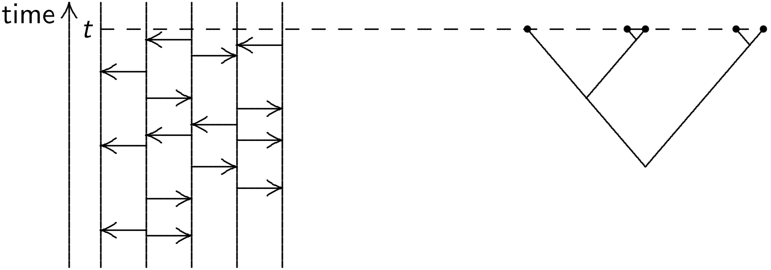
Path-properties of the tree-valued Fleming-Viot process

Peter Pfaffelhuber

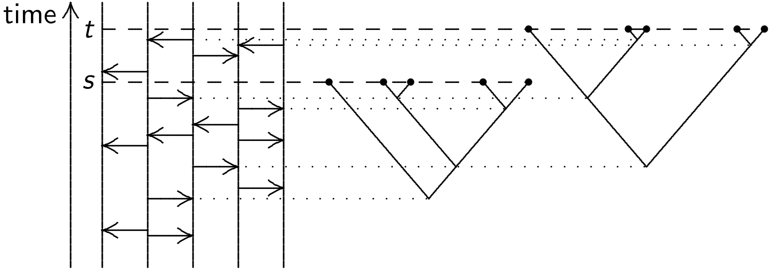
Joint work with Andrej Depperschmidt and Andreas Greven

Luminy, 4.9.2012

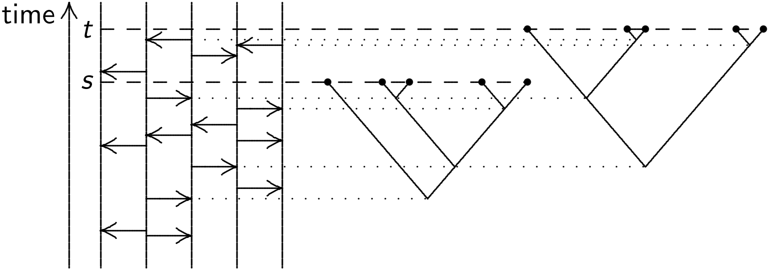
The Moran model



The Moran model



The Moran model



- As every population model, the Moran model (of size N) gives rise to a **tree-valued process** $\mathbf{(X_t^N)_{t\geq 0}}$
- Without proof, we assume that $(X_t^N)_{t\geq 0} \xrightarrow{N\to\infty} (X_t)_{t\geq 0}$ in an appropriate sense
- We call $(X_t)_{t\geq 0}$ the **tree-valued Fleming-Viot dynamics**

The tree-valued Fleming-Viot dynamics

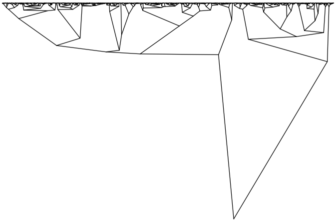
Theorem

- 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 $\{(X, r, \mu) : \mu \in \mathcal{M}_1(X)\}$, 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 Φ, 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_∞ be the **Kingman coalescent**, a random tree with:
  - Start with ∞ many lines
  - If there are k lines left, wait $S_k \sim \exp\binom{k}{2}$ and merge two randomly chosen lines
  - Stop upon reaching one line



- Then, $\mathbf{X_t} \xRightarrow{\mathbf{t\to\infty}} \mathbf{X_\infty}$

- Important property:

subtree with n leaves $\sim$ tree started with n lines

Goal

- **Lift properties of Kingman coalescent $\mathbf{X}_\infty$ to the paths of $(\mathbf{X}_\mathbf{t})_{\mathbf{t}\geq \mathbf{0}}$** (when started in equilibrium)

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

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

Is it also true that almost surely

$$\sup_{t\geq 0} \left| \varepsilon N_\varepsilon^t - 2 \right| \xrightarrow{\varepsilon \downarrow 0} 0?$$

A law of large numbers for the number of ancestors

Theorem

- Let $\mathbf{N}_\varepsilon^\infty$ be the the **number of ancestors of the Kingman coalescent $\mathbf{X}_\infty$** at time ε. 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 $S_k \sim \exp\binom{k}{2}$ is the time the coalescent has k lines. The assertion is the same as

$$(\underbrace{S_{n+1} + S_{n+2} + \ldots}_{=: T_n})n - 2 \xrightarrow{n\to\infty} 0.$$

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

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

The result follows from the Borel-Cantelli-Lemma.

