A dual process for the spatial Moran model with mutation and selection carrying the family decomposition with founding fathers

Peter Seidel
University of Erlangen

Mini-Workshop of the DFG priority programme SPP 1590
Berlin, 6 November 2014

The Moran model

- **Migration:** Each individual performs a continuous time random walk on the set of sites in geographic space.
- **Mutation:** The type of each individual evolves as a continuous time Markov chain on the type space.
- **Resampling:** Each pair of individuals located at the same site dies after an exponential holding time and gives birth to a new pair of individuals, called descendants, which randomly choose an ancestor from the dying pair from which both descendants inherit the type.
- **Selection:** The ancestor is chosen according to a probability distribution depending on the fitness of the individuals in the dying pair, that is, the fitter an individual is the more likely it is chosen as an ancestor.

General goal [PhD thesis, submitted September '14]

The evolution of genealogical information of a locally finite Moran population forward in time, in particular, the evolution of ancestral lines and genealogical distances.

This includes:

- Refined description of the Moran Model
- Definition of an appropriate state space
- Analytical characterization

The results:

- Dual process to express fixed time probabilities of a subpopulation
- Strong duality for the conditional law of genealogical information of a subpopulation given the site-type information of the subpopulation

Background and motivation

- **Historical processes:**
  [*Dawson* & *Perkins* '91] for branching populations,
  [*Greven*, *Limic* & *Winter* '05] spatial Moran model without mutation and selection

- **Tree-valued Moran model:**
  [*Depperschmidt*, *Greven* & *Pfaffelhuber* '12 & '13] genealogical distances are given by a marked ultrametric measure space

- **Different approaches:**
  [*Krone* & *Neuhauser* '97] ancestral selection graph,
  [*Donnelly* & *Kurtz* '96 & '99] look-down construction

Outline of this talk

1. **MMFF:** The evolution of both the site-type information and the family decomposition with founding fathers of a finite Moran population on a finite set of sites

2. **Duality:** Feynman-Kac duality, introduction of the state space, main features of the dual process

3. **Applications:** Fixation probabilities, conditioned family decomposition with founding fathers given the site-type information

MMFF: Ingredients

Let

- I be a finite set (life-sites), e.g. $I = \{1, \ldots, N\}$
- G be a finite Abelian group describing the sites (or locations)
- K be a finite set describing the types
- $\chi : K \to [0, 1]$ be a function determining the fitness

At each time t the population can be modeled by an element in

$$(G \times K)^I \, .$$

MMFF: Ingredients

Let

- I be a finite set (life-sites), e.g. $I = \{1, \dots, N\}$
- G be a finite Abelian group describing the sites (or locations)
- K be a finite set describing the types
- $\chi : K \to [0, 1]$ be a function determining the fitness

At each time t the population can be modeled by an element in

$$(G \times K)^I \, .$$

Problem: If two life-sites have the same type, their states do not change in a resampling event.

MMFF: Refined description

Solution: Extend the state space and use

$$(G \times K \times \textsc{Names})^I \ !$$

Concept:

- Take a copy $\tilde{I}$ of I and assign

  at time 0 to each $i \in I$ its associated element $\tilde{i} \in \tilde{I}$,

  that is,

  $\tilde{I}$ models the first generation of individuals.

- Assign in each resampling event to both life-sites a new NAME, namely in such a way that each NAME contains the NAMES of all its ancestors.

MMFF: Focus on the founding fathers

Instead of the complete NAME consider **only** the ancestor alive at time 0 which is a member of

the first generation of individuals $=:$ founding fathers.

This means:
If we consider $i \in I$ at time t, then we do not assign the NAME of the individual occupying i but

its founding father which is an element in $\tilde{I}$.

MMFF: The process $(X_t)_{t\geq 0}$

At each time t

$$X_t = \left((X_t)_{i,G}\,,(X_t)_{i,K}\,,(X_t)_{i,\tilde{I}}\right)_{i\in I} \;\in\; \left(G\times K\times \tilde{I}\right)^{I},$$

where

- $(X_t)_{i,G}$ = the location of the individual occupying i
- $(X_t)_{i,K}$ = the type of the individual occupying i
- $(X_t)_{i,\tilde{I}}$ = the founding father of the individual occupying i.

For $J\subset I$ use the abbreviation

$$(X_t)^J_{G\times K} = \left((X_t)_{j,G}\,,(X_t)_{j,K}\right)_{j\in J} \text{ as well as } (X_t)^J_{\tilde{I}} = \left((X_t)_{j,\tilde{I}}\right)_{j\in J}\ .$$

MMFF: Information contained in $(X_t)_{\tilde{I}}^{J}$

Family decomposition of J with founding fathers:

- For $\tilde{i} \in \tilde{I}$ the set

$$\left\{ j \in J : (X_t)_{j,\tilde{I}} = \tilde{i} \right\}$$

  describes the life-sites in J which are occupied by the descendants of $\tilde{i}$ at time t and hence a family of the subpopulation occupying J at time t.

