Branching processes

Sophie Hautphenne

University of Melbourne & EPFL

Stochastic Modelling Meets Phylogenetics
16–18 November 2015
Branching processes are stochastic processes describing the dynamics of a population of individuals which reproduce and die independently, according to some specific probability distributions.

Branching processes have numerous applications in population biology and phylogenetics.
There are many types of branching processes:

- Discrete time (Galton-Watson branching processes)
- Continuous time,
  - with exponential lifetime distributions (Markovian branching process), or
  - general lifetime distributions (age-dependent, Bellman-Harris branching process)
- Single type, or multitype (with finitely or $\infty$ many types)
- Individuals reproduction rules may depend on the actual size of the population (population size-dependent branching process)
- Branching processes can undergo catastrophes or live in a random environment
- ...
Outline

1. Introduction
2. Galton-Watson branching processes
3. Markovian branching processes
4. Markovian binary trees
The Galton-Watson branching process

- Time is **discrete** and represents **successive generations**
- Each individual has a **unit lifetime**, at the end of which it might give birth to one or more offsprings simultaneously
- The **offspring distribution** is described by a random variable $\xi$ taking non-negative integer values with corresponding probabilities
  \[ p_k = \mathbb{P}[\xi = k], \quad k \geq 0. \]
- All individuals behave **independently** of each other
A realisation of a GW process through 3 generations starting with a single individual at generation 0:
The Galton-Watson branching process

- The population size at generation $n$ is denoted by $Z_n$
- We have the branching process equation
  \[ Z_n = \sum_{i=1}^{Z_{n-1}} \xi_i, \quad n \geq 1, \]
  where $\xi_1, \xi_2, \ldots$ are i.i.d. copies of $\xi$.
- The process $\{Z_n, n \geq 0\}$ is a discrete-time Markov chain on the state space $\{0, 1, 2, 3, \ldots\}$ where state 0 is absorbing and all other states are transient.
The probability generating function (p.g.f.) of $\xi$ is

$$P(s) := \mathbb{E}[s^\xi] = \sum_{k=0}^{\infty} \mathbb{P}[\xi = k]s^k, \quad s \in [0, 1].$$

If $Z_0 = 1$, then $P(s)$ corresponds to the p.g.f of $Z_1$.

Let $P_n(s)$ denote the p.g.f. of $Z_n$,

$$P_n(s) := \sum_{k=0}^{\infty} \mathbb{P}[Z_n = k]s^k.$$
Define the random variable $S_N$ as

$$S_N := \sum_{i=1}^{N} X_i,$$

where $X_i$ are i.i.d. with p.g.f. $G_X(s)$, and $N$ is an independent random variable taking values in $\mathbb{Z}^+$, with p.g.f. $G_N(s)$.

Then the p.g.f. of $S_N$ is given by

$$G_{S_N} = G_N(G_X(s)).$$
Population size distribution

Recall that for all $n \geq 1$,

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_i$$

where $\xi_1, \xi_2, \ldots$ are i.i.d. copies of $\xi$.

**Proposition**

*Conditionally on $Z_0 = 1$, the p.g.f. of $Z_n$ satisfies*

$$P_n(s) = P_{n-1}(P(s))$$

$$= P(P(\ldots P(s)))$$

$$= \underbrace{P(P(\ldots P(s)))}_{n}$$

$$= P(P_{n-1}(s)), \quad n \geq 1,$$

with $P_1(s) = P(s)$. 

