq i}$ のサイトに伝わる．そのレートは意見に依存し，$\eta_t(x) = k$ のとき $\lambda_N c(1-(-1)^k\beta)$ である．

3. 意見が伝わるとき，確率 a で意見を変更する．

このモデルは，グラフ内の意見の伝搬には偏りがないが，グラフ間の意見の伝搬には偏りがある状況を表している．$\lambda_N = N/2 \to \infty$ の拡散極限において，$a, c \to 0, \alpha \equiv Na > 0$, $\gamma \equiv Nc > 0$ とすると，各グラフにおける意見 0 の割合 (x_i) の拡散極限は $[0,1]^m$ 上の Wright-Fisher 拡散であり，その生成作用素は

$$
\begin{aligned}
L = & \sum_{i=1}^{m} \frac{x_i(1-x_i)}{2}\frac{\partial^2}{\partial x_i^2} + \frac{\alpha}{2}\sum_{i=1}^{m}(1-2x_i)\frac{\partial}{\partial x_i} + \frac{\gamma}{2}\sum_{i=1}^{m}\left(\sum_{j=1}^{m} x_j - m x_i\right)\frac{\partial}{\partial x_i} \\
& - \beta\frac{\gamma}{2}\sum_{i=1}^{m}\left\{(1-2x_i)\sum_{j\neq i} x_j + (m-1)x_i\right\}\frac{\partial}{\partial x_i}
\end{aligned}
$$

である．分割の確率函数を明示的に得ることはできないが，[3] では，分枝を伴うコアレセントを双対として導出し，母数 (α, β, γ) を推定するためのインポータンスサンプラーを構成している．

参考文献

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

[2] Mano, S.: Partition structure and the A-hypergeometric distribution associated with the rational normal curve. Electron. J. Statist. **11**, 4452–4487 (2017)

[3] Mano, S.: Ancestral graph with bias in gene conversion. J. Appl. Probab. **50**, 239–255 (2013)

[4] Mano, S.: *Partitions, Hypergeometric Systems, and Dirichlet Processes in Statistics*. JSS Research Series in Statistics, SpringerBriefs in Statistics (2018)

[5] Gel'fand, I.M., Zelevinsky, A.V., Kapranov, M.M.: Generalized Euler Integrals and A-hypergeometric functions. Adv in Math. **84**, 255–271 (1990)

[6] Ohara, K., Takayama, N.: Pfaffian systems of A-hypergeometric systems II – holonomic gradient method. arXiv: 1505.02947

[7] Takayama N., Kuriki, S., Takemura, A.: A-hypergeometric distributions and Newton polytope. Adv. in Appl. Math. **99**, 109–133 (2018)

[8] Pitman, J.: Exchangeable and partially exchangeable random partitions. Probab. Theory Related Fields **102**, 145–158 (1995)

[9] Gnedin, A., Pitman, J.: Exchangeable Gibbs partitions and Stirling triangles. Zap. Nauchn. Sem. S. POMI **325**, 83–102 (2005) English translation: J. Math. Sci. **138**, 5674–5685 (2006)

[10] De Iorio, M., Griffiths, R.C.: Importance sampling on coalescent histories. I. Adv. Appl. Probab. **36**, 417–433 (2004)