Ewens分布を不変にする２つのMarkov連鎖

半田賢司（佐賀大学）

題材

遺伝子変換モデル

・各個体のタイプが世代交代に伴い変化（ランダムサンプリング／突然変異）

・同じタイプの個体からなるグループサイズの時間発展を記述

凝結・分裂モデル

・２つのクラスターが合体／１つのクラスターが２つのクラスターに分裂

・クラスターサイズの時間発展を記述

内容

- ２つのモデル共に**Ewens分布**を定常分布として持つ.
- それらの連続版：**無限個アレルモデル**と**区間のsplit-merge過程**
- 連続版は共に**Poisson-Dirichlet分布**を定常分布として持つ.

整数分割とEwens分布

N:自然数，$\mathcal{P}_N$: N の分割全体

$\boxed{\text{記法}}$ $\mathcal{P}_N \ni \lambda = (1^{\lambda_1} 2^{\lambda_2} \cdots)$

$$\longleftrightarrow N = \overbrace{1 + \cdots + 1}^{\lambda_1} + \overbrace{2 + \cdots + 2}^{\lambda_2} + \cdots + \overbrace{N}^{\lambda_N} = \sum i\lambda_i$$

$$=: n_1 + \cdots + n_k \ (n_1 \geq \ldots \geq n_k) \text{ ただし } k = \sum \lambda_i =: |\lambda|$$

$$\longleftrightarrow \underline{n} = (n_1, \ldots, n_k, 0, 0, \ldots) \in \mathbf{Z}_+^\infty$$

<u>Ewens分布</u> μ_θ **on** $\mathcal{P}_N$ **(**$\theta > 0$**: パラメータ)**

$$\mu_\theta(\lambda) = \frac{N!}{(\theta)_N} \prod_{i=1}^{N} \left(\left(\frac{\theta}{i}\right)^{\lambda_i} \frac{1}{\lambda_i!} \right) = \frac{\theta^{|\lambda|}}{(\theta)_N} \prod_{i=1}^{N} \frac{i}{i^{\lambda_i} \lambda_i!}$$

ただし，$(\theta)_N = \theta(\theta + 1) \cdots (\theta + N - 1)$

$\boxed{\text{注意}}$ $\theta \to \infty$ のとき $\mu_\theta \to \delta_{(1^N 2^0 \ldots)}$，$\theta \downarrow 0$ のとき $\mu_\theta \to \delta_{(1^0 \ldots (N-1)^0 N^1)}$.

Ewens分布とサイクル長

$\mathcal{S}_N$: $\{1,\ldots,N\}$ の置換全体

$\mathcal{S}_N \ni \sigma$ は，互いに排反な「台」を持つ巡回置換の積で表せる：

$$\sigma = \tau_1 \circ \cdots \circ \tau_k \quad (k =: |\sigma|)$$

τ_i の「台 **(support)**」の大きさ（不動点は大きさ１）に着目：

$$N = \sharp\mathbf{supp}(\tau_1) + \cdots + \sharp\mathbf{supp}(\tau_k) \longrightarrow \lambda \in \mathcal{P}_N$$

（結局）「ランダム置換 σ はランダム分割 λ を導く.」

$\boxed{\text{定理}}$ **[Cauchy,18??]** **(for** $\theta = 1$**)**

$\theta > 0$ に対し，$\mathcal{S}_N$ 上の分布 m_θ を $m_\theta(\sigma) \propto \theta^{|\sigma|}$ で定めるとき，

$$\sigma \sim m_\theta \Longrightarrow \lambda \sim \mu_\theta$$

遺伝子変換モデル

状態空間 $\mathcal{P}_N$

・個体数 N の集団において，個体の‘タイプ’に基づくグループ分け

・グループサイズに着目 $\longrightarrow$ $\lambda \in \mathcal{P}_N$

（結局）「ランダムなタイプ変化はランダム分割の時間発展を導く.」

タイプの時間発展（概要） **(Wight-Fisher 型) cf.** 清水昭信『確率過程論』

(1) random sampling

各個体が独立に，N 個体から当確率で選んだ個体と同じタイプになる.