A law of large numbers for the number of ancestors

Theorem

- Let $\mathbf{N}^{\mathbf{t}}_{\varepsilon}$ be the the **number of ancestors of the time-**t **population** $\mathbf{X_t}$, 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

$$\sup_{t\geq 0}\left|\varepsilon N_\varepsilon^t - 2\right| \xrightarrow{\varepsilon\downarrow 0} 0$$

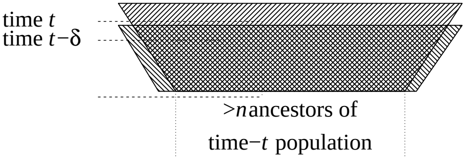
- **Proof**: Let $T_n^t = S_n^t + S_{n+1}^t + \ldots$ and S_n^t is the time the time-t tree spends with n lines.
- It suffices to show $\sup_{0\leq t\leq 1}|T_n^t n - 2| \xrightarrow{n\to\infty} 0$.
- Using moment calculations,

$$\mathbb{P}\Big(\sup_{k=0,\ldots,n^2}|T_n^{k/n^2} n - 2| > \varepsilon\Big) \lesssim \frac{1}{\varepsilon^8 n^2}.$$

A law of large numbers for the number of ancestors

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

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



$$\Rightarrow \mathbb{P}\Big(\sup_{0\le t\le 1} T_n^t n > 2+\varepsilon\Big)$$

$$\le \mathbb{P}\Big(\sup_{0\le t\le 1} T_n^{\lfloor tn^2\rfloor/n^2} > \frac{2+\varepsilon}{n} - \underbrace{\Big(t-\frac{\lfloor tn^2\rfloor}{n^2}\Big)}_{\le 1/n^2}\Big) \le \ldots \lesssim \frac{1}{\varepsilon^8 n^2}.$$

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

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

- **Open problem**:
  Is it also true, that

$$\sup_{t\ge 0}\sup_{0\le x<\infty}\left|\varepsilon N^t(x\varepsilon,\varepsilon)-2e^{-2x}\right| \xrightarrow{\varepsilon\downarrow 0} 0?$$

A law of large numbers for the tree metric

Theorem

Let $F_1^t(\varepsilon), ..., F_{N_\varepsilon}^t(\varepsilon)$ be the family sizes of the ancestors $1, ..., N_\varepsilon^t$ in X_t. Then, almost surely,

$$\frac{1}{\varepsilon}\sum_{i=1}^{N_\varepsilon^t}(F_i^t(\varepsilon))^2 = \lim_{N\to\infty}\frac{1}{\varepsilon N^2}\sum_{\substack{u,v=1\\ \text{leaves in } X_t}}^{N} 1_{\{r_t(u,v)<\varepsilon\}} \xrightarrow{\varepsilon\downarrow 0} 1.$$

A law of large numbers for the tree metric

- **Lemma**

$$\lim_{N\to\infty} \frac{\lambda}{N^2} \sum_{\substack{u,v=1 \\ \text{leaves in } X}}^{N} 1_{\{r_t(u,v)<1/\lambda\}} \xrightarrow{\lambda\to\infty} 1 \iff \Psi_\lambda(X) \xrightarrow{\lambda\to\infty} 1.$$

with

$$\Psi_\lambda(X) := (\lambda+1)\cdot \lim_{N\to\infty} \frac{1}{N^2} \sum_{\substack{u,v=1 \\ \text{leaves in } X}}^{N} e^{-\lambda r_t(u,v)},$$

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 Ψ_λ as above, almost surely,

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

- **Proof**: For random leaves U, V, U', V' from X_∞,

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

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

$= \cdots$ some **calculations on tree with 4 leaves** $\cdots$

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

$$\mathbb{E}[\big(\Psi_\lambda(X_\infty) - 1\big)^4] = \cdots \overset{\lambda\to\infty}{\approx} \frac{3}{4\lambda^2}$$

