The coalescence problem in branching processes and its applications

Jyy-I Hong

National Sun Yat-sen University

July, 17, 2017

Outline

Outline

Galton-Watson Branching Processes

Consider a population starting with one individual (the 0th generation) and evolves over time.

Each individual, upon its death, produces the offspring according to the offspring distribution $\{P_j\}_{j\geq 0}$, where

$$P_j = \text{the probability of producing } j \text{ children.}$$

Assume that the reproduction is independent of the history of the population and of other individuals existing at the present.

We record the number of individuals in each generation

Galton-Watson Branching Processes

.
Let

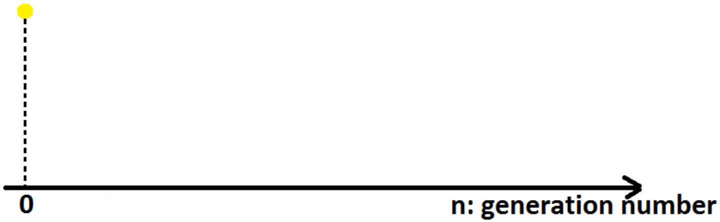
$$Z_n \quad = \quad \text{the population size of the } n\text{th generation.}$$

Then the processes

$$\{Z_n\}_{n\geq 0}$$

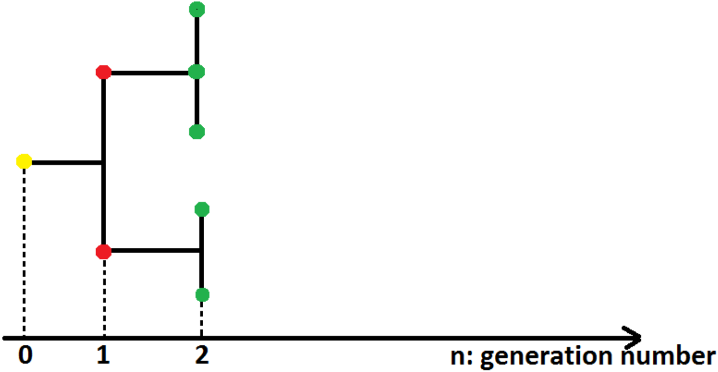
is called a discrete-time single-type Galton-Watson branching process.

Galton-Watson Branching Processes

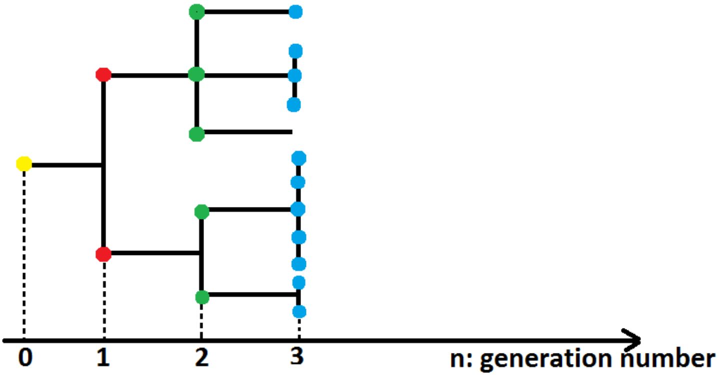


Galton-Watson Branching Processes

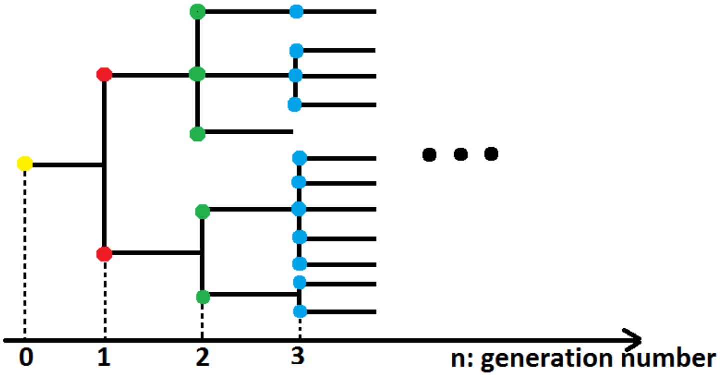
Galton-Watson Branching Processes



Galton-Watson Branching Processes



Galton-Watson Branching Processes



GWBP: Cases

Let $m = \sum_{j=0}^{\infty} jP_j$ be the offspring mean (expectation), then the process $\{Z_n\}_{n \geq 0}$ is said to be

(1) subcritical, if $0 < m < 1$

(2) critical, if $m = 1$;

(3) supercritical, if $1 < m < \infty$;

(4) explosive, if $m = \infty$

Aspects to investigate the population

Questions:

(1) What happens to the population size Z_n when n grows? Will the population extinct, explore, or become stabilized?

(2) If the population has a chance to die out, what is the probability of extinction?

(3) If the population has a chance to grow big, what is the growth rate?

Aspects to investigate the population

Questions:

(1) What happens to the population size Z_n when n grows? Will the population extinct, explore, or become stabilized?

(2) If the population has a chance to die out, what is the probability of extinction?

(3) If the population has a chance to grow big, what is the growth rate?

One may look at the history of the population:

(1) What happens to the certain characteristic of the family line of a randomly chosen individual?

(2) How "close" are any two randomly chosen individuals?

Outline

The Coalescence Problem

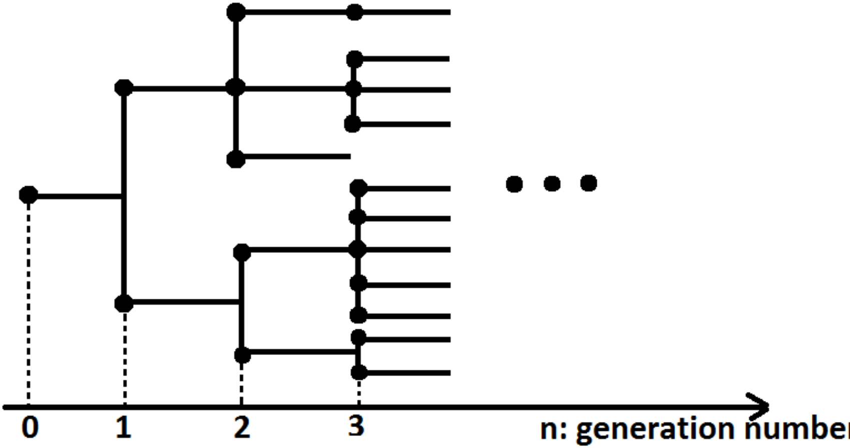
Consider a Galton-Watson branching process $\{Z_n : n \geq 0\}$ with $Z_0 = 1$.

Pick two individuals from the population in the nth generation by simple random sampling without replacement.

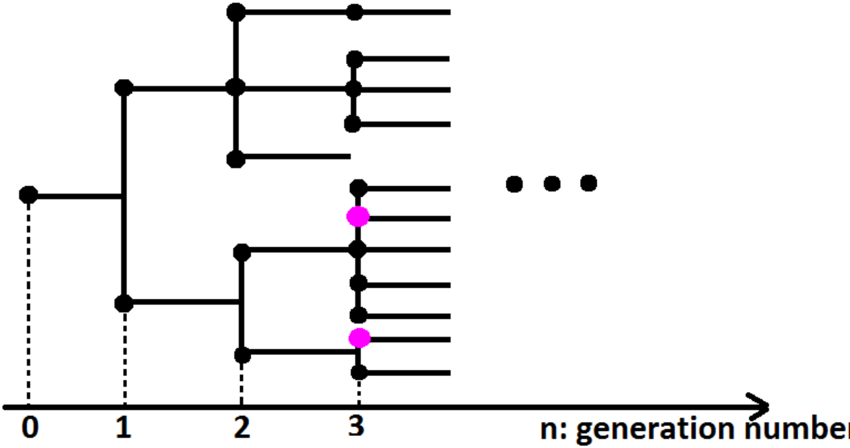
Trace their lines of descent backward in time till they meet.

We call the common ancestor in the generation where the lines meet for the first time the most recent common ancestor of these two chosen individuals.