Examples of progeny distributions

- **Binary case**: \( \xi = \begin{cases} 0 & \text{w.p. } 1/3, \\ 2 & \text{w.p. } 2/3 \end{cases} \)
  
  \[ P(s) = \sum_k P[\xi = k] s^k = \frac{1}{3} + \frac{2}{3}s^2 \]

- **Geometric case**: \( \xi \sim \text{Geom}(p) \),
  
  \[ P[\xi = k] = (1 - p)^k p \quad \rightarrow \quad P(s) = \frac{p}{1 - (1 - p)s} \]

- **Poisson case**: \( \xi \sim \text{Poisson}(\lambda) \),
  
  \[ P[\xi = k] = e^{-\lambda} \lambda^k / k! \quad \rightarrow \quad P(s) = e^{\lambda(s-1)} \]
Mean progeny and criticality

The mean progeny is

\[ m := \mathbb{E}[\xi] = P'(1-) \in (0, \infty). \]

**Corollary**

The mean population size at generation \( n \), conditional on \( Z_0 = 1 \), is given by

\[ M_n := \mathbb{E}[Z_n | Z_0 = 1] = P'_n(1-) = m^n. \]

- \( m < 1 \): subcritical case, \( M_n \downarrow 0 \) as \( n \to \infty \)
- \( m = 1 \): critical case \( M_n = 1 \) for all \( n \)
- \( m > 1 \): supercritical case, \( M_n \uparrow \infty \) as \( n \to \infty \)
Extinction probability

- Let $q_n = \mathbb{P}[Z_n = 0] = P_n(0)$ be the probability that the $n$th generation is empty.
  
  If $Z_0 = 1$, then $q_0 = 0$ and
  
  $$P_n(s) = P(P_{n-1}(s)) \Rightarrow q_n = P(q_{n-1}), \quad n \geq 1.$$  

- The probability of extinction of the branching process is
  
  $$q = \lim_{n \to \infty} \mathbb{P}[Z_n = 0] = \mathbb{P}[\lim_{n \to \infty} Z_n = 0] \Rightarrow q = P(q)$$

**Theorem**

The extinction probability $q$ is the minimal nonnegative solution of the fixed point equation

$$s = P(s).$$
In practice, when \( s = P(s) \) cannot be solved explicitly, \( q \) is obtained as the limit of the sequence \( q_n \) computed using the functional iteration \( q_n = P(q_{n-1}) \) with \( q_0 = 0 \).

Examples:

- **Binary case**: \( s = (1/3) + (2/3)s^2 \rightarrow q = 1/2 \)

- **Geometric case**: \( s = \frac{p}{1 - (1 - p)s} \rightarrow q = \min\left(\frac{p}{1 - p}, 1\right) \), i.e.
  \[
  q = \begin{cases} 
  1 & \text{if } p \geq 1/2 \\
  p/(1 - p) & \text{if } p < 1/2 
  \end{cases}
  \]

- **Poisson case**: \( s = e^{\lambda(s-1)} \rightarrow \) the functional iteration is useful here!
A closer look at the progeny generating function

$P(s)$ is an increasing, convex function such that $P(1) = 1$

→ $P(s)$ has at most two fixed points in $[0, 1]$: $P''(1) \leq 1$

$m = P'(1) \leq 1$

→ $P(s)$ has a fixed point $q < 1$ if and only if $m > 1$
**Extinction criterion**

**Proposition (Extinction criterion)**

\[ q < 1 \iff m > 1 \]

Recall

- in the supercritical case, \( m > 1 \implies q < 1 \)
- in the critical case, \( m = 1 \implies q = 1 \)
- in the subcritical case, \( m > 1 \implies q = 1 \)
Dichotomy: extinction or explosion!

Theorem

Regardless the value of $m$, any state $k \neq 0$ is transient, so that

$$\lim_{n \to \infty} \mathbb{P}[Z_n = k] = 0 \quad \text{for any } k = 1, 2, \ldots.$$ 

Moreover, 

$$\mathbb{P}[Z_n \to 0] + \mathbb{P}[Z_n \to \infty] = 1.$$ 

In the supercritical case, conditionally on non-extinction, 

$$\lim_{n \to \infty} Z_n = +\infty \text{ a.s.}$$

Can we tell more about the growth rate of $Z_n$?
Let $W_n = Z_n/m^n$. The sequence $\{W_n, n \geq 0\}$ is a martingale with $E[W_n] = 1$ for all $n$, so it converges a.s. to a nonnegative random variable $W$,

$$W := \lim_{n \to \infty} \frac{Z_n}{m^n} \to Z_n \sim W m^n$$

**Theorem (Kesten-Stigum)**

Either $P[W = 0] = q$ or $P[W = 0] = 1$.

