Surfing in population genetics

Emmanuel Schertzer
University of Vienna

joint work with ...

Figure: Julie Tourniaire (U. Wien/ISTA), Felix Foutel–Rodier (Oxford)

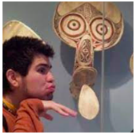


Figure: Alejandro Wences (Toulouse), Florin Boenkost (U. Wien), Zsofia Talyigàs (U. Wien)

A (very very brief) introduction to population genetics

- Aim of population genetics: Modeling the effect of evolutionary forces on today's observed genetic diversity.
- Difficult task:
  1. Natural selection (e.g., balancing, directional selection etc.)
  2. Genetic drift.
  3. Recombination (e.g., hitch-hiking effect)
  4. Demographic and geographic effects (e.g., bottlenecks, migration)

What do we observe today ? Mutational patterns

- Mutational patterns in a sequence alignment.
- SNP. Single Nucleotide Polymorphism.
- Summary statistics:
  1. Number of SNP's
  2. Site frequency spectrum (counting the number of SNP's carried by k individuals)
  3. Allele frequency spectrum (population is partioned into 4 haplotype blocks $\{1\}\{2\}\{3,4\}$).

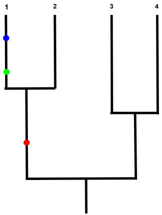
Sequence 1	A	T	C	C	T	A	$\cdots$
Sequence 2	A	T	C	T	T	T	$\cdots$
Sequence 3	T	T	C	T	T	T	$\cdots$
Sequence 4	T	T	C	T	T	T	$\cdots$

Table: Sequence alignment of size $n = 4$ with 3 SNP's.

The genealogical approach

- Which genealogies could explain the observed mutational pattern ?

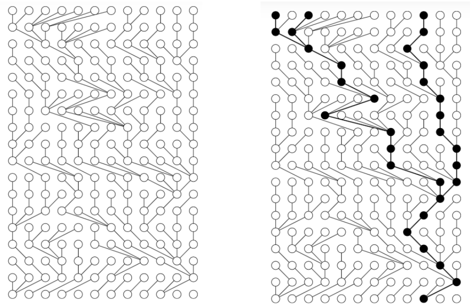
Sequence 1	A	T	C	C	T	A	...
Sequence 2	A	T	C	T	T	T	...
Sequence 3	T	T	C	T	T	T	...
Sequence 4	T	T	C	T	T	T	...



Deeper genealogies generate more genetic diversity

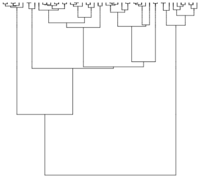
The Wright-Fisher model

- Fixed population size of N.
  1. Assign allelic types at generation 0.
  2. Each individual at generation $t+1$ picks a parent from generation t uniformly at random and inherits the type of its parent.
- Genealogy. Pairs of ancestral lineages coalesce with probability $1/N$ in a single generation.



Kingman coalescent

- Genealogy. Kingman coalescent*:
  1. At $t = 0$, start with n lineages.
  2. Construct the tree from the leaves to the root.
  3. Pairs of lineages coalesce at rate 1.



Kingman (82)

A first glimpse at universality (Cannings models)

A1: At each generation t, we have an independent offspring vector

$$(\nu_1^{(t)}, \cdots, \nu_N^{(t)}), \quad \sum_{i=1}^{N} \nu_i^{(t)} = N$$

and the entries of the vector are exchangeable (neutral assumptions).

A2: Under the assumptions

$$\mathrm{Var}(\nu_1^{(t)}) = o(N), \;\; \mathbb{E}[(\nu_1^{(t)} - 1)^3] \; = \; o(N\sigma_N^2)$$

the discrete genealogy of a sample of size n converges to the the n-Kingman coalescent*.

A2': “Skewed" offspring distribution: convergence to Λ-coalescent.

*Kingman (82), Sagitov, Möhle (03), **Pitman (99), Sagitov (99), Schweinsberg (03)

Λ-coalescent*

- Let $\Lambda(dx)$ be a probability measure on $[0,1]$.
- Ultrametric tree generated from leaves to root in a Markov way.
- If n lineages alive at time t (from the leaves) any k-uplet of lineages merge into a single lineage at rate

$$k \in \{2, \cdots, n\}, \quad \lambda^n(k) \ := \ \int_0^1 x^k (1-x)^{n-k} \frac{\Lambda(dx)}{x^2}$$

- Poisson construction:
  1. At rate $dt \times \frac{\Lambda(dx)}{x^2}$, mark each lineage with probability x.
- $\Lambda(dx) = \delta_0$, Kingman
- $\Lambda(dx) = dx$, Bolthausen-Sznitman coalescent.
- $\Lambda(dx) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}dx$, Beta$(a,b)$ coalescent $(a, b > 0)$.

Pitman, Sagitov (99)

Universality

- Random walks with finite second moments converge to Brownian motion.
- In general, random walks converge to Lévy processes. Discontinuous process characterized by a jump measure Λ.
- For exchangeable (Cannings) models (no space-no selection)
  1. Kingman $=$ Brownian motion
  2. Λ-coalescent $=$ Λ-Lévy processes
- One of the challenge of Mathematical Biology: going beyond exchangeable models ...
- A common belief in Mathematical Biology is that the universality class of Λ-coalescents can be found in unexpected places,

Outline

1. A stochastic population model.
2. A general approach to the convergence of genealogies
   2.1 The marked Gromov-weak topology and the methods of moments*
   2.2 spinal decomposition and many-to-few
3. Two applications
   3.1 Multitype Branching processes
   3.2 BBM of Tourniaire (22).
4. Fitness waves

see also *Gonzales, Harris, Horton, Kyprianou, Powell, Wang (20,21,22)

Noisy F-KPP equation with Allee effect

Let $\varepsilon > 0$. Noisy F-KPP with demographic noise.

$$\partial_t u = \frac{1}{2}\partial_{xx} u + u(1-u)(1+\varepsilon u) + \sqrt{\frac{u(1-u)}{N}}\,\eta$$

where η is a space-time white noise. In the deterministic regime $(N = \infty)$, existence and uniqueness $(\varepsilon > 0)$ of a travelling wave solution* $\psi(x - ct)$.

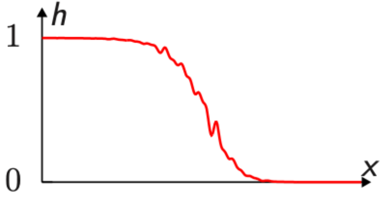


Figure: Stochastic front

*Roques (12)

Local reproduction rate

Local reproduction rate $r(u) = \underbrace{(1-u)}_{\text{saturation}} \quad \underbrace{(1+\varepsilon u)}_{\text{Allee effect}}$

- $(1-u)$: saturation,
- $(1+\varepsilon u)$: Allee effect (cooperation, vortex of extinction).

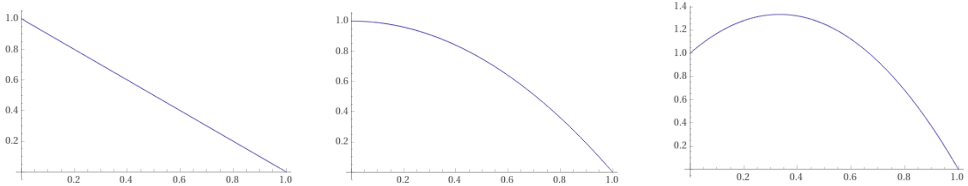


Figure: $r(u)$ for $\varepsilon = 0, 1, 3$.

For $\varepsilon < 1$, r is optimized at frequency 0 (tip)
For $\varepsilon > 1$, r is optimized at intermediary frequency (bulk)

Deterministic equation: Pulled vs Pushed

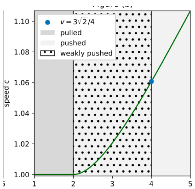


Figure: Speed of the wave as a function of ε

Pulled.$(\varepsilon < 2)$ the speed does not vary with ε.

Pushed.$(\varepsilon > 2)$ the speed increases with ε.

Analytic interpretation. In the pulled case, the speed (and the shape of the profile at the tip) is predicted by linearizing the profile close to the edge of the front.
Probabilistic interpretation. Sample $n = 2$ individuals at the edge then the MRCA is located

- Pulled. At the edge.
- Pushed. In the bulk.

*Roques et al.(12)

Stochastic effects: semi vs fully pushed waves

Recent simulations and studies * investigated the effect of demographic fluctuations on the fronts.

	pulled	semi-pushed	fully-pushed
strength of the Allee effect	$\varepsilon \in (0,2)$	$\varepsilon \in (2,4)$	$\varepsilon > 4$
time scale fluctuations	$\log(N)^3$	$N^{\alpha-1}$, $\alpha \in (1,2)$	N
Genealogy	Bothausen-Sznitman	Beta(α,2-α)	Kingman

Recent rigorous results in the fully pushed regime**.

*Birzu et al.(18), **Etheridge, Pennington (21)

BBM with inhomogeneous branching rate*

- Let $\mu > 0$, $f \geq 0$ with $\mathrm{Supp}(f) \subseteq [0,1]$.
- $\varepsilon \geq 0$ (Allee effect).
- BBM with
  1. Inhomogeneous branching rate
  $$\forall x > 0, \quad r(x) = \frac{1}{2} + \frac{\varepsilon}{2} f(x)$$
  2. Drift $-\mu$ (speed of the front)

     μ is chosen in such a way that the system is critical, i.e., the average number of particles remains roughly constant (stable front).
  3. Killing at 0 (particles inside the front are killed)
- $\varepsilon = 0$: Standard BBM**.

