Genealogy of catalytic branching models

Anita Winter, Erlangen

(with Andreas Greven & Lea Popovic)

"Probalistic models in evolutionary biology"

C.I.R.M. Marseille, 25.Mai 2009

Motivation

Branching Models - encoding the evolution of a population:

−> **population size**

−> **family forest** (genealogy/ancestral relationships)

Scaling Limits
- many individuals, each of small mass, rapid branching

Invariance Principles - general classes of models with same limit:

−> population - **continuous state branching processes (CSBP)**

−> genealogy - **continuum random trees (CRTs)**

Example: Independent Branching

critical Galton-Watson processes whose **offspring distribution** ...

- ... has **finite variance**:

  - population $\Rightarrow$ **Feller diffusion**   Feller '51

  - genealogy $\Rightarrow$ **Aldous's CRT**   Aldous '93

- ... is in the domain of attraction of an α**-stable distribution**:

  - population $\Rightarrow$ α**-stable CSBP**   Lamperti '67

  - genealogy $\Rightarrow$ α**-stable CRT**   Duquesne-LeGall '02

Our model: catalytic branching

$->$ the **catalyst** $(\eta_t)_{t\geq 0}$ is a critical binary GW with constant branching rate $= 1$

$->$ the **reactant** $(\xi_t)_{t\geq 0}$ is a critical binary GW with time-inhomogeneous branching rate **depending on** the current population size of the **catalyst**.

... and the scaling limit

In the nth-rescaling step $(\tilde{\eta}^n, \tilde{\xi}^n)$ is a continuous time MC with $(\tilde{\eta}_0^n, \tilde{\xi}_0^n) = (1, 1)$ and

$$(\tilde{\eta}^n, \tilde{\xi}^n) \mapsto \begin{cases} (\tilde{\eta}^n \pm \frac{1}{n}, \tilde{\xi}^n), & \text{at rate } \frac{1}{2} n^2 \tilde{\eta}^n, \\ (\tilde{\eta}^n, \tilde{\xi}^n \pm \frac{1}{n}), & \text{at rate } \frac{1}{2} n^2 \tilde{\eta}^n \tilde{\xi}^n. \end{cases}$$

Fact. Greven, Klenke & Wakolbinger '99

$(\tilde{\eta}^n, \tilde{\xi}^n) \Rightarrow (X, Y)$ where

$$\mathrm{d}X_t = \sqrt{X_t}\, \mathrm{d}W_t^X$$

$$\mathrm{d}Y_t = \sqrt{X_t Y_t}\, \mathrm{d}W_t^Y$$

A remark on extinction times

$$\mathrm{d}X_t = \sqrt{X_t}\,\mathrm{d}W_t^X$$

$$\mathrm{d}Y_t = \sqrt{X_t Y_t}\,\mathrm{d}W_t^Y$$

and

$$\rho^0 := \inf\{t \geq 0 : X_t = 0\}, \qquad \tau^0 := \inf\{t \geq 0 : Y_t = 0\}.$$

Basic Facts.

- $\rho^0 < \infty$, almost surely.

- The reactant gets absorbed in Y_{ρ^0} at time ρ^0.

- Penssel '03

$$\mathbb{P}\{\rho^0 < \tau^0\} = \tfrac{1}{\sqrt{5}} \in (0,1).$$

Problems we would like to answer ...

- from a **quenched view** (given the **catalyst** mass process):

  – establish **existence** of a limit reactant genealogy

  – **analytic characterizations** of the limit

- from an **annealed view:**

  – **joint convergence** of catalyst and reactant genealogies

  – **differences** in catalyst forest from Aldous's CRF

Existence of limit reactant genealogies

Rooted real trees

A **rooted real tree** (T, r, ρ) is a

- **metric space** (T, r),

- **path-connected**,

- 0-**hyperbolic**, i.e. $\forall A, B, C \in T \; \exists P \in [A, B]$ with
  $P \in [B, C] \cap [A, C]$
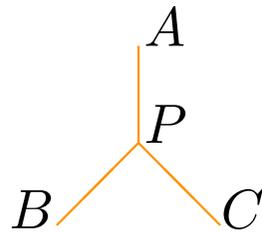
- with the **root** $\rho \in T$ distinguished.

