Path-properties of the tree-valued Fleming-Viot process

Peter Pfaffelhuber

Joint work with Andrej Depperschmidt and Andreas Greven

Luminy, 4.9.2012
As every population model, the Moran model (of size \( N \)) gives rise to a tree-valued process \((X_{Nt})_{t \geq 0}\).

Without proof, we assume that \((X_{Nt})_{t \geq 0}\) converges in an appropriate sense as \(N \to \infty\) to \((X_{t})_{t \geq 0}\), which we call the tree-valued Fleming-Viot dynamics.

Path-properties of the tree-valued Fleming-Viot process.
As every population model, the Moran model (of size $N$) gives rise to a tree-valued process $(X_{N,t})_{t \geq 0}$. Without proof, we assume that $(X_{N,t})_{t \geq 0}$ converges in an appropriate sense as $N \to \infty$ to $(X_{t})_{t \geq 0}$, which we call the tree-valued Fleming-Viot dynamics.

**The Moran model**

Path-properties of the tree-valued Fleming-Viot process
The Moran model

As every population model, the Moran model (of size $N$) gives rise to a tree-valued process $\left( X_t^N \right)_{t \geq 0}$

Without proof, we assume that $\left( X_t^N \right)_{t \geq 0} \xrightarrow{N \to \infty} \left( X_t \right)_{t \geq 0}$ in an appropriate sense

We call $\left( X_t \right)_{t \geq 0}$ the tree-valued Fleming-Viot dynamics
The tree-valued Fleming-Viot dynamics

**Theorem**

- The process \((X_t)_{t \geq 0}\) exists as limit of tree-valued Moran models and is unique. Its state space is the set of (equivalence classes of) metric measure spaces \(\{(X, r, \mu) : \mu \in M_1(X)\}\), equipped with the Gromov-Prohorov topology. It can be described by a martingale problem.

- Almost surely,
  - \((X_t)_{t \geq 0}\) has **continuous** sample paths.
  - \((X_t)_{t \geq 0}\) is **compact** for all \(t > 0\).
  - For many functions \(\Phi\), the **quadratic variation** of \((\Phi(X_t))_{t \geq 0}\) can be computed.
The Kingman coalescent as the equilibrium of \((X_t)_{t \geq 0}\)

**Theorem**

- Let \(X_\infty\) be the **Kingman coalescent**, a random tree with:
  - Start with \(\infty\) many lines
  - If there are \(k\) lines left, wait \(S_k \sim \exp \left( \frac{k}{2} \right)\) and merge two randomly chosen lines
  - Stop upon reaching one line

Then, \(X_t \xrightarrow{t \to \infty} X_\infty\)

- Important property:
  
  subtree with \(n\) leaves \(\sim\) tree started with \(n\) lines
Goal

‑ Lift properties of Kingman coalescent $X_\infty$ to the paths of $(X_t)_{t \geq 0}$ (when started in equilibrium)

‑ Example:
  ‑ Let $N^t_\varepsilon$ be the number of ancestors of the time-$t$ population at time $t - \varepsilon$
  ‑ It is well-known that almost surely

$$\varepsilon N^\infty_\varepsilon - 2 \xrightarrow{\varepsilon \downarrow 0} 0$$

Is it also true that almost surely

$$\sup_{t \geq 0} \left| \varepsilon N^t_\varepsilon - 2 \right| \xrightarrow{\varepsilon \downarrow 0} 0?$$
A law of large numbers for the number of ancestors

**Theorem**

Let $N_\varepsilon^\infty$ be the number of ancestors of the Kingman coalescent $X^\infty$ at time $\varepsilon$. Then, almost surely,

$$\varepsilon N_\varepsilon^\infty - 2 \xrightarrow{\varepsilon \downarrow 0} 0$$
A law of large numbers for the number of ancestors

\[\varepsilon N_\varepsilon^\infty - 2 \xrightarrow{\varepsilon \downarrow 0} 0\]