*Tourniaire (22), **Berestycki, Berestycki, Schweinsberg (13)

Criticality

- Define $(t, y) \to p_t(x, y)$ as the density of particles at time t starting with a unique particle at x, i.e., for every test function f

$$\mathbb{E}_x \left(\sum_{u \in N_t} f(X_u(t)) \right) = \int_{\mathbb{R}_+} p_t(x, y) f(y) dy.$$

- $(t, y) \to p_t(x, y)$ is solution of the fundamental equation

$$\partial_t u = \frac{1}{2}\partial_{xx} u + \mu \partial_x u + r(x) u$$
$$\text{with} \quad u(t, 0) = 0.$$

- Criticality: Choose μ to keep the number of particles "under control" (stable front, μ is interpreted as the speed of the front).

Criticality

Let $L >> 1$ and consider the modified equation

$$\partial_t u = \frac{1}{2}\partial_{xx} u + \mu \partial_x u + r(x) u$$
$$\text{with} \quad u(t,0) = u(t,L) = 0$$

BBM killed at 0 and L.

Write

$$p_t(x,y) = e^{\mu(x-y)+\frac{1}{2}(1-\mu^2)t} q_t(x,y) \quad \text{(Girsanov transformation).}$$

Then $(t,y) \to q_t(x,y)$ is solution of

$$\partial_t u = \underbrace{\frac{1}{2}\partial_{xx} u + (r(x) - \frac{1}{2}) u}_{\text{Self-adjoint}} \quad \text{with} \quad u(t,0) = u(t,L) = 0$$

Criticality

Consider

$$\partial_t u = \frac{1}{2}\partial_{xx} u + (r(x) - \frac{1}{2})u \quad \text{with} \quad u(t,0) = u(t,L) = 0$$

Sturm-Liouville: The eigenvalues of the operator can be numbered $\lambda_1 > \lambda_2 > \cdots > \lambda_k \cdots \to -\infty$ and

$$q_t(x,y) \;=\; \sum_{k\geq 1} e^{\lambda_k t} v_k(x) v_k(y),$$

where the v_k's are the eigenfunctions with $||v_k||_2 = 1$, i.e.,

$$v_k''(x) + (r(x) - \frac{1}{2})v(x) = \lambda_k v_k(x), \quad v_k(0) = v_k(L) = 0.$$

Criticality

By reversing the Girsanov transformation

$$p_t(x,y) \;=\; e^{\mu(x-y)+(\frac{1}{2}(1-\mu^2)+\lambda_1)t}\sum_{k\geq 1} e^{(\lambda_k-\lambda_1)t} v_k(x)v_k(y).$$

As $t\to\infty$, $\lim_{L\to\infty}\lambda_1-\lambda_2>0$ (spectral gap) and $\lambda_1\uparrow\lambda_1^\infty<\infty$* as $L\to\infty$,

$$p_t(x,y) \;\approx_{t\to\infty}\; e^{\mu(x-y)+(\frac{1}{2}(1-\mu^2)+\lambda_1^\infty)t} v_1(x)v_1(y)$$

The system is critical when

$$\mu=\sqrt{1+\beta^2}, \quad \text{with } \beta:=\sqrt{2\lambda_1^\infty}.$$

where $\lambda_1\uparrow\lambda_1^\infty$ is the principal eigenvalue of the self-adjoint operator

$$\mathcal{L}u \;=\; \frac{1}{2}u''+(r(x)-\frac{1}{2})u(x), \quad u(0)=u(L)=0.$$

Pinsky (95)

Pushed vs Pulled

Recall that

$$r(x) = \frac{1}{2} + \frac{\varepsilon}{2} f(x)$$

By Pinsky (95), there exists $\varepsilon_1 > 0$ such that

1. For $\varepsilon < \varepsilon_1$, $\lambda_1^\infty = 0$.
2. For $\varepsilon > \varepsilon_1$, $\lambda_1^\infty > 0$.

As a consequence,

Theorem (Tourniare (21), S., Tourniare (22+))

There exists $\varepsilon_1 > 0$ such that

(Pullled) *For $\varepsilon < \varepsilon_1$, $\mu = 1$.*

(Pushed) *For $\varepsilon > \varepsilon_1$, $\mu > 1$.*

What about the genealogical structure of the population ?

Conjectures

There exists $0 < \varepsilon_1 < \varepsilon_2$ such that

BBM	pulled	semi pushed	fully pushed
strength of the Allee effect	$\varepsilon \in (0, \varepsilon_1)$	$\varepsilon \in (\varepsilon_1, \varepsilon_2)$	$\varepsilon > \varepsilon_2$
time scale fluctuations	$\log(N)^3$	$N^{\alpha-1}$, $\alpha \in (1,2)$	N
limiting CSBP	Neveu	α-stable	Feller

to be compared with the stochastic PDE

$$\partial_t u = \frac{1}{2}\partial_{xx} u + u(1-u)\underbrace{(1+\varepsilon u)}_{\text{Allee effect}} + \sqrt{\frac{u(1-u)}{N}}\eta$$

in the PDE	pulled	semi pushed	fully pushed
strength of the Allee effect	$\varepsilon \in (0,2)$	$\varepsilon \in (2,4)$	$\varepsilon > 4$
time scale fluctuations	$\log(N)^3$	$N^{\alpha-1}$, $\alpha \in (1,2)$	N
Genealogy	Bothausen-Sznitman	Beta(α,2-α)	Kingman

Outline

1. A stochastic population model and its branching approximation.

2. A general approach to the convergence of genealogies

   2.1 The marked Gromov-weak topology and the methods of moments*

   2.2 spinal decomposition and many-to-few

3. Two applications

   3.1 Multitype Branching processes

   3.2 BBM of Tourniaire (22).

see also *Gonzales, Harris, Horton, Kyprianou, Wang (20,21,22)

General (discrete time) branching processes

Let E be a general type space, and consider a collection of random point measures

$$\forall x \in E, \quad \Xi(x) = \sum_{i=1}^{K(x)} \delta_{\xi_i(x)}$$

An individual with type x gives birth to $K(x)$ children with types $(\xi_1(x), \xi_2(x), \dots)$

This constructs a random tree T and a collection of types $(X_u \in E,\ u \in T)$

It corresponds to:

- a Galton-Watson tree if $\mathrm{Card}\, E = 1$
- a multitype branching process if E is finite (countable)
- a branching random walk if $E = \mathbb{R}^d$ and $\Xi(x) = x + \Xi$

General (discrete time) branching processes

Let E be a general type space, and consider a collection of random point measures

$$\forall x \in E, \quad \Xi(x) = \sum_{i=1}^{K(x)} \delta_{\xi_i(x)}$$

An individual with type x gives birth to $K(x)$ children with types $(\xi_1(x), \xi_2(x), \dots)$

This constructs a random tree T and a collection of types $(X_u \in E,\ u \in T)$

It corresponds to:

- a Galton-Watson tree if $\operatorname{Card} E = 1$
- a multitype branching process if E is finite (countable)
- a branching random walk if $E = \mathbb{R}^d$ and $\Xi(x) = x + \Xi$

General (discrete time) branching processes

Let E be a general type space, and consider a collection of random point measures

$$\forall x \in E, \quad \Xi(x) = \sum_{i=1}^{K(x)} \delta_{\xi_i(x)}$$

An individual with type x gives birth to $K(x)$ children with types $(\xi_1(x), \xi_2(x), \dots)$

This constructs a random tree T and a collection of types $(X_u \in E,\ u \in T)$

It corresponds to:

- a Galton-Watson tree if $\mathrm{Card}\, E = 1$
- a multitype branching process if E is finite (countable)
- a branching random walk if $E = \mathbb{R}^d$ and $\Xi(x) = x + \Xi$

General (discrete time) branching processes

Let E be a general type space, and consider a collection of random point measures

$$\forall x \in E, \quad \Xi(x) = \sum_{i=1}^{K(x)} \delta_{\xi_i(x)}$$

An individual with type x gives birth to $K(x)$ children with types $(\xi_1(x), \xi_2(x), \dots)$

This constructs a random tree T and a collection of types $(X_u \in E,\ u \in T)$

It corresponds to:

- a Galton-Watson tree if $\mathrm{Card}\, E = 1$
- a multitype branching process if E is finite (countable)
- a branching random walk if $E = \mathbb{R}^d$ and $\Xi(x) = x + \Xi$

General (discrete time) branching processes

Let E be a general type space, and consider a collection of random point measures

$$\forall x \in E, \quad \Xi(x) = \sum_{i=1}^{K(x)} \delta_{\xi_i(x)}$$

An individual with type x gives birth to $K(x)$ children with types $(\xi_1(x), \xi_2(x), \dots)$

This constructs a random tree T and a collection of types $(X_u \in E,\ u \in T)$

It corresponds to:

- a Galton-Watson tree if $\mathrm{Card}\, E = 1$
- a multitype branching process if E is finite (countable)
- a branching random walk if $E = \mathbb{R}^d$ and $\Xi(x) = x + \Xi$

Some background

Q: Can we study the scaling limit of the genealogy and distribution of types in the population?

The tree structure has already been studied for

- Galton-Watson processes: Aldous (91); Duquesne, Le Gall (02); Popovic (04)
- limits of branching random walks: Le Gall snake (93)
- multi-type Galton-Watson processes: Miermont (07); Popovic, Rivas (14)
- Branching diffusion in a bounded domain: Powell (19)

Most works on the tree structure of branching processes rely on height and contour processes

We will present a different approach involving spinal decompositions

Some background

Q: Can we study the scaling limit of the genealogy and distribution of types in the population?

The tree structure has already been studied for

- Galton-Watson processes: Aldous (91); Duquesne, Le Gall (02); Popovic (04)
- limits of branching random walks: Le Gall snake (93)
- multi-type Galton-Watson processes: Miermont (07); Popovic, Rivas (14)
- Branching diffusion in a bounded domain: Powell (19)

Most works on the tree structure of branching processes rely on height and contour processes

We will present a different approach involving spinal decompositions

Some background

Q: Can we study the scaling limit of the genealogy and distribution of types in the population?