(2) mutation

一定の率で，これまでになかった新たなタイプへと変化する.

連続時間Markov過程としての定式化（１）

$\theta > 0$：パラメータ（突然変異率）

$\lambda \in \mathcal{P}_N$ からの推移先と推移率 **(rate)**：$i, j \in \{1, \ldots, N\}$

(1) random sampling

- $i \neq j$ に対し，$\lambda \mapsto \lambda^{i \to j}$ ；

「大きさ i のグループの１個体が大きさ j のグループへ移動」**rate** $i\lambda_i j\lambda_j$

- 各 i に対し，$\lambda \mapsto \lambda^{i \to i}$ ；

「大きさ i のグループの１個体が大きさ i の別のグループへ移動」**rate** $i\lambda_i i(\lambda_i - 1)$

(2) mutation

各 i に対し，$\lambda \mapsto \lambda^{i \to *}$ ；

「大きさ i のグループの１個体が大きさ１のグループを形成」**rate** $\theta i\lambda_i$

$\boxed{\text{記号}}$ **rate**$_{WF}(\lambda \mapsto \lambda')$ ：「$\lambda \in \mathcal{P}_N$ から $\lambda' \in \mathcal{P}_N$ への推移率」

例）**rate**$_{WF}(\lambda \mapsto \lambda^{i \to *}) = \theta i\lambda_i$

Ewens分布の不変性 – 遺伝子変換モデルの場合

生成作用素

$$G_N f(\lambda) = \sum_{\lambda'} \mathbf{rate}_{WF}(\lambda \mapsto \lambda') \left[f(\lambda') - f(\lambda) \right]$$

定理　パラメータ θ の遺伝子変換モデルは，**Ewens**分布 μ_θ を唯一の定常分布として持つ．実際，次の詳細釣合条件 **(DBC)** が満たされる：

$$\mathbf{rate}_{WF}(\lambda \mapsto \lambda')\mu_\theta(\lambda) = \mathbf{rate}_{WF}(\lambda' \mapsto \lambda)\mu_\theta(\lambda'), \quad \forall \lambda, \lambda' \in \mathcal{P}_N$$

すなわち，μ_θ は（もっと強く）可逆分布である．

（略証）**DBC**から次が示される．

$$\sum_{\lambda} G_N f(\lambda) g(\lambda) \mu_\theta(\lambda) = \sum_{\lambda} f(\lambda) G_N g(\lambda) \mu_\theta(\lambda).$$

凝結分裂モデル

状態空間 $\mathcal{P}_N$

・N 個の粒子がいくつかのクラスターを成す.

・クラスターサイズに着目 $\longrightarrow \lambda \in \mathcal{P}_N$

（結局）「クラスターのランダムな合体・分裂はランダム分割の時間発展を導く.」

クラスター群の時間発展（概要）

cf. [Diaconis, Mayer-Wolf, Zeitouni, Zerner, 2004]

(0) N 粒子から**2**粒子を非復元抽出

(1) coagulation

2粒子が異なるクラスターに属する場合，２つのクラスターが合体

(2) fragmentation

2粒子が同じクラスターに属する場合，クラスターが無作為の大きさで**2**つに分裂

[**e.g.** 大きさ**10**のクラスターの分裂でできる小クラスターサイズ$\in \{1, 2, \ldots, 9\}$]

連続時間Markov過程としての定式化（２）

$\widehat{\beta} > 0$：凝結率，$\breve{\beta} > 0$：分裂率

$\lambda \in \mathcal{P}_N$ からの推移先と推移率 **(rate)**：$i, j \in \{1, \ldots, N-1\}$

(1) coagulation

- $i \neq j$ に対し，$\lambda \mapsto \mathbf{Coag}_{i,j}\lambda$ ；

「大きさ i のクラスター１個と大きさ j のクラスター１個が合体」**rate** $\widehat{\beta}\frac{i\lambda_i j\lambda_j}{N(N-1)}$