- Doing this for each $\tilde{i} \in \tilde{I}$ one gets a partition of J which is enriched by the information on the founding fathers.

MMFF: Evolution of $(X_t)_{t\geq 0}$, initial state

Set

$$(X_0)^I_{\tilde{I}} = \mathbb{I} \in \tilde{I}^I, \text{ where } \mathbb{I}_i = \tilde{i} \text{ for all } i \in I$$

and

$$\mathcal{L}\left[(X_0)^I_{G\times K}\right] = \mu \in \mathcal{M}_1\left((G \times K)^I\right) ,$$

where μ is the initial site-type distribution.

MMFF: Evolution of $(X_t)_{t\geq 0}$, initial state

Set

$$(X_0)^I_{\tilde{I}} = \mathbb{I} \in \tilde{I}^I, \text{ where } \mathbb{I}_i = \tilde{i} \text{ for all } i \in I$$

and

$$\mathcal{L}\left[(X_0)^I_{G\times K}\right] = \mu \in \mathcal{M}_1\left((G \times K)^I\right) ,$$

where μ is the initial site-type distribution.

For the jump process description let

$$\eta = \left(\eta_{i,G}, \eta_{i,K}, \eta_{i,\tilde{I}}\right)_{i\in I} \in \left(G \times K \times \tilde{I}\right)^I$$

be a typical element of the state space.

MMFF: Evolution of $(X_t)_{t\geq 0}$, migration

If

$$\left(\eta_{i,G}, \eta_{i,K}, \eta_{i,\tilde{I}}\right)$$

is the current state of the life-site $i \in I$, then

$$\eta_{i,G} \to x \;\text{ at rate }\; a(\eta_{i,G}, x) = a(0, x - \eta_{i,G}) \;,$$

where $a(\cdot,\cdot)$ is a stochastic and homogeneous matrix.

MMFF: Evolution of $(X_t)_{t\geq 0}$, mutation

If

$$\left(\eta_{i,G}, \eta_{i,K}, \eta_{i,\tilde{I}}\right)$$

is the current state of the life-site $i \in I$, then

$$\eta_{i,K} \to u \quad \text{at rate} \quad b(\eta_{i,K}, u) \ ,$$

where $b(\cdot, \cdot)$ is a general stochastic matrix.

MMFF: Evolution of $(X_t)_{t\geq 0}$, resampling and selection

If

$$\left(\eta_{i,G}, \eta_{i,K}, \eta_{i,\tilde{I}}\right) \text{ and } \left(\eta_{j,G}, \eta_{j,K}, \eta_{j,\tilde{I}}\right)$$

are the current states of the life-sites $i, j \in I$, then

$$\left(\eta_{j,G}, \eta_{j,K}, \eta_{j,\tilde{I}}\right) \to \left(\eta_{j,G}, \eta_{i,K}, \eta_{i,\tilde{I}}\right)$$

at rate

$$\mathbb{1}\left\{\eta_{i,G} = \eta_{j,G}\right\} \frac{1}{2}\left[1 + \chi\left(\eta_{i,K}\right) - \chi\left(\eta_{j,K}\right)\right] \ .$$

MMFF: Aims concerning the law at a fixed time T

For tagged $J \subset I$ consider probabilities of the form

$$P_{(\mu,\mathbb{I})}\left((X_T)^J_{G\times K} = \xi^J_{G\times K}, (X_T)^J_{\tilde{I}} = \xi^J_{\tilde{I}}\right) \, ,$$

where

$$\xi^J_{G\times K} \in (G \times K)^J \text{ and } \xi^J_{\tilde{I}} \in \tilde{I}^J \, .$$

The plan is to express these probabilities in terms of a dual process that

starts in $\xi^J_{G\times K}$ as well as in $\mathbb{J} \in \tilde{I}^J$ given by $\mathbb{J}_j = \tilde{j}$ for all $j \in I$

and

is restricted to reach $\xi^J_{\tilde{I}}$.

Note: First of all there are no restrictions on μ!

MMFF: Aims concerning the law at a fixed time T

Use the Feynman-Kac duality to obtain a stochastic representation (**strong duality**) for the conditional law

$$P_{(\mu,\mathbb{I})}\left((X_T)^J_{\tilde{I}} \in \cdot \;\middle|\; (X_T)^J_{G\times K} = \xi^J_{G\times K}\right) \in \mathcal{M}_1\left(\tilde{I}^J\right)$$

for the case where μ is **nice**, i.e.