The following are equivalent:

(i) $E[\xi \log^+ \xi] < \infty$

(ii) $P[W = 0] = q$

(iii) $W_n$ converges in mean ($L^1$)

Quasi-stationary distribution in the subcritical case $m < 1$

**Theorem (Yaglom)**

For each $j = 1, 2, \ldots$, 

$$ \lim_{n \to \infty} \mathbb{P}[Z_n = j \mid Z_n \neq 0] = b_j $$

exists, and $\sum_j b_j = 1$.

Moreover, the p.g.f. $G(s) = \sum_k b_k s^k$ satisfies the equation

$$ G(P(s)) = m G(s) + 1 - m. $$

In addition, the vector $\mathbf{b} = (b_1, b_2, \ldots)$ satisfies

$$ \mathbf{b}Q = m \mathbf{b}, $$

where $Q$ is the truncated probability transition matrix of the GW restricted to the transient states.
The critical case $m = 1$

**Theorem (Kesten-Ney-Spitzer)**

Assume $\sigma^2 := \text{Var}(\xi) < \infty$. Then we have

- **Kolmogorov’s estimate**
  \[
  \lim_{n \to \infty} n \mathbb{P}[Z_n \neq 0] = \frac{2}{\sigma^2}.
  \]

- **Yaglom’s universal limit law**
  \[
  \lim_{n \to \infty} \mathbb{P}[Z_n/n \geq x \mid Z_n \neq 0] = \exp(-2x/\sigma^2), \quad x > 0.
  \]
Multitype Galton-Watson branching process

Suppose now there are \( r > 1 \) types of individuals, each type having its own reproduction law.

Example with \( r = 2 \):

![Diagram of a multitype Galton-Watson branching process]

0 1 2 3 Generation n
Multitype Galton-Watson branching process

Population size vector: $Z_n = (Z_{n1}, Z_{n2}, \ldots, Z_{nr})$, $n \in \mathbb{N}$, where

$Z_{ni}$: # of individuals of type $i$ at the $n$th generation

$\{Z_n\}$ is an $r$-dimensional Markov process with state space $\mathbb{N}^r$ and an absorbing state $0 = (0, 0, \ldots, 0)^T$

Progeny distribution: $p_{ij}$: $j = (j_1, j_2, \ldots, j_r)$, where

$p_{ij} =$ probability that a type $i$ gives birth to $j_1$ children of type 1, $j_2$ children of type 2, \ldots, $j_r$ children of type $r$. 
Multitype Galton-Watson branching processes

Progeny generating vector $\mathbf{P}(s) = (P_1(s), P_2(s), \ldots, P_r(s))$, where $P_i(s)$ is the progeny generating function of an individual of type $i$

\[ P_i(s) = \sum_{j \in \mathbb{N}^r} p_{ij} s^j = \sum_{j \in \mathbb{N}^r} p_{ij} \prod_{k=1}^r s_{jk}^k, \quad s_i \in [0, 1] \]

Mean progeny matrix $M$ with elements

\[ M_{ij} = \frac{\partial P_i(s)}{\partial s_j} \bigg|_{s=1} = \text{expected number of direct offsprings of type } j \text{ born to a parent of type } i \]

Irreducible branching process $\equiv M$ is irreducible
Extinction probability

As usual, we assume that the process starts with a single individual

**Extinction probability vector** \( \mathbf{q} = \left( q_1, q_2, \ldots, q_r \right)^\top \), with

\[
q_i = \mathbb{P} \left[ \lim_{n \to \infty} |Z_n| = 0 \mid \phi_0 = i \right]
\]

\[
\mathbf{q} = \mathbb{P} \left[ \lim_{n \to \infty} |Z_n| = 0 \mid \phi_0 \right],
\]

where \( \phi_0 \) is the type of the first individual in generation 0.