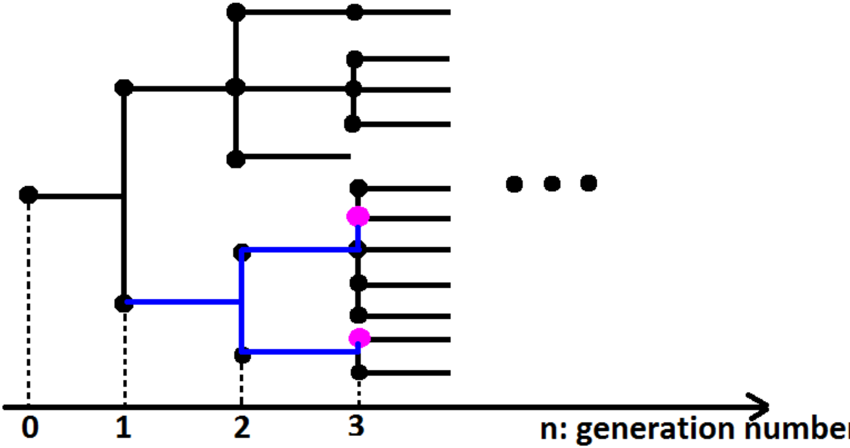
The Coalescence Problem



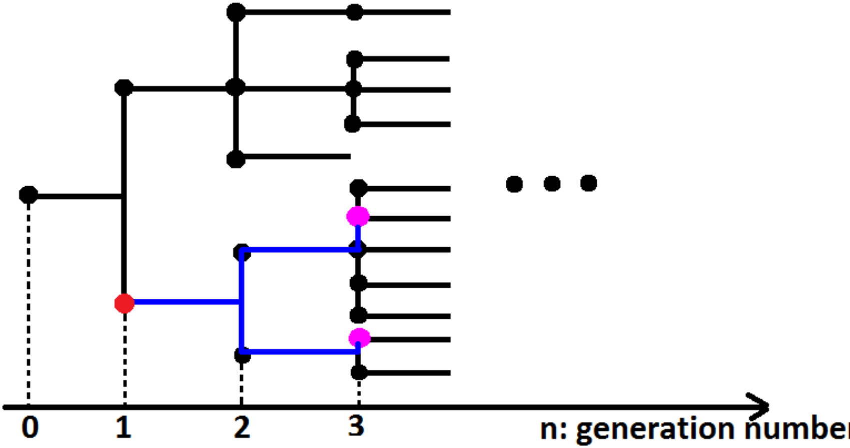
The Coalescence Problem



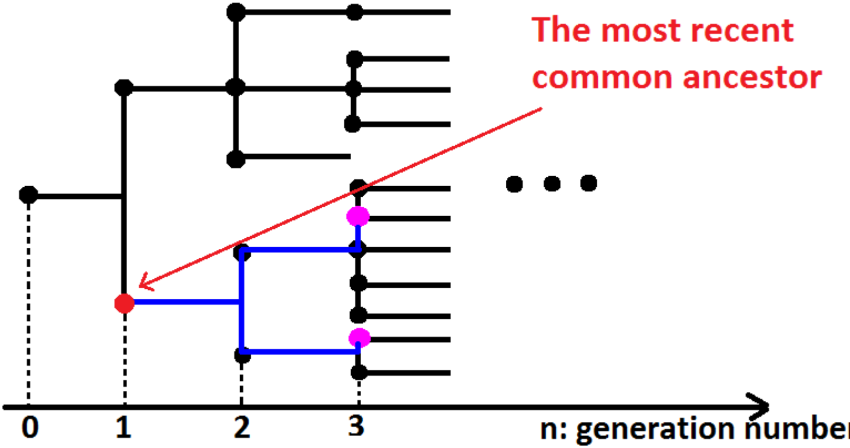
The Coalescence Problem



The Coalescence Problem



The Coalescence Problem



The Coalescence Problem

Let $\tau_{n,2}$ be the number of the generation which their most recent common ancestor belonged to.

When the population is big, we can expect that the most recent common ancestor would be found way back to the beginning of the family tree.

When the population is small, we may expect that the most recent common ancestor would not be too far away from the current generation.

The Coalescence Problem

Let $\tau_{n,2}$ be the number of the generation which their most recent common ancestor belonged to.

When the population is big, we can expect that the most recent common ancestor would be found way back to the beginning of the family tree.

When the population is small, we may expect that the most recent common ancestor would not be too far away from the current generation.

Question: How does $\tau_{n,2}$ behave as $n \to \infty$?

Results of the coalescence problem

Athreya (2012) has the following results:

(1) The supercritical case ($1 < m < \infty$):

$$\tau_{n,2} \big| Z_n \geq 2 \xrightarrow{d} \text{a proper distribution on } \{0, 1, 2, \cdots\}$$

$\Rightarrow$ Coalescence is **near the beginning of the tree**

Results of the coalescence problem

Athreya (2012) has the following results:

(1) The supercritical case $(1 < m < \infty)$:

$$\tau_{n,2} \Big| Z_n \geq 2 \xrightarrow{d} \text{a proper distribution on } \{0, 1, 2, \cdots\}$$

$\Rightarrow$ Coalescence is **near the beginning of the tree**

(2) The critical case $(m = 1)$:

$$\frac{\tau_{n,2}}{n} \Bigg| Z_n \geq 2 \xrightarrow{d} \text{a proper distribution on } (0, 1)$$

$\Rightarrow$ Coalescence is **of order** n

Summary of the coalescence problem

(3) The subcritical case ($0 < m < 1$):

$n - \tau_{n,2} \big| Z_n \geq 2 \xrightarrow{d}$ a proper distribution on $\{0, 1, \cdots\}$

$\Rightarrow$ Coalescence is **near the present**

Summary of the coalescence problem

(3) The subcritical case ($0 < m < 1$):

$n - \tau_{n,2} \big| Z_n \geq 2 \xrightarrow{d}$ a proper distribution on $\{0, 1, \cdots\}$

$\Rightarrow$ Coalescence is **near the present**

(4) The explosive case ($m = \infty$):

$n - \tau_{n,2} \big| Z_n \geq 2 \xrightarrow{d}$ a proper distribution on $\{0, 1, \cdots\}$

$\Rightarrow$ Coalescence is **near the present**

Outline

1. Galton-Watson Branching Processes
   - Galton-Watson Model
   - The Coalescence Problem
   - Application to Branching Random Walks
2. The Coalescence Problem in Other Branching Processes
   - Coalescence in Multitype Galton-Watson Processes
   - The Coalescence in Bellman-Harris Processes
3. Open Questions

Branching Random Walks

We impose the following movement structure to the family tree of the process $\{Z_n\}_{n\geq 0}$.

If an individual is located at $x \in \mathbb{R}$ and, upon death, produces k children, then these k children will move to

$$x + X_{k1}, \quad x + X_{k2}, \quad \cdots, \quad x + X_{kk}$$

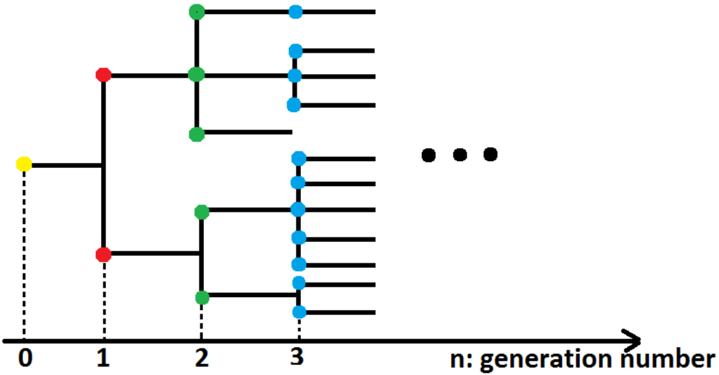
where $X_k \equiv (X_{k1}, X_{k2}, \cdots, X_{kk})$ is a random vector with a joint distribution π_k on $\mathbb{R}^k$ for each k.

Assume that the random vector X_k is stochastically independent of the history up to that generation and the movement of the offspring of other individuals.

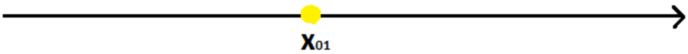
Let x_{ni} be the position of the ith individual in the nth generation.

