Coalescent Structure of Galton-Watson trees in varying environment

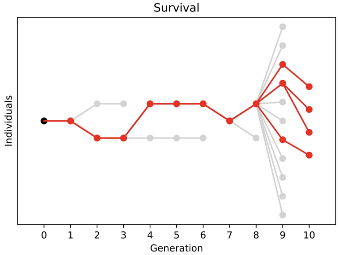
Sandra Palau (UNAM, México)

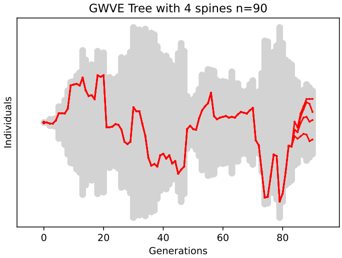
Joint work with: Simon Harris (University of Auckland, N.Z.)
and Juan Carlos Pardo (CIMAT, México)
Simulations: Enrique Moctezuma (UNAM, México)

Probability Seminar.
Mathematics Institute, Federal University of Rio de Janeiro.

Goal

Take a critical Galton-Watson process in a varying environment. If the system survives until time N, consider the genealogical tree of a sample of $k \geq 1$ particles chosen uniformly without replacement from those alive. How does their ancestral tree look like? What happens when $N \to \infty$?





Galton-Watson in varying environment

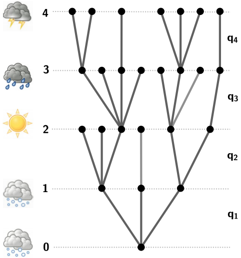
A *varying environment* $e = (q_1, q_2, \ldots)$ is a sequence of probability measures on $\mathbb{N}_0 = \{0, 1, 2, \cdots\}$. We define a *Galton-Watson process* $Z = \{Z_n, n \geq 0\}$ *in a varying environment e* as

$$Z_0 = 1 \quad \text{and} \quad Z_n := \sum_{i=1}^{Z_{n-1}} \chi_i^{(n)}, \qquad n \geq 1,$$

where $\{\chi_i^{(n)}; i, n \geq 1\}$ is a sequence of independent random variables

$$\mathbb{P}(\chi_i^{(n)} = k) = q_n(k), \qquad k \in \mathbb{N}_0.$$

$\chi_i^{(n)}$ is the offspring of the i-th particle in the $(n-1)$-th generation.



We denote by $(Z_n; \mathbb{P}^{(e)})$ the law of the process.

Galton-Watson in varying environment

Let f_i be the generating function associated with q_i. By applying the branching property recursively, we deduce that

$$\mathbb{E}^{(e)}\left[s^{Z_n}\right] = f_1 \circ \cdots \circ f_n(s), \qquad \text{for} \quad 0 \le s \le 1 \quad \text{and} \quad n \ge 1,$$

where $f \circ g$ denotes the composition of f with g. Let $f_{1,n}(s) := f_1 \circ \cdots \circ f_n(s)$. By differentiating, we obtain the mean and second factorial moment

$$\mathbb{E}^{(e)}[Z_n] = \mu_n, \quad \text{and} \quad \frac{\mathbb{E}^{(e)}[Z_n(Z_n - 1)]}{\mathbb{E}^{(e)}[Z_n]^2} = \sum_{k=0}^{n-1} \frac{\nu_{k+1}}{\mu_k} =: \rho_n, \quad n \ge 1,$$

where $\mu_0 := 1$ and for any $n \ge 1$,

$$\mu_n := f_1'(1) \cdots f_n'(1), \quad \text{and} \quad \nu_n := \frac{f_n''(1)}{f_n'(1)^2}.$$

Hypothesis in the environment

GWVEs can behave in an strange manner as possessing multiple rates of growth; see [MacPhee and Schuh 1983]. Kersting, (2020) showed that these exotic possibilities can be precluded by the following condition:

Condition (⋆)

For every $\epsilon > 0$, there is a finite constant c_ϵ such that for all $n \geq 1$

$$\mathbb{E}\left[\left(\chi_1^{(n)}\right)^2 \mathbf{1}_{\{\chi_1^{(n)} > c_\epsilon (1+\mathbb{E}[\chi_1^{(n)}])\}}\right] \leq \epsilon \mathbb{E}\left[\left(\chi_1^{(n)}\right)^2 \mathbf{1}_{\{\chi_1^{(n)} > 2\}}\right]. \qquad (\star)$$

We say that a GWVE is *regular* if it satisfies Condition (⋆).

