Kallenberg Symposium 2013

Evolution of genealogies in populations

A. Greven

Universität Erlangen-Nürnberg

17. Juli 2013

Joint work with

- Anita Winter (Essen)
- Andrej Depperschmidt (Freiburg)
  Peter Pfaffelhuber (Freiburg)
- Patric Glöde (Erlangen)
  Thomas Rippl (Erlangen)
- Chiranjib Mukherjee (München)

Goal

Present framework for modelling of genealogical information of a stochastically evolving population as a strong Markov process, i.e. the *forward* picture is described.
Apply this to obtain properties of process, equilibria, parameter dependence, path properties.

Content

1. Population models
2. Description of genealogies
3. Topologies on genealogies
4. Martingale problems for genealogical processes and applications

1. Population models
   - Moran model (with migration)
   - Logistic branching random walk
   - Diffusion models
2. Description of genealogies: Basics
   - Genealogy
   - Moran modell size-N genealogies
   - Logistic branching genealogies: extended up-space
   - Population with types and locations: marked up-space
   - Extension to infinite populations
3. Topologies on genealogies: Analytic aspects
   - Polynomials
   - Topology
4. Martingale problems for genealogical processes and applications
   - Martingale problems
   - Applications

1. Population models
   - Moran model (with migration)
   - Logistic branching random walk
   - Diffusion models
2. Description of genealogies: Basics
   - Genealogy
   - Moran modell size-N genealogies
   - Logistic branching genealogies: extended up-space
   - Population with types and locations: marked up-space
   - Extension to infinite populations
3. Topologies on genealogies: Analytic aspects
   - Polynomials
   - Topology
4. Martingale problems for genealogical processes and applications
   - Martingale problems
   - Applications

Moran dynamics

- N-individuals with types, type space $\mathbb{K}$
- (N-individuals at each of the sites of $\mathbb{G}$)
- pairs of individuals resample if at same site
- (Each individual performs independently continuous time rate $a(\cdot,\cdot)$ random walks on $\mathbb{G}$.)
- We can add mutation and selection.
- We could work with Cannings models.

Logistic branching random walk

- particles give at rate b birth to two or 0 particles and die
- particles give birth at rate s to a new particle
- particles die at rate $d(k)$, if we have k particles at the site
- migration of individuals according to independent continuous time rate $a(\cdot,\cdot)$ random walks

Assumption: $x \to F(x) = sx - d(x)x$, concave, $F(x) < 0$ for $x \geq K$.

One wants to take the number of individuals (per site) to infinity and obtain in suitable scalings:
Diffusion processes of type-frequences, respectively *masses*
examples:

$$dy_t = \sqrt{y_t(1-y_t)}dB_t \qquad \text{Fisher-Wright}$$

$$dx_t = F(x_t)dt + \sqrt{bx_t}dB_t \qquad \text{Logistic branching}$$

1. Population models
   - Moran model (with migration)
   - Logistic branching random walk
   - Diffusion models
2. Description of genealogies: Basics
   - Genealogy
   - Moran modell size-N genealogies
   - Logistic branching genealogies: extended up-space
   - Population with types and locations: marked up-space
   - Extension to infinite populations
3. Topologies on genealogies: Analytic aspects
   - Polynomials
   - Topology
4. Martingale problems for genealogical processes and applications
   - Martingale problems
   - Applications

The genealogical tree

Birth event
$\iota \longrightarrow \iota', \iota''$
ι', ι'' are the descendents of ι, ι is ancestor of ι', ι''.
This defines:
(1) a "genealogical tree", a rooted labelled tree.
(2) Genealogical distance: $2 \cdot$ time to MRCA

Of interest is a structure abstracting from labels.

We are interested to model the stochastic evolution of the genealogical structure as solution to a **martingale problem** *on a* **Polish space,** in particular for the diffusion limit.

Our model for a Moran population genealogy:

$\mathcal{U} = \overline{(U, r, \mu)},$

U =set of individuals
$r : U \times U \longrightarrow [0, \infty)$, quasi ultrametric
$r(\iota, \iota') =$ genealogical distance of ι and $\iota' =$ twice time to MRCA
$\mu = \frac{1}{|U|} \cdot \sum_{\iota \in U} \delta_\iota \in \mathcal{M}_1(U)$, sampling measure
$\overline{}$: equivalence class under isometries and measure-preserving maps of supp(μ) : $\mathcal{U}$.

$\mathcal{U}$ is called an *ultrametric probability space*
$\mathbb{U} := \{\mathcal{U} | \mathcal{U}$ is *um*-space$\}$
$\mathbb{U}$ is a Polish space for topology introduced later.