- initial locations are chosen uniformly and independently
- initial types are exchangeable

Duality: Result

For tagged $J \subset I$ there is a dual process $(\bar{X}_t)_{t\geq 0}$ with

$$\bar{X}_t = \left(\bar{X}_t^J, \bar{X}_t^{\tilde{I}}\right) \in \left(G \times K \times \tilde{I}\right)^J \times \left(G \times 2^K\right)^{\tilde{I}} ,$$

a family of duality functions

$$\left\{H_{\xi_{\tilde{I}}^J} : \xi_{\tilde{I}}^J \in \tilde{I}^J\right\}$$

and a Feynman-Kac function V such that for each $\xi_{\tilde{I}}^J$,

$$E_\eta\left[H_{\xi_{\tilde{I}}^J}(X_t, \bar{\eta})\right] = \bar{E}_{\bar{\eta}}\left[H_{\xi_{\tilde{I}}^J}\left(\eta, \bar{X}_t\right) \exp\left(\int_0^t V(\bar{X}_s)ds\right)\right]$$

holds for all $\eta \in (G \times K \times \tilde{I})^I$, all $\bar{\eta} \in \left(G \times K \times \tilde{I}\right)^J \times \left(G \times 2^K\right)^{\tilde{I}}$, and all $t \geq 0$.

Duality: Main features of the dual process

Reversal of migration and mutation:

- Use $\bar{a}(x,y) := a(y,x)$
- Use $\bar{b}(u,v) := b(v,u)$, but observe that in general $\bar{b}(\cdot,\cdot)$ is not a stochastic matrix

Marked coalescent structure in $\bar{X}_t^J$:

- $(\bar{X}_t^J)_{i,\tilde{I}} = (\bar{X}_t^J)_{j,\tilde{I}}$ implies that $(\bar{X}_t^J)_{i,G\times K} = (\bar{X}_t^J)_{j,G\times K}$
- Partition of J is given by $i \sim j \iff (\bar{X}_t^J)_i = (\bar{X}_t^J)_j$
- Partition elements located at the same site can merge together if they have the same type

Duality: Main features of the dual process

Interactions between $\bar{X}_t^J$ and $\bar{X}_t^{\tilde{I}}$:

- For $\tilde{i} \in \tilde{I}$, $(\bar{X}_t^{\tilde{I}})_{\tilde{i}}$ only evolves if $\tilde{i}$ is not occupied by a partition element.
- Due to selection partition elements create subsets of K when they merge together.

The case without selection:

- There are still interactions between $\bar{X}_t^J$ and $\bar{X}_t^{\tilde{I}}$.
- If we drop the information on the founding fathers, then $(\bar{X}_t^J)_{G \times K}$ evolves as a $(G \times K)$-marked J-coalescent.

Duality: Relation to fixed time T probabilities

Recall

$$P_{(\mu,\mathbb{I})}\left((X_T)^J_{G\times K} = \xi^J_{G\times K}, (X_T)^J_{\tilde{I}} = \xi^J_{\tilde{I}}\right)$$

and

$$E_\eta\left[H_{\xi^J_{\tilde{I}}}(X_T, \bar{\eta})\right] = \bar{E}_{\bar{\eta}}\left[H_{\xi^J_{\tilde{I}}}\left(\eta, \bar{X}_T\right)\exp\left(\int_0^T V(\bar{X}_s)ds\right)\right] .$$

Remember also that we want to start the dual process with the information

contained in $\xi^J_{G\times K}$ and $\mathbb{J} \in \tilde{I}^J$ given by $\mathbb{J}_j = \tilde{j}$ for all $j \in J$.

Duality: Relation to fixed time T probabilities

For $\mathbf{y} = (y_i)_{i \in I \setminus J} \in G^{I \setminus J}$ and $\xi^J_{G \times K} \in (G \times K)^J$ let

$$\bar{\eta}\left(\xi^J_{G \times K}, \mathbb{J}, \mathbf{y}\right) := (\bar{\eta}^J, \bar{\eta}^I) \in \left(G \times K \times \tilde{I}\right)^J \times \left(G \times 2^K\right)^{\tilde{I}}$$

be the element given by

$$\bar{\eta}^J_{G \times K} = \xi^J_{G \times K}, \; \bar{\eta}^J_{\tilde{I}} = \mathbb{J} \;\text{ and }\; \bar{\eta}^{\tilde{I}}_{\tilde{i}} = (y_i, K) \;\text{ for all } i \in I \setminus J\,,$$

where $\mathbb{J}_j = \tilde{j}$ for all $j \in J$.