$\mathbb{T} :=$ (root invariant) isometry classes of **compact** rooted $\mathbb{R}$-trees

Coding finite forest as rooted $\mathbb{R}$-trees

$$T := \bigcup_{t \in [0, \text{height}]} \text{all individuals alive time } t$$

$$d(\iota, \iota') := \text{generation of}(\iota) + \text{generation of}(\iota') - 2 \cdot \text{generation of}(\iota' \wedge \iota)$$

$$\rho := \text{glueing together all the trees' ancestors}$$
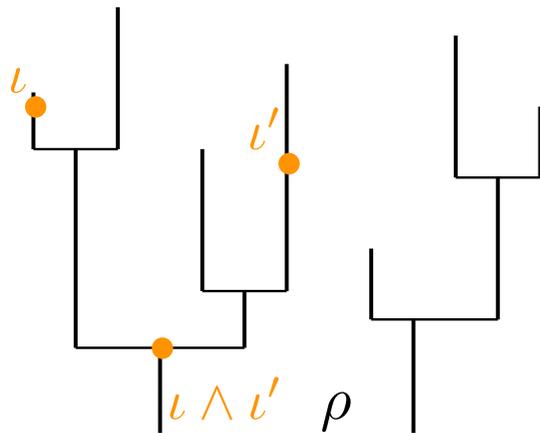
Rule. At the height where the catalyst dies **cut off** the reactant forest

Hausdorff distance

Let (X, d) be a metric space.

Hausdorff distance:

For $A_1, A_2 \subseteq_{\mathrm{closed}} X$,

$$d_{\mathrm{H}}(A_1, A_2) := \inf\{\varepsilon > 0 : A_1 \subseteq A_2^\varepsilon \text{ \textbf{and} } A_2 \subseteq A_1^\varepsilon\},$$

where A^ε is the ε-neighborhood of A.

Gromov strong topology

$$(\mathbf{T_N}, \mathbf{d_N}, \rho_{\mathbf{N}}) \underset{\mathbf{N}\to\infty}{\longrightarrow} (\mathbf{T}, \mathbf{r}, \rho)$$

iff $\{(T_N, d_N); \ N \in \mathbb{N}\}$ and (T, d) can be embedded via isometries $\{\varphi_N; \ N \in \mathbb{N}\}$ and φ, resp., in one and the same compact metric space $(\mathbf{Z}, \mathbf{d_Z})$ on which

$$\varphi_{\mathbf{N}}(\mathbf{T_N}) \underset{\mathbf{N}\to\infty}{\longrightarrow} \varphi(\mathbf{T}), \qquad \varphi_{\mathbf{N}}(\rho_{\mathbf{N}}) \underset{\mathbf{N}\to\infty}{\longrightarrow} \varphi(\rho)$$

in the **Hausdorff topology on** $(\mathbf{Z}, \mathbf{d})$.

Existence of reactant limit forest

- let $(\tilde{\eta}^{\mathrm{for},n}, \tilde{\xi}^{\mathrm{for},n})$ be the pairs of $\mathbb{R}$-trees corresponding to the catalyst and reactant population in the nth-approximation step

-

  $(\cdot \, ; \cdot) :=$ conditional law of reactant **given** a realization of the catalyst

Proposition. Assume that $\sup_{t \leq T} \left| \tilde{\eta}^n_t - X_t \right| \to 0$ a.s. Then the family $\{ (\tilde{\xi}^{\mathrm{for},n} ; \tilde{\eta}^n); \, n \in \mathbb{N} \}$ is relatively compact.

Characterization via excursions

Main example: Tree "below" an excursion

$\varphi \in C([0, L(\varphi)])$, $L(\varphi) \in (0, \infty)$, $\varphi\big|_{\{0, L(\varphi)\}} \equiv 0$, $\varphi\big|_{(0, L(\varphi))} > 0$