Our model for a logistic branching genealogy with *varying population size requires* $\mu \in \mathcal{M}_{\text{fin}}(U)$.

$$\mathcal{U} = (\bar{\mu}, \overline{((U, r, \widehat{\mu}))},$$

$$\widehat{\mu} \in \mathcal{M}_1(U), \quad \bar{\mu} \in \mathbb{R}^+,$$
$$\mu = \sum_{\iota \in U} \delta_\iota \quad , \quad \bar{\mu} = \mu(U) \quad , \quad \widehat{\mu}(\cdot) = \frac{\mu(\cdot)}{\mu(U)}.$$

$\mathbb{U}^* = \mathbb{R}^+ \times \mathbb{U}$ (extended ultrametric probability measures)

$\mathbb{U}^*$ is a Polish space

Individuals with types and locations require:

$V = \mathbb{K} \times \mathbb{G}$, $\mathbb{K}$ = type space, $\mathbb{G}$ = geographic space.
Assume that $\mathbb{K}$ is a compact Polish space, $\mathbb{G}$ finite or compact.
Mark function is added to description $\kappa : U \longrightarrow V$.
Consider (U, r, κ, μ).
Form equivalence class including mark preservation.
Obtain (extended) V-marked ultrametric probability measure spaces:

$$\mathcal{U} = \overline{(U, r, \kappa, \mu)}$$

Allow : κ : kernel $U \times V$, $\nu = \mu \otimes \kappa : (U, r, \nu)$

State space: $\mathbb{U}_V$, respectively $\mathbb{U}_V^*$ ultrametric probability measure spaces.

$\mathbb{U}_V, U_V^*$ are Polish spaces.

It is necessary to consider spaces $\mathbb{G}$ which are infinite and allow infinite total populatin size, populations which are only "locally" finite.

Assume: $\mathbb{G}$ is a metric (Polish) space. $\mathbb{G} = \mathbb{Z}^d, \quad \mathbb{R}^d$ etc.

Infinite populations are modelled as:

$$\mathcal{U} = (\bar{\nu}, \overline{(U, r, \widehat{\nu})})$$

$$\bar{\nu}(\cdot) = \nu(U \times (\mathbb{K} \times \{\cdot\})), \quad \bar{\nu} \in \mathcal{M}(\mathbb{G})$$

$$\widehat{\nu}(\cdot \times A) = \sum_{i \in A} \widehat{\nu}_i(\cdot) \quad , \quad \widehat{\nu}_i(\cdot) = \frac{\nu(\cdot x\{i\})}{\nu(U \times \mathbb{K} \times \{i\})}, \quad \forall\, i \in \mathbb{G}, \quad \widehat{\nu}_i \in \mathcal{M}_1(U \times \mathbb{K}).$$

Gives as set of all isomorphy classes:

$\mathbb{U}_V^*$ (Polish space)

.

1. Population models
   - Moran model (with migration)
   - Logistic branching random walk
   - Diffusion models
2. Description of genealogies: Basics
   - Genealogy
   - Moran modell size-N genealogies
   - Logistic branching genealogies: extended up-space
   - Population with types and locations: marked up-space
   - Extension to infinite populations
3. **Topologies on genealogies: Analytic aspects**
   - Polynomials
   - Topology
4. Martingale problems for genealogical processes and applications
   - Martingale problems
   - Applications

Polynomials are functions of the form:

$$\Phi : \mathbb{U} \longrightarrow \mathbb{R}, \quad \mathbb{U}^* \longrightarrow \mathbb{R}, \quad \mathbb{U}^*_V \longrightarrow \mathbb{R}$$

$$\Phi(\mathcal{U}) = \int\limits_{U^n} \varphi((r(u_i, u_j))_{1 \leq i < j \leq n}) \mu(du_1) \cdots \mu(du_n) \tag{1}$$

$$\varphi \in C_b^1((\mathbb{R})^{\binom{n}{2}}, \mathbb{R}) \quad , \quad \mu \in \mathcal{M}_1(U).$$

(2) $\Phi(\mathcal{U}) = \bar{\Phi}(\bar{\mu})\widehat{\Phi}((U,r,\widehat{\mu})) \quad , \quad \bar{\Phi} \in C_b(\mathbb{R},\mathbb{R})$

(3)

$$\Phi(\mathcal{U}) = \bar{\Phi}(\bar{\nu}) \int_{(U \times V)^n} \varphi((r(u_i,u_j))_{1 \leq i < j \leq n})\chi((v_1,v_2,\cdots,v_n))\widehat{\nu}^{\otimes n}(d(\underline{u},\underline{v}))$$

$$\nu = \mu \otimes \kappa$$

$$\chi \in C_b(V,\mathbb{R})(\chi \in C_{bb}(V,\mathbb{R}))$$

We call the generated algebra of functions by Π, Π^*, Π_V.

Define topology via convergence of sequences.

