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

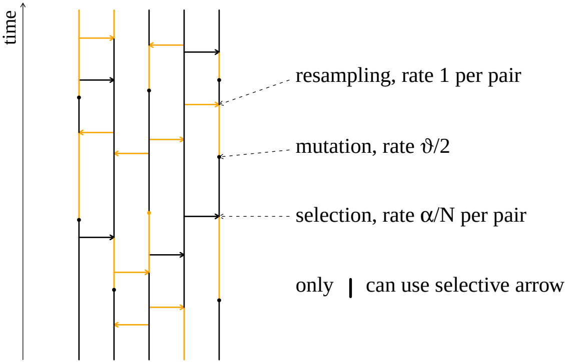
Populations of constant size have been **modelled** by

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

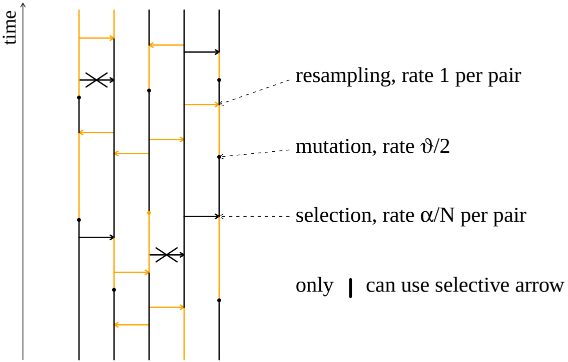
$$dX = \alpha X(1-X)dt + \sqrt{X(1-X)}dW$$

  or **Measure-valued diffusions** (Fleming-Viot superprocess)
- **New:** Extend Fleming-Viot process by genealogical information $\rightarrow$ **Tree-valued Fleming-Viot process**

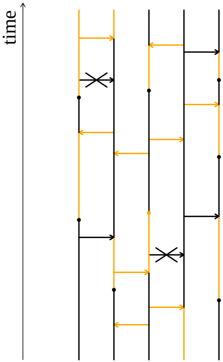
The Moran model with mutation and selection



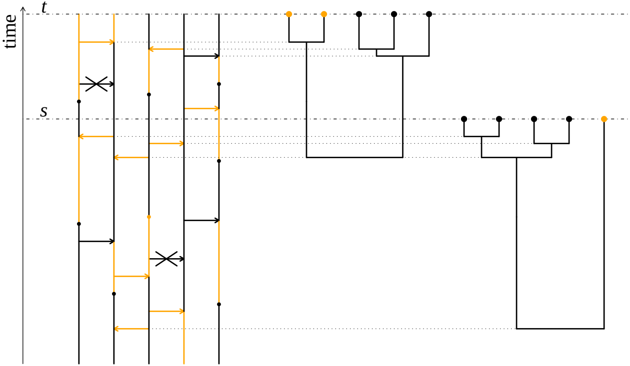
The Moran model with mutation and selection



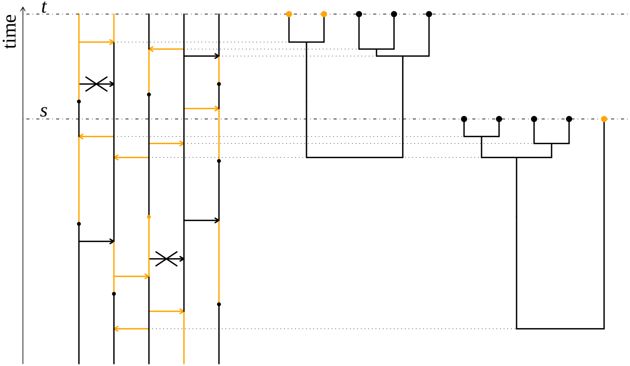
The Moran model with mutation and selection



The Moran model 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 (Ω, Π)-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 α 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}^{α}, 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) 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) complete and separable **metric** space, $\mu \in \mathcal{P}(X \quad)$

- $r(x_1, x_2)$ defines the genealogical distance of individuals x_1 and x_2
- μ 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:

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

- $r(x_1, x_2)$ defines the genealogical distance of individuals x_1 and x_2
- μ 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 $\mathcal{U}$:

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

- $r(x_1, x_2)$ defines the genealogical distance of individuals x_1 and x_2
- μ marks currently living individuals

Martingale Problem

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

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

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

$$\Big(\Phi(\mathcal{X}_t) - \int_0^t \Omega\Phi(\mathcal{X}_s)ds\Big)_{t\geq 0}$$

is a martingale. The MP is **well-posed** if there is exactly one such process.

Polynomials on $\mathcal{P}(I)$

Π: functions of the form (polynomials)

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

for $\underline{u} = (u_1, u_2, \dots), \phi \in \mathcal{C}_b(I^{\mathbb{N}})$ depending on finitely many coordinates

- Π **separates points** in $\mathcal{P}(I)$

Polynomials on $\mathbb{U}$

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

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

for $(\underline{x}, \underline{u}) = ((x_1, u_1), (x_2, u_2), \ldots), \phi \in \mathcal{C}_b(\mathbb{R}^{\binom{\mathbb{N}}{2}} \times I^{\mathbb{N}})$ depending on finitely many coordinates

- Π **separates points** in $\mathbb{X}$

Generator for the Fleming-Viot process: measure-valued

$$\Omega := \Omega^{\mathrm{res}} + \Omega^{\mathrm{mut}} + \Omega^{\mathrm{sel}}$$