For $u, u' \in [0, L(\varphi)]$ put $\mathbf{u} \equiv_\varphi \mathbf{u}'$ iff **no minima** in between them.
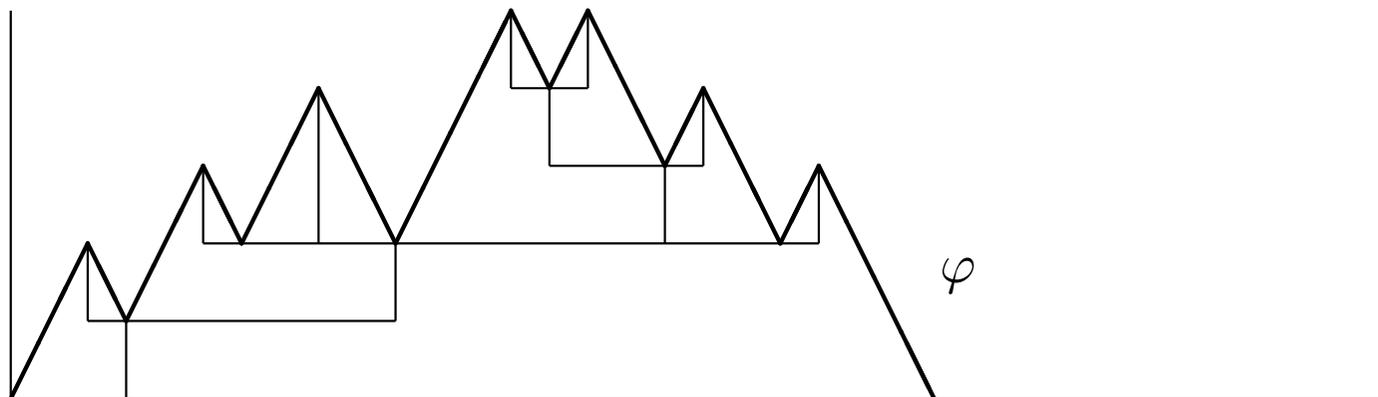
$$\mathbf{T}_\varphi := \{[u]; u \in [0, L(\varphi)]\}, \qquad \rho_\varphi := [0],$$

$$\mathbf{d}_\varphi([u], [u']) := \varphi(u) + \varphi(u') - 2 \inf_{s \in [u \wedge u', u \vee u']} \varphi(s).$$

Fact. $T\big|_\varphi$ is **compact** real tree.

Example. Aldous's CRT $=$ tree "below" Brownian excursion

From trees to forests

several trees in a forest = several aligned excursions

trees $=$ # **excursions** $=$ # **of zeros** -1.

Planar embedding and linear order

Planar embedding requires a **linear ordering** which

- extends the partial order, and

- if x, y, x', y' are s.t. $x \leq^{\text{lin}} y$ and $x \wedge y \leq x' \wedge x$ and $x \wedge y \leq y' \wedge y$ then $x' \leq^{\text{lin}} y'$.

Example.

$+$ **lexicographic order**
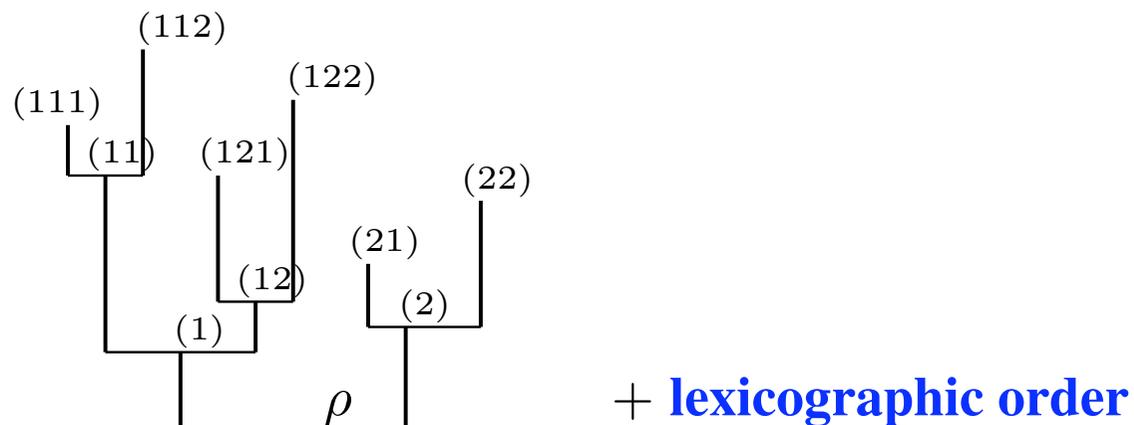
e.g., $x = (111)$, $y = (122)$, $x' = (112)$ and $y' = (12)$