It can be difficult to verify. A easier condition, which it is satisfied by most common probability distributions, is: There exists $c > 0$ such that

$$f_n'''(1) \leq c f_n''(1)(1 + f_n'(1)), \qquad \text{for any } n \geq 1.$$

Asymptotic behaviour

Kersting showed that under Condition ($\star$), the behaviour of a GWVE is dictated by the two sequences

$$\mu_n := \mathbb{E}^{(e)}[Z_n] \qquad \text{and} \qquad \rho_n := \frac{\mathbb{E}^{(e)}[Z_n(Z_n-1)]}{\mathbb{E}^{(e)}[Z_n]^2}, \qquad n \geq 1.$$

Specifically, $\lim_{n\to\infty} \frac{\rho_n}{2}\mathbb{P}^{(e)}[Z_n > 0] = 1$ and $\mathbb{E}^{(e)}[Z_n \mid Z_n > 0] \sim \frac{\mu_n \rho_n}{2}$ as $n \to \infty$.

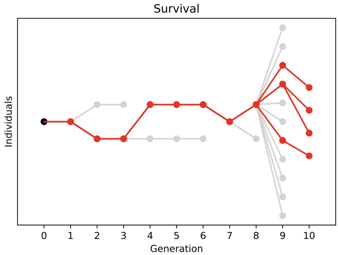
A regular GWVE is *critical* if and only if

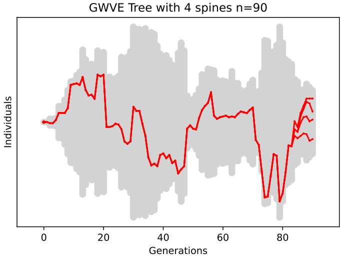
$$\lim_{n\to\infty} \rho_n = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \mu_n \rho_n = \infty.$$

In this case, $\lim_{n\to\infty} \mathbb{P}^{(e)}[Z_n > 0] = 0$ and Yaglom's limit exists: $\frac{2}{\mu_n \rho_n} Z_n$ conditioned on $\{Z_n > 0\}$ converges in distribution to a standard exponential random variable.

Goal

Take a critical Galton-Watson process in a varying environment. It has extinction a.s. Conditional on survival, take the genealogical tree of a sample of $k \geq 1$ particles chosen uniformly without replacement from those alive. How does it look like? What happens when $N \to \infty$?





Rooted trees: Ulam-Harris labeling

Let $\mathcal{U} := \{\varnothing\} \cup \bigcup_{n=1}^{\infty} \mathbb{N}^n$ be the set of finite sequences of positive integers.

- We define the *length* of u by $|u| = n$, if $u = (u_1, \cdots, u_n) \in \mathbb{N}^n$ and $|\varnothing| = 0$.
- The concatenation of $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_m)$ is denote by $uv := (u_1, \ldots, u_n, v_1, \ldots, v_m)$, with the convention that $u\varnothing = \varnothing u = u$.
- We say that v is an ancestor of u and write $v \preccurlyeq u$ if there exists $w \in \mathcal{U}$ such that $u = vw$.
- For $u \in \mathcal{U}$, we define the *genealogical line of u* as $[\![\varnothing, u]\!] := \{w \in \mathcal{U} : \varnothing \preccurlyeq w \preccurlyeq u\}$.