- 各 i に対し，$\lambda \mapsto \mathbf{Coag}_{i,i}\lambda$ ；

「大きさ i の２つの異なるクラスター同士が合体」**rate** $\widehat{\beta}\frac{i\lambda_i i(\lambda_i-1)}{N(N-1)}$

(2) fragmentation

i, j に対し，$\lambda \mapsto \mathbf{Frag}_{i,j}\lambda$ ；

「大きさ $i+j$ のクラスターが大きさ i, j の２つのクラスターに分裂」

$$\textbf{rate}\ \breve{\beta}\frac{(i+j)\lambda_{i+j}}{N(N-1)} = \breve{\beta}\frac{i+j}{N}\cdot\frac{i+j-1}{N-1}\cdot\frac{1}{i+j-1}\cdot\lambda_{i+j}$$

Ewens分布の不変性 – 凝結分裂モデルの場合

<u>生成作用素</u>

$$L_N f(\lambda) = \sum_{\lambda'} \mathbf{rate}_{CF}(\lambda \mapsto \lambda') \left[f(\lambda') - f(\lambda) \right]$$

$\boxed{\text{定理}}$　凝結率 $\widehat{\beta}$，分裂率 $\breve{\beta}$ の凝結分裂モデルは，パラメータ $\theta = \breve{\beta}/\widehat{\beta}$ の **Ewens** 分布 μ_θ を唯一の定常分布として持つ．実際，次の<u>詳細釣合条件 **(DBC)**</u> が満たされる：

$$\mathbf{rate}_{CF}(\lambda \mapsto \lambda')\mu_\theta(\lambda) = \mathbf{rate}_{CF}(\lambda' \mapsto \lambda)\mu_\theta(\lambda'), \quad \forall \lambda, \lambda' \in \mathcal{P}_N$$

すなわち，μ_θ は（もっと強く）可逆分布である．
（略証）**DBC** から次が示される．

$$\sum_{\lambda} L_N f(\lambda) g(\lambda) \mu_\theta(\lambda) = \sum_{\lambda} f(\lambda) L_N g(\lambda) \mu_\theta(\lambda).$$

$\boxed{\text{注意}}$　$\breve{\beta} = \widehat{\beta} = 1$ のとき，$\mathbf{rate}_{CF}(\lambda \mapsto \cdot)$ は $\mathcal{P}_N$ 上の推移確率でもあり,「ランダム互換 **(transposition)** モデル」が導く推移確率と同じ.

連続極限 as $N \to \infty$（定式化）

$\mathcal{P}_N \ni \lambda \longleftrightarrow N = n_1 + \cdots + n_k \ (n_1 \geq \ldots \geq n_k > 0)$
$\longleftrightarrow \underline{n} = (n_1, \ldots, n_k, 0, 0, \ldots)$ を正規化：

$$\frac{1}{N}\underline{n} = \left(\frac{n_1}{N}, \ldots, \frac{n_k}{N}, 0, 0, \ldots\right) \in \nabla_\infty$$

新たな状態空間は

$$\nabla_\infty = \left\{ \underline{p} = (p_1, p_2, \ldots) : p_1 \geq p_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} p_i = 1 \right\}$$

テスト関数　$\phi(\underline{p}) = \prod_{i=1}^{\infty} f(p_i)$

ただし，$f \geq 0$，$f - 1 \in C_c^\infty(\mathbf{R}_+)$，$\mathrm{supp}(f-1) \subset (0, \infty)$.

注意　このとき，$f(0) = 1$ であり，$\prod_{i=1}^{\infty} f(p_i)$ は有限積.