The tree structure has already been studied for

- Galton-Watson processes: Aldous (91); Duquesne, Le Gall (02); Popovic (04)
- limits of branching random walks: Le Gall snake (93)
- multi-type Galton-Watson processes: Miermont (07); Popovic, Rivas (14)
- Branching diffusion in a bounded domain: Powell (19)

Most works on the tree structure of branching processes rely on height and contour processes

We will present a different approach involving spinal decompositions

Outline

1. A stochastic population model and its branching approximation.

2. A general approach to the convergence of genealogies

   2.1 The marked Gromov-weak topology

   2.2 spinal decomposition and many-to-few

3. Two applications

   3.1 Multitype Branching processes

   3.2 BBM of Tourniaire (22).

see also *Gonzales, Harris, Horton, Kyprianou, Wang (20,21,22)

Genealogies as Marked Metric Space

Recall the tree T and collection of marks $(X_u;\ u \in T)$ of the branching process

We define

- U_N the set of individuals alive at generation N
- the tree distance
$$\forall u, v \in U_N, \quad d_T(u, v) = N - |u \wedge v|$$
- the mark measure on $U_N \times E$ as
$$\mu_N = \sum_{u \in U_N} \delta_{(u, X_u)}$$

The triple $[U_N, d_T, \mu_N]$ represents the genealogy and types at generation N

Genealogies as Marked Metric Space

Recall the tree T and collection of marks $(X_u;\ u \in T)$ of the branching process

We define

- U_N the set of individuals alive at generation N
- the tree distance

$$\forall u, v \in U_N, \quad d_T(u,v) = N - |u \wedge v|$$

- the mark measure on $U_N \times E$ as

$$\mu_N = \sum_{u \in U_N} \delta_{(u, X_u)}$$

The triple $[U_N, d_T, \mu_N]$ represents the genealogy and types at generation N

Genealogies as Marked Metric Space

Recall the tree T and collection of marks $(X_u;\ u \in T)$ of the branching process

We define

- U_N the set of individuals alive at generation N
- the tree distance
$$\forall u, v \in U_N, \quad d_T(u,v) = N - |u \wedge v|$$
- the mark measure on $U_N \times E$ as
$$\mu_N = \sum_{u \in U_N} \delta_{(u, X_u)}$$

The triple $[U_N, d_T, \mu_N]$ represents the genealogy and types at generation N

Genealogies as Marked Metric Space

Recall the tree T and collection of marks $(X_u;\ u \in T)$ of the branching process

We define

- U_N the set of individuals alive at generation N
- the tree distance
$$\forall u, v \in U_N, \quad d_T(u,v) = N - |u \wedge v|$$
- the mark measure on $U_N \times E$ as
$$\mu_N = \sum_{u \in U_N} \delta_{(u, X_u)}$$

The triple $[U_N, d_T, \mu_N]$ represents the genealogy and types at generation N

Genealogies as Marked Metric Space

Recall the tree T and collection of marks $(X_u;\ u \in T)$ of the branching process

We define

- U_N the set of individuals alive at generation N
- the tree distance
$$\forall u, v \in U_N, \quad d_T(u, v) = N - |u \wedge v|$$
- the mark measure on $U_N \times E$ as
$$\mu_N = \sum_{u \in U_N} \delta_{(u, X_u)}$$

The triple $[U_N, d_T, \mu_N]$ represents the genealogy and types at generation N

Marked Metric Spaces (mmm)*

Let (U, d) be a separable metric space.

E be a Polish space called the mark space ($\mathbb{R}_+$ in our case).

A marked metric measure space* (mmm) is a triple (U, d, μ) where μ is a finite measure on $U \times E$

*Depperschmidt, Greven, Pfaffelhuber (11)

A continuous example: the Brownian coalescent point process

- Let $Y \sim Exp(1)$. m a finite measure on $\mathbb{R}_+$.
- Consider a Poisson point process P on $(0,1) \times (0, Y)$ with intensity $\mathrm{d}t \otimes \frac{1}{x^2}\mathrm{d}x$, and define

$$\forall y < z, \quad d_P(y,z) = \sup\{x : (t,x) \in P,\ y < t < z\}$$

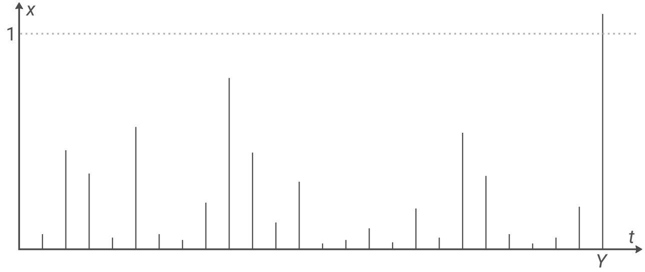
The marked Brownian CPP* is the ultrametric space $[(0, Y), d_P, \mathrm{Leb} \otimes m]$

*Popovic (04)

A continuous example: the Brownian coalescent point process

- Let $Y \sim Exp(1)$. m a finite measure on $\mathbb{R}_+$.
- Consider a Poisson point process P on $(0,1) \times (0, Y)$ with intensity $\mathrm{d}t \otimes \frac{1}{x^2}\mathrm{d}x$, and define

$$\forall y < z, \quad d_P(y,z) = \sup\{x : (t,x) \in P,\ y < t < z\}$$



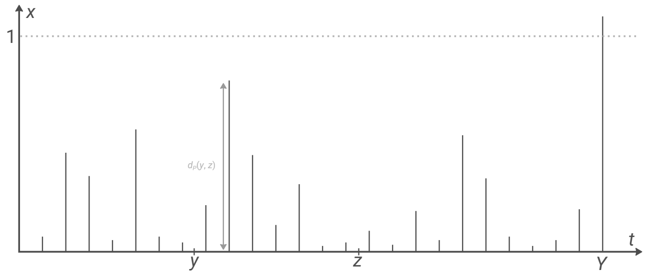
The marked Brownian CPP* is the ultrametric space $[(0, Y), d_P, \mathrm{Leb} \otimes m]$

*Popovic (04)

A continuous example: the Brownian coalescent point process

- Let $Y \sim Exp(1)$. m a finite measure on $\mathbb{R}_+$.
- Consider a Poisson point process P on $(0,1) \times (0, Y)$ with intensity $\mathrm{d}t \otimes \frac{1}{x^2}\mathrm{d}x$, and define

$$\forall y < z, \quad d_P(y,z) = \sup\{x : (t,x) \in P,\ y < t < z\}$$



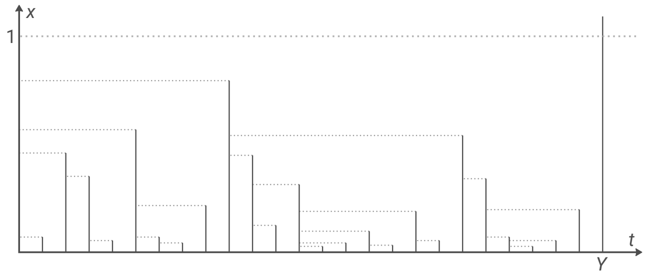
The marked Brownian CPP* is the ultrametric space $[(0, Y), d_P, \mathrm{Leb} \otimes m]$

*Popovic (04)

A continuous example: the Brownian coalescent point process

- Let $Y \sim Exp(1)$. m a finite measure on $\mathbb{R}_+$.
- Consider a Poisson point process P on $(0,1) \times (0, Y)$ with intensity $\mathrm{d}t \otimes \frac{1}{x^2}\mathrm{d}x$, and define

$$\forall y < z, \quad d_P(y,z) = \sup\{x : (t,x) \in P,\ y < t < z\}$$



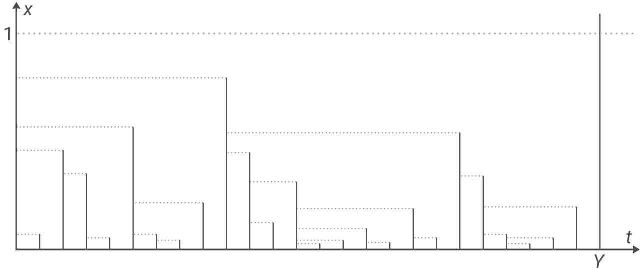
The marked Brownian CPP* is the ultrametric space $[(0, Y), d_P, \mathrm{Leb} \otimes m]$

*Popovic (04)

A continuous example: the Brownian coalescent point process

- Let $Y \sim Exp(1)$. m a finite measure on $\mathbb{R}_+$.
- Consider a Poisson point process P on $(0,1) \times (0, Y)$ with intensity $\mathrm{d}t \otimes \frac{1}{x^2}\mathrm{d}x$, and define

$$\forall y < z, \quad d_P(y,z) = \sup\{x : (t,x) \in P,\ y < t < z\}$$



The marked Brownian CPP* is the ultrametric space $[(0, Y), d_P, \mathrm{Leb} \otimes m]$

*Popovic (04)

CRT perspective

Consider a Brownian excursion e conditioned on $h(e) \geq 1$.

Define

$$d_e(x, y) = e(x) + e(y) - 2 \inf_{[x\wedge y, x\vee y]} e, \quad T = [0, \infty)/\sim$$

where $x \sim y$ iff $d(x, y) = 0$. The CRT* is the random metric space (T, d).

When $e(x) = e(y) = 1$

$$\forall x, y \in \mathbb{R}_+, \quad d_e(x, y) = 2(1 - \inf_{[x\wedge y, x\vee y]} e)$$

so that the distance is given by the depth of the deepest excursion between x and y.

The Brownian CPP results from the encoding the points at level 1 by their local time.

*Aldous (91)

The (marked) Gromov-weak topology

A polynomial* is a functional

$$\begin{aligned}\Phi[U,d,\mu] &= \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\mu(dv_i\otimes dx_i)\\ &= |U|^k \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\frac{\mu(dv_i\otimes dx_i)}{|U|}\end{aligned}$$

for some k and continuous bounded $\psi_{i,j}, \varphi_i$ and $|U| := \mu(U\times E)$.