Example:
If $u = 1\,1\,2$, then $|u| = 3$ and $[\![\varnothing, u]\!] = \{\varnothing, 1, 1\,1, 1\,1\,2\}$.
If $v = 3\,2$, then $uv = 1\,1\,2\,3\,2$.

ROOTED TREES

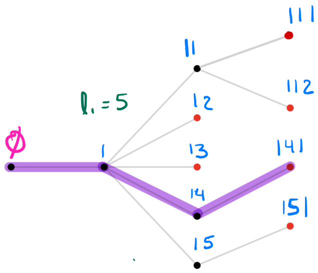
A *rooted tree* $\mathbf{t}$ is a subset of $\mathcal{U}$ that satisfies

- $\varnothing \in \mathbf{t}$.
- $[\![\varnothing, u]\!] \subset \mathbf{t}$ for any $u \in \mathbf{t}$.
- For every $u \in \mathbf{t}$ there exists a number $l_u(\mathbf{t})$ such that $uj \in \mathbf{t}$ if and only if $1 \le j \le l_u(\mathbf{t})$.

The empty string $\varnothing$ is called *the root*.

The integer $l_u(\mathbf{t})$ represents the number of offspring of $u \in \mathbf{t}$.

A *leaf* is a $u \in \mathbf{t}$ such that $l_u(\mathbf{t}) = 0$. Its genealogical line $\mathbf{u} := [\![\varnothing, u]\!]$ is called a *spine*.



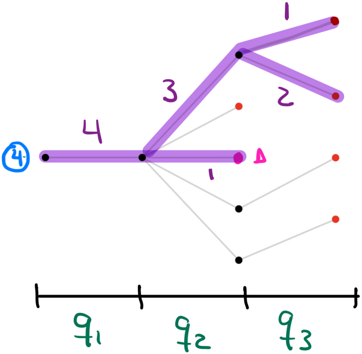
The *set of trees with k spines* is denote by

$$\mathcal{T}^k := \{(\mathbf{t}; \mathbf{v}_1, \dots, \mathbf{v}_k) : \ \mathbf{t} \text{ is a tree and } \mathbf{v}_i \text{ is a spine on } \mathbf{t} \text{ for all } i \le k\}.$$

Tree in varying environment with $k \geq 1$ spines

Construction:

- Start with one particle with k marks.
- Particles in generation m gives birth according to q_{m+1}.
- If a particle with j marks gives birth to $a > 0$ particles, then, each mark chooses the line to follow uniformly.
- If a particle with j marks give birth to 0 particles, then its marks are transferred to a graveyard.



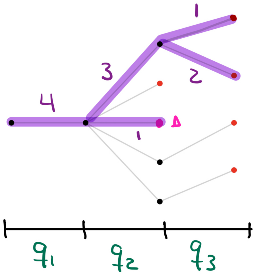
Tree in varying environment with $k \geq 1$ spines

A *Galton-Watson tree in environment e with $k \geq 1$ spines*, $(\mathbf{T}; \mathbf{V}_1, \ldots, \mathbf{V}_k)$ is a $\mathcal{T}^k$-valued r.v. with distribution

$$\mathbb{P}_n^{(e,k)}((\mathbf{t}; \mathbf{v}_1, \ldots, \mathbf{v}_k)) = \prod_{u \in \mathbf{t}:\, |u|<n} q_{|u|+1}(l_u(\mathbf{t})) \prod_{i=1}^{k} \prod_{u \in \mathbf{v}_i:\, |u|<|\mathbf{v}_i| \wedge n} \frac{1}{l_u(\mathbf{t})},$$

for any $n \geq 0$ and $(\mathbf{t}; \mathbf{v}_1, \ldots, \mathbf{v}_k) \in \mathcal{T}^k$.

