The tree-valued Fleming-Viot process with mutation and selection

Peter Pfaffelhuber
University of Freiburg

Joint work with Andrej Depperschmidt and Andreas Greven

March 29th, 2011
Population genetic models

Populations of constant size have been modelled by

- **Markov Chains** (Wright-Fisher-model, Moran model)
- **Diffusion approximations** (Fisher-Wright diffusion)

\[
dX = \alpha X(1 - X)dt + \sqrt{X(1 - X)}dW
\]

or **Measure-valued diffusions** (Fleming-Viot superprocess)

- **New**: Extend Fleming-Viot process by genealogical information → **Tree-valued Fleming-Viot process**

The tree-valued Fleming-Viot process with mutation and selection
The Moran model with mutation and selection

The tree-valued Fleming-Viot process with mutation and selection
The Moran model with mutation and selection

Goal: construct a tree-valued stochastic process \( U = (U_t)_{t \geq 0} \) that describes genealogical relationships dynamically and makes the forward and backward picture implicit.

The tree-valued Fleming-Viot process with mutation and selection
The Moran model with mutation and selection
The Moran model with mutation and selection

The tree-valued Fleming-Viot process with mutation and selection
The Moran model with mutation and selection

**Goal:** construct a tree-valued stochastic process $\mathcal{U} = (\mathcal{U}_t)_{t \geq 0}$

- describe genealogical relationships *dynamically*
- make *forward* and *backward* picture implicit
Summary: The tree-valued Fleming-Viot process

- **Theorem:** The \((\Omega, \Pi)\)-martingale problem is well-posed. Its solution – the **tree-valued Fleming-Viot process** – arises as weak limit of tree-valued Moran models.

- **Theorem:** Tree-valued processes for different \(\alpha\) are absolutely **continuous** with respect to each other.

- **Theorem:** The **measure-valued** Fleming-Viot process is **ergodic** iff the **tree-valued** Fleming-Viot process is **ergodic**.

- **Theorem:** The distribution of \(R_{12}^\alpha\), the distance of **two** randomly sampled points in equilibrium, can be computed.
Formalizing genealogical trees

- **Leaves in genealogical trees** form a metric space; leaves are marked by elements of \( I \) (compact)

A tree is given by:

\[
(X, r) \text{ complete and separable metric space}
\]

- \( r(x_1, x_2) \) defines the genealogical distance of individuals \( x_1 \) and \( x_2 \)
Formalizing genealogical trees

- **Leaves in genealogical trees** form a metric space; leaves are marked by elements of $I$ (compact)

A tree is given by:

$$(X, r) \text{ complete and separable metric space, } \mu \in \mathcal{P}(X)$$

- $r(x_1, x_2)$ defines the genealogical distance of individuals $x_1$ and $x_2$
- $\mu$ marks currently living individuals
Formalizing genealogical trees

- **Leaves in genealogical trees** form a metric space; leaves are marked by elements of $I$ (compact)

A tree is given by:

$$(X, r) \text{ complete and separable metric space, } \mu \in \mathcal{P}(X \times I)$$

- $r(x_1, x_2)$ defines the genealogical distance of individuals $x_1$ and $x_2$
- $\mu$ marks currently living individuals
Formalizing genealogical trees

- **Leaves in genealogical trees** form a metric space; leaves are marked by elements of $I$ (compact)

State space of $\mathcal{U}$:

$\mathbb{X} := \{ \text{isometry class of } (X, r, \mu) : 
(X, r) \text{ complete and separable metric space, } \mu \in \mathcal{P}(X \times I) \}$

- $r(x_1, x_2)$ defines the genealogical distance of individuals $x_1$ and $x_2$
- $\mu$ marks currently living individuals
Martingale Problem

- **Given:** Markov process $\mathcal{X} = (\mathcal{X}_t)_{t \geq 0}$. The generator is

  \[ \Omega \Phi(x) := \lim_{h \to 0} \frac{1}{h} \mathbb{E}_x [\Phi(\mathcal{X}_h) - \Phi(x)]. \]

- **Given:** Operator $\Omega$ on $\Pi$. A solution of the \((\Omega, \Pi)\)-martingale problem is a process $\mathcal{X} = (\mathcal{X}_t)_{t \geq 0}$ if for all $\Phi \in \Pi$,

  \[ \left( \Phi(\mathcal{X}_t) - \int_0^t \Omega \Phi(\mathcal{X}_s) ds \right)_{t \geq 0} \]

  is a martingale. The MP is **well-posed** if there is exactly one such process.
Polynomials on $\mathcal{P}(I)$