The vector \( \mathbf{q} \) is the (componentwise) minimal nonnegative solution of

\[
\mathbf{P}(\mathbf{s}) = \mathbf{s}, \quad s_i \in [0, 1], \quad i \in S.
\]

A before, \( \mathbf{q} = \lim_{n \to \infty} \mathbf{q}_n \) where \( \mathbf{q}_0 = \mathbf{0} \) and \( \mathbf{q}_n = \mathbf{P}(\mathbf{q}_{n-1}) \) for all \( n \geq 1 \).
Mean population size and extinction criterion

**Proposition**

The mean number of individuals of type $j$ at generation $n$, starting with a single individual of type $i$, is given by

$$\mathbb{E} \left[ Z_{nj} \mid \varphi_0 = i \right] = (M^n)_{ij}$$

The extinction criterion simply relies on the Perron-Frobenius eigenvalue $\rho$ of the mean progeny matrix $M$: in the irreducible case,

- If $\rho > 1$, then $q < 1$ (supercritical case)
- If $\rho = 1$, then $q = 1$ (critical case)
- If $\rho < 1$, then $q = 1$ (subcritical case)
Stable type distribution (irreducible case)

Let $\mathbf{u}$ be the (strictly positive) Perron-Frobenius left eigenvector of $M$ corresponding to $\rho$, normalized s.t. $\mathbf{u} \mathbf{1} = 1$.

Then, in the long run, we can show that

$$\lim_{n \to \infty} \frac{\mathbb{E}[Z_{ni}]}{\mathbb{E}[|Z_n|]} = u_i.$$ 

The vector $\mathbf{u}$ is called the stable type distribution.
When $r = \infty$, things become more tricky and very interesting behaviours can occur.

For instance, there are several types of extinction events to consider.

$M$ becomes an infinite matrix and the concept of Perron-Frobenius eigenvalue is replaced by the concept of convergence norm.

More details in the next talk given by Peter B.!
Markovian branching processes

In a Markovian branching process,

- the lifetime of an individual is *exponentially* distributed with parameter $\mu$;

- during its lifetime, the individual reproduces according to a *Poisson process* with rate $\lambda$, giving birth to one child at a time.

- all new individuals behave *independently of each other*, following the same rules as their parent.
Markovian branching processes
Evolution of a population over time
**Markovian branching processes**

Population size at time $t$

- $Z(t)$ population size in the branching process at time $t$

  $\rightarrow \{Z(t), t \in \mathbb{R}^+\}$ is a **continuous-time Markov chain** with state space $\{0, 1, 2, 3, \ldots\}$, where state 0 is absorbing, and with generator

  $$Q = \begin{bmatrix}
  0 & 0 & 0 & 0 & \ldots \\
  \mu & -(\mu + \lambda) & \lambda & 0 & \ldots \\
  0 & 2\mu & -2(\mu + \lambda) & 2\lambda & \ldots \\
  0 & 0 & 3\mu & -3(\mu + \lambda) & \ldots \\
  \vdots & \vdots & \vdots & \vdots & \ddots
  \end{bmatrix}.$$

- Such a Markov chain is also called a **linear birth and death process**
Markovian branching processes
Population size at time $t$

- $F(s, t)$: probability generating function (p.g.f.) of $Z(t)$:

$$F(s, t) = \sum_{k \geq 0} \mathbb{P}[Z(t) = k] s^k, \quad s \in [0, 1].$$

- We assume that $\mathbb{P}[Z_0 = 1] = 1$. Therefore, $F(s, 0) = s$.

**Theorem**

$F(s, t)$ satisfies the differential equation

$$\frac{\partial F(s, t)}{\partial t} = \mu - (\mu + \lambda) F(s, t) + \lambda F^2(s, t)$$

with initial condition $F(s, 0) = s$.