Example:



$$\mathbb{P}_3^{(e,4)}((\mathbf{t}; \mathbf{v}_1, \ldots, \mathbf{v}_k))$$

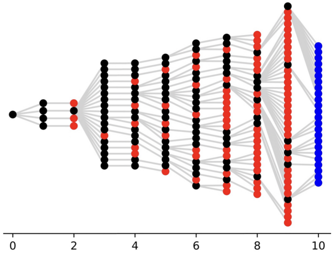
$$= q_1(1) q_2(5) q_3(2) q_3(0)^2 q_3(1)^2 \left(\frac{1}{5}\right)^4 \left(\frac{1}{2}\right)^3.$$

Change of measure.

Let $\theta \in [0,1)$ and define the function $g_{n,\theta} : \mathcal{T}^k \to \mathbb{R}$, as follows

$$g_{n,\theta}(\mathbf{t}; \mathbf{v}_1, \ldots, \mathbf{v}_k) := e^{-\theta X_n(\mathbf{t})} \mathbf{1}_{\{\mathbf{v}_i \neq \mathbf{v}_j, i \neq j\}} \mathbf{1}_{\{|\mathbf{v}_i| = n, i \leq k\}} \prod_{i=1}^{k} \prod_{u \in \mathbf{v}_i : |u| < n} l_u(\mathbf{t}).$$

$g_{n,\theta}(\mathbf{t}; \mathbf{v}_1, \ldots, \mathbf{v}_k)$ is non negative if the k spines are different, alive at time n and uniformly chosen.



Observe that

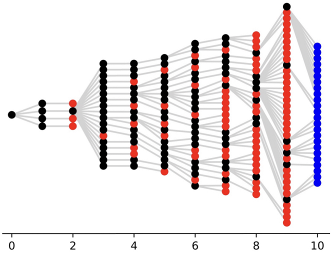
$$\mathbb{E}_n^{(e,k)}[g_{n,\theta}(\mathbf{T}; \mathbf{V}_1, \ldots, \mathbf{V}_k)] = \mathbb{E}^{(e)}[Z_n(Z_n - 1)\cdots(Z_n - k + 1)e^{-\theta Z_n}] < \infty.$$

Change of measure $\mathbf{Q}_n^{(e,k,\theta)}$

We define the change of measure

$$\mathbf{Q}_n^{(e,k,\theta)}((\mathbf{t};\mathbf{v}_1,\ldots,\mathbf{v}_k)) := \frac{g_{n,\theta}(\mathbf{t};\mathbf{v}_1,\ldots,\mathbf{v}_k)\mathbb{P}_n^{(e,k)}((\mathbf{t};\mathbf{v}_1,\ldots,\mathbf{v}_k))}{\mathbb{E}_n^{(e,k)}[g_{n,\theta}(\mathbf{T};\mathbf{V}_1,\ldots,\mathbf{V}_k)]},$$

for $(\mathbf{t};\mathbf{v}_1,\ldots,\mathbf{v}_k) \in \widehat{\mathcal{T}}_n^k$.



Spine splitting time

Denote by ψ_1 and $\hat{\psi}_1$ the last time where all spines are together and the first spine splitting time, respectively.

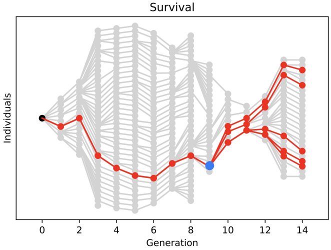
Let $0 \leq m < n$, then

$$\mathbf{Q}_n^{(e,k,\theta)}(\psi_1 \geq m)$$
$$= \frac{\mathbb{E}^{(e)}\left[e^{-\theta Z_n} Z_n\right]}{\mathbb{E}^{(e)}\left[e^{-\theta Z_n} Z_n^{[k]}\right]} \frac{\mathbb{E}^{(e_m)}\left[e^{-\theta Z_{n-m}} Z_{n-m}^{[k]}\right]}{\mathbb{E}^{(e_m)}\left[e^{-\theta Z_{n-m}} Z_{n-m}\right]},$$

where $x^{[k]}$ its k-th factorial, that is $x^{[k]} := x(x-1) \times \cdots \times (x-k+1)$.