$\Pi$: functions of the form (polynomials)

$$\Phi(\mu) := \langle \mu^N, \phi \rangle := \int \phi(u) \mu^N(d\mu)$$

for $u = (u_1, u_2, \ldots), \phi \in C_b(I^N)$ depending on finitely many coordinates

- $\Pi$ separates points in $\mathcal{P}(I)$
Polynomials on $\mathbb{U}$

$\Pi$: functions on $\mathbb{U}$ of the form (polynomials)

$$\Phi(X, r, \mu) := \langle \mu^N, \phi \rangle := \int \phi(r(x, x), u) \mu^N(d(x, u))$$

for $(x, u) = ((x_1, u_1), (x_2, u_2), \ldots), \phi \in C_b(\mathbb{R}^N_2 \times I^N)$ depending on finitely many coordinates

- $\Pi$ separates points in $\mathbb{X}$
Generator for the Fleming-Viot process: measure-valued

\[ \Omega := \Omega^{\text{res}} + \Omega^{\text{mut}} + \Omega^{\text{sel}} \]

- \( \Omega^{\text{res}} \): resampling
- \( \Omega^{\text{mut}} \): mutation
- \( \Omega^{\text{sel}} \): selection
Generator for the Fleming-Viot process: tree-valued

\[ \Omega := \Omega^{\text{grow}} + \Omega^{\text{res}} + \Omega^{\text{mut}} + \Omega^{\text{sel}} \]

- \( \Omega^{\text{grow}} \): tree growth
- \( \Omega^{\text{res}} \): resampling
- \( \Omega^{\text{mut}} \): mutation
- \( \Omega^{\text{sel}} \): selection
Tree Growth

When no resampling occurs the tree grows

Distances in the sample grow at speed 2

\[ \Omega^{\text{grow}} \Phi(X, r, \mu) = 2 \cdot \left\langle \mu^N, \sum_{i<j} \frac{\partial}{\partial r_{ij}} \phi \right\rangle. \]
Resampling: measure-valued

$$\Omega^\text{res} \Phi(\mu) := \sum_{k < l} \langle \mu^\mathbb{N}, \phi \circ \theta_{k,l} - \phi \rangle$$

with

$$(\theta_{k,l}(u))_i := \begin{cases} u_i, & i \neq l \\ u_k, & i = l \end{cases}$$
Resampling: tree-valued

$$\Omega_{\text{res}} \Phi(\mathbf{X}, r, \mu) := \sum_{k < l} \langle \mu^N, \phi \circ \theta_{k,l} - \phi \rangle$$

with

$$(\theta_{k,l}(u))_i := \begin{cases} u_i, & i \neq l \\ u_k, & i = l \end{cases}$$

In addition,

$$(\theta_{k,l}(r(x,x)))_{i,j} := \begin{cases} r(x_i, x_j), & \text{if } i, j \neq l, \\ r(x_i, x_k), & \text{if } j = l, \\ r(x_k, x_j), & \text{if } i = l, \end{cases}$$
Mutation: measure-valued

- $\vartheta$: total mutation rate
- $\vartheta \cdot \beta(u, dv)$: mutation rate from $u$ to $v$

$$\Omega^\text{mut} \Phi(\mu) = \vartheta \cdot \sum_k \langle \mu^N, \beta_k \phi - \phi \rangle$$

with $\beta_k(u, dv)$ acting on $k$th variable
Mutation: tree-valued

- $\vartheta$: total mutation rate
- $\vartheta \cdot \beta(u, dv)$: mutation rate from $u$ to $v$

\[
\Omega^{\text{mut}} \Phi(X, r, \mu) = \vartheta \cdot \sum_{k} \langle \mu^N, \beta_k \phi - \phi \rangle
\]

with $\beta_k(u, dv)$ acting on $k$th variable
Selection: measure-valued

- $\alpha$: selection coefficient
- $\chi(u) \in [0, 1]$: fitness of type $u$ (continuous)

$$\Omega_{\text{sel}}^\Phi(\mu) := \alpha \cdot \sum_{k=1}^{n} \langle \mu^N, \chi_k \cdot \phi - \chi_{n+1} \cdot \phi \rangle$$

where $\phi$ only depends on sample of size $n$
with $\chi_k$ acting on $k$th variable
Selection: tree-valued

- $\alpha$: selection coefficient
- $\chi(u) \in [0, 1]$: fitness of type $u$ (continuous)

\[
\Omega_{\text{sel}}^{\Phi}(X, r, \mu) := \alpha \cdot \sum_{k=1}^{n} \langle \mu^N, \chi_k \cdot \phi - \chi_{n+1} \cdot \phi \rangle
\]

where $\phi$ only depends on sample of size $n$
with $\chi_k$ acting on $k$th variable
Selection: tree-valued