The contour process of finite trees

$(\mathbf{T}, \rho)$: finite (linearly ordered) forest

σ : speed of traversal

$\mathcal{C}(T, \sigma)$: records height of the depth-first search around the forest
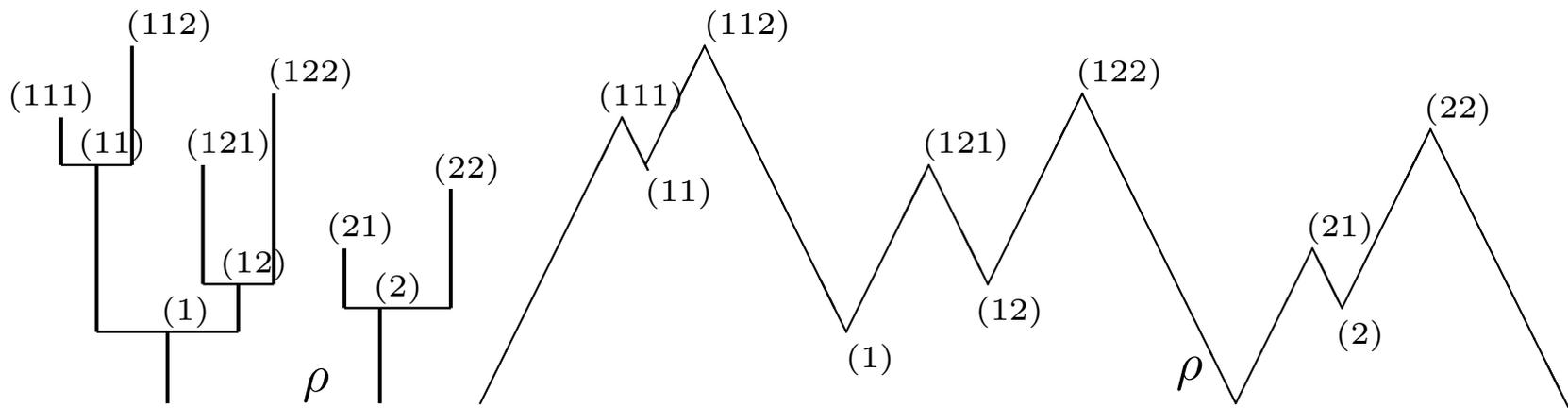
Lemma. In continuous time binary GW with **given** branching rate, $\mathcal{C}(T, \sigma)$ is a linear interpolation of an alternating random walk.

Re-scaling the contour

- Speeding up branching by n = rescaling **edges** by a factor $\frac{1}{\mathbf{n}}$,

- a GW-trees has height of order $\mathcal{O}(n)$ with probability $\mathcal{O}(\frac{1}{n})$

$\Rightarrow$ if we start initially with n-trees yields a **Poisson number** of trees each having height $\mathcal{O}(n)$

- given a GW-tree has height of order $\mathcal{O}(n)$ its **number of edges** is of order $\mathcal{O}(\mathbf{n^2})$

Since edges are of length $\mathcal{O}(\frac{1}{n})$, we choose in the nth-approximation step

$$\sigma := \mathbf{n}.$$

Obvious. [Brownian rescaling]

If $|\beta|$ is **reflected Brownian motion**, $\ell(|\beta|)$ its **local time** process, i.e.,

$$\ell(|\beta|)_t := \lim_{\varepsilon \to 0} \varepsilon^{-1} \int_0^t \mathrm{d}u \, \mathbf{1}\{|\beta|_u \in [0, \varepsilon]\},$$

and $\ell(|\beta|)^{-1}$ its **inverse**, i.e.,

$$\ell(|\beta|)^{-1}(s) := \inf \left\{ t \geq 0 : \ell(|\beta|)_t = s \right\},$$

then

$$\mathcal{C}(\tilde{\eta}^{\mathrm{for}, \mathbf{n}}; \mathbf{n}) \underset{\mathbf{n} \to \infty}{\Longrightarrow} \mathbf{2} \cdot |\beta|_{\cdot \wedge \ell_0^{-1}(\mathbf{1})}.$$

$\delta > 0$,

$$\tilde{T}^{\delta,n} := \inf\left\{t \geq 0 : \tilde{\eta}^n_t \leq \delta\right\}.$$

Lemma.