We also have explicitly

$$F(m; k_1, \cdots k_g)$$
$$:= \mathbf{Q}_n^{(e,k,\theta)}\left(\text{At } \hat{\psi}_1 = m \text{ spines split into } g \text{ groups with } k_1, \ldots, k_g \text{ marks}\right).$$



$k = 5$, $\psi_1 = 9$, there are 3 groups with 1, 1, and 3 marks.

Construction of a GWTVE under $\mathbf{Q}_n^{(e,k,\theta)}$

- Start with one particle with k marks.
- If $k = 1$, we consider $\psi_1 = n$. Otherwise, select ψ_1, the number of spine groups g and their sizes $k_1, \ldots, k_g$ according to $F(m; k_1, \cdots k_g)$.
- An unmarked particle in gen. $m \in \{0, \ldots, n-1\}$ gives birth to unmarked particles with probability

$$q_{m+1}^{(0,\theta)}(\ell) = q_{m+1}(\ell) \frac{f_{m+1,n}(e^{-\theta})^{\ell}}{\mathbb{E}^{(e_m)}\left[e^{-\theta Z_{n-m}}\right]}.$$

Construction of a GWTVE under $\mathbf{Q}_n^{(e,k,\theta)}$

- Start with one particle with k marks.
- If $k = 1$, we consider $\psi_1 = n$. Otherwise, select ψ_1, the number of spine groups g and their sizes $k_1, \ldots, k_g$ according to $F(m; k_1, \cdots k_g)$.
- An unmarked particle in gen. $m \in \{0, \ldots, n-1\}$ gives birth to unmarked particles with probability

$$q_{m+1}^{(0,\theta)}(\ell) = q_{m+1}(\ell) \frac{f_{m+1,n}(e^{-\theta})^{\ell}}{\mathbb{E}^{(e_m)}\left[e^{-\theta Z_{n-m}}\right]}.$$

- A marked particle in gen. $m \in \{0, \ldots, \psi_1 - 1\}$ gives birth to particles accordingly to

$$q_{m+1}^{(1,\theta)}(\ell) = \frac{\ell q_{m+1}(\ell)\left(f_{m+1,n}(e^{-\theta})\right)^{\ell-1}}{f'_{m+1}(f_{m+1,n}(e^{-\theta}))}$$

Uniformly, select one of the particles to carry the k marks. All the other particles remain unmarked.

Construction of a GWTVE under $\mathbf{Q}_n^{(e,k,\theta)}$

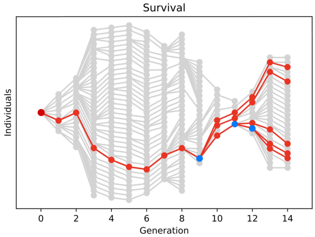
- Start with one p. with k marks.
- Select ψ_1, g and $k_1, \ldots, k_g$ according to $F(m; k_1, \cdots k_g)$.
- Unmarked particles in gen. $m \in \{0, \ldots, n-1\}$ gives birth to unmarked particles according to $q_{m+1}^{(0,\theta)}$.
- Marked particle in gen. $m \in \{0, \ldots, \psi_1 - 1\}$ gives birth as $q_{m+1}^{(1,\theta)}$. Select one to carry the k marks. All the other particles remain unmarked.
- The marked particle at generation $m = \psi_1$ gives birth accordingly to

$$q_{m+1}^{(g,\theta)}(\ell) = \frac{\ell^{[g]} q_{m+1}(\ell) (f_{m+1,n}(e^{-\theta}))^{l-g}}{\frac{\partial^g}{\partial s^g} f_{m+1}(s)\Big|_{s=f_{m+1,n}(e^{-\theta})}}$$

  Uniformly, select g of them as marked with $k_1, \ldots, k_g$ marks, respectively. All the other particles remain unmarked.
- Repeat steps for each of the g marked particles.