- Why is selection the same as for measure-valued case?
- $\Omega^\text{sel}_N$: generator for finite model of size $N$
- $\phi$: only depends on first $n \ll N$ individuals

$$
\Omega^\text{sel}_N \Phi(X, r, \mu) \approx \frac{\alpha}{N} \sum_{k,l=1}^{N} \langle \mu^N, \chi_k(\phi \circ \theta_k, l - \phi) \rangle
$$

$$
\approx \alpha \cdot \sum_{l=1}^{n} \langle \mu^N, \chi_{n+1}(\phi \circ \theta_{n+1}, l - \phi) \rangle
$$

$$
= \alpha \cdot \sum_{l=1}^{n} \langle \mu^N, \chi_l \cdot \phi - \chi_{n+1} \cdot \phi \rangle
$$
The tree-valued Fleming-Viot process

**Theorem:** The $(\Omega, \Pi)$-martingale problem is well-posed. Its solution $\mathcal{X} = (\mathcal{X}_t)_{t \geq 0}$, $\mathcal{X}_t = (X_t, r_t, \mu_t)$ – the **tree-valued Fleming-Viot process** – arises as weak limit of tree-valued Moran models and satisfies:

- $\mathbf{P}(t \mapsto \mathcal{X}_t \text{ is continuous}) = 1$,
- $\mathbf{P}((\mathcal{X}_t, r_t) \text{ is compact for all } t > 0) = 1$,
- $\mathcal{X}$ is Feller (hence strong Markov)
Girsanov transform: measure-valued

**Theorem:** Let $\alpha, \alpha' \in \mathbb{R}$, $\mathcal{X}$ solution of $(\Omega, \Pi)$-MP for selection coefficient $\alpha$,

$$\Psi(\mu) := (\alpha' - \alpha) \cdot \langle \mu^N, \chi_1 \rangle$$

and

$$\mathcal{M} = \left( \psi(\mu_t) - \psi(\mu_0) - \int_0^t \Omega \psi(\mu_s) ds \right)_{t \geq 0}.$$

Then, $Q$, defined by

$$\frac{dQ}{dP}\bigg|_{\mathcal{F}_t} = e^{\mathcal{M}_t - \frac{1}{2}[\mathcal{M}]_t}$$

solves $(\Omega, \Pi)$-MP for selection coefficient $\alpha'$. 

The tree-valued Fleming-Viot process with mutation and selection
Girsanov transform: tree-valued

**Theorem:** Let $\alpha, \alpha' \in \mathbb{R}$, $X$ solution of $(\Omega, \Pi)$-MP for selection coefficient $\alpha$,

$$
\Psi(X, r, \mu) := (\alpha' - \alpha) \cdot \langle \mu^N, \chi_1 \rangle
$$

and

$$
M = \left( \Psi(X_t, r_t, \mu_t) - \Psi(X_0, r_0, \mu_0) - \int_0^t \Omega \Psi(X_s, r_s, \mu_s) ds \right)_{t \geq 0}.
$$

Then, $Q$, defined by

$$
\frac{dQ}{dP}\bigg|_{\mathcal{F}_t} = e^{M_t - \frac{1}{2}[M]_t}
$$

solves $(\Omega, \Pi)$-MP for selection coefficient $\alpha'$.
Long-time behavior

**Theorem:**

The *tree-valued* Fleming-Viot process is *ergodic* iff the *measure-valued* Fleming-Viot process is *ergodic*.
Application: distances is equilibrium

▶ **Theorem:**

- \( I = \{\bullet, \bullet\}, \chi(u) = 1_{\{u=\bullet\}} \) (\( \bullet \) is fit, \( \bullet \) is unfit)
- \( \frac{\vartheta}{2} \): mutation rate \( \bullet \rightarrow \bullet \) and \( \bullet \rightarrow \bullet \)
- \( R_{12}^{\alpha} \): **distance of two randomly sampled points** in equilibrium

\[
E[e^{-\lambda R_{12}^{\alpha}/2}] = \frac{1}{1 + \lambda} + \frac{4\vartheta(2 + \lambda + 2\vartheta)\lambda}{(1 + \vartheta)(1 + \lambda + \vartheta)(6 + \lambda + \vartheta)(1 + \lambda)(6 + 2\lambda + \vartheta)} \alpha^2 + O(\alpha^3)
\]

▶ **Proof:** Use

\[
E[\Omega\langle \mu_\infty^N, e^{-\lambda r(x_1,x_2)/2} \rangle] = 0
\]
Summary and outlook

Once the **right state-space** is chosen, construction of tree-valued Fleming-Viot process straight-forward

Genealogical distances can be **computed** using generators

Next step: Include **recombination**