連続極限 – 遺伝子変換モデルの場合

- $\phi(\underline{p}) = \prod_{i=1}^{\infty} f(p_i)$ に対して，$\phi_N(\underline{n}) = \prod_{i=1}^{\infty} f\left(\frac{n_i}{N}\right)$ とおくと，

$$G_N\phi_N(\underline{n}) = \sum_{i\neq j} n_i n_j \left[f\left(\frac{n_i-1}{N}\right) f\left(\frac{n_j+1}{N}\right) - f\left(\frac{n_i}{N}\right) f\left(\frac{n_j}{N}\right)\right] \prod_{l\neq i,j} f\left(\frac{n_l}{N}\right)$$
$$+\theta \sum_{i} n_i \left[f\left(\frac{n_i-1}{N}\right) f\left(\frac{1}{N}\right) - f\left(\frac{n_i}{N}\right)\right] \prod_{l\neq i} f\left(\frac{n_l}{N}\right).$$

- $N \to \infty$ のとき $N^{-1}\underline{n}^{(N)} \to \underline{p}$ **(i.e.,** $N^{-1}n_i^{(N)} \to p_i$ $(\forall i \in \mathbf{N})$**)** ならば，

$$G_N\phi_N(\underline{n}^{(N)}) \to \sum_{i,j=1}^{\infty} p_i(\delta_{ij}-p_j)\frac{\partial^2\phi}{\partial p_i \partial p_j}(\underline{p}) - \theta\sum_{i=1}^{\infty} p_i \frac{\partial\phi}{\partial p_i}(\underline{p})$$
$$=: \mathcal{G}\phi(\underline{p})$$ 「無限個アレルモデル」の生成作用素

cf. [Shimizu, 1987] 「n-locus モデル」の場合 ($i, j \in \mathcal{P}_n$**?)**

連続極限 – 凝結分裂モデルの場合

- $\phi(\underline{p}) = \prod_{i=1}^{\infty} f(p_i)$ に対して，$\phi_N(\underline{n}) = \prod_{i=1}^{\infty} f\left(\frac{n_i}{N}\right)$ とおくと，

$$\begin{aligned} L_N \phi_N(\underline{n}) &= \hat{\beta} \sum_{i \neq j} \frac{n_i n_j}{N(N-1)} \left[f\left(\frac{n_i + n_j}{N}\right) - f\left(\frac{n_i}{N}\right) f\left(\frac{n_j}{N}\right) \right] \prod_{l \neq i,j} f\left(\frac{n_l}{N}\right) \\ &\quad + \check{\beta} \sum_i \frac{n_i}{N(N-1)} \sum_{m=1}^{n_i - 1} \left[f\left(\frac{m}{N}\right) f\left(\frac{n_i - m}{N}\right) - f\left(\frac{n_i}{N}\right) \right] \prod_{l \neq i} f\left(\frac{n_l}{N}\right). \end{aligned}$$

- $N \to \infty$ のとき $N^{-1}\underline{n}^{(N)} \to \underline{p}$ **(i.e.,** $N^{-1} n_i^{(N)} \to p_i$ **(**$\forall i \in \mathbf{N}$**))** ならば，

$$\begin{aligned} L_N \phi_N(\underline{n}^{(N)}) &\to \hat{\beta} \sum_{i \neq j} p_i p_j \left[f\left(p_i + p_j\right) - f\left(p_i\right) f\left(p_j\right) \right] \prod_{l \neq i,j} f\left(p_l\right) \\ &\quad + \check{\beta} \sum_i p_i^2 \int_0^1 du \left[f\left(u p_i\right) f\left((1-u) p_i\right) - f\left(p_i\right) \right] \prod_{l \neq i} f\left(p_l\right) \\ &=: \mathcal{L}\phi(\underline{p}) \end{aligned}$$

「区間の split-merge 過程」の生成作用素

cf. [Mayer-Wolf, Zeitouni, Zerner, 2002]

Poisson-Dirichlet分布（定義）

$\sum \delta_{Z_i}$: $(0,\infty)$ 上の **Poisson** 点過程で，平均測度 $\theta z^{-1} e^{-z} dz$

定義

$Z_{(1)} \geq Z_{(2)} \geq \cdots$: $Z_1, Z_2, \ldots$ を大きい順に並べたもの
$Z := Z_1 + Z_2 + \cdots (< \infty$ **a.s.**$)$
このとき，正規化された無限列 $\left(Z_{(1)}/Z, Z_{(2)}/Z, \ldots\right)$ が導く

$$\nabla_\infty = \left\{ (p_1, p_2, \ldots) : p_1 \geq p_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} p_i = 1 \right\}$$

上の分布をパラメータ θ の **Poisson-Dirichlet分布** とよぶ.