Then

$$\begin{aligned} & H_{\xi^J_{\tilde{I}}}\left(X_T, \bar{\eta}\left(\xi^J_{G \times K}, \mathbb{J}, \mathbf{y}\right)\right) \\ = \;\; & \mathbb{1}\left\{(X_T)^J_{G \times K} = \xi^J_{G \times K}, (X_T)^J_{\tilde{I}} = \xi^J_{\tilde{I}}, (X_T)^{I \setminus J}_G = \mathbf{y}\right\} \end{aligned}$$

Duality: Relation to fixed time T probabilities

$$
\begin{aligned}
& P_{(\mu,\mathbb{I})}\left((X_T)^J_{G\times K} = \xi^J_{G\times K}, (X_T)^J_{\tilde{I}} = \xi^J_{\tilde{I}}\right) \\
= \quad & \bar{E}_{\bar{\eta}(\xi^J_{G\times K},\mathbb{J},\bar{\mu}^{\tilde{I}})}\left[\mathbb{1}\left\{\left(\bar{X}^J_T\right)_{\tilde{I}} = \xi^J_{\tilde{I}}\right\} h(\bar{X}_T) \exp\left(\int_0^T V(\bar{X}_s)ds\right)\right],
\end{aligned}
$$

where

$$h(\bar{X}_T) \text{ depends on } \mu \, .$$

Remark:

If μ is **nice**, there is **no selection** and the information on **founding fathers** is **dropped**, then the right hand side of this formula can be expressed in terms of a $(G \times K)$-**marked** J-**coalescent**.

Applications: Site-type information, strong selection

Let $|G| = 1$, $K = \{0, 1\}$, $I = \{1, \ldots, N\}$, $b(1, 1) = b(0, 0) = 1$, $\mu = \nu^I$ with $\nu \in \mathcal{M}_1(K)$ and $s := \chi(1) - \chi(0) > 0$.

Applications: Site-type information, strong selection

Let $|G| = 1$, $K = \{0, 1\}$, $I = \{1, \ldots, N\}$, $b(1,1) = b(0,0) = 1$, $\mu = \nu^I$ with $\nu \in \mathcal{M}_1(K)$ and $s := \chi(1) - \chi(0) > 0$.

For each $J \subset I$,

$$P_\mu \left((X_T)_K^J = \mathbf{0}^J \right) = \bar{E}_{|J|} \left[\nu(\{0\})^{\bar{Y}_T^N} \right] ,$$

where $(\bar{Y}_t^N)_{t \geq 0}$ is a **birth** and **death** process on I in which

$$n \to n+1 \text{ at } \textbf{rate } s(N-n)n$$

and

$$n \to n-1 \text{ at } \textbf{rate } (1-s)\binom{n}{2} .$$

Applications: Fixation probabilities, strong selection

Let $|G| = 1$, $K = \{0,1\}$, $I = \{1,\dots,N\}$, $b(1,1) = b(0,0) = 1$, $\mu = \nu^I$ with $\nu \in \mathcal{M}_1(K)$ and $s := \chi(1) - \chi(0) > 0$.

For each $J \subset I$,

$$\lim_{T\to\infty} P_\mu\left((X_T)_K^J = \mathbf{0}^J\right) = \frac{1-\left(1+\frac{2s}{1-s}\nu(\{0\})\right)^N}{1-\left(1+\frac{2s}{1-s}\right)^N}\,.$$

For fixed finite $J \subset \mathbb{N}$,

$$\lim_{N\to\infty} P_\mu\left((X_T)_K^J = \mathbf{0}^J\right) = \begin{cases} 1, & \nu(\{0\}) = 1 \\ 0, & \nu(\{0\}) \in [0,1) \end{cases}$$

for all $T > 0$.

Applications: Site-type information, weak selection

Let $|G| = 1$, $K = \{0, 1\}$, $I = \{1, \ldots, N\}$, $b(1,1) = b(0,0) = 1$, $\mu = \nu^I$, $\chi(0) = 0$ and $\chi(1) = \frac{S}{N}$ with fixed $S \in [0, N]$.

Applications: Site-type information, weak selection

Let $|G| = 1$, $K = \{0, 1\}$, $I = \{1, \dots, N\}$, $b(1,1) = b(0,0) = 1$, $\mu = \nu^I$, $\chi(0) = 0$ and $\chi(1) = \frac{S}{N}$ with fixed $S \in [0, N]$.