**Proof:** Recall \(S_k \sim \exp \left(\frac{k}{2}\right)\) is the time the coalescent has \(k\) lines. The assertion is the same as

\[\left(\sum_{n=1}^{\infty} S_n + S_{n+2} + \ldots\right)n - 2 \xrightarrow{n \to \infty} 0.\]

With \(\mathbb{E}[T_n] = 2/n\) and \(\mathbb{E}\left[(T_n - \mathbb{E}[T_n])^4\right] \lesssim \frac{1}{n^6}\), we find

\[\mathbb{P}\left(|T_n n - 2| > \varepsilon\right) \leq \frac{n^4 \mathbb{E}\left[(T_n - \mathbb{E}[T_n])^4\right]}{\varepsilon^4} \lesssim \frac{1}{\varepsilon^4 n^2}.\]

The result follows from the Borel-Cantelli Lemma.
A law of large numbers for the number of ancestors

Theorem

Let $N^t_\varepsilon$ be the number of ancestors of the time-$t$ population $X_t$, at time $t - \varepsilon$. Then, almost surely,

$$\sup_{t \geq 0} \left| \varepsilon N^t_\varepsilon - 2 \right| \xrightarrow{\varepsilon \downarrow 0} 0$$
A law of large numbers for the number of ancestors

\[
\sup_{t \geq 0} \left| \varepsilon N^t_\varepsilon - 2 \right| \xrightarrow{\varepsilon \downarrow 0} 0
\]

**Proof:** Let \( T^t_n = S^t_n + S^t_{n+1} + \ldots \) and \( S^t_n \) is the time the time-\( t \) tree spends with \( n \) lines.

\( \blacktriangleright \) It suffices to show \( \sup_{0 \leq t \leq 1} |T^t_n n - 2| \xrightarrow{n \to \infty} 0 \).

\( \blacktriangleright \) Using moment calculations,

\[
P \left( \sup_{k=0, \ldots, n^2} |T^{k/n^2}_n n - 2| > \varepsilon \right) \lesssim \frac{1}{\varepsilon^8 n^2}.
\]
A law of large numbers for the number of ancestors

It suffices to show \( \sup_{0 \leq t \leq 1} \left| T_n^t n - 2 \right| \xrightarrow{n \to \infty} 0 \).

We claim that \( \{ T_n^t > s \} \subseteq T_n^{t-\delta} > s - \delta \} \).

\[
\mathbb{P} \left( \sup_{0 \leq t \leq 1} T_n^t n > 2 + \varepsilon \right) \\
\leq \mathbb{P} \left( \sup_{0 \leq t \leq 1} T_n^{\lfloor tn^2 \rfloor / n^2} > \frac{2 + \varepsilon}{n} - \left( t - \frac{\lfloor tn^2 \rfloor}{n^2} \right) \right) \leq \ldots \lesssim \frac{1}{\varepsilon^8 n^2}.
\]

Path-properties of the tree-valued Fleming-Viot process
Small family sizes

▶ **Theorem**

Let $N^t(x, \varepsilon)$ be the number of ancestors of the time-$t$ population with families of size at least $x$. Then, almost surely

$$\sup_{0 \leq x < \infty} \left| \varepsilon N^\infty(x \varepsilon, \varepsilon) - 2e^{-2x} \right| \xrightarrow{\varepsilon \downarrow 0} 0.$$ 

▶ **Open problem:**

Is it also true, that

$$\sup_{t \geq 0} \sup_{0 \leq x < \infty} \left| \varepsilon N^t(x \varepsilon, \varepsilon) - 2e^{-2x} \right| \xrightarrow{\varepsilon \downarrow 0} 0?$$
A law of large numbers for the tree metric

Theorem
Let \( F^t_1(\varepsilon), \ldots, F^t_{N^t_\varepsilon}(\varepsilon) \) be the family sizes of the ancestors \( 1, \ldots, N^t_\varepsilon \) in \( X_t \). Then, almost surely,

\[
\frac{1}{N^t_\varepsilon} \sum_{i=1}^{N^t_\varepsilon} (F^t_i(\varepsilon))^2 = \lim_{N \to \infty} \frac{1}{\varepsilon N^2} \sum_{u,v=1}^{N^t_\varepsilon} 1\{r_t(u,v) < \varepsilon\} \xrightarrow{\varepsilon \downarrow 0} 1.
\]
A law of large numbers for the tree metric