Idea: Convergence $\longleftrightarrow$ convergence of sampled (marked) subtrees and population sizes.

$\mathcal{U}_n \Longrightarrow \mathcal{U}$ as $n \to \infty$ iff

$$\Phi(\mathcal{U}_n) \underset{n\to\infty}{\longrightarrow} \Phi(\mathcal{U}), \quad \forall\, \Phi \in \Pi \text{ resp. } \Pi^*, \Pi_V, \Pi_V^*.$$

In particular polynomials are bounded continuous functions on :

$$\mathbb{U}, \mathbb{U}_V, \mathbb{U}^*, \mathbb{U}_V^*.$$

1. Population models
   - Moran model (with migration)
   - Logistic branching random walk
   - Diffusion models
2. Description of genealogies: Basics
   - Genealogy
   - Moran modell size-N genealogies
   - Logistic branching genealogies: extended up-space
   - Population with types and locations: marked up-space
   - Extension to infinite populations
3. Topologies on genealogies: Analytic aspects
   - Polynomials
   - Topology
4. Martingale problems for genealogical processes and applications
   - Martingale problems
   - Applications

Consider a linear operator L on $C_b(E, \mathbb{R})$, for $E = \mathbb{U}, \mathbb{U}_V, \cdots$ with domain a subalgebra of Π(resp. (Π_v, Π^*_V) of smooth functions and the

$$(L, \Pi, \delta_{\mathcal{U}})\text{-martingale problem}$$

for a $\mathbb{U}^*_V$-*valued diffusion process* ($\mathbb{U}_V$ resp. $\mathbb{U}$).

Key operator

$$L_{\mathrm{res}}\widehat{\Phi}(\mathcal{U}) = \sum_{\substack{k,\ell=1 \\ k<\ell U^n}}^{n} (\varphi \circ \theta_{k,\ell}(\underline{\underline{r}}(\underline{u})) - \varphi(\underline{\underline{r}}(\underline{u})))\mu^{\otimes n}(d\underline{u})$$

$$\underline{\underline{r}} = (r(u_i, u_j))_{1\le i<j\le n}, \quad \underline{u} = (u_1, \cdots, u_n)$$

$$\theta_{k,\ell}(\underline{\underline{r}}(\underline{u})) = \underline{\underline{r}}(\underline{u}_{k,\ell}) \quad ,$$

$$\underline{u}_{k,\ell} = \quad (u_1, \cdots, u_{\ell-1}, u_k, u_{\ell+1}, \cdots, u_n).$$

In case of varying population size:

product rule form

$$L(\bar{\Phi}\widehat{\Phi}) = (L^{\text{mass}}\bar{\Phi})\widehat{\Phi} + \bar{\Phi} \cdot L^{\text{gen}}(\widehat{\Phi})$$

Connects with fixed mass case (Fleming-Viot).

L^{gen}: time-inhomogeneous resampling/migration operators.

Conditional duality: time-inhomogeneous coalescent.

We get

- wellposed martingale problems.
- Genealogical diffusion models arise as many individual per site limits.
- Convergence into equilibrium states, equilibrium genealogy.
- Duality theory, conditioned duality.
- Coalescent representation of equilibria.
- Spatial continuum limit of genealogy processes.
- Almost sure path properties for qualitative features.

Example of path property:
Consider the number of covering balls of radius 2ε of U_t:

$$N_t^\varepsilon .$$

Then in path space

$$(\frac{\varepsilon}{2} N_t^\varepsilon)_{t>0} \Longrightarrow (1)_{t>0}, \quad \text{as } \varepsilon \to 0.$$

Bibliography

- A. Greven, P. Pfaffelhuber and A. Winter:
  Tree-valued resampling dynamics: Martingale problems and applications.
  PTRF, Vol. 155, Nr. 3-4, S. 789–838 (2013).
- A. Depperschmidt, A. Greven and P. Pfaffelhuber:
  Tree-valued Fleming-Viot dynamics with mutation and selection.
  Annals of Appl. Probability, Vol. 22(6), S. 2560–2615 (2012).
- A. Depperschmidt, A. Greven and P. Pfaffelhuber:
  Marked metric measure spaces.
  Elect. Comm. in Probab., Vol. 16, S. 174–188 (2011).

- A. Greven, Lea Popovic and A. Winter:
  Genealogy of catalytic branching models.
  Annals of Applied Probability, Vol. 19, Nr. 3, S. 1232–1272 (2009).
- A. Greven, P. Pfaffelhuber and A. Winter:
  Convergence in distribution of random metric measure spaces (The Λ-coalescent measure tree).
  PTRF, Vol. 145, Issue 1, S. 285–322 (2009).
- A. Greven and P. Glöde and T. Rippl:
  Stable features in genealogies of logistic branching diffusions.
  In preparation 2013.