The marked Gromov-weak topology is that smallest topology such that each Φ is continuous.

The moments of a random marked ultrametric space are

$$\forall\Phi,\quad \mathbb{E}[\Phi[U,d,\mu]]$$

*Greven, Pfaffelhuber, Winter (09); Depperschmidt, Greven, Pfaffelhuber (11)

The (marked) Gromov-weak topology

A polynomial* is a functional

$$\begin{aligned}\Phi[U,d,\mu] &= \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\mu(dv_i \otimes dx_i) \\ &= |U|^k \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\frac{\mu(dv_i \otimes dx_i)}{|U|}\end{aligned}$$

for some k and continuous bounded $\psi_{i,j}, \varphi_i$ and $|U| := \mu(U \times E)$.

The marked Gromov-weak topology is that smallest topology such that each Φ is continuous.

The moments of a random marked ultrametric space are

$$\forall \Phi, \quad \mathbb{E}[\Phi[U,d,\mu]]$$

*Greven, Pfaffelhuber, Winter (09); Depperschmidt, Greven, Pfaffelhuber (11)

The (marked) Gromov-weak topology

A polynomial* is a functional

$$\Phi[U,d,\mu] = \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\mu(dv_i \otimes dx_i)$$
$$= |U|^k \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\frac{\mu(dv_i \otimes dx_i)}{|U|}$$

for some k and continuous bounded $\psi_{i,j}, \varphi_i$ and $|U| := \mu(U \times E)$.

The marked Gromov-weak topology is that smallest topology such that each Φ is continuous.

The moments of a random marked ultrametric space are

$$\forall \Phi, \quad \mathbb{E}[\Phi[U,d,\mu]]$$

*Greven, Pfaffelhuber, Winter (09); Depperschmidt, Greven, Pfaffelhuber (11)

Moments and size biasing

$$\begin{aligned}\Phi[U,d,\mu] &= \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\mu(dv_i \otimes dx_i) \\ &= |U|^k \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\frac{\mu(dv_i \otimes dx_i)}{|U|}\end{aligned}$$

$$\mathbb{E}[\Phi[U,d,\mu]] = \mathbb{E}(|U|^k)\ \mathbb{E}\left(\frac{|U|^k}{\mathbb{E}(|U|^k)} \int \prod_{i,j=1,i\neq j}^{k} \psi_{i,j}(d(v_i,v_j)) \prod_{i=1}^{k} \varphi_i(x_i)\frac{\mu(dv_i \otimes dx_i)}{|U|}\right)$$

The moments are obtained by biasing the population size by $|U|^k$ and then sampling k individuals according to $\frac{\mu(dv\otimes dx)}{|U|}$.

Marked Gromov Prokhorov (MGP) metric

Define

$$d_{MGP}([U_1, d_1, \mu_1], [U_2, d_2, \mu_2]) = \inf_{Z, \varphi_1, \varphi_2} d_{Pr}(\varphi_1 \star \mu_1, \varphi_2 \star \mu_2)$$

where φ_i is an isometric embedding from U_i to Z, d_{Pr} is the Prokhorov distance. The GW topology is metrizable by d_{MGP}.

The method of moments

Theorem (Method of moments)

Let $(X_n;\ n \geq 1)$ be a sequence of real r.v. such that

$$\forall k \geq 1, \quad \lim_{n\to\infty} \mathbb{E}[X_n^k] = m_k, \quad \limsup_{k\to\infty} \frac{m_k^{1/k}}{k} < \infty$$

then there exists X such that $X_n \to X$ in distribution.

The method of moments for trees

Theorem (Greven, Pfaffelhuber, Winter (09); Depperschmidt, Greven, Pfaffelhuber (11))

Let $[U_n, d_n, \mu_n;\ n \geq 1]$ *be a sequence of random marked ultrametric spaces such that*

$$\forall \Phi, \quad \lim_{n \to \infty} \mathbb{E}[\Phi[U_n, d_n, \mu_n]] = \mathbb{E}[\Phi[U, d, \mu]],$$

for some $[U, d, \mu]$ *verifying*

$$\limsup_{k \to \infty} \frac{\mathbb{E}[\mu(U \times E)^k]^{1/k}}{k} < \infty.$$

Then $[U_n, d_n, \mu_n] \to [U, d, \mu]$ *in distribution.*

The method of moments for trees (II)

Theorem (Foutel--Rodier, Lambert, S (21); Foutel--Rodier, S. (22+))

Let $[U_n, d_n, \mu_n;\ n \geq 1]$ *be a sequence of random marked ultrametric spaces such that*

$$\forall \Phi, \quad \lim_{n\to\infty} \mathbb{E}[\Phi[U_n, d_n, \mu_n]]$$

exists and

$$\limsup_{k\to\infty} \lim_{n\to\infty} \frac{\mathbb{E}[\mu_n(U_n \times E)^k]^{1/k}}{k} < \infty.$$

Then there exists a (possibly non-separable) marked ultrametric space $[U, \mathcal{U}, d, \mu]$ *such that* $[U_n, d_n, \mu_n] \to [U, \mathcal{U}, d, \mu]$ *in distribution.*

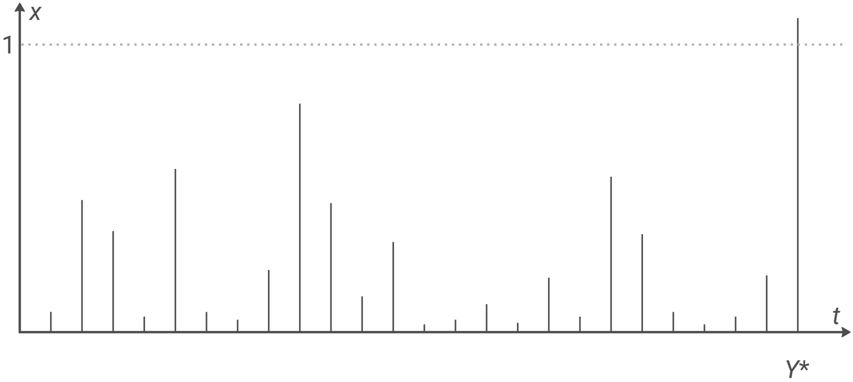
Moments of random trees: the Brownian CPP

The moment of order k of the Brownian CPP is

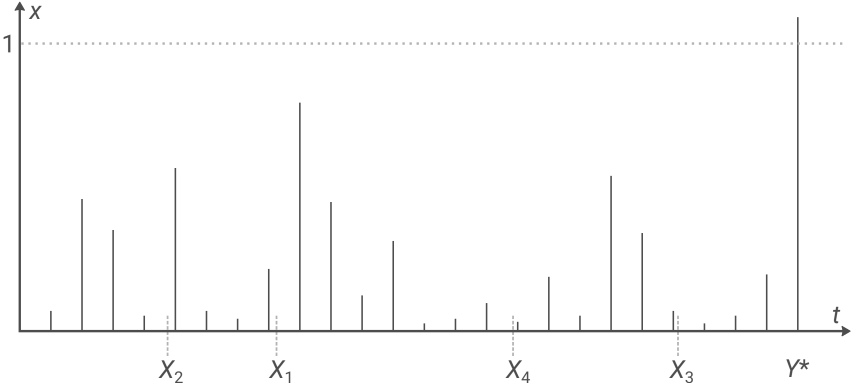
$$\begin{aligned}\mathbb{E}[\Phi[(0,Y),d_P,\mathrm{Leb}\times m]] &= \mathbb{E}\Big[\int_{((0,Y)\times\mathbb{R}_+)^k}\prod_{i,j}\psi_{i,j}\big(d(v_i,v_j)\big)\prod_i \varphi_i(x_i)\mathrm{d}v_i\otimes m(dx_i)\Big]\\ &= k!\,T^k\,\mathbb{E}\Big[\int_{(0,Y^*)^k}\prod_{i,j}\varphi\big(d_P(v_i,v_j)\big)\prod_i dv_i\Big]\prod_i\int_{\mathbb{R}_+}\varphi_i(x_i)m(dx_i)\end{aligned}$$

where $Y^*\sim\mathrm{Gamma}(k+1,1)$

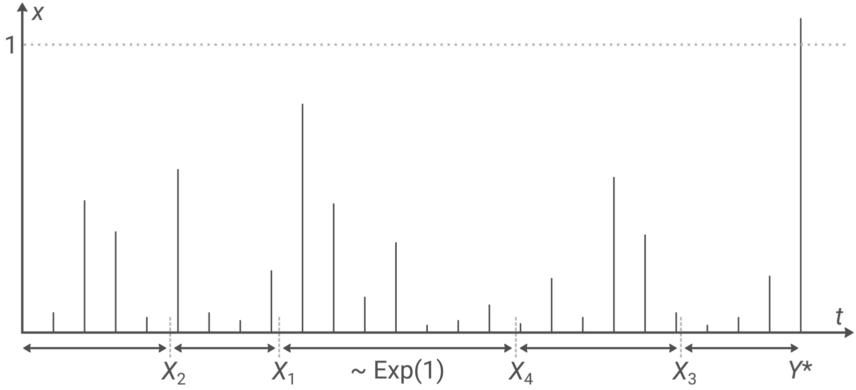
Moments of random trees: the Brownian CPP



Moments of random trees: the Brownian CPP



Moments of random trees: the Brownian CPP



Discrete CPP

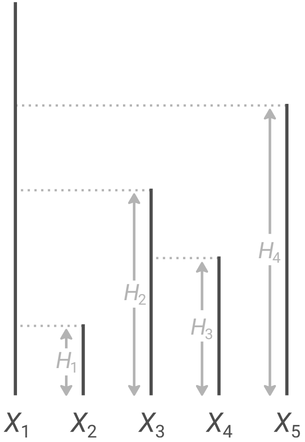
By sampling k points in the size biased Brownian CPP, we obtain a discrete CPP:

i.i.d. branch lengths $(H_1, \ldots, H_{k-1}) \in \{1, \ldots, N\}$:

$$\forall i < j, \quad d(V_i, V_j) = \max\{H_i, \ldots, H_{j-1}\}$$

where (H_i) are i.i.d. uniform r.v.'s.