Lemma

\[ \lim_{N \to \infty} \frac{\lambda}{N^2} \sum_{u,v=1}^{N} 1\{r_t(u,v) < 1/\lambda\} \xrightarrow{\lambda \to \infty} 1 \iff \Psi(\lambda)(X) \xrightarrow{\lambda \to \infty} 1. \]

with

\[ \Psi(\lambda)(X) := (\lambda + 1) \cdot \lim_{N \to \infty} \frac{1}{N^2} \sum_{u,v=1}^{N} e^{-\lambda r_t(u,v)}, \]

where \( r_t(u, v) \) is the time to the most recent common ancestor of \( u \) and \( v \) in \( X \).
A law of large numbers for the tree metric

**Theorem**

For $\Psi_\lambda$ as above, almost surely,

$$\Psi_\lambda(X_\infty) - 1 \xrightarrow{\lambda \to \infty} 0. \quad (*)$$

**Proof:** For random leaves $U, V, U', V'$ from $X_\infty$,

$$\mathbb{E}[\Psi_\lambda(X_\infty)] = \mathbb{E}[(\lambda + 1)e^{-\lambda r(U,V)}] = 1,$$

$$\mathbb{E}[(\Psi_\lambda(X_\infty) - 1)^2] = (\lambda + 1)^2 (\mathbb{E}[e^{-\lambda(r(U,V)+r(U',V'))}] - 1)$$

$$= \cdots \text{some calculations on tree with 4 leaves} \cdots$$

$$= \frac{2\lambda^2}{(\lambda + 3)(2\lambda + 1)(2\lambda + 3)} \xrightarrow{\lambda \to \infty} \frac{1}{2\lambda},$$

$$\mathbb{E}[(\Psi_\lambda(X_\infty) - 1)^4] = \cdots \xrightarrow{\lambda \to \infty} \frac{3}{4\lambda^2}$$
A law of large numbers for the tree metric

**Theorem**

- For $\Psi_\lambda$ as above, in probability,

\[
\sup_{0 \leq t \leq T} \left| \Psi_\lambda(X_t) - 1 \right| \xrightarrow{\lambda \to \infty} 0. \quad (**)
\]

- $(*) \Rightarrow fdd$-convergence in $(**)$

**Lemma**

\[
\sup_{\lambda > 0} \mathbb{E}[(\Psi_\lambda(X_t) - \Psi_\lambda(X_0))^4] \lesssim t^2.
\]

$\Rightarrow$ tightness in $C_\mathbb{R}[0, \infty)$
A law of large numbers for the tree metric

**Proof of Lemma**: Recall

\[
\mathbb{E}[\Psi_\lambda(X_t)] = (\lambda + 1)\mathbb{E}[e^{-\lambda r_t(U,V)}]
\]

By the dynamics of the Moran model,

\[
\frac{d}{dt}\mathbb{E}[\Psi_\lambda(X_t)] = \frac{\lambda + 1}{dt} \left( \mathbb{E}[e^{-\lambda \cdot 0} - e^{-\lambda r_t(U,V)}] \right)
\]

\[
= \frac{\lambda + 1}{dt} \left( \mathbb{E}[e^{-\lambda r_t(U,V)+dt} - e^{-\lambda r_t(U,V)}] \right)
\]

\[
= (\lambda + 1)(1 - \mathbb{E}[(\lambda + 1)e^{-\lambda r_t(U,V)}])
\]

\[
= -(\lambda + 1)(\mathbb{E}[\Psi_\lambda(X_t)] - 1).
\]

Path-properties of the tree-valued Fleming-Viot process
A law of large numbers for the tree metric

Proof of Lemma: Recall

\[
\frac{d}{dt}(\mathbb{E}[\psi_\lambda(X_t)] - 1) = -(\lambda + 1)(\mathbb{E}[\psi_\lambda(X_t)] - 1)
\]