- Ω^{res} : resampling
- Ω^{mut} : mutation
- Ω^{sel} : selection

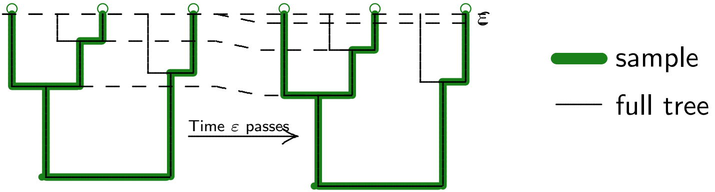
Generator for the Fleming-Viot process: tree-valued

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

- Ω^{grow} : tree growth
- Ω^{res} : resampling
- Ω^{mut} : mutation
- Ω^{sel} : selection

Tree Growth

When no resampling occurs the tree **grows**



Distances in the sample grow at speed 2

$$\Omega^{\text{grow}}\Phi(X,r,\mu) = 2 \cdot \left\langle \mu^{\mathbb{N}}, \sum_{i<j} \frac{\partial}{\partial r_{ij}} \phi \right\rangle.$$

Resampling: measure-valued

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

with

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

Resampling: tree-valued

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

with

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

In addition,

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

Mutation: measure-valued

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

$$\boldsymbol{\Omega}^{\mathsf{mut}}\boldsymbol{\Phi}(\quad \mu) = \vartheta \cdot \sum_{\mathbf{k}} \langle \mu^{\mathbb{N}}, \beta_{\mathbf{k}}\phi - \phi \rangle$$

with $\beta_k(\underline{u}, dv)$ acting on kth variable

Mutation: tree-valued

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

$$\boldsymbol{\Omega}^{\mathsf{mut}}\boldsymbol{\Phi}(\mathbf{X}, \mathbf{r}, \boldsymbol{\mu}) = \boldsymbol{\vartheta} \cdot \sum_{\mathbf{k}} \langle \boldsymbol{\mu}^{\mathbb{N}}, \boldsymbol{\beta}_{\mathbf{k}}\boldsymbol{\phi} - \boldsymbol{\phi} \rangle$$

with $\beta_k(\underline{u}, dv)$ acting on kth variable

Selection: measure-valued

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

$$\boldsymbol{\Omega^{\mathrm{sel}}\Phi(\qquad \mu)} := \boldsymbol{\alpha} \cdot \sum_{\mathbf{k=1}}^{\mathbf{n}} \langle \boldsymbol{\mu}^{\mathbb{N}}, \boldsymbol{\chi}_{\mathbf{k}} \cdot \phi - \boldsymbol{\chi}_{\mathbf{n+1}} \cdot \phi \rangle$$

where ϕ only depends on sample of size n
with χ_k acting on kth variable

Selection: tree-valued

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

$$\boldsymbol{\Omega}^{\mathrm{sel}}\boldsymbol{\Phi}(\mathbf{X},\mathbf{r},\boldsymbol{\mu}) := \boldsymbol{\alpha} \cdot \sum_{\mathbf{k}=1}^{\mathbf{n}} \langle \boldsymbol{\mu}^{\mathbb{N}}, \boldsymbol{\chi}_{\mathbf{k}} \cdot \boldsymbol{\phi} - \boldsymbol{\chi}_{\mathbf{n+1}} \cdot \boldsymbol{\phi} \rangle$$

where ϕ only depends on sample of size n
with χ_k acting on kth variable

Selection: tree-valued

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

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

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

$$= \alpha\cdot\sum_{l=1}^{n}\langle\mu^N, \chi_l\cdot\phi-\chi_{n+1}\cdot\phi\rangle$$

The tree-valued Fleming-Viot process

- **Theorem:** The (Ω, Π)-martingale problem is well-posed. Its solution $\mathcal{X} = (\mathcal{X}_t)_{\geq 0}$, $\mathcal{X}_t = (X_t, r_t, \mu_t)$ – the **tree-valued Fleming-Viot process** – arises as weak limit of tree-valued Moran models and satisfies:
  - $\mathbf{P}(t \mapsto \mathcal{X}_t \text{ is continuous}) = 1$,
  - $\mathbf{P}((X_t, r_t) \text{ is compact for all } t > 0) = 1$,
  - $\mathcal{X}$ is Feller (hence strong Markov)

Girsanov transform: measure-valued

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

$$\boldsymbol{\Psi(\quad \mu) := (\alpha' - \alpha) \cdot \langle \mu^{\mathbb{N}}, \chi_1 \rangle}$$

and

$$\mathcal{M} = \Big(\Psi(\quad \mu_t) - \Psi(\quad \mu_0) - \int_0^t \Omega\Psi(\quad \mu_s) ds \Big)_{t \geq 0}.$$

Then, $\mathbf{Q}$, defined by

$$\frac{d\mathbf{Q}}{d\mathbf{P}}\Big|_{\mathcal{F}_t} = e^{\mathcal{M}_t - \frac{1}{2}[\mathcal{M}]_t}$$

solves (Ω, Π)-MP **for selection coefficient α'**

Girsanov transform: tree-valued

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

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

and

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

Then, $\mathbf{Q}$, defined by

$$\frac{d\mathbf{Q}}{d\mathbf{P}}\bigg|_{\mathcal{F}_t} = e^{\mathcal{M}_t - \frac{1}{2}[\mathcal{M}]_t}$$

solves (Ω, Π)-MP **for selection coefficient** α'

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 $\bullet \to \bullet$ and $\bullet \to \bullet$
  - R^{α}_{12}: **distance of two randomly sampled points** in equilibrium

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

- Proof: Use

$$\mathbf{E}[\Omega\langle \mu^{\mathbb{N}}_{\infty}, e^{-\lambda r(x_1,x_2)/2}\rangle] = 0$$

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**