The genealogy of k sampled point in a Brownian CPP is a discrete CPP.



Moments of random trees: the Brownian CPP

Proposition

The moments of the Brownian CPP are

$$\mathbb{E}[\Phi[(0,Y), d_P, \mathrm{Leb}\times m]] = k!\,\mathbb{E}\Big[\prod_{i,j} \varphi_{i,j}\big(H_{\sigma(i),\sigma(j)}\big)\Big] \prod_i \int \varphi_i(x)m(dx)$$

where

$$\forall i<j, \quad H_{i,j} = H_{j,i} = \max\{H_i, \dots, H_{j-1}\}$$

for an i.i.d. collection $(H_i;\ i<k)$ of uniform r.v. on $(0,1)$ and a uniform permutation σ of $\{1,\dots,k\}$.

Convergence criterium

Proposition

Let $[U_N, d_N, \mu_N]$ be a sequence of random mmm. If

$$\mathbb{E}\left(\Phi([U_N, d_N, \mu_N])\right) \to k!\,\mathbb{E}\Big[\prod_{i,j} \varphi_{i,j}\big(H_{\sigma(i),\sigma(j)}\big)\Big] \prod_i \int \varphi_i(x) m(dx)$$

Then $[U_N, d_N, \mu_N]$ converges in distribution to a Brownian CPP (with measure m).

Sampling from a CPP

Proposition

Sample K points (v_i, X_{v_i}) in a Brownian CPP(m). Then

$$(d(v_i, v_j), x_{v_i}) = (H^\theta_{\sigma(i)\sigma(j)}, W_i)$$

where $\mathcal{L}(\theta) = \frac{K}{(1+\theta)^2}(\frac{\theta}{1+\theta})^{K-1} d\theta$. Further, cond. on θ, (H^θ_i) i.i.d. with

$$\mathbb{P}(H^\theta \leq s) = \frac{(1+\theta)\mathbb{P}(U \leq s)}{1+\theta\mathbb{P}(U \leq s)}, \quad U \sim \textit{Uniform}$$

and $\mathcal{L}(W_i)$ are distributed according to m.

see also Lambert (18), Johnston (19), Harris, Johnston, Roberts (21)

Yaglom's law and limiting genealogy

If $[U_N, d_N, \mu_N]$ converges to the a Brownian CPP(T, m). Then

- μ_N converges in distribution to mExp(1).
- Convergence of the genealogy of a K-sampling to the K-sampling of the Brownian CPP.

- Conclusion. Convergence of the moments to the moments of a Brownian CPP implies the generalized Yaglom's law and the convergence of the genealogy.
- Conjecture. Convergence of the moments at fixed time horizon implies convergence of the whole genealogy to the CRT.

Moments of random trees: discrete branching processes

Recall the tree T and collection of marks $(X_u;\ u \in T)$ of the branching process
Let $[U_N, d_N, \mu_N]$ be the mmm generated at time N where

$$\mu_N \;=\; \sum_{u \in U_N} \delta_{u,X_u}$$

The moments of $[U_N, d_T, \mu_N]$ are

$$\mathbb{E}[\Phi[U_N, d_T, \mu_N]] = \mathbb{E}\Big[\sum_{u_1,\dots,u_k \in T_N} \psi_{i,j}\big(d_T(u_i, u_j)\big) \prod_i \varphi_i(X_{u_i})\Big]$$

Moments of random trees: discrete branching processes

We rather work with factorial moments:

$$\begin{aligned}
\mathbb{E}[\Phi[T_N, d_T, \mu_N]] &= \mathbb{E}\Big[\sum_{u_1 \neq \dots \neq u_k \in T_N} \psi_{i,j}\big(d_T(u_i, u_j)\big) \prod_i \varphi_i(X_{u_i})\Big] \\
&= \mathbb{E}(Z_N^{(k)})\mathbb{E}\Big[\frac{Z_N^{(k)}}{\mathbb{E}(Z_N^{(k)})} \sum_{u_1 \neq \dots \neq u_k \in T_N} \psi_{i,j}\big(d_T(u_i, u_j)\big) \prod_i \varphi_i(X_{u_i})\Big]
\end{aligned}$$

with

- Z_N the population size at time N
- $Z_N^{(k)} = Z_N(Z_N - 1)\dots(Z_N - k + 1)$.
- Let S_k be the marked tree by (1) biasing by $Z_N^{(k)}$, and (2) picking k leaves without replacement.

Outline

1. A stochastic population model and its branching approximation.

2. A general approach to the convergence of genealogies

   2.1 The marked Gromov-weak topology and the methods of moments*

   2.2 Many-to-few and spinal decomposition

3. Two applications

   3.1 Multitype Branching processes

   3.2 BBM of Tourniaire (22).

see also *Gonzales, Harris, Horton, Kyprianou, Wang (20,21,22)

Multiple spines and many-to-few formula

The aim is to define a law $\mathbf{Q}_x^{N,k}$ on the tree with k leaves

$$\mathbb{E}\Big[\frac{Z_N^{(k)}}{\mathbb{E}(Z_N^{(k)})} \cdot \varphi(S_k)\Big] = \mathbf{Q}_x^{N,k}\big[\Delta_k \cdot \varphi(S_k)\big]$$

- under $\mathbf{Q}_x^{N,k}$ the tree S_k has a simple law
- the bias term Δ_k remains tractable and vanishes at the limit.

There are multiple ways of constructing $\mathbf{Q}_x^{N,k}$:

- in the original work of Harris and Roberts*, $\mathbf{Q}_x^{N,k}$ is defined forward in time as a system of branching particles
- in our work, $\mathbf{Q}_x^{N,k}$ is a obtained using a discrete CPP

*Harris, Roberts (2017)

Multiple spines and many-to-few formula

The aim is to define a law $\mathbf{Q}_x^{N,k}$ on the tree with k leaves

$$\mathbb{E}\Big[\frac{Z_N^{(k)}}{\mathbb{E}(Z_N^{(k)})} \cdot \varphi(S_k)\Big] = \mathbf{Q}_x^{N,k}\big[\Delta_k \cdot \varphi(S_k)\big]$$

- under $\mathbf{Q}_x^{N,k}$ the tree S_k has a simple law
- the bias term Δ_k remains tractable and vanishes at the limit.

There are multiple ways of constructing $\mathbf{Q}_x^{N,k}$:

- in the original work of Harris and Roberts*, $\mathbf{Q}_x^{N,k}$ is defined forward in time as a system of branching particles
- in our work, $\mathbf{Q}_x^{N,k}$ is a obtained using a discrete CPP

*Harris, Roberts (2017)

Construction of the 1-spine

Suppose that there exists a nonnegative harmonic function h:

$$\forall x \in E, \quad h(x) = \mathbb{E}[\langle \Xi(x), h \rangle] = \mathbb{E}\Big(\sum_{i=1}^{K(x)} h(\xi_i(x))\Big)$$

In particular, $W_N = \sum_{u \in T_N} h(X_u)$ is an additive martingale

We can define a Markov chain $(\zeta(n);\ n \geq 0)$ with transition

$$\mathbb{E}[\varphi(\zeta(1)) \mid \zeta(0) = x] = \frac{1}{h(x)} \mathbb{E}[\langle \Xi(x), h\varphi \rangle].$$

called the spine process.

For critical multitype branching processes

- $q(x, y) = m_{x,y} \frac{h(y)}{h(x)}$.
- h is the right eigenvector of $(m_{x,y})$.

Construction of the 1-spine

Suppose that there exists a nonnegative harmonic function h:

$$\forall x \in E, \quad h(x) = \mathbb{E}[\langle \Xi(x), h \rangle] = \mathbb{E}\Big(\sum_{i=1}^{K(x)} h(\xi_i(x))\Big)$$

In particular, $W_N = \sum_{u \in T_N} h(X_u)$ is an additive martingale

We can define a Markov chain $(\zeta(n);\ n \geq 0)$ with transition

$$\mathbb{E}[\varphi(\zeta(1)) \mid \zeta(0) = x] = \frac{1}{h(x)} \mathbb{E}[\langle \Xi(x), h\varphi \rangle].$$

called the spine process.

For critical multitype branching processes

- $q(x, y) = m_{x,y} \frac{h(y)}{h(x)}$.
- h is the right eigenvector of $(m_{x,y})$.

Construction of the 1-spine

Suppose that there exists a nonnegative harmonic function h:

$$\forall x \in E, \quad h(x) = \mathbb{E}[\langle \Xi(x), h \rangle] = \mathbb{E}(\sum_{i=1}^{K(x)} h(\xi_i(x)))$$

In particular, $W_N = \sum_{u \in T_N} h(X_u)$ is an additive martingale

We can define a Markov chain $(\zeta(n);\ n \geq 0)$ with transition

$$\mathbb{E}[\varphi(\zeta(1)) \mid \zeta(0) = x] = \frac{1}{h(x)} \mathbb{E}[\langle \Xi(x), h\varphi \rangle].$$

called the spine process.

For critical multitype branching processes

- $q(x, y) = m_{x,y} \frac{h(y)}{h(x)}$.
- h is the right eigenvector of $(m_{x,y})$.