This Riccati o.d.e corresponds to the backward Kolmogorov equation for $\{Z(t)\}$, and can be solved explicitly.
Markovian branching processes

Population size at time $t$

- $M(t)$: mean population size at time $t$:

$$M(t) = \left. \frac{\partial F(s, t)}{\partial s} \right|_{s=1}$$

$$\rightarrow \frac{dM(t)}{dt} = (\lambda - \mu) M(t), \quad M(0) = 1.$$ 

Therefore

$$M(t) = e^{(\lambda-\mu)t}.$$ 

→ Three types of behaviour:

- if $\lambda > \mu$, then $\lim_{t \to \infty} M(t) = \infty$ (supercritical case),
- if $\lambda = \mu$, then $M(t) = 1$ for all $t$ (critical case),
- if $\lambda < \mu$, then $\lim_{t \to \infty} M(t) = 0$ (subcritical case).
Markovian branching processes

Time until extinction

- $T_e$: time until extinction of the branching process
- $F(t) = P[T_e < t] = P[Z(t) = 0] = F(0, t)$, the distribution of the time until extinction

Theorem

$F(t)$ satisfies the differential equation

$$\frac{\partial F(t)}{\partial t} = \mu - (\mu + \lambda) F(t) + \lambda F^2(t)$$

with initial condition $F(0) = 0$.

This equation can also be solved explicitly.
Markovian branching processes
Extinction probability

- $q = P[T_e < \infty] = \lim_{t \to \infty} F(t)$: the probability that the branching process eventually becomes extinct

- $q$ satisfies the quadratic equation

$$0 = \mu - (\mu + \lambda) s + \lambda s^2,$$

or equivalently

$$s = \sum_{n \geq 0} \left( \frac{\lambda}{\mu + \lambda} \right)^n \left( \frac{\mu}{\mu + \lambda} \right) s^n P(s)$$

where $P(s)$ is the progeny generating function of an individual

(it can be shown that the progeny is geometrically distributed with parameter $\lambda/(\lambda + \mu)$)
Markovian branching processes
Extinction probability

Theorem

The extinction probability $q$ is the minimal nonnegative solution of the fixed-point equation

$$s = P(s)$$

where $P(s)$ is the progeny generating function of the individuals.

In our case, we can solve the fixed point equation explicitly, and we find

$$q = \min(1, \mu/\lambda).$$

We see that

$$q < 1 \Leftrightarrow \lambda > \mu,$$

which again shows that the process has a positive probability to explode only when $\lambda > \mu$. 
Extinction criterion

**Theorem**

*Extinction is almost sure if and only if the mean progeny is less than or equal to one, or in mathematical notation,*

\[ q = 1 \iff m = P'(1) \leq 1. \]

For the Markovian branching process, the mean progeny is 

\[ m = (1 - p)/p \]

where \( p = \mu/(\lambda + \mu) \), that is, \( m = \lambda/\mu \).

We thus again have that \( q = 1 \iff \lambda \leq \mu \).
Example of multitype Markovian branching process
Yule’s evolutionary process

Two levels of individuals are considered: species and genus (group of species)

- within a single genus, the collection of species evolve as a pure birth process with birth rate $\lambda_s$;

- the collection of genera evolve according to a pure birth process with birth rate $\lambda_g$.

At each ramification, a new genus evolves which has exactly one species, and the old genus continues unchanged.

The process is a branching process within a branching process.
Example of multitype Markovian branching process

Yule’s evolutionary process

A sample path of Yule’s evolutionary process:

Branching of species is represented by *continuous lines*. Boundaries of genera are represented by *dotted line* ‘tubes’, and branching of genera by *arrows*.

Figure taken from Kimmel and Axelrod, 2002
Example of multitype Markovian branching process

Yule’s evolutionary process

Question: what is the rate of evolution of new genera compared to the evolution of new species, that is, what is $\lambda_g/\lambda_s$?

Two sample paths of the Yule’s process: low and high value of the $\lambda_g/\lambda_s$ ratio.

Figure taken from Kimmel and Axelrod, 2002
The genera process can be seen as a multitype Markovian branching process with countably infinitely many types, where the type of a genus is the number of species living in that genus.