Similarly,

\[
\mathbb{E}[\psi_\lambda(X_t) - 1 | \mathcal{F}_s] = e^{-(\lambda+1)(t-s)}(\psi_\lambda(X_s) - 1)
\]

and

\[
\left( e^{(\lambda+1)t} \left( \psi_\lambda(X_t) - 1 \right) \right)_{t \geq 0} \text{ is a martingale}
\]
A law of large numbers for the tree metric

Proof of Lemma:

\[ \mathbb{E}[(\Psi_\lambda(X_t) - \Psi_\lambda(X_0))^2] \]
\[ = -\mathbb{E}[2\Psi_\lambda(X_0)(\Psi_\lambda(X_t) - \Psi_\lambda(X_0))] \]
\[ = -2\mathbb{E}\left[ \Psi_\lambda(X_0) \left( e^{-(\lambda+1)t} \mathbb{E}\left[ e^{(\lambda+1)t} \left( \Psi_\lambda(X_t) - 1 \right) \right| \mathcal{F}_0 \right) \right. \]
\[ \quad \left| \text{martingale} \right. \]
\[ \quad - \left. \left( \Psi_\lambda(X_0) - 1 \right) \right] \]
\[ = 2\mathbb{E}\left[ \Psi_\lambda(X_0)(\Psi_\lambda(X_0) - 1)(1 - e^{-(\lambda+1)t}) \right] \]
\[ \lambda \rightarrow \infty \quad 1 - e^{-(\lambda+1)t} \leq t \]
A law of large numbers for the tree metric

Proof of Lemma:

\[ \mathbb{E}[(\psi_{\lambda}(X_t) - \psi_{\lambda}(X_0))^4] \]

\[ \cdots \]

\[ \cdots \]

\[ \cdots \]

\[ \cdots \]

\[ \lambda \to \infty \approx 1 - e^{-(\lambda + 1)t} \]

\[ 2 \lambda \lesssim t^2 \]

Path-properties of the tree-valued Fleming-Viot process
A Brownian motion in the Fleming-Viot dynamics

**Theorem**

Let $\mathcal{W}_\lambda = (W_\lambda(t))_{t \geq 0}$ be given by

$$W_\lambda(t) := \lambda \int_0^t (\Psi_\lambda(X_s) - 1)ds.$$ 

Then,

$$\mathcal{W}_\lambda \xrightarrow{\lambda \to \infty} \mathcal{W},$$

where $\mathcal{W} = (W_t)_{t \geq 0}$ is a Brownian motion.
A Brownian motion in the Fleming-Viot dynamics

\[ \mathcal{W}_\lambda = \lambda \int_0^\cdot \left( \frac{\Psi_\lambda(X_s) - 1}{\lambda} \right) ds \xrightarrow{\lambda \to \infty} \mathcal{W} \]

mean = 0

variance \approx \frac{1}{2\lambda}

**Proof** (part):

\[ \mathbb{E}[\mathcal{W}_\lambda(t)^2] = 2\lambda^2 \int_0^t \int_0^s \mathbb{E}\left[\mathbb{E}[\Psi_\lambda(X_s) - 1|\mathcal{F}_r](\Psi_\lambda(X_r) - 1)] dr \right] ds \]

\[ = 2\lambda^2 \int_0^t \int_0^s e^{-(\lambda+1)(s-r)} \mathbb{E}[\Psi_\lambda(X_r) - 1)^2] dr \right] ds \]

\[ \xrightarrow{\lambda \to \infty} \lambda \int_0^t \int_0^s e^{-(\lambda+1)r} dr \right] ds \xrightarrow{\lambda \to \infty} t \]
More about **formalising trees** (and Gromov-Prohorov convergence) and **construction of tree-valued processes** (via well-posed martingale problem) can be said.

All result also hold in models with mutation and **selection** (individuals also carry types which are (dis)favored to get offspring).

All Theorems affect properties near the tree top → do similar properties hold for **branching trees**?