Let $V_u := \frac{1}{2}\mathsf{slope}(C_u)$ then $(C_u, V_u)_{u \geq 0}$ is a random evolution with

$$\tilde{\mathbf{A}}^{\delta,\mathbf{n}}\mathbf{f}(\mathbf{c},\mathbf{v}) = \mathbf{nv}\frac{\partial}{\partial \mathbf{c}}\mathbf{f}(\mathbf{c},\mathbf{v}) + \mathbf{n^2}\tilde{\eta}^{\mathbf{n}}_{\mathbf{c}}\left[\mathbf{f}(\mathbf{c},-\mathbf{v}) - \mathbf{f}(\mathbf{c},\mathbf{v})\right]$$

and domain

$$\mathcal{D}(\tilde{\mathbf{A}}^{\delta,\mathbf{n}}) = \left\{\mathbf{f} \in \mathbf{C^{1,0}}(\mathbf{0}, \tilde{\mathbf{T}}^{\delta,\mathbf{n}} \times \{-\mathbf{1},\mathbf{1}\}) : \partial_{\mathbf{c}}\mathbf{f}\big|_{\{\mathbf{0},\tilde{\mathbf{T}}^{\delta,\mathbf{n}}\}\times\{\mathbf{0},\mathbf{1}\}} \equiv \mathbf{0}\right\}.$$

Limiting contour process

For $\delta > 0$, let $\tilde{\xi}^{\mathrm{for},\delta,\mathbf{n}} :=$ reactant tree **cut off** in height $\tilde{T}^{\delta,n}$ and

$$A^\delta f(c) := 2\left(\frac{1}{X_c} f'\right)'(c)$$

on the domain

$$\mathcal{D}(A^\delta) := \left\{ f \in \mathcal{C}^2([0,\tau^\delta]) : \frac{1}{X_.} f' \in \mathcal{C}^1([0,\tau^\delta]), f'\big|_{\{0,\tau^\delta\}} = 0 \right\}.$$

Theorem The $(A^\delta, \mathcal{D}(A^\delta))$-martingale problem is well-posed. If ζ^δ is the solution of the $(A^\delta, \mathcal{D}(A^\delta))$-martingale problem, then

$$\left(\mathcal{C}(\tilde{\xi}^{\mathrm{for},\delta,n}); \tilde{\eta}^n\right) \underset{n\to\infty}{\Longrightarrow} \left(\zeta^\delta; X\right).$$

.... but what if $\delta \downarrow 0$?

Yet another useful representation

Genealogical point process (Popovic '04)

$\mathcal{P}(\mathbf{T}, \zeta) = \{\pi_\mathbf{t}\}_{\mathbf{t} \geq \mathbf{0}}$: collection of processes indexed by t; each giving the ancestry of the population alive at time t
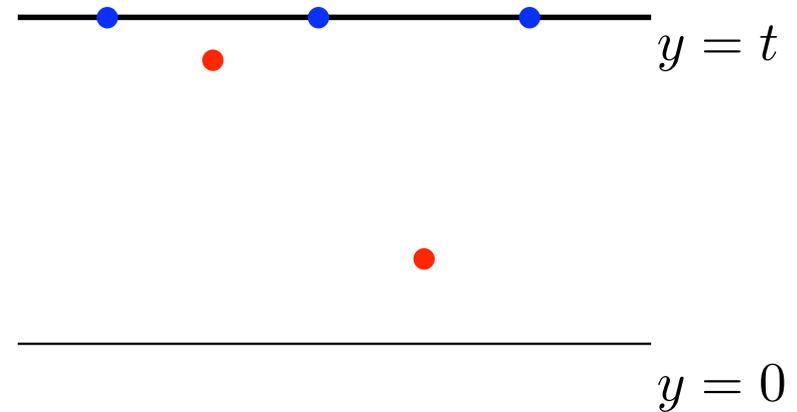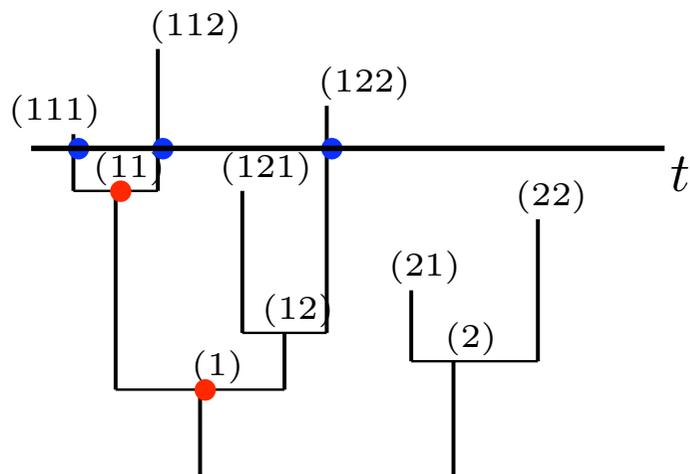
$\zeta :=$ spacing