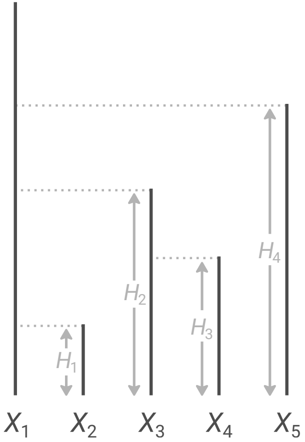
Construction of the k-spine

The measure $\mathbf{Q}_x^{N,k}$ is the probability measure:

- the tree structure is the CPP with i.i.d. branch lengths $(H_1, \dots, H_{k-1}) \in \{1, \dots, N\}$:

$$\forall i < j, \quad d(V_i, V_j) = \max\{H_i, \dots, H_{j-1}\}$$

- the marks along the branches evolve as the spinal process $(\zeta(n);\ n \geq 1)$
- at each branch point the process is duplicated and then evolve independently (!)



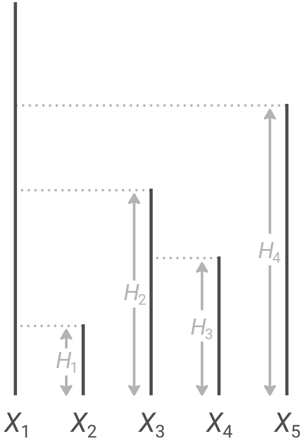
Construction of the k-spine

The measure $\mathbf{Q}_x^{N,k}$ is the probability measure:

- the tree structure is the CPP with i.i.d. branch lengths $(H_1, \dots, H_{k-1}) \in \{1, \dots, N\}$:

$$\forall i < j, \quad d(V_i, V_j) = \max\{H_i, \dots, H_{j-1}\}$$

- the marks along the branches evolve as the spinal process $(\zeta(n);\ n \geq 1)$
- at each branch point the process is duplicated and then evolve independently (!)



The biasing term

With this construction of $\mathbf{Q}_x^N$, if H_1 is uniform on $\{1, \ldots, N\}$ the bias is

$$\Delta_k = h(x) N^{k-1} k! \cdot \prod_{u \in B} \frac{m_{d_u}(\zeta_u)}{d_u!\, h(\zeta_u)} \cdot \prod_{u \in L} \frac{1}{h(\_u)}$$

where d_u is the (out-)degree of u, and

$$\forall x \in E, \quad m_d(x) = \mathbb{E}[\langle \Xi^{(d)}(x), h^{\otimes d} \rangle] := \mathbb{E}\Big[\sum_{\substack{i_1, \ldots, i_d = 1 \\ i_1 \neq \cdots \neq i_d}}^{K(x)} \prod_{j=1}^{d} h(\xi_{i_j}) \Big]$$

When $d = 2$,

$$\forall x \in E, \quad m_2(x) = \mathbb{E}\Big[\sum_{\substack{i_1, i_2 = 1 \\ i_1 \neq \cdots \neq i_d}}^{K(x)} h(\xi_{i_1}) h(\xi_{i_2}) \Big]$$

Theorem (Many to few, Foutel--Rodier, S (22+))

$$\mathbb{E}\left(\Phi([U_N, d_N, \mu_N])\right) = Q_x^{N,k}\left(\Delta \prod_{i,j} \psi_{i,j}(H_{\sigma(i),\sigma(j)}) \prod_i \varphi_i(\zeta_{V_{\sigma(i)}})\right)$$

where V_i *is the* i^{th} *leaf in the discrete CPP and*

$$\Delta_k = h(x)N^{k-1}k! \cdot \prod_{u\in B} \frac{m_{d_u}(\zeta_u)}{d_u!\, h(\zeta_u)} \cdot \prod_{u\in L} \frac{1}{h(\zeta_u)}$$

and

$$\forall x \in E, \quad m_d(x) := \mathbb{E}\Big[\sum_{\substack{i_1,\dots,i_d=1 \\ i_1\neq\cdots\neq i_d}}^{K(x)} \prod_{j=1}^{d} h(\xi_{i_j})\Big]$$

Outline

1. A stochastic population model and its branching approximation.
2. A general approach to the convergence of genealogies
    2.1 The marked Gromov-weak topology and the methods of moments*
    2.2 spinal decomposition and many-to-few
3. Two applications
    3.1 Multitype Branching processes
    3.2 BBM of Tourniaire (22).

---

see also *Gonzales, Harris, Horton, Kyprianou, Wang (20,21,22)

Application to multi-type branching processes

Suppose that E is finite, so that

$$\forall x \in E, \quad \Xi(x) = \sum_{y \in E} L_{xy} \cdot \delta_y$$

and assume that the mean reproduction matrix $M = (m_{xy})$ defined as

$$\forall x, y \in E, \quad m_{xy} = \mathbb{E}[L_{xy}]$$

is irreducible and aperiodic

By the Perron-Frobenius theorem we can find for M

- a leading eigenvalue λ, and we assume $\lambda = 1$ (criticality)
- a corresponding stationary measure $\tilde{h}$ (the left eigenvector)
- a corresponding harmonic function h (the right eigenvector) with $\langle \tilde{h}, h \rangle = 1$

Application to multi-type branching processes

Suppose that E is finite, so that

$$\forall x \in E, \quad \Xi(x) = \sum_{y \in E} L_{xy} \cdot \delta_y$$

and assume that the mean reproduction matrix $M = (m_{xy})$ defined as

$$\forall x, y \in E, \quad m_{xy} = \mathbb{E}[L_{xy}]$$

is irreducible and aperiodic

By the Perron-Frobenius theorem we can find for M

- a leading eigenvalue λ, and we assume $\lambda = 1$ (criticality)
- a corresponding stationary measure $\tilde{h}$ (the left eigenvector)
- a corresponding harmonic function h (the right eigenvector) with $\langle \tilde{h}, h \rangle = 1$

Application to multi-type branching processes

Theorem

Consider a critical multi-type branching process with offspring distribution (L_{xy}) satisfying

$$\forall k \geq 1, \quad \mathbb{E}[L_{xy}^k] < \infty.$$

Then, started from any initial condition and conditional on survival at generation N,

$$\lim_{N\to\infty} \left[T_N, \tfrac{d_T}{N}, \tfrac{\mu_N}{N}\right] = [(0, Y), d_P, \tfrac{\Sigma}{2} \,\mathrm{Leb} \otimes \tilde{h}]$$

in distribution for the Gromov-weak topology, where $[(0, Y), d_P]$ is a Brownian CPP and

$$\Sigma = \langle \tilde{h}(\mathrm{d}x), m_2(x) \rangle, \;\; m_2(x) \; = \; \mathbb{E}_x\Big(\sum_{i_1 \neq i_2} h(\xi_{i_1}) h(\xi_{i_2}) \Big).$$

Proof:cutoff

Proposition

Let $[T_n, d_n, \mu_n]$ be a sequence of random mmm and let $T'_n \subset T_n$ be a closed set. Assume that

1. $|U'_n|$ *converges in distribution to a positive r.v.*
2. $\mathbb{E}(|U_n| - |U'_n|) \to 0$

If $[T_n, d_n, \mu_n]$ converges in distribution then M_n converges to the same limit.

Without loss of generality, we can restrict ourself to the case where $\mathbb{E}(L^k_{x,y}) < \infty$ for every $k \in \mathbb{N}$, and apply the method of moments.

Proof: Many-to-few

According to the Gromov-weak topology, we need to compute

$$\mathbb{E}\left[\Phi(T_N, \tfrac{d_T}{N}, \tfrac{\mu_N}{N}) \mid Z_N > 0\right] = \frac{1}{N^k}\mathbb{E}\Big[\sum_{u_1,\dots,u_k \in T_N} \prod_{i,j} \varphi_{i,j}\big(\tfrac{d_T(v_i,v_j)}{N}\big) \prod_i \varphi_i(X_{v_i}) \Big| Z_N > 0\Big].$$

and show that it converges to the moments of a CPP

$$k!\, \mathbb{E}\Big[\prod_{i,j} \varphi_{i,j}\big(H_{\sigma(i),\sigma(j)}\big)\Big] \prod_i \int \varphi_i(x) \frac{\Sigma}{2} \tilde{h}(x)\, dx$$

By the (rescaled) many-to-few formula,

$$\begin{aligned} &\mathbb{E}\left[\Phi(T_N, \tfrac{d_T}{N}, \tfrac{\mu_N}{N}) \mid Z_N > 0\right] \\ &= \frac{1}{N\mathbb{P}(Z_N > 0)} \bar{\mathbf{Q}}_x^{1,k}\Big(\Delta_k \cdot \prod_{i,j} \varphi_{i,j}\big(H_{\sigma(i),\sigma(j)}\big) \prod_i \varphi_i(\zeta_{V_\sigma(i)})\Big) \end{aligned}$$

where $\bar{\mathbf{Q}}_x^{1,k}$ is the rescaled spine (time rescaled by $\frac{1}{N}$).