Galton-Watson Branching Processes

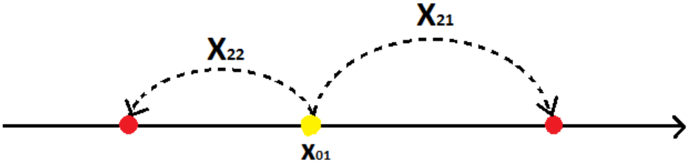
Recall a family tree in the Galton-Watson branching process:



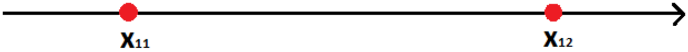
Branching Random Walks



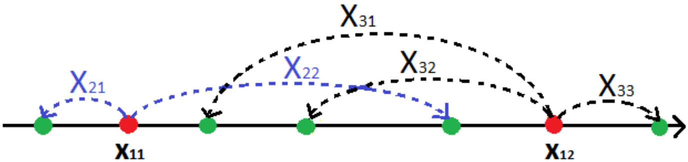
Branching Random Walks



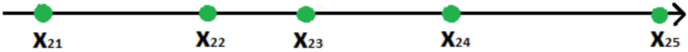
Branching Random Walks



Branching Random Walks



Branching Random Walks



Branching Random Walks

Let $\zeta_n = \{x_{ni} : 1 \leq i \leq Z_n\}$ be the position of the Z_n individuals in the nth generation. Then the sequence of pairs

$$\{Z_n, \zeta_n\}_{n \geq 0}$$

is called a branching random walk.

Branching Random Walks

Let $\zeta_n = \{x_{ni} : 1 \leq i \leq Z_n\}$ be the position of the Z_n individuals in the nth generation. Then the sequence of pairs

$$\{Z_n, \zeta_n\}_{n \geq 0}$$

is called a branching random walk.

Question: Where are individuals?

(1) Let Y_n be the position of a random chosen individual in the nth generation. How does Y_n behave as $n \to \infty$?

Branching Random Walks

Let $\zeta_n = \{x_{ni} : 1 \leq i \leq Z_n\}$ be the position of the Z_n individuals in the nth generation. Then the sequence of pairs

$$\{Z_n, \zeta_n\}_{n \geq 0}$$

is called a branching random walk.

Question: Where are individuals?

(1) Let Y_n be the position of a random chosen individual in the nth generation. How does Y_n behave as $n \to \infty$?

(2) Let $Z_n(x)$ be the number of individuals in ζ_n whose locations are less than or equal to x. What happens to the proportion $\dfrac{Z_n(x)}{Z_n}$ as $n \to \infty$?

Branching Random Walks

Let $\zeta_n = \{x_{ni} : 1 \leq i \leq Z_n\}$ be the position of the Z_n individuals in the nth generation. Then the sequence of pairs

$$\{Z_n, \zeta_n\}_{n \geq 0}$$

is called a branching random walk.

Question: Where are individuals?

(1) Let Y_n be the position of a random chosen individual in the nth generation. How does Y_n behave as $n \to \infty$?

(2) Let $Z_n(x)$ be the number of individuals in ζ_n whose locations are less than or equal to x. What happens to the proportion $\dfrac{Z_n(x)}{Z_n}$ as $n \to \infty$? How can we choose x such that the proportion will converge to a proper distribution?

(1) Let Y_n be the position of a random chosen individual in the nth generation. How does Y_n behave as $n \to \infty$?

It is clear that the movement along the line of descent of any individual is that of a classical random walk.

In addition, the location Y_n of any individual in the nth generation is the sum of the location x_{01} of the first ancestor and the movements the ancestors along the line of decent.

So, if X_{ki} are identically distributed with mean μ and finite variance σ^2, then, by the central limit theorem, as $n \to \infty$,

$$\frac{Y_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} \text{standard normal distribution}$$

i.e., the location Y_n of an individual of the nth generation should be approximately Normal with mean $n\mu$ and variance $n\sigma^2$ when n is large.

(2) Let $Z_n(x)$ be the number of individuals in ζ_n whose locations are less than or equal to x. What happens to the proportion $\frac{Z_n(x)}{Z_n}$ as $n \to \infty$? How can we choose x such that the proportion will converge to a proper distribution?

First,

$$\begin{aligned} P(Y_n \le n\mu + \sqrt{n}\sigma y) &= E(P(Y_n \le n\mu + \sqrt{n}\sigma y | Z_n)) \\ &= E\left(\frac{Z_n(n\mu + \sqrt{n}\sigma y)}{Z_n}\right) \end{aligned}$$

This suggests that if $Z_n \to \infty$ as $n \to \infty$, and if $x_n = \sqrt{n}\sigma y + n\mu$, then

$$\frac{Z_n(x_n)}{Z_n}$$

may be close to the standard normal cumulative distribution function $\Phi(y)$, where $\Phi(y) = P(N \le y)$ and N is a standard normal random variable.

Main Results for the supercritical case

Athreya and Hong (2010, 2013) have the following results:

Theorem 1.3.1.

Let $p_0 = 0$ and $1 < m < \infty$.
Let π_k be such that $\{X_{ki} : i = 1, 2, \cdots, k\}$ are identically distributed with $EX_{k1} = 0$ and $EX_{k1}^2 = \sigma^2 < \infty$.
Then, for any $y \in \mathbb{R}$,

(a) $P(Y_n \leq \sqrt{n}\sigma y) \to \Phi(y)$, *as* $n \to \infty$;

(b) $\frac{Z_n(\sqrt{n}\sigma y)}{Z_n} \to \Phi(y)$ *in mean square*, *as* $n \to \infty$.

Main Results for the explosive case

Surprisingly, the result for the explosive case does not support our conjecture:

Theorem 1.3.2.

Let $p_0 = 0$, $m = \infty$ *and* $\{p_j\}_{j\geq 0} \in D(\alpha)$, $0 < \alpha < 1$.
Let π_k *be such that* $\{X_{ki} : i = 1, 2, \cdots, k\}$ *are identically distributed with* $EX_{k1} = 0$ *and* $EX_{k1}^2 = \sigma^2 < \infty$.
Then, for any $y \in \mathbb{R}$,

(a) $P(Y_n \leq \sqrt{n}\sigma y) \to \Phi(y)$, *as* $n \to \infty$;

(b) $\dfrac{Z_n(\sqrt{n}\sigma y)}{Z_n} \to \delta_y$ *in distribution*, *as* $n \to \infty$, *where* δ_y *has the Bernoulli distribution with parameter* $\Phi(y)$, *i.e.*,

$$P(\delta_y = 1) = \Phi(y) = 1 - P(\delta_y = 0).$$

Application of the Coalescence Problem

To prove (b) for the supercritical case, we need to show that

$$E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n} - \Phi(y)\right)^2 \to 0, \text{ as } n \to \infty.$$

First,

$$E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n} - \Phi(y)\right)^2$$

$$= E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^2 - 2\Phi(y)\cdot E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right) + \Phi(y)^2$$

we only need to prove that

$$E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^2 \to \Phi(y)^2, \text{ as } n \to \infty.$$

Application of the Coalescence Problem

For the explosive case, since, by (a), as $n \to \infty$,

$$E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right) = P(Y_n \leq \sqrt{n}\sigma y) \to \Phi(y),$$

so, if we can show that

$$E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^2 \to \Phi(y), \text{ as } n \to \infty,$$

then, since the random variable $\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}$ takes values in $[0, 1]$ and the limits of its first and second moments both equal to $\Phi(y)$, its limit distribution is a Bernoulli distribution with parameter $\Phi(y)$.

One can derive the following:

$$\lim_{n\to\infty} E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^2$$

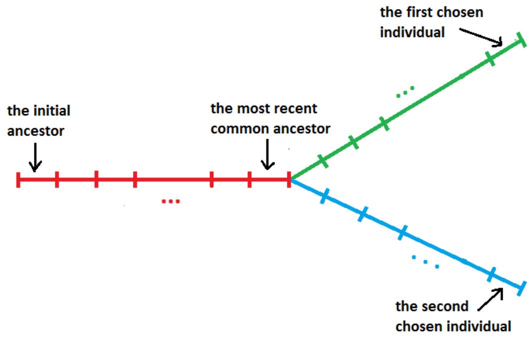
$$= \lim_{n\to\infty} P(x_{n1} \leq \sqrt{n}\sigma y, x_{n2} \leq \sqrt{n}\sigma y)$$

where $x_{n,i}$ is the position of $i$the individual in the $n$the generation.