For each $J \subset I$,

$$P_\mu\left((X_T)^J_K = \mathbf{0}^J\right) = \bar{E}_{|J|}\left[\nu(\{0\})^{\bar{Y}^N_T}\right] \ ,$$

where $(\bar{Y}^N_t)_{t\geq 0}$ is a **birth** and **death** process on I in which

$$n \to n+1 \ \text{ at } \textbf{rate} \ \ Sn - \frac{S}{N}n^2$$

and

$$n \to n-1 \ \text{ at } \textbf{rate} \ \ \binom{n}{2} - \frac{S}{N}\binom{n}{2} \ .$$

Applications: Fixation probabilities, weak selection

Let $|G| = 1$, $K = \{0, 1\}$, $I = \{1, \ldots, N\}$, $b(1, 1) = b(0, 0) = 1$, $\mu = \nu^I$, $\chi(0) = 0$ and $\chi(1) = \frac{S}{N}$ with fixed $S \in [0, N]$.

For each $J \subset I$,

$$\lim_{T \to \infty} P_\mu \left((X_T)_K^J = \mathbf{0}^J \right) = \frac{1 - \left(1 + \frac{2S}{N-S} \nu(\{0\})\right)^N}{1 - \left(1 + \frac{2S}{N-S}\right)^N} \, .$$

This implies that for fixed finite $J \subset \mathbb{N}$,

$$\lim_{N \to \infty} \lim_{T \to \infty} P_\mu \left((X_T)_K^J = \mathbf{0}^J \right) = \frac{1 - e^{2S\nu(\{0\})}}{1 - e^{2S}} \, .$$

Applications: Strong duality

Let μ be nice. There is a change of measure

$$\bar{P}_{\bar{\eta}(\xi^J_{G\times K},\mathbb{J},\bar{\mu}^{\tilde{I}})} \rightsquigarrow \bar{Q}_{\bar{\eta}(\xi^J_{G\times K},\mathbb{J},\bar{\mu}^{\tilde{I}})}$$

such that

$$P_{(\mu,\mathbb{I})}\left((X_T)^J_{\tilde{I}} \in \cdot \;\Big|\; (X_T)^J_{G\times K} = \xi^J_{G\times K}\right) = \bar{Q}_{\bar{\eta}(\xi^J_{G\times K},\mathbb{J},\bar{\mu}^{\tilde{I}})}\left(\left(\bar{X}^J_T\right)_{\tilde{I}} \in \cdot\right)$$

where under

$$\bar{Q}_{\bar{\eta}(\xi^J_{G\times K},\mathbb{J},\bar{\mu}^{\tilde{I}})} \text{ the dual process } \left(\bar{X}^J_t, \bar{X}^{\tilde{I}}_t\right)_{t\in[0,T]}$$

is again Markov process but time-inhomogeneous.

Applications: Strong duality, no selection

If there is **no selection**, then

$$P_{(\mu,\mathbb{I})}\left(\textbf{Partition of}\left((X_T)^J_{\tilde{I}}\right) \in \cdot \;\middle|\; (X_T)^J_{G\times K} = \xi^J_{G\times K}\right)$$

is equal to

the law of a

time-inhomogeneous $(G \times K)$-marked J-coalescent at time T,

where

- the trivial partition of J into singletons is the initial partition
- $\xi^J_{G\times K}$ are the initial marks

Applications: Strong duality, no selection, example

Consider the case where $|G| = 1$, $b(v, v) = 1\,\forall v \in K$, $\mu = \nu^I$, $J = \{i, j\}$ and $u, w \in K$ with $\nu(\{u\}), \nu(\{w\}) > 0$.

Applications: Strong duality, no selection, example

Consider the case where $|G| = 1$, $b(v,v) = 1\,\forall v \in K$, $\mu = \nu^I$, $J = \{i,j\}$ and $u, w \in K$ with $\nu(\{u\}), \nu(\{w\}) > 0$.

Recall that

$$P_{(\mu,\mathbb{I})}\left(\textbf{Partition of}\left((X_T)_{\tilde{I}}^{J}\right) = \{\{i\},\{j\}\}\right) = e^{-T}\,.$$

In contrast to that,

$$P_{(\mu,\mathbb{I})}\left(\textbf{Partition of}\left((X_T)_{\tilde{I}}^{J}\right) = \{\{i\},\{j\}\}\;\middle|\;(X_T)_K^J = (u,w)\right)$$

$$= \begin{cases} e^{-T}\frac{\nu(\{u\})}{1-e^{-T}[1-\nu(\{u\})]} & \text{if } u = w \\ 1 & \text{if } u \neq w \end{cases}.$$

Thank you for your attention