We can simplify the model by considering two classes of genera only:

- class 1 genera contain one species only (monotype), and
- class 2 genera contain more than one species (polytype).
Using results on multitype Markovian branching processes, we obtain that the asymptotic expected proportion of monotypic genera is equal to

$$p = \frac{1}{(\lambda_s/\lambda_g) + 1}$$

\(\rightarrow \lambda_s/\lambda_g\) can be computed from estimated values of \(p\).
1. Introduction

2. Galton-Watson branching processes

3. Markovian branching processes

4. Markovian binary trees
Trade-off between realism and tractability

Simplest case:

- Lifetimes follow an **exponential** distribution
- Reproduction occurs according to a **Poisson process**

Not realistic enough!

It is well known in evolutionary biology that **molecular evolution cannot be described by a time-homogeneous Poisson process** (Gillespie and Langley, 1979).
Trade-off between realism and tractability

Bellman-Harris branching processes:
- Lifetimes follow an arbitrary distribution
- Reproduction occurs according to a more general process

Not tractable enough!

We introduce a type of branching process, called the Markovian binary tree (MBT), which is at the same time very general and tractable.

In an MBT, individuals’ lifetime is structured into phases.
Phase-structured lifetime

Example 1

Human’s lifetime decomposed into age-classes (e.g. of 5 years)

1 = 0 – 4 years,
2 = 5 – 9 years,
3 = 10 – 14 years,
...
22 = over 100 years,
0 = dead
Phase-structured lifetime

Example 2

The right whale’s life cycle decomposed into stages

1 = calf,
2 = immature,
3 = mature,
4 = mother,
5 = dead
Phase-structured lifetime

3 types of transitions:
- "Hidden" transitions
- Birth
- Death
The individuals’ lifetime in an MBT

- Lifetime controlled by an underlying Markov process with \( n \) transient phases and one absorbing phase;

\[ i \rightarrow j \rightarrow k \rightarrow 0 \]

- \( \alpha_i \): initial phase distribution (\( 1 \times n \) vector);
- \( D_0 \): hidden phase transition rates (\( n \times n \) matrix);
- \( B \): transition rates associated with a birth (\( n \times n^2 \) matrix);
- \( d \): transition rates associated with the death (\( n \times 1 \) vector).
The lifetime of the individuals is distributed according to a phase-type (PH) distribution.

PH distributions are dense in the class of all distributions with non-negative support.

The reproduction process is a generalisation of the Poisson process called a (transient) Markovian arrival process. In such a point process, the inter-arrival times are not necessarily independent nor identically distributed.

MAPs are dense in the class of all stochastic point-processes.
MBTs and matrix analytic methods

MBTs are matrix generalisation of Markovian branching processes → they are mathematical objects at the intersection of branching processes and matrix analytic methods.

- We use techniques inspired by the matrix analytic methods to compute the extinction probability of branching processes;
- All the algorithms have a probabilistic interpretation.
Gillespie (1986) suggested that molecular evolution may be an episodic process, with bursts of substitutions followed by long periods with no substitution.

→ MBT is an excellent candidate for modelling evolution; indeed, the transient MAP controlling the lifetime and the reproduction of individuals offers enough flexibility to account for non-constant speciation and extinction rates.

Kontoleon (2006) showed that the MBT not only subsumes many macroevolutionary models but is also amenable to simpler analysis.
Population size at time \( t \) in an MBT

- \( \mathbf{Z}(t) = (Z_1(t), \ldots, Z_n(t)) \) the pop. size vector at time \( t \), where \( Z_i(t) \) : \# of individuals in phase \( i \) at time \( t \)

- \( F_i(s, t) \) : the conditional p.g.f. of the pop. size at time \( t \), given that the MBT starts with a first individual in phase \( i \)

\[
F_i(s, t) = \sum_{k \in \mathbb{N}^n} \mathbb{P}[\mathbf{Z}(t) = k \mid \varphi_0 = i] s^k,
\]

where \( s = (s_1, \ldots, s_n) \trans , \ |s_i| \leq 1 \) and \( s^k = s_1^{k_1} \cdots s_n^{k_n} \).