Distribution of the point process

π_t are points of maximal depths of i.i.d. excursions below t
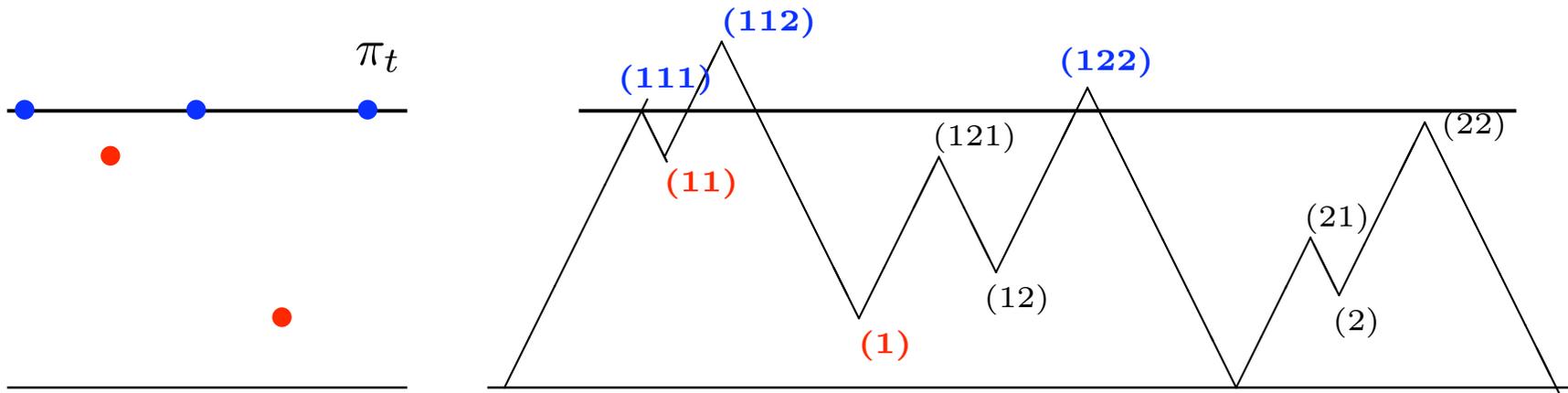
Lemma. For all $t > 0$, given a medium η, π_t is a simple point-process, $\{\tau_i\}_{i=0,\dots,(\xi_t-1)}$ are i.i.d. with

$$\mathbf{P}(\tau_{\mathbf{i}} \in d\mathbf{h}) = \frac{\eta_{\mathbf{h}}}{(1 + \int_{\mathbf{h}}^{\mathbf{t}} \eta_{\mathbf{s}} d\mathbf{s})^{\mathbf{2}}} \frac{1 + \int_{\mathbf{0}}^{\mathbf{t}} \eta_{\mathbf{s}} d\mathbf{s}}{\int_{\mathbf{0}}^{\mathbf{t}} \eta_{\mathbf{s}} d\mathbf{s}}, \quad \mathbf{h} \in (\mathbf{0}, \mathbf{t})$$

Rescaling the point process

Since after the rescaling of edge lengths we find at any height $t > 0$ a **Poisson number** of trees **each** having family size of order $\mathcal{O}(\mathbf{n})$, we choose in the nth-approximation step

$$\zeta := \frac{\mathbf{1}}{\mathbf{n}}.$$

Limit of the reactant genealogical point-process

Theorem. For any $t > 0$,

$$\left(\mathcal{P}^{\mathbf{t}}(\tilde{\xi}^{\mathrm{for},\mathbf{n}}; \frac{\mathbf{1}}{\mathbf{n}}); \tilde{\eta}^{\mathbf{n}}\right) \Rightarrow \left(\pi^{\zeta,\mathbf{t}}; \mathbf{X}\right),$$

where for given path of X the point process $\pi^{\zeta,t}$ consists of:

- Poisson point process on $\mathbb{R}^{+} \times \mathbb{R}^{+}$ whose intensity measure is

$$\aleph^{\zeta,\mathbf{t}}\left[\mathrm{d}\ell \times \mathrm{d}\mathbf{h}\right] = \mathbf{1}_{[\mathbf{0},\mathbf{Y_t}]}(\ell)\mathrm{d}\ell \otimes \mathbf{1}_{(\mathbf{0},\mathbf{t} \wedge \tau^{\mathbf{o}} \wedge \rho^{\mathbf{o}})}(\mathbf{h})\frac{\mathbf{X_h}}{(\int_{\mathbf{h}}^{\mathbf{t}} \mathbf{X_s}\mathrm{d}\mathbf{s})^{\mathbf{2}}}\mathrm{d}\mathbf{h}.$$

- rate $\left(\int_{\mathbf{0}}^{\mathbf{t}} \mathbf{X_s}\mathrm{d}\mathbf{s}\right)^{-1}$ Poisson point process at height 0 whose points separate distinct trees in the forest.

Remark. Result is δ-free!!!!

Properties of the Reactant Forest

Comparison with the constant rate branching forest:

$->$ differences in tree structure due to inhomogeneity of the random environment (evolving branching rates)

- stretching of the tree metric

$->$ behavior of the forest Y^{for} at time τ^0 of extinction of the random environment (in the event the catalyst dies first)

- infinite ℓ^2-length of the tree tips!

Differences in the reactant forest due to inhomogeneous rates

$$\partial Q_t(Y, d, \rho) := \big\{ u \in Y : \ d(\rho, u) = t \big\}.$$

Stretching Lemma. Let $(Z^{\mathrm{for}}, d_{Z^{\mathrm{for}}}, \rho)$ be the Brownian CRF and Y^{for} the reactant forest, and let $X : [0, \tau^0) \to \mathbb{R}_+$ be a given continuous function. For a fixed $t > 0$ define an increasing function $s_t^X : [0, t] \to [0, \int_0^t X_s \mathrm{d}s]$ by

$$s_t^X(h) := \int_{t-h}^t X_s \mathrm{d}s,$$

and let $(s_t^X)^{-1} : [0, \int_0^t X_s \mathrm{d}s] \to [0, t]$ be its inverse. Then

$$\Big((\partial Q_t(Y^{\mathrm{for}}), d_{Y^{\mathrm{for}}}, \rho); X \Big) \overset{d}{=} \big(\partial Q_{s_t^X(t)}(Z^{\mathrm{for}}), 2(s_t^X)^{-1}(\tfrac{1}{2} d_{Z^{\mathrm{for}}}), \rho \big)$$

Application of the stretching lemma

Theorem. [Comparing the probability to belong to different families]

Let Y^{for} be the reactant CRF and Z^{for} be Brownian CRF with the **same expected number of trees of a given height** $t > 0$. If $\mu^{t,Y}$ and $\mu^{t,Z}$ are the "uniform" distributions on $\partial Q_t(Y^{\text{for}})$ and $\partial Q_t(Z^{\text{for}})$, respectively, then

$$\mathbb{E}\left[\int_{(\partial Q_t(Y^{\text{for}}))^2} (\mu^{t,Y})^{\otimes 2}(\mathrm{d}u, \mathrm{d}u')\, \mathbf{1}\{d_{Y^{\text{for}}}(u, u') = 2t\}\right]$$

$$\leq \mathbb{E}\left[\int_{(\partial Q_t(Z^{\text{for}}))^2} (\mu^{t,Z})^{\otimes 2}(\mathrm{d}u, \mathrm{d}u')\, \mathbf{1}\{d_{Z^{\text{for}}}(u, u') = 2t\}\right];$$

Case. $\rho^0 < \tau^0$, i.e., reactant dies first

Proposition.
$$\lim_{\delta \downarrow 0} \langle \zeta^\delta, \zeta^\delta \rangle_{\ell^{-1}(\zeta^\delta)(1)} = \infty, \qquad \mathbb{P}\text{-a.s.}$$

Note. With positive probability there does **not exist** a limiting diffusion ζ^0 describing the contour of the full forest Y^{for} including its highest tips.

Many thanks