Note that x_{ni} is the sum of the position of the ancestor in the 0th generation and the movements of the rest of its ancestors along its line of descent.

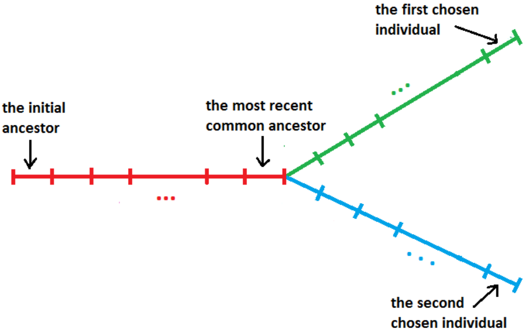
Application of the Coalescence Problem

When we look at the lines of descent of these two individuals in the nth generation, the random walks along the lines are the same up to their most recent common ancestor and become independent after the movement of this common ancestor.



Application of the Coalescence Problem

When we look at the lines of descent of these two individuals in the nth generation, the random walks along the lines are the same up to their most recent common ancestor and become independent after the movement of this common ancestor.



So, the coalescence time when these two lines of descent met in the past is crucial.

Conclusion

Therefore, we have shown that

(1) In supercritical case,

$$\lim_{n\to\infty} E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^2$$

$$= \lim_{n\to\infty} P(x_{n1} \le \sqrt{n}\sigma y, x_{n2} \le \sqrt{n}\sigma y) = \Phi(y)^2$$

(2) In explosive case,

$$\lim_{n\to\infty} E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^2$$

$$= \lim_{n\to\infty} P(x_{n1} \le \sqrt{n}\sigma y, x_{n2} \le \sqrt{n}\sigma y) = \Phi(y)$$

and hence the theorems are proved.

Outline

Multitype Galton-Watson Branching Processes

Consider a population consists individuals of various types, say d types in total. Assume that this population starts with one individual and evolves in time.

Each individual lives a unit of time and, upon its death, produces children of all types according to the offspring distribution

$$\left\{ P^{(i)}(j_1, j_2, \cdots, j_d) : j_1, j_2, \cdots, j_d \in \mathbb{N}^d \right\}_{1 \le i \le d}$$

independently of other individuals in the population where $\mathbb{N}$ is the set of all the nonnegative integers and

Multitype Galton-Watson Branching Processes

$$
\begin{aligned}
& P^{(i)}(j_1, j_2, \cdots, j_d) \\
= \ & \text{the probability that an individual of type } i \text{ produces} \\
& \ j_1 \text{ children of type } 1,\ j_2 \text{ children of type } 2, \\
& \ \cdots,\ j_d \text{ children of type } d.
\end{aligned}
$$

Multitype Galton-Watson Branching Processes

Let $Z_{n,i}$ be the number of the individuals of type i in the nth generation, $i = 1, 2, \cdots, d$. This vector

$$\mathbf{Z}_n = (Z_{n,1}, Z_{n,2}, \cdots, Z_{n,d})$$

shows the type composition of the population in the nth generation. Then, the process $\{\mathbf{Z}_n\}_{n\geq 0}$ is called a multitype (d-type) Galton-Watson branching process.

Let $m_{ij} = E\left(Z_{1,j}\middle|\mathbf{Z}_0 = \mathbf{e}_i\right)$ be the expected number of children of type j produced by a single individual of type i. Then

$$\mathbf{M} = \left\{m_{ij} : i, j = 1, 2, \cdots, d\right\}$$

is called the offspring mean matrix.

Multitype Galton-Watson Branching Processes

We impose the following conditions on the process $\{\mathbf{Z}_n\}_{n\geq 0}$:

1. Non-singularity, i.e. for every $i = 1, 2, \cdots, d$,
$$P\left(\mathbf{Z}_1 = \mathbf{e}_i \middle| \mathbf{Z}_0 = \mathbf{e}_i\right) < 1.$$
2. Positive regularity, i.e. there exists an n such that for all $i, j = 1, 2, \cdots, d$,
$$m_{ij}^{(n)} > 0$$
where $m_{ij}^{(n)}$ is the entry in $\mathbf{M}^n$.

Multitype Galton-Watson Branching Processes

By the Perron-Frobenius theorem, the matrix $\mathbf{M}$ has a maximal eigenvalue ρ which is positive, simple and has associated strictly positive normalized right and left eigenvectors $\mathbf{u}$ and $\mathbf{v}$ such that

$$\mathbf{u} \cdot \mathbf{1} = 1 \quad \text{and} \quad \mathbf{u} \cdot \mathbf{v} = 1.$$

We say that the branching process $\{\mathbf{Z}_n\}_{n \geq 0}$ is

(a) subcritical, if $\rho < 1$;

(b) critical, if $\rho = 1$;

(c) supercritical, if $1 < \rho < \infty$.

The Coalescence Problem

Let $X_{n,k}$ be the number of the generation which the most recent common ancestor of these k chosen individuals belonged to.

Question 1 : How does $X_{n,k}$ behave as $n \to \infty$?

Let T_n be the generation number of the most recent common ancestor of the whole population in the nth generation. We call T_n the total coalescence time.

Question 2 : How does T_n behave as $n \to \infty$?

The Type Problem

Moreover, in a multitype process, let $\zeta_{n,i}$ be the type of the ith chosen individual from the nth generation.

Let η_n be the type of the most recent common ancestor of the chosen individuals.

Question 3 : What happens to the joint distribution of

$$(X_{n,k}, \eta_n, \zeta_{n,1}, \cdots, \zeta_{n,k}),$$

as $n \to \infty$?

Main Results in the Supercritical Case

Theorem 2.1.1 (Hong, 2015).

Let $\rho > 1$, $\mathbf{Z}_0 = \mathbf{e}_{i_0}$ *and* $E\big(\|Z_1\| \log \|Z_1\| \big| \mathbf{Z}_0 = \mathbf{e}_i\big) < \infty$ *for all* $1 \le i \le d$. *Then*

(a) *(Quenched version) for almost all trees* $\mathcal{T}$ *and* $r = 1, 2, \cdots$, *there exists positive real-valued random variables* $W_{r,i}^{(l)}$, $i = 1, 2, \cdots, Z_r^{(l)}$, $l = 1, 2, \cdots, d$ *such that*

$$P(X_{n,k} < r|\mathcal{T}) \to \phi_k(r, \mathcal{T}) \equiv 1 - \frac{\sum_{l=1}^{d} \sum_{i=1}^{Z_r^{(l)}} (W_{r,i}^{(l)})^k}{\left(\sum_{l=1}^{d} \sum_{i=1}^{Z_r^{(l)}} W_{r,i}^{(l)}\right)^k}$$

as $n \to \infty$.

Main Results in the Supercritical Case

Theorem 2.1.1 (cont.).

(b) *(Annealed version) there exist random variable* $\tilde{X}_k$ *such that* $X_{n,k} \xrightarrow{d} \tilde{X}_k$ *as* $n \to \infty$*, where*

$$P(\tilde{X}_k < r) \equiv \phi_k(r) = 1 - E\left(\frac{\sum_{l=1}^{d} \sum_{i=1}^{Z_r^{(l)}} (W_{r,i}^{(l)})^k}{\left(\sum_{l=1}^{d} \sum_{i=1}^{Z_r^{(l)}} W_{r,i}^{(l)} \right)^k} \right)$$

for any $r = 1, 2, \cdots$. *Further,* $\lim_{r \to \infty} \phi_k(r) = 1$ *so that* $\tilde{X}_k$ *is a proper random variable.*

Main Results in the Supercritical Case

The next theorem tells us the existence of the limit distribution of $\tilde{X}_k$ as $k \to \infty$ and says that the coalescence time goes away back to one generation before the first time when this process began to split when more individuals are chosen.

Theorem 2.1.2 (Hong, 2015).

Let $\rho > 1$ and $E\|Z_1\| \log \|Z_1\| < \infty$.
Let $U = \min\left\{n \geq 1 : |\mathbf{Z}_n| \geq 2\right\}$ be the first time when the population exceeds 1. Then

$$\tilde{X}_k \xrightarrow{d} U - 1$$

as $k \to \infty$.