- \( \mathbf{F}(s, t) = (F_1(s, t), \ldots, F_n(s, t)) \trans \)
Kronecker products

If $A$ is an $m \times n$ matrix and $B$ is a $p \times q$ matrix, then the Kronecker product $A \otimes B$ is the $mp \times nq$ block matrix defined by

$$A \otimes B = \begin{bmatrix}
A_{11} B & \cdots & A_{1n} B \\
\vdots & \ddots & \vdots \\
A_{m1} B & \cdots & A_{mn} B
\end{bmatrix}.$$
Population size at time $t$ in an MBT

**Theorem**

$\mathbf{F}(s, t)$ satisfies the following matrix differential equation

$$\frac{\partial \mathbf{F}(s, t)}{\partial t} = \mathbf{d} + D_0 \mathbf{F}(s, t) + B (\mathbf{F}(s, t) \otimes \mathbf{F}(s, t)),$$

with $\mathbf{F}(s, 0) = s$

This differential equation corresponds to the *backward Kolmogorov equation* for the Markov chain $\{\mathbf{Z}(t), t \in \mathbb{R}^+\}$.

It has no explicit solution in general!
Population size at time $t$ in an MBT

- $M_{ij}(t) = E[Z_j(t) | \varphi_0 = i]$ : mean population size in phase $j$ at time $t$ starting with one individual in phase $i$ at time $t = 0$

  $\rightarrow M(t) = (M_{ij}(t))$ with

  $$M_{ij}(t) = \left( \frac{\partial F_i(s, t)}{\partial s_j} \right) \bigg|_{s=1}.$$

- We obtain the matrix differential equation for $M(t)$ :

  $$\frac{\partial M(t)}{\partial t} = \Omega M(t), \quad M(0) = I,$$

  where $\Omega = D_0 + B (1 \otimes I + I \otimes 1)$.

  $\rightarrow M(t) = e^{\Omega t}.$

- $\omega = P-F$ eigenvalue of $\Omega$, plays a similar role as $\lambda - \mu$.
Time until extinction

- \( F_i(t) \): conditional probability that the population becomes extinct before time \( t \), given that it started at time \( t = 0 \) with one individual in phase \( i \),

\[ \rightarrow \mathbf{F}(t) = (F_1(t), \ldots, F_n(t))^T = \mathbf{F}(0, t) \]

**Theorem**

\( \mathbf{F}(t) \) satisfies the following matrix differential equation

\[
\frac{\partial \mathbf{F}(t)}{\partial t} = d + D_0 \mathbf{F}(t) + B (\mathbf{F}(t) \otimes \mathbf{F}(t)),
\]

with \( \mathbf{F}(0) = 0 \).

As for \( \mathbf{F}(s, t) \), no explicit solution in general!
Extinction probability

- $q_i$: conditional probability that the population eventually becomes extinct, given that it started at time $t = 0$ with one individual in phase $i$

$$q = (q_1, \ldots, q_n)^T = \lim_{t \to \infty} F(t)$$

$F(t)$ is the minimal non-negative solution of

$$0 = d + D_0 s + B (s \otimes s).$$

- Let $\theta = (-D_0)^{-1} d$ and $\Psi = (-D_0)^{-1} B$.

$$s = \theta + \Psi (s \otimes s),$$

or equivalently

$$s = [I - \Psi (s \otimes l)]^{-1} \theta \quad \text{or} \quad s = [I - \Psi (l \otimes s)]^{-1} \theta.$$
Extinction probability

\[ s = \theta + \Psi(s \otimes s) \]

\[ \equiv \]

\[ s = [I - \Psi(I \otimes s)]^{-1} \theta \]

\[ \equiv \]

\[ s = [I - \Psi(s \otimes I)]^{-1} \theta \]