A law of large numbers for the tree metric

Theorem

- For Ψ_λ as above, in probability,

$$\sup_{0\le t\le T}\Big|\Psi_\lambda(X_t)-1\Big| \xrightarrow{\lambda\to\infty} 0. \qquad (**)$$

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

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

$\Rightarrow$ tightness in $\mathcal{C}_{\mathbb{R}}[0,\infty)$

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,

$$\begin{aligned}
&\frac{d}{dt}\mathbb{E}[\Psi_\lambda(X_t)] \\
&\quad = \frac{\lambda+1}{dt}\Big(\underbrace{dt\cdot\mathbb{E}[e^{-\lambda\cdot 0} - e^{-\lambda r_t(U,V)}]}_{\textbf{resampling}\text{ between } U \text{ and } V} \\
&\qquad\qquad + \underbrace{(1-dt)\cdot\mathbb{E}[e^{-\lambda(r_t(U,V)+dt)} - e^{-\lambda r_t(U,V)}]}_{\text{tree }\textbf{growth}}\Big) \\
&\quad = (\lambda+1)(1-\mathbb{E}[(\lambda+1)e^{-\lambda r_t(U,V)}]) \\
&\quad = -(\lambda+1)(\mathbb{E}[\Psi_\lambda(X_t)]-1).
\end{aligned}$$

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

$$\left(\mathbf{e}^{(\boldsymbol{\lambda}+\mathbf{1})\mathbf{t}}\Big(\boldsymbol{\Psi}_{\boldsymbol{\lambda}}(\mathbf{X}_\mathbf{t}) - \mathbf{1}\Big)\right)_{\mathbf{t}\geq\mathbf{0}} \textbf{is a martingale}$$

A law of large numbers for the tree metric

- **Proof of Lemma**:

$$\begin{aligned}
&\mathbb{E}[(\Psi_\lambda(X_t) - \Psi_\lambda(X_0))^2] \\
&\quad = -\mathbb{E}[2\Psi_\lambda(X_0)(\Psi_\lambda(X_t) - \Psi_\lambda(X_0))] \\
&\quad = -2\mathbb{E}\Big[\Psi_\lambda(X_0) \\
&\qquad \Big(e^{-(\lambda+1)t}\mathbb{E}\Big[\underbrace{e^{(\lambda+1)t}\Big(\Psi_\lambda(X_t) - 1\Big)}_{\text{martingale}}\Big|\mathcal{F}_0\Big] \\
&\qquad\qquad - \Big(\Psi_\lambda(X_0) - 1\Big)\Big] \\
&\quad = 2\mathbb{E}\Big[\Psi_\lambda(X_0)(\Psi_\lambda(X_0) - 1)(1 - e^{-(\lambda+1)t})\Big] \\
&\quad \overset{\lambda\to\infty}{\approx} \frac{1 - e^{-(\lambda+1)t}}{\lambda} \lesssim t
\end{aligned}$$

A law of large numbers for the tree metric

- **Proof of Lemma**:

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

...

...

...

...

... $\lesssim t^2$

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

$$\mathcal{W}_\lambda = \lambda \int_0^\bullet \big(\underbrace{\Psi_\lambda(X_s) - 1}_{\text{mean } = 0} \big) ds \xRightarrow{\lambda\to\infty} \mathcal{W}$$

variance $\approx 1/2\lambda$

- **Proof** (part):

$$\begin{aligned}\mathbb{E}[W_\lambda(t)^2] &= 2\lambda^2 \int_0^t \int_0^s \mathbb{E}\big[\mathbb{E}[\Psi_\lambda(X_s) - 1|\mathcal{F}_r](\Psi_\lambda(X_r) - 1)\big] dr\, ds \\ &= 2\lambda^2 \int_0^t \int_0^s e^{-(\lambda+1)(s-r)} \mathbb{E}[(\Psi_\lambda(X_r) - 1)^2] dr\, ds \\ &\overset{\lambda\to\infty}{\approx} \lambda \int_0^t \int_0^s e^{-(\lambda+1)r} dr\, ds \overset{\lambda\to\infty}{\approx} t\end{aligned}$$

Summary and outlook

- 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
  $\rightarrow$ do similar properties hold for **branching trees?**