Main Results in the Supercritical Case

The next theorem is the result on the joint distribution of $(X_{n,2}, \eta_n, \zeta_{n,1}, \zeta_{n,2})$:

Theorem 2.1.3 (Hong, 2015).

Let $\rho > 1$, $\mathbf{Z}_0 = \mathbf{e}_{i_0}$, *and* $E\Big(\|Z_1\| \log \|Z_1\| \Big| \mathbf{Z}_0 = \mathbf{e}_i\Big) < \infty$ *for all* $1 \leq i \leq d$. *Then*

$$\lim_{n\to\infty} P(X_{n,2} = r, \eta_n = j, \zeta_{n,1} = i_1, \zeta_{n,2} = i_2) \equiv \varphi_2(r, j, i_1, i_2)$$

exists and $\displaystyle\sum_{(r,j,i_1,i_2)} \varphi_2(r, j, i_1, i_2) = 1.$

Main Results in the Critical Case

Theorem 2.1.4 (Hong, 2016).

Let $\rho = 1$ *and* $E\|\mathbf{Z}_1\|^2 < \infty$. *Then, for* $k = 2, 3, \cdots$, *there exists a random variable* $\tilde{X}_k$ *such that*

$$\frac{X_{n,k}}{n}\Big| |\mathbf{Z}_n| \geq k \xrightarrow{d} \tilde{X}_k$$

as $n \to \infty$ *and, for any* $\alpha \in (0, 1)$,

$$P(\tilde{X}_k < \alpha) = 1 - (1 - \alpha)F(1, 2; k + 1; \alpha) \equiv H_k(\alpha)$$

where F *is a hypergeometric function. Further,* $\lim_{\alpha \uparrow 1} H_k(\alpha) = 1$.

Main Results in the Critical Case

This the result on the total coalescence time:

Theorem 2.1.5 (Hong, 2016).

Let $\rho = 1$ *and* $E\|\mathbf{Z}_1\|^2 < \infty$. *Then*

$$\frac{T_n}{n}\Big|\,|\mathbf{Z}_n| > 0 \xrightarrow{d} \tilde{T}$$

as $n \to \infty$, *where* $\tilde{T}$ *has a uniform distribution in* $(0, 1)$.

Main Results in the Subcritical Case

The following result in the subcritical case can be extended to the case when k individuals are chosen from the nth generation.

Theorem 2.1.6 (Hong, 2016).

Let $0 < \rho < 1$ and $E\|Z_1\| \log \|Z_1\| < \infty$. Then there exists a random variable $\tilde{X}_2$ such that

$$n - X_{n,2}\Big||Z_n| \geq 2 \xrightarrow{d} \tilde{X}_2$$

as $n \to \infty$, and, for any $r = 0, 1, 2, \cdots$,

$$\begin{aligned} P(\tilde{X}_2 \leq r) &= 1 - \frac{1}{\rho^r P\big(|\mathbf{Y}| \geq 2\big)} E\Big(\phi\big(Y^{(1)}, Y^{(2)}, \cdots, Y^{(d)}, r\big)\Big) \\ &\equiv H_2(r) \end{aligned}$$

Main Results in the Subcritical Case

Theorem 2.1.6 (cont.).

where

$$\phi(t_1, t_2, \cdots, t_d, r) = E\left(\frac{\sum_{l=1}^{d} \sum_{i \neq j=1}^{t_l} |\tilde{\mathbf{Z}}_{r,i}^{(l)}||\tilde{\mathbf{Z}}_{r,j}^{(l)}| + \sum_{l \neq p=1}^{d} \sum_{i=1}^{t_l} \sum_{j=1}^{t_p} |\tilde{\mathbf{Z}}_{r,i}^{(l)}||\tilde{\mathbf{Z}}_{r,j}^{(p)}|}{\left(\sum_{l=1}^{d} \sum_{i \neq j=1}^{t_l} |\tilde{\mathbf{Z}}_{r,i}^{(l)}|\right)\left(\sum_{l=1}^{d} \sum_{i \neq j=1}^{t_l} |\tilde{\mathbf{Z}}_{r,i}^{(l)}| - 1\right)} I_{\left\{\sum_{l=1}^{d} \sum_{i \neq j=1}^{t_l} |\tilde{\mathbf{Z}}_{r,i}^{(l)}| \geq 2\right\}} \right)$$

and $\left\{ \tilde{Z}_{r,i}^{(l)} : i \geq 1 \right\}_{r \geq 0}$ *are i.i.d copies of* $\{\mathbf{Z}_r^{(l)}\}_{r \geq 0}$. *Further,* $\lim_{r \to \infty} H_2(r) = 1$.

Main Results in the Subcritical Case

Theorem 2.1.7 (Hong, 2016).

Let $0 < \rho < 1$ *and* $E\|Z_1\| \log \|Z_1\| < \infty$. *Then there exists a random variable* $\tilde{T}$ *such that* $n - T_n \Big| |\mathbf{Z}_n| > 0 \xrightarrow{d} \tilde{T}$ *as* $n \to \infty$, *and, for any* $r = 0, 1, 2, \cdots$,

$$\begin{aligned} & P(\tilde{T} \le r) \\ = \;& \rho^{-r} E\left(\sum_{l=1}^{d} Y^{(l)} \left(1 - f_r^{(l)}(\mathbf{0})\right) \left(f_r^{(l)}(\mathbf{0})\right)^{Y^{(l)}-1} \prod_{p \neq l} \left(f_r^{(l)}(\mathbf{0})\right)^{Y^{(p)}} \right) \\ \equiv \;& \pi(r) \end{aligned}$$

where $\mathbf{Y}$ *is the random vector with distribution* $\{b(\mathbf{j})\}_{\mathbf{j} \in \mathbb{R}_+^d}$ *defined as in Theorem* **??** *(a). Also,* $\lim_{r \uparrow \infty} \pi(r) = 1$, *i.e.,* $\tilde{T}$ *is a proper random variable.*

Main Results in the Subcritical Case

Theorem 2.1.8 (Hong, 2016).

Let $0 < \rho < 1$ *and* $E\|Z_1\| \log \|Z_1\| < \infty$. *Then*

$$\lim_{n\to\infty} P\left(X_{n,2} = r, \eta_n = j, \zeta_{n,1} = i_1, \zeta_{n,2} = i_2 \middle| |\mathbf{Z}_n| \geq 2\right) \equiv \psi_2(r, j, i_1, i_2)$$

exists and $\sum_{(r,j,i_1,i_2)} \psi_2(r, j, i_1, i_2) = 1$.

Outline

Bellman-Harris Branching Processes

We consider a population starting with an individual

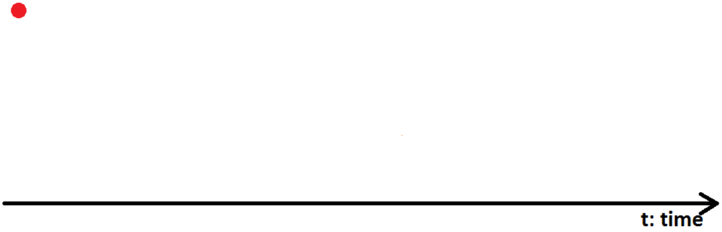
Each individual in this population lives for a random amount of time, L, with the life distribution function G.

Upon the death, each individual produces a random number ξ of children according the offspring distribution $\{p_j\}_{j\geq 0}$ where

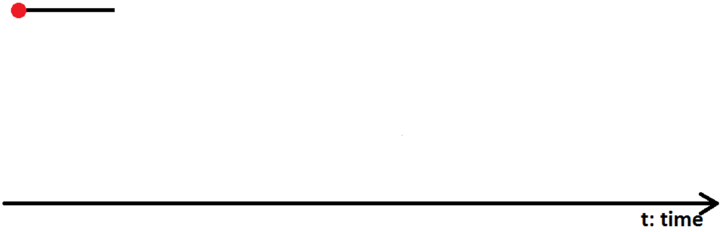
$$p_j = \text{the probability of producing } j \text{ children.}$$

We assume that the life time and the reproduction of each individual are independent and are also independent of those of other individuals.