K-SAMPLE TREE

For each $i \leq k-1$, denote by ψ_i the last time where there are at most i marked particles.



$\psi_1 = 9 = \psi_2$, $\psi_3 = 11$ and $\psi_4 = 12$.

We know the joint distribution of the spines split times $(\psi_1, \ldots, \psi_{k-1})$ and the splitting groups under $\mathbf{Q}_n^{(e,k,\theta)}$. Now, for a CRITICAL GWVE, we want to know its asymptotic behaviour as $n \to \infty$.

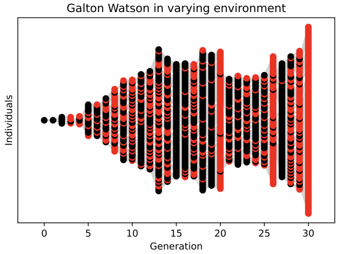
Time rescaling

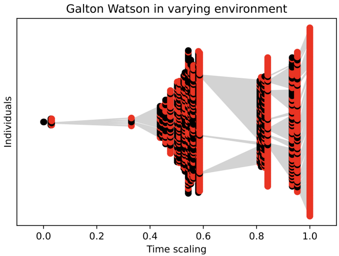
Recall that

$$\rho_0 = 0, \quad \text{and} \quad \rho_n = \frac{\mathbb{E}^{(e)}[Z_n(Z_n-1)]}{\mathbb{E}^{(e)}[Z_n]^2} = \sum_{k=0}^{n-1} \frac{\nu_{k+1}}{\mu_k}, \qquad n \geq 1.$$

Observe that $\{\rho_m/\rho_n; m \geq n\}$ can be thought as a cumulative probability distribution. We define its right-continuous generalised inverse as

$$\tau_n(t) = \max\{k \geq 0 : \rho_k \leq t\rho_n\}, \qquad t \in [0,1].$$

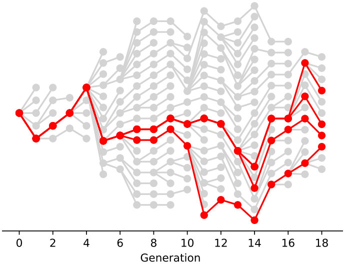


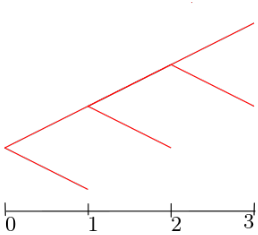


Binary trees

Let $\mathcal{B}^k$ be the set of full binary trees with k leaves, i.e. $\mathbf{b} \in \mathcal{B}^k$ if and only if $l_u(\mathbf{b}) \in \{0, 2\}$ for all $u \in \mathbf{b}$ and $k = |\{u \in \mathbf{b} : l_u(\mathbf{b}) = 0\}|$. We endow $\mathcal{B}^k$ with the $\mathbb{P}_{\mathcal{B}^k}$ probability measure of choosing uniformly a binary branching tree with k leaves.

Denote by $\mathcal{O} : \{(\mathbf{t}; \mathbf{v}_1, \ldots, \mathbf{v}_k) \in \mathcal{T}^k : \text{spine splittings are binary}\} \longrightarrow \mathcal{B}^k$, the operation that squeeze or stretch each line of the tree $\bigcup_{i=1}^{k} \mathbf{v}_i$ in a way that we obtain a tree in $\mathcal{B}^k$.