記号　**PD**(θ)

Ewens分布とPoisson-Dirichlet分布

$\boxed{\text{事実}}$ **(1)** 各$\lambda \in \mathcal{P}_N$に対して

$$\mu_\theta(\lambda) = \frac{N!}{\prod_{i=1}^{N} \left((i!)^{\lambda_i} \lambda_i!\right)} E^{\mathrm{PD}(\theta)} \left[\sum_{\substack{i_1, \ldots, i_k \\ \textbf{distinct}}} P_{i_1}^{n_1} \cdots P_{i_k}^{n_k} \right]$$

ただし，$\underline{n} = (n_1, \ldots, n_k, 0, 0, \ldots)$は$\lambda$から定まるものとする．

(2) ランダム分割$\lambda^{(N)} \in \mathcal{P}_N$は**Ewens**分布$\mu_\theta = \mu_\theta^{(N)}$に従うとする．
$\lambda^{(N)}$から定まる$\underline{n}^{(N)} = (n_1^{(N)}, \ldots, n_k^{(N)}, 0, 0, \ldots)$について，$N \to \infty$のとき

$$\frac{1}{N}\underline{n}^{(N)} \Rightarrow (P_1, P_2, \ldots) \sim \mathrm{PD}(\theta)$$（法則収束）

定常分布としてのPoisson-Dirichlet分布

− 無限個アレルモデルの場合

$\boxed{\text{定理}}$ **[Ethier-Kurtz, 1981]**

PD(θ)は，生成作用素

$$\mathcal{G}\phi(\underline{p}) = \sum_{i,j=1}^{\infty} p_i(\delta_{ij} - p_j)\frac{\partial^2 \phi}{\partial p_i \partial p_j}(\underline{p}) - \theta \sum_{i=1}^{\infty} p_i \frac{\partial \phi}{\partial p_i}(\underline{p})$$

に付随する∇_∞上の**Markov**過程の一意な定常分布である．また，可逆分布でもある．

$\boxed{\text{注意}}$ $t \to \infty$のときの定常分布への収束の速さについては，ある意味で指数的であることが知られている．(2パラメータ版も含めて **[Petrov, 2009]**）

定常分布としてのPoisson-Dirichlet分布

－ 区間のsplit-merge過程の場合

$\boxed{\text{定理}}$ **[Mayer-Wolf, Zeitouni, Zerner, 2002]**

$\theta = \check{\beta}/\hat{\beta}$ とおく．$\mathbf{PD}(\theta)$ は，生成作用素

$$
\begin{aligned}
\mathcal{L}\phi(\underline{p}) \;&=\; \hat{\beta}\sum_{i\neq j} p_i p_j \left[f\left(p_i+p_j\right) - f\left(p_i\right) f\left(p_j\right)\right] \prod_{l\neq i,j} f\left(p_l\right) \\
&\quad + \check{\beta}\sum_{i} p_i^2 \int_0^1 du\left[f\left(u p_i\right) f\left((1-u)p_i\right) - f\left(p_i\right)\right] \prod_{l\neq i} f\left(p_l\right)
\end{aligned}
$$

に付随する ∇_∞ 上のMarkov過程の可逆定常分布である．

$\boxed{\text{注意}}$ **(1)** 定常分布は一意的であろうとの予想があるが，今の所，$\theta = 1$ の場合のみが示されている．**[Diaconis, Mayer-Wolf, Zeitouni, Zerner, 2004]**

(2) $t \to \infty$ のときの定常分布への収束の速さについては，指数的でないことが示せる．