Proof: convergence of the tree

We construct $\bar{Q}_x^{1,k}$ with uniform branch lengths on $\{\frac{1}{N},\dots,1\}$

The distance between leaf V_i and $V_j > V_i$ is

$$d(V_i, V_j) = \max\{\tfrac{H_i}{N},\dots,\tfrac{H_{j-1}}{N}\} \xrightarrow[N\to\infty]{} \max\{\tilde{H}_i,\dots,\tilde{H}_{j-1}\}$$

for $(\tilde{H}_1,\dots,\tilde{H}_{k-1})$ i.i.d. uniform on $(0,1)$

In the limit:

- the tree is binary and is the discrete CPP constructed from uniforms (as in the size biased CPP)
- the number of vertices between two branch points is large (of the order of N)

Proof: convergence of the bias

The Perron-Frobenius theorem ensures that

- the spine has stationary measure $h\tilde{h}$
- it converges in distribution to it

The bias then becomes

$$\Delta_k = h(x)k! \cdot \prod_{u\in B} \frac{m_{d_u}(\zeta_u)}{d_u!\,h(\zeta_u)} \cdot \prod_{u\in L} \frac{1}{h(\zeta_u)}$$

$$\xrightarrow[N\to\infty]{\bar{\mathbb{Q}}_x^{1,k}} h(x)k! \cdot \prod_{i=1}^{k-1} \frac{m_2(Y_i)}{2h(Y_i)} \cdot \prod_{i=1}^{k} \frac{1}{h(Y_i')}$$

for an i.i.d. collection (Y_i, Y_i') of r.v. distributed as $h\tilde{h}$ (invariant distribution of the 1-spine)

Proof: convergence of the bias

The Perron-Frobenius theorem ensures that

- the spine has stationary measure $h\tilde{h}$
- it converges in distribution to it

The bias then becomes

$$\Delta_k = h(x)k! \cdot \prod_{u \in B} \frac{m_{d_u}(\zeta_u)}{d_u!\, h(\zeta_u)} \cdot \prod_{u \in L} \frac{1}{h(\zeta_u)}$$

$$\xrightarrow[N \to \infty]{\bar{\mathbf{Q}}_x^{1,k}} h(x)k! \cdot \prod_{i=1}^{k-1} \frac{m_2(Y_i)}{2h(Y_i)} \cdot \prod_{i=1}^{k} \frac{1}{h(Y_i')}$$

for an i.i.d. collection (Y_i, Y_i') of r.v. distributed as $h\tilde{h}$ (invariant distribution of the 1-spine)

Proof: end

The previous convergences prove that

$$\bar{\mathbf{Q}}_x^{1,k}\big(\Delta_k \cdot \prod_{i,j} \psi_{i,j}\big(H_{\sigma_i,\sigma_j}\big) \prod_i \varphi_i(\zeta_{V_{\sigma(i)}})\big)$$

$$\xrightarrow[N\to\infty]{} h(x)\Big(\frac{\langle \pi, m_2\rangle}{2}\Big)^{k-1} k!\, \mathbb{E}[\prod_{i,j} \varphi_{i,j}(H_{\sigma(i),\sigma(j)})] \prod_i \mathbb{E}(\frac{\varphi_i(Y)}{h(Y)})$$

where $X \sim h\tilde{h}$, (H_i) are i.i.d. uniform on $(0,1)$ and

$$\forall i < j, \quad H_{i,j} = \max\{H_i, \dots H_{j-1}\}$$

Using the Kolmogorov estimate,

$$\lim_{N\to\infty} N\mathbb{P}(Z_N > 0) = \frac{2h(x)}{\langle \pi, m_2\rangle}$$

We recover the moments of a marked Brownian CPP with measure $\frac{\Sigma}{2}\tilde{h}$!

Proof: end

The previous convergences prove that

$$\bar{\mathbf{Q}}_x^{1,k}\Big(\Delta_k \cdot \prod_{i,j} \psi_{i,j}\big(H_{\sigma_i,\sigma_j}\big) \prod_i \varphi_i(\zeta_{V_{\sigma(i)}})\Big)$$

$$\xrightarrow[N\to\infty]{} h(x)\Big(\frac{\langle \pi, m_2\rangle}{2}\Big)^{k-1} k!\,\mathbb{E}[\prod_{i,j} \varphi_{i,j}(H_{\sigma(i),\sigma(j)})] \prod_i \mathbb{E}(\frac{\varphi_i(Y)}{h(Y)})$$

where $X \sim h\tilde{h}$, (H_i) are i.i.d. uniform on $(0,1)$ and

$$\forall i < j, \quad H_{i,j} = \max\{H_i, \dots H_{j-1}\}$$

Using the Kolmogorov estimate,

$$\lim_{N\to\infty} N\mathbb{P}(Z_N > 0) = \frac{2h(x)}{\langle \pi, m_2\rangle}$$

We recover the moments of a marked Brownian CPP with measure $\frac{\Sigma}{2}\tilde{h}$!

Outline

1. A stochastic population model and its branching approximation.

2. A general approach to the convergence of genealogies

   2.1 The marked Gromov-weak topology and the methods of moments*

   2.2 spinal decomposition and many-to-few

3. Two applications

   3.1 Multitype Branching processes

   3.2 BBM of Tourniaire (22).

---

see also *Gonzales, Harris, Horton, Kyprianou, Wang (20,21,22)

BBM with inhomogeneous branching rate*

- Let $\mu > 0$, $f \geq 0$ with $\mathrm{Supp}(f) \subseteq [0,1]$.
- $\varepsilon \geq 0$ (Allee effect).
- BBM with
  1. Inhomogeneous branching rate
  $$\forall x > 0, \quad r(x) = \frac{1}{2} + \frac{\varepsilon}{2} f(x)$$
  2. Drift $-\mu$

     μ is chosen in such a way that the system is critical, i.e., the average number of particles remains roughly constant (see later).
  3. Killing at 0.
- $\varepsilon = 0$: Standard BBM**.

*Tourniaire (22), **Berestycki, Berestycki, Schweinsberg (13)

Criticality

- Define $(t, y) \to p_t(x, y)$ as the density of particles at time t starting with a unique particle at x, i.e., for every test function f

$$\mathbb{E}_x\left(\sum_{u\in N_t} f(X_u(t))\right) = \int_{\mathbb{R}_+} p_t(x, y)f(y)\,dy.$$

- $(t, y) \to p_t(x, y)$ is solution of the fundamental equation

$$\partial_t u = \frac{1}{2}\partial_{xx}u + \mu\partial_x u + r(x)u$$
$$\text{with} \quad u(t, 0) = 0.$$

- Criticality: Choose μ to keep the number of particles "under control" (stable front, μ is interpreted as the speed of the front).

Criticality

Let $L >> 1$ and consider the modified equation

$$\partial_t u = \frac{1}{2}\partial_{xx} u + \mu \partial_x u + r(x) u$$
$$\text{with} \quad u(t,0) = u(t,L) = 0$$

BBM killed at 0 and L.

Write $p_t(x,y) = e^{\mu(x-y)+\frac{1}{2}(1-\mu^2)t} q_t(x,y)$ (Girsanov transformation). Then $(t,y) \to q_t(x,y)$ is solution of

$$\partial_t u = \underbrace{\frac{1}{2}\partial_{xx} u + (r(x) - \frac{1}{2}) u}_{\text{Self-adjoint}} \quad \text{with} \quad u(t,0) = u(t,L) = 0$$

Criticality

Consider

$$\partial_t u = \frac{1}{2}\partial_{xx} u + (r(x) - \frac{1}{2})u \quad \text{with} \quad u(t,0) = u(t,L) = 0$$

Sturm-Liouville: The eigenvalues of the operator can be numbered $\lambda_1 > \lambda_2 > \cdots > \lambda_k \cdots \to -\infty$ and

$$q_t(x,y) \;=\; \sum_{k\geq 1} e^{\lambda_k t} v_k(x) v_k(y),$$

where the v_k's are the eigenfunctions with $||v_k||_2 = 1$. As $t \to \infty$

$$p_t(x,y) \quad \approx_{t\to\infty} \quad e^{\mu(x-y)+(\frac{1}{2}(1-\mu^2)+\lambda_1)t} v_1(x) v_1(y)$$

$\lambda_1 \uparrow \lambda_1^\infty < \infty$* as $L \to \infty$, the system is critical when

$$\mu = \sqrt{1+\beta^2}, \quad \text{with } \beta := \sqrt{2\lambda_1}.$$

Pinsky (95)

Left-right principal eigenfunctions

As $t \to \infty$

$$
\begin{aligned}
p_t(x,y) \quad \approx_{t\to\infty} \quad & e^{\mu(x-y)+(\frac{1}{2}(1-\mu^2)+\lambda_1)t} v_1(x) v_1(y) \\
\approx \quad & h(x)\tilde{h}(y), \text{ at criticality}
\end{aligned}
$$

where

$$
h(x) \;=\; \frac{1}{\tilde{c}} e^{\mu x} v_1(x), \tilde{h}(y) = \tilde{c} e^{-\mu y} v_1(y)
$$

and $\tilde{c}$ is a renormalization constant such that $\int_0^L \tilde{h}(x)dx = 1$.

Analogously to multitype GW, we think of h (resp., $\tilde{h}$) as the right (resp, left) eigenfunction of the mean operator

$$
\mathcal{L}^* f \;=\; \frac{1}{2}\partial_{xx} f - \mu \partial_x f + r(x) f.
$$

Pulled, semi/fully Pushed

Define

$$\alpha = \frac{\mu + \beta}{\mu - \beta}, \quad \text{where } \beta := \sqrt{2\lambda_1^\infty}$$

Definition

1. Pulled regime: $\lambda_1^\infty = 0 \iff \alpha = 1.$
2. Semi Pushed: $\lambda_1^\infty \in (0, \frac{1}{16}) \iff \alpha \in (1, 2).$
3. Fully Pushed : $\lambda_1^\infty > \frac{1}{16} \iff \alpha > 2.$

Conjecture

1. Pulled regime: Neveu CSBP.
2. Semi Pushed: α-stable CSBP
3. Fully Pushed : Feller

Pulled, semi/fully Pushed

Take

$$r(x) = \frac{1}{2} + \varepsilon f(x)$$

Proposition (Pinsky (95))

There exists $0 < \varepsilon_1 < \varepsilon_2 < \infty$ *such that*

- $\varepsilon < \varepsilon_1$: *pulled*
- $\varepsilon \in (\varepsilon_1, \varepsilon_2)$: *semi-pushed*
- $\varepsilon \in (\varepsilon_2, \infty)$: *fully-pushed*

Pulled vs Pushed

In the PDE, we observe the following transition. Let c be the speed of the front

- Pulled. $c = 1$
- Pushed. $c > 1$

In the BBM, let μ be negative drift

- Pulled. $(\varepsilon < \varepsilon_1)$ $\mu = 1$
- Pushed. $(\varepsilon > \varepsilon_1)$ $\mu > 1$

Semi vs Fully Pushed

In multitype GW, the limit involves

$$\frac{\Sigma}{2} = \left\langle m_2, \tilde{h} \right\rangle, \quad \text{with } m_2(x) = \mathbb{E}_x \left(\sum_{i_1 \neq i_2} h(\xi_{i_1}) h(\xi_2) \right).$$

The the BBM, the analogous quantity is given by

$$\frac{\Sigma}{2} = \int_0^L r(x) h^2(x) \tilde{h}(x) dx = \int_0^L r(x) e^{\mu x} v_1^3(x) dx$$

where

$$\frac{1}{2} v_1'' + r(x) - \frac{1}{2} = \lambda_1 v_1, \quad \text{with} \quad v_1(0) = v_1(L) = 0$$

so that when $x > 1$

$$\frac{1}{2} v_1'' = \lambda_1 v_1 \quad \text{and} \quad v_1(x) \approx O(e^{-\beta x})$$

Σ remains finite only when $\alpha > 2$ (i.e., $\mu < 3\beta$)

Fully Pushed

Let $\mathcal{N}_t$ the set of particles alive at time t

Theorem (Tourniaire, S., 22+)

In the fully pushed regime ($\alpha > 2$),

$$\mathbb{P}_x(|\mathcal{N}_t| > 0) \sim_{2t \to \infty} \frac{h(x)}{t\Sigma}$$

where

$$\frac{\Sigma}{2} = \int_0^\infty cr(x)\tilde{h}(x)h^2(x)dx < \infty.$$

Fully Pushed

Let $\mu_t = \sum_{u \in \mathcal{N}_t} \delta_{u, X_u}$. d is the genealogical distance.