Asymptotic limit under $\mathbf{Q}_n^{(e,k,\theta_n)}$

Proposition (Harris, Palau, Pardo)

Consider a critical GWVE. Let $0 \le s_1 \le t_1 \le s_2 \le \cdots \le s_{k-1} \le t_{k-1} \le 1$ and $\mathbf{b} \in \mathcal{B}^k$. Then

$$\lim_{n\to\infty} \mathbf{Q}_n^{(e,k,\theta_n)}\Big(\psi_1 \in (\tau_n(s_1), \tau_n(t_1)), \ldots, \psi_{k-1} \in (\tau_n(s_{k-1}), \tau_n(t_{k-1})), \mathcal{O} = \mathbf{b}\Big)$$
$$= \mathbb{P}_{\mathcal{B}^k}(\mathbf{b})(k-1)!\,(1+\theta)^{k-1} \prod_{i=1}^{k-1} \Big(\tfrac{1-s_i}{1+\theta(1-s_i)} - \tfrac{1-t_i}{1+\theta(1-t_i)}\Big).$$

Observations:

- Asymptotically, all the splitting are binary.
- The tree topology $\{\mathcal{O} = \mathbf{t}, \mathbf{t} \in \mathcal{B}^k\}$ and the split times $\{\psi_i, i = 1, \ldots, k-1\}$ are asymptotically independent.

Asymptotic limit under $\mathbf{Q}_n^{(e,k,\theta_n)}$

Observations:

- $\{\psi_i, i \le k-1\}$ converges in distribution to an ordered sample of $k-1$ random variables. Let $\{\widetilde{\psi}_i, i \le k-1\}$ be a uniformly random permutation of $\{\psi_i, i \le k-1\}$.
- If we start with i groups of spines of sizes $a_1, \ldots, a_i$, in the limit, the split times for any group j will be distributed like $a_j - 1$ independent random variables, all with the same distribution. In particular, this implies that the first group to split will be group j with probability proportional to $a_j - 1$, that is, with probability $(a_j - 1)/(k-i)$.

Asymptotic behaviour under $\mathbb{P}^{(e)}$

Theorem (Harris, Palau, Pardo)

Consider a Critical GWVE and $\{t_1, \ldots, t_{k-1}\} \subset (0,1)$ with $t_i \neq t_j$. Then,

$$\lim_{n\to\infty} \mathbb{P}^{(e)}\left(\widetilde{\psi}_1 \geq \tau_n(t_1), \ldots, \widetilde{\psi}_{k-1} \geq \tau_n(t_{k-1}) \mid Z_n \geq k\right)$$
$$= k\left(\prod_{i=1}^{k-1} \frac{1-t_i}{t_i} - \sum_{j=1}^{k-1} \frac{1-t_j}{t_j} \prod_{i=1, i\neq j}^{k-1} \frac{t_i}{t_i - t_j} \log(1-t_j)\right).$$

The times $\{\widetilde{\psi}_i, i \leq k-1\}$ are asymptotically independent of the sample tree topology. The partition process $(P_1(n), \ldots, P_{k-1}(n))$ that describes the tree topology satisfies:

- If a block of size a splits, it creates 2 blocks whose sizes are ℓ and $a-\ell$ with probability converging to $1/(a-1)$, for $1 \leq \ell \leq a-1$.
- If $P_i(n)$ contains blocks of sizes $a_1, \ldots, a_{i+1}$, the probability that block j is the next to split converges to $(a_j - 1)/(k-i-1)$.

THANK YOU

- Boenkost, Foutel–Rodier, Schertzer. The genealogy of a nearly critical branching processes in varying environment. Arxiv: 2207. 11612.
- Harris, Johnston, and Roberts. The coalescent structure of continuous -time Galton–Watson trees. *The Annals of Applied Probability*, 30(3):1368–1414, 2020.
- Harris, Palau and Pardo. The coalescent structure of Galton-Watson trees in varying environments. ArXiv: 2207.10923v2.
- Kersting. A unifying approach to branching processes in a varying environment. *J. Appl. Probab.* 57(1), 196–220, 2020.
- Kersting. On the genealogical structure of critical branching processes in a varying environment. *Proceedings of the Steklov Institute of Mathematics*, 316(1), 209–219, 2022.
- Kersting and Vatutin. *Discrete time branching processes in random environment*. Wiley Online Library, 2017.