1. The Depth and the Order algorithms (linear) (Bean et al., 2008)
2. The Thicknesses algorithm (linear) (Hautphenne et al., 2011).
3. The Newton algorithms (quadratic) (Hautphenne et al., 2008; Hautphenne and van Houdt, 2010)
The Depth algorithm

\[ s_0 = \theta \]
\[ s_k = \theta + \psi (s_{k-1} \otimes s_{k-1}), \quad k \geq 1 \]

For \( k \geq 0 \),
- \( \mathcal{M}_k \) = the set of MBTs considered at stage \( k \)
- \( s_k = \mathbb{P}[\mathcal{M}_k \mid \varphi_0] \) = the \( k \)th approximation of \( q \).

\[ \mathcal{M}_0 = \]

\[ \mathcal{M}_k = \begin{array}{c} \mathcal{M}_k-1 \\ \cup \\ \mathcal{M}_k-1 \end{array} \]

for \( k \geq 1 \).
Probabilistic interpretation of the Depth algorithm

\[ s_0 = \theta \]
\[ s_k = \theta + \psi (s_{k-1} \otimes s_{k-1}), \quad k \geq 1 \]

**Depth** of an MBT = number of branching points along the longest branch

For \( k \geq 0 \),

- \( \mathcal{M}_k = \) the set of **extinct MBTs with a depth \( \leq k \)** (constraint on the shape of the tree)
- \( \mathcal{M}_k \subseteq \mathcal{M}_{k+1} \subseteq \cdots \subseteq \mathcal{M} = \) the set of all extinct MBTs.
- \( s_k = \mathbb{P}[\mathcal{M}_k \mid \varphi_0] \uparrow \mathbf{q} \) as \( k \to \infty \).
The Order algorithm

\[
\begin{align*}
    s_0 &= \theta \\
    s_k &= [I - \Psi (s_{k-1} \otimes I)]^{-1} \theta, \quad k \geq 1
\end{align*}
\]

For \( k \geq 0 \),
- \( \mathcal{M}_k = \) the set of MBTs considered at stage \( k \)
- \( s_k = \mathbb{P}[\mathcal{M}_k | \varphi_0] = \) the \( k \)th approximation of \( q \).

\( \mathcal{M}_0 = \)

\[ \xymatrix{ & & \mathcal{M}_k \\
\mathcal{M}_k & & \mathcal{M}_k-1 \ar[lu] \ar[rr] & & \mathcal{M}_k-1 \ar[lu] \ar[rr] & & \cdots \ar[ll] \quad \text{for } k \geq 1.} \]
Probabilistic interpretation of the Order algorithm

\begin{align*}
  s_0 &= \theta \\
  s_k &= [I - \Psi (s_{k-1} \otimes I)]^{-1} \theta, \quad k \geq 1
\end{align*}

Order of an MBT = total number of children generations

For \( k \geq 0 \),

\begin{itemize}
  \item \( \mathcal{M}_k = \) the set of **extinct** MBTs with an order \( \leq k \)
    (constraint on the shape of the tree)
  \item \( \mathcal{M}_k \subseteq \mathcal{M}_{k+1} \subseteq \cdots \subseteq \mathcal{M} = \) the set of all extinct MBTs.
  \item \( s_k = \mathbb{P}[\mathcal{M}_k \mid \varphi_0] \uparrow q \) as \( k \to \infty \).
\end{itemize}
Branching processes have numerous applications in population biology and in demography.

Here, we use the MBT to model the dynamics of female families in several countries.

Since our model is asexual, we look at women only.
We structure the lifetime of women in age classes.

**Age-specific** fertility and mortality rates on UN and WHO mainly correspond to 5 years age classes → 22 age classes or **phases**.

We assume **exponential** sojourn time in each phase (time unit = 1 year).

\[
D_0 = \begin{bmatrix}
* & 1 & * & 1/4 & 1/5 & \cdots & * \\
* & * & * & 1/5 & * & \cdots & * \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
* & 1/5 & * & * & * & \cdots & *
\end{bmatrix}
\]
Age-specific fertility rate in age class $i = \#$ living births during the year, according to the age class $i$ of the