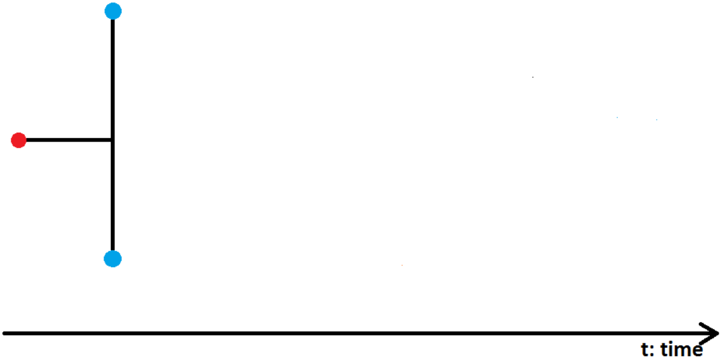
Bellman-Harris Branching Processes



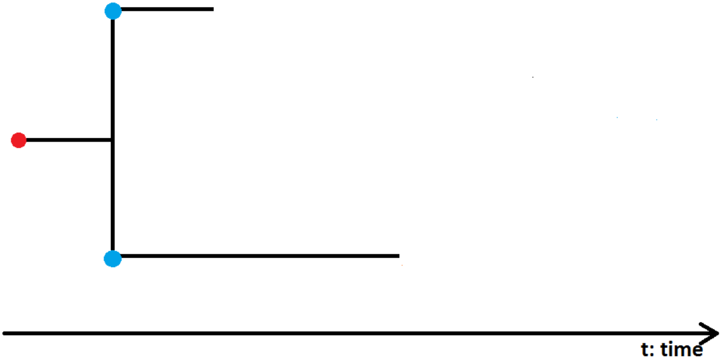
Bellman-Harris Branching Processes



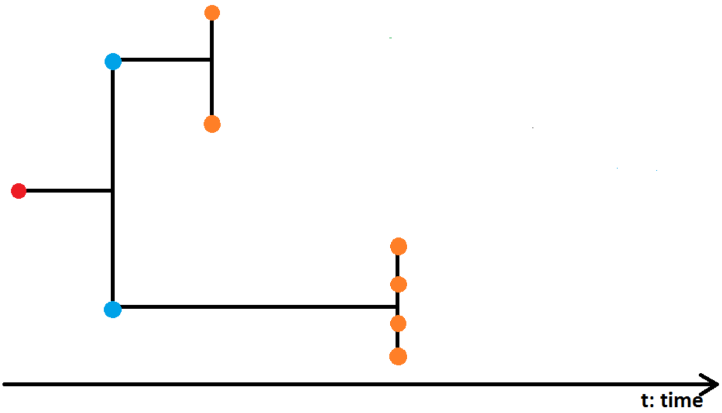
Bellman-Harris Branching Processes



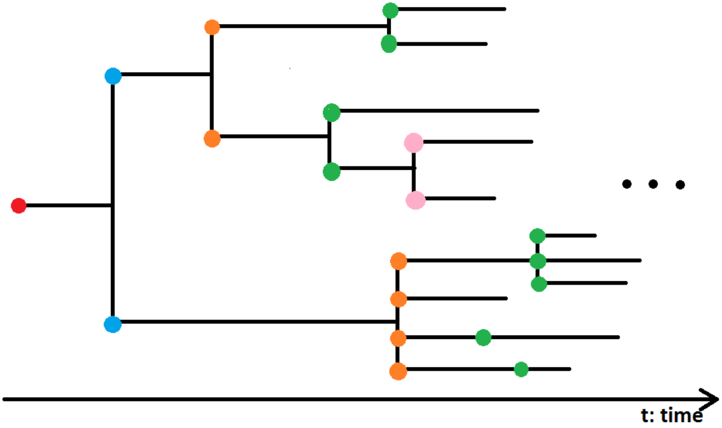
Bellman-Harris Branching Processes



Bellman-Harris Branching Processes



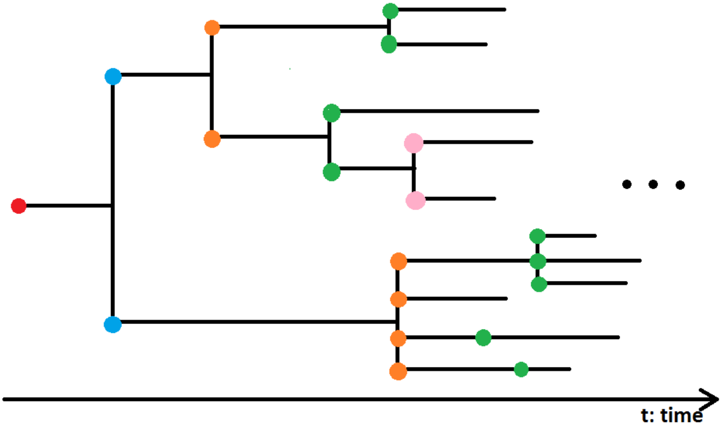
Bellman-Harris Branching Processes



Bellman-Harris Branching Processes

Let $Z(t)$ be the population size at time t, i.e.,

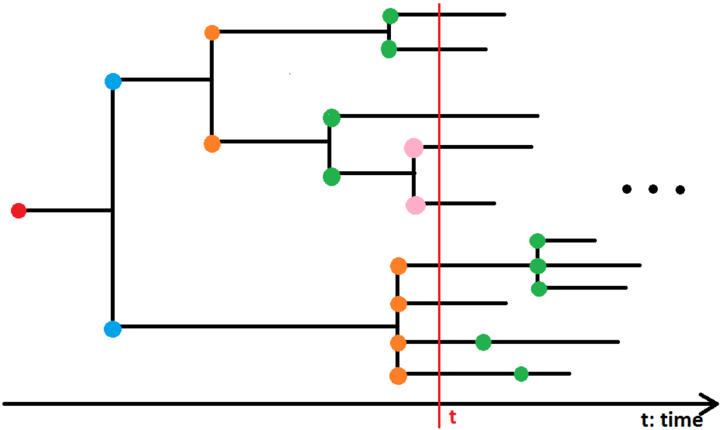
$$Z(t) \quad = \quad \text{the number of individuals alive at time } t.$$



Bellman-Harris Branching Processes

Let $Z(t)$ be the population size at time t, i.e.,

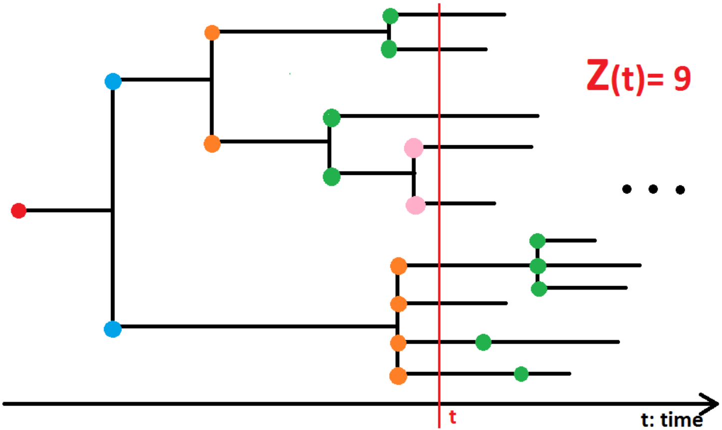
$$Z(t) \quad = \quad \text{the number of individuals alive at time } t.$$



Bellman-Harris Branching Processes

Let $Z(t)$ be the population size at time t, i.e.,

$$Z(t) \quad = \quad \text{the number of individuals alive at time } t.$$



Bellman-Harris Branching Processes

The process $\{Z(t) : t \geq 0\}$ is called a *Bellman-Harris branching process* with the lifetime distribution $G(\cdot)$ and the offspring distribution $\{p_j\}_{j \geq 0}$.

Let

$$m \equiv \sum_{j=1}^{\infty} j p_j \qquad \text{(the offspring mean)}$$

then

(a) $m < 1 \quad \Rightarrow \quad$ the subcritical case

(b) $m = 1 \quad \Rightarrow \quad$ the critical case

(c) $1 < m < \infty \quad \Rightarrow \quad$ the supercritical case

(c) $m = \infty \quad \Rightarrow \quad$ the explosive case

The Coalescence Problem

Consider a Bellman-Harris branching process $\{Z(t) : t \geq 0\}$.

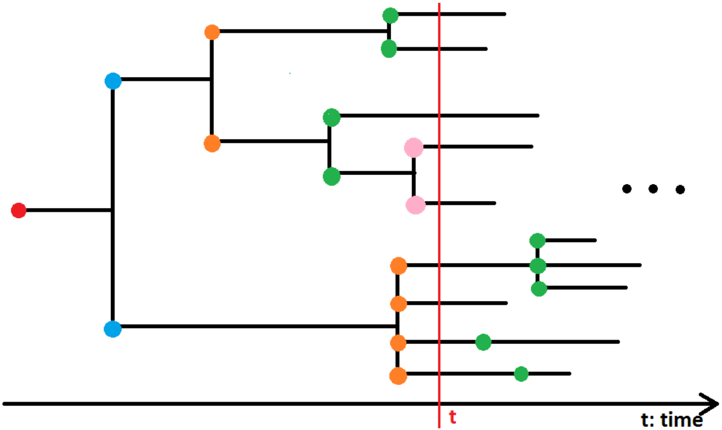
Pick k individuals from the population alive at time t by simple random sampling without replacement.

Trace their lines of descent backward in time till they meet.

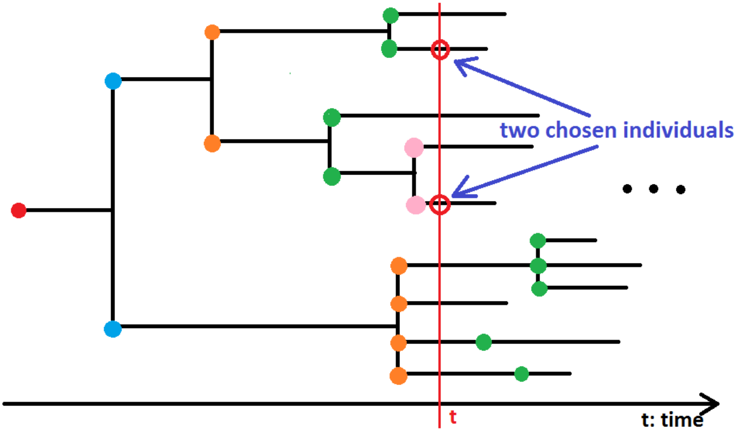
We call the time when the lines of descent met the coalescent time.

We call the common ancestor alive at this coalescent time the most recent common ancestor of these k chosen individuals.

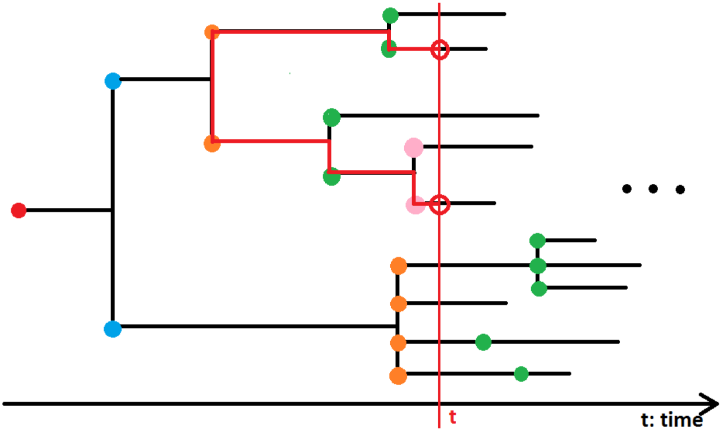
Bellman-Harris Branching Processes



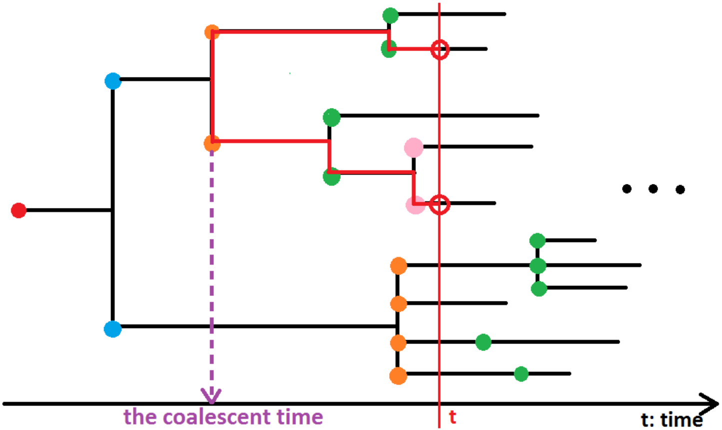
Bellman-Harris Branching Processes



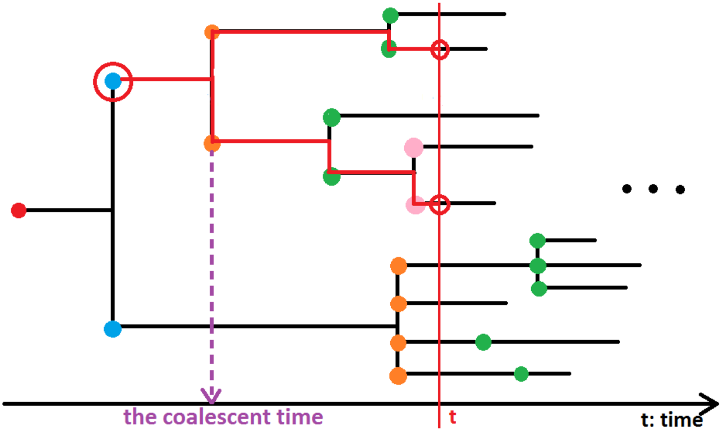
Bellman-Harris Branching Processes



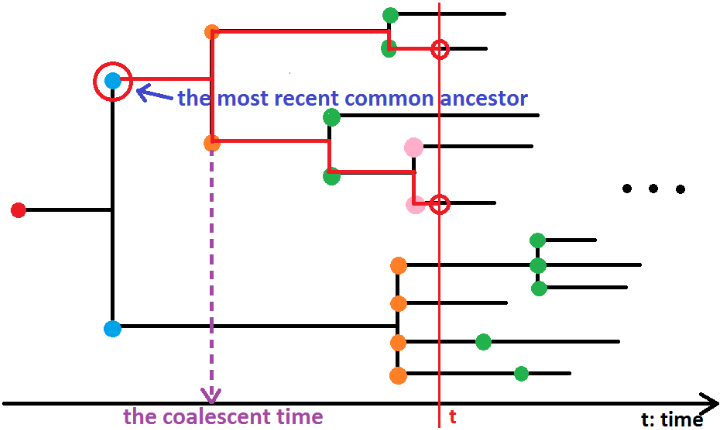
Bellman-Harris Branching Processes



Bellman-Harris Branching Processes



Bellman-Harris Branching Processes



The Coalescence Problem

Let $D_k(t)$ be the coalescence time, then $D_k(t)$ is also the death time of the most recent common ancestor.

Let $X_k(t)$ be the generation number of the most recent common ancestor.

Questions:

(1) What are the distributions of $D_k(t)$ and $X_k(t)$?

(2) What happens to $D_k(t)$ and $X_k(t)$ as $t \to \infty$?

Theorem 2.2.1 (Hong, 2013, Subcritical, Age Chart).

Let $0 < m < 1$ *and* $\sum_{j=1}^{\infty}(j \log j)p_j < \infty$. *Assume that the lifetime distribution* G *is non-lattice,* $G(0+) = 0$ *and such that the Malthusian parameter* α *exists and* $\int_0^{\infty} te^{-\alpha t} dG(t) < \infty$. *Then, conditioned on the event* $\{Z(t) > 0\}$, *the point process*

$$A(t) \equiv \left\{a_{t,i} : 1 \leq i \leq Z(t)\right\}$$

converges in distribution, as $t \to \infty$, *to a point process*

$$\tilde{A} \equiv \left\{\tilde{a}_i : 1 \leq i \leq Y\right\}$$

Theorem 2.2.1 (cont.).

where Y *is the random variable with the distribution* $\{b_j\}_{j\geq 0}$ *as defined in the classical limit theorem. The distribution of* $\tilde{A}$ *is determined by its Laplace functional* $\varphi(s)$.

Theorem 2.2.2 (Hong, 2013, Subcritical).

Let $0 < m < 1$ *and* $\sum_{j=1}^{\infty}(j\log j)p_j < \infty$. *Assume that the lifetime distribution* G *is non-lattice,* $G(0+) = 0$, *the Malthusian parameter* α *exists and* $\int_0^{\infty} te^{-\alpha t}dG(t) < \infty$. *Let* $D_k(t)$ *be defined in Section 1.3. Then, conditioned on* $\{Z(t) \geq 2\}$,

$$t - D_2(t) \xrightarrow{d} \tilde{D}_2 \quad as\ t \to \infty,$$

where $\tilde{D}_2$ *is a positive random variable such that* $P(0 < \tilde{D}_2 < \infty) = 1$. *For any* $u \geq 0$,

$$P\big(\tilde{D}_2 \leq u\big) = 1 - \frac{1}{e^{\alpha u}P(Y \geq 2)}E\Big(\phi(\tilde{A}, u)\Big) \equiv H_2(u)$$

Theorem 2.2.2 (Cont.).

where $\tilde{A}$ and Y are as defined in Theorem 2.2.1, where

$$\phi((a_1, a_2, \cdots, a_k), u) = E\left(\frac{\sum_{i \neq j=1}^{k} \tilde{Z}_i(a_i + u)\tilde{Z}_j(a_j + u)}{\left(\sum_{i=1}^{k} \tilde{Z}_i(a_i + u)\right)\left(\sum_{i=1}^{k} \tilde{Z}_i(a_i + u) - 1\right)} I_{\left(\sum_{i=1}^{k} \tilde{Z}_i(a_i+u) \geq 2\right)} \right)$$

for any positive integer k and any positive real numbers $a_1, a_2, \cdots, a_k$ and $\left\{\tilde{Z}_i(t) : t \geq 0\right\}_{i \geq 1}$ are i.i.d. copies of $\{Z(t) : t \geq 0\}$ with new born initial ancestors.

Theorem 2.2.3 (Athreya and Hong, Accepted by TJM, Supercritical, Generation Number).

Let $1 < m < \infty$, $\sum_{j=1}^{\infty}(j \log j)p_j < \infty$, $p_0 = 0$ *and the life time distribution* G *be non-lattice with* $G(0+) = 0$. *Then, for any integer* $k \geq 2$,

$$X_k(t) \xrightarrow{d} \tilde{X}_k \text{ as } t \to \infty$$

with

$$P(\tilde{X}_k < r) = 1 - E\left(\frac{\sum_{i=1}^{Y_r}\left(e^{-\alpha S_{r,i}} W_{r,i}\right)^k}{\left(\sum_{i=1}^{Y_r} e^{-\alpha S_{r,i}} W_{r,i}\right)^k}\right) \equiv \phi_k(r)$$

for $r = 0, 1, 2, \cdots$ *and* $\lim_{r \uparrow \infty} \phi_k(r) = 1$.

Theorem 2.2.4 (Athrey and Hong, Accepted by TJM, Supercritical, Generation Number).

Let $1 < m < \infty$ *and* $U = \min\left\{n \geq 1 : Y_n \geq 2\right\}$. *Under the same hypotheses of Theorem 2.2.3, then* $\tilde{X}_k \xrightarrow{d} U - 1 \quad \text{as } k \to \infty$.

Theorem 2.2.5 (Athreya and Hong, Accepted by TJM, Supercritical, Death Time).

Let $1 < m < \infty$, $\sum_{j=1}^{\infty}(j \log j)p_j < \infty$, $p_0 = 0$ *and the life time distribution* G *be non-lattice with* $G(0+) = 0$. *Then, for any integer* $k \geq 2$, *there exists a nonnegative real-valued random variable* $\tilde{D}_k$ *such that*

$$D_k(t) \xrightarrow{d} \tilde{D}_k \text{ as } t \to \infty$$

and, for any $s \geq 0$

$$P\big(\tilde{D}_k \leq s\big) = 1 - E\left(\frac{\sum_{i=1}^{Z(s)} \left(e^{-\alpha R_{s,i}} \tilde{W}_{s,i}\right)^k}{\left(\sum_{i=1}^{Z(s)} e^{-\alpha R_{s,i}} \tilde{W}_{s,i}\right)^k}\right) \equiv H_k(s).$$

The random variables $\tilde{W}_{r,i}$, $i = 1, 2, \cdots, Z(s)$, are the i.i.d copies of the sum $\sum_{j=1}^{\xi} W_j$ where ξ is a random variable with the offspring distribution $\{p_j\}_{j\geq 0}$ and $\left\{W_j\right\}_{j\geq 0}$ are i.i.d. copies and ξ and $\{W_j\}_{j\geq 1}$ are independent.

Theorem 2.2.6 (Athreya and Hong, Accepted by TJM, Supercritical, Death time).

Let $1 < m < \infty$ *and* $U = \min\left\{n \geq 1 : Y_n \geq 2\right\}$ *where* $\{Y_n\}_{n\geq 0}$ *is the embedded Galton-Watson branching process. Under the hypotheses of Theorem 2.2.5, there exist a random variable* $\tilde{D}$ *such that*

$$\tilde{D}_k \xrightarrow{d} \tilde{D} \text{ as } k \to \infty$$

and, for any $s \geq 0$,

$$P\left(\tilde{D} \leq s\right) = P(L_0 + L_1 + \cdots + L_{U-1} \leq s)$$

where $\left\{L_i\right\}_{i\geq 0}$ *are i.i.d. random variables with distribution* G *and independent of* U.

Outline

Open Questions

(1) The explosive case in multitype Galton-Watson branching processes and Bellman-Harris branching Processes

(2) The coalescence in the critical Bell-Harris branching process

(3) The positions in branching random work in critical and subcritical cases

(4) $\cdots$

Thank You!!!

Idea of Proofs

Now, let

$$\delta_{ni} = \begin{cases} 1, & \text{if } x_{ni} \leq \sqrt{n}\sigma y \\ 0, & \text{otherwise} \end{cases}$$

Idea of Proofs

Now, let

$$\delta_{ni} = \begin{cases} 1, & \text{if } x_{ni} \le \sqrt{n}\sigma y \\ 0, & \text{otherwise} \end{cases}$$

then

$$E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^2$$

$$= \quad E\left(\frac{1}{Z_n}\sum_{i=1}^{Z_n}\delta_{ni}\right)^2$$

$$= \quad E\left(\frac{1}{Z_n^2}\sum_{i=1}^{Z_n}\delta_{ni}\right) + E\left(\frac{1}{Z_n^2}\sum_{i\neq j}\delta_{ni}\delta_{nj}\right)$$

Idea of Proofs

and

$$E\left(\frac{1}{Z_n^2}\sum_{i\neq j}\delta_{ni}\delta_{nj}\right)$$

Idea of Proofs

and

$$E\left(\frac{1}{Z_n^2}\sum_{i\neq j}\delta_{ni}\delta_{nj}\right)$$

$$= E\left(\frac{1}{Z_n^2}\sum_{i\neq j}E\left(\delta_{ni}\delta_{nj}\middle|Z_n\right)\right) = E\left(\frac{1}{Z_n^2}\sum_{i\neq j}E\left(\delta_{n1}\delta_{n2}\middle|Z_n\right)\right)$$

$$= E\left(\frac{Z_n(Z_n-1)}{Z_n^2}E\left(\delta_{n1}\delta_{n2}\right)\right) = E\left(\frac{Z_n(Z_n-1)}{Z_n^2}\right)\cdot E\left(\delta_{n1}\delta_{n2}\right)$$

$$= E\left(\frac{Z_n(Z_n-1)}{Z_n^2}\right)\cdot P(x_{n1}\le\sqrt{n}\sigma y, x_{n2}\le\sqrt{n}\sigma y)$$

Idea of Proofs

In supercritical case, we want to show that

$$P(x_{n1} \le \sqrt{n}\sigma y, x_{n2} \le \sqrt{n}\sigma y) \to \Phi(y)^2, \text{ as } n \to \infty.$$

In explosive case, we want to show that

$$P(x_{n1} \le \sqrt{n}\sigma y, x_{n2} \le \sqrt{n}\sigma y) \to \Phi(y), \text{ as } n \to \infty.$$