Theorem (Tourniaire, S., 22+)

In the fully pushed regime ($\alpha > 2$), the random mmm

$$\left(\mathcal{N}_N, \frac{d}{N}, \frac{\mu_N}{N}\right)$$

conditioned on $|\mathcal{N}_N| > 0$ converges to the Brownian CPP with measure

$$m(dx) \;=\; \frac{\Sigma}{2}\tilde{h}(x)\,dx.$$

This implies Yaglom's law and convergence of the genealogy to the genealogy of a critical branching process.

Elements of the proof

GW with heavy tailed distribution

Consider a GW $(Z_t; t \in \mathbb{N})$ with

$$\mathbb{P}(\xi > x) \approx \frac{c}{x^\alpha}, \quad \alpha \in (1,2).$$

Theorem (Duquesnes, Le Gall (03))

Define $\beta_N = N^{\frac{1}{\alpha-1}}$. *The process* $(\frac{1}{\beta_N} Z_{tN}; t \geq 0)$ *converges in* $D(0,\infty)$ *to a* α*-stable CSBP.*

Let $A > 0$ and consider the GW process with offspring distribution

$$\bar{\xi}_N = \xi 1(\xi \leq A\beta_N).$$

For $p \geq 2$,

$$\mathbb{E}(\bar{\xi}_N^p) = \beta_N^{\alpha-p} \int_0^\infty \Pi(dx), \text{ where } \Pi(dx) = 1(x \leq A)\, dx$$

GW with heavy tailed distribution

The cutoff procedure allows to use the method of moments for random mmm spaces. This provides an alternative proof for

Theorem (Duquesnes, Le Gall (03))

Assume that $Z_0 = x\beta_N$. Let U_n be the individuals at generation n.

1. $(\frac{1}{\beta_N} Z_{tN}; t \geq 0)$ *converges to an* α*-stable CSBP*
2. *For every* $t > 0$*, there exists a limiting random mmm* $[U_\infty^{(t,x)}, d_\infty^{(t,x)}, \mu_\infty^{(t,x)}]$*. such that*

$$[U_{Nt}, \frac{1}{N} d_{Nt}, \frac{1}{\beta_N} \mu_{Nt}] \ = \ [U_\infty^{(t,x)}, d_\infty^{(t,x)}, \textit{Leb}].$$

The limiting metric space can defined using reduced processes* or flows of bridges**

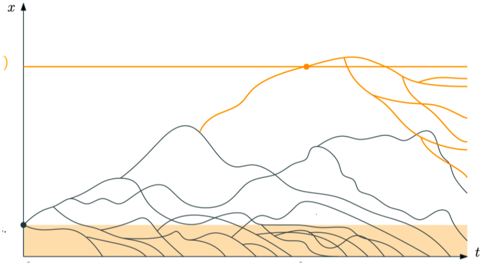
*Duquesnes, Le Gall (03), ** Bertoin, Le Gall (03)

Semi-pushed regime

Imagine a particle performing a large excursion to level $L >> 1$.

The particle is pushed to 0 by the drift $-\mu$.

During this relaxation period, the particle generates a large number of particles.



Cutoff. Kill the particles for a well chosen level $L >> 1$

The 1-spine satisfies

$$d\zeta_t = \frac{v_1'(\zeta_t)}{v_1(\zeta_t)} dt + dw_t$$

The invariant distribution is given by $v_1^2 \sim O(e^{-2\beta x})$.

Take $L_A = \frac{1}{2\beta} \log(N) + \frac{1}{\mu - \beta} \log(A)$. Start N particles at x. By the many-to-one formula

$$\begin{aligned} \beta_N \mathbb{E}\left(\int_0^{L_A} p_t(x,y)\, dy \right) &\approx \beta_N \mathbb{E}(\int_0^{L_A} v_1^2(y) \frac{h(x)}{h(y)} dy) \\ &\approx 1 - O(\frac{1}{A}) \end{aligned}$$

Theorem (Tourniaire (21), Foutel--Rodier, S. Tourniaire (22++))

Assume that the existence of a density such that

$$\frac{1}{\beta_N} \sum_{u \in \mathcal{N}_0} \delta_{X_u} \Longrightarrow f(x)\,dx$$

Let U_n be the individuals at generation n.

1. $(\frac{1}{\beta_N} Z_{tN}; t \geq 0)$ *converges to an α-stable CSBP*
2. *For every $t > 0$, there exists $c > 0$ such that*

$$[U_{Nt}, \frac{1}{N} d_{Nt}, \frac{1}{\beta_N} \mu_{Nt}] \;=\; [U_\infty^{(t,x)}, d_\infty^{(t,x)}, c\,\mathrm{Leb} \otimes \tilde{h}].$$

where $[U_\infty^{(t,x)}, d_\infty^{(t,x)}]$ is the genealogy of an α-stable CSBP (see above).

Related work and perspective

- [Powell 15] considers branching Brownian motions in a bounded domain (with no drift). Yaglom law and convergence of the the genealogy to the CRT (CV via contour processes).
- [Kyrpianou, Horton, Roberts et al 21] consider the moments of branching processes (no genealogy) under the assumption that h is bounded + Perron Froebenius-like assumptions (convergence of the spine to the invariance measure exponentially fast).
- In our case $||h||_\infty = \infty$. This goes beyond pure technicality: the pushed and semi-pushed regimes are determined by the integrability condition
$$\int h(x)\Pi(x)\,dx < \infty.$$
- Spinal decomposition in the semi-pushed regime ? Yes $\cdots$ but some cutoff is needed ([Tourniaire, Berestycki, Berestycki, Schweinseberg]).
- Going beyond the toy model [Birzu et al. 18], [Etheridge, Pennington 20] .

Summary and perspective

The approach we propose to study the scaling limit of genealogies relies on

- using the Gromov-weak topology to work with moments
- using spinal decomposition results to compute the moments

This approach seems successful for critical processes that have "finite variance" such as:

- multitype branching processes
- more general processes with "ergodic" spines
- the two examples of Emmanuel's talk
- branching process in varying environment: Florin Boenkost's work

Some future directions:

- extend it to multiple mergers
- consider some subcritical and supercritical processes

Summary and perspective

The approach we propose to study the scaling limit of genealogies relies on

- using the Gromov-weak topology to work with moments
- using spinal decomposition results to compute the moments

This approach seems successful for critical processes that have "finite variance" such as:

- multitype branching processes
- more general processes with "ergodic" spines
- the two examples of Emmanuel's talk
- branching process in varying environment: Florin Boenkost's work

Some future directions:

- extend it to multiple mergers
- consider some subcritical and supercritical processes

Summary and perspective

The approach we propose to study the scaling limit of genealogies relies on

- using the Gromov-weak topology to work with moments
- using spinal decomposition results to compute the moments

This approach seems successful for critical processes that have "finite variance" such as:

- multitype branching processes
- more general processes with "ergodic" spines
- the two examples of Emmanuel's talk
- branching process in varying environment: Florin Boenkost's work

Some future directions:

- extend it to multiple mergers
- consider some subcritical and supercritical processes

References

F. Foutel-Rodier, E. Schertzer. Convergence of genealogies through spinal decomposition with an application to population genetics. ArXiv: 2201.12412.

A. Greven, P. Pfafflehuber, A. Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probability Theory and Related Fields* **145** (2009), 285-322.

A. Depperschmidt, A. Greven, P. Pfaffelhuber. Marked metric measure spaces. *Electronic Communications in Probability* **16** (2011), 174-188.

S.C. Harris, M.I. Roberts. The many-to-few lemma and multiple spines. *Annales de l'Institut Henri Poincaré Probabilités et Statistique* **53** (2017), 226-242.

L. Popovic. Asymptotic genealogy of a critical branching process. *Annals of Applied Probability* **14** (2004), 2120-2148.

References

D. Aldous. The Continuum Random Tree. I. *The Annals of Probability* **19** (1991), 1-28.

T. Duquesnes, J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque* **281** (2002), 153pp.

G. Miermont. Invariance principles for spatial multitype Galton–Watson trees. *Annales de l'Institut Henri Poincaré Probabilités et Statistique* **44** (2008), 1128-1161.

J.-F. Le Gall. A class of path-valued Markov processes and its applications to superprocesses. *Probability Theory and Related Fields* **95** (1993), 25-46.

L. Popovic, M. Rivas. The coalescent point-process of multi-type branching trees. *Stochastic Processes and their Applications* **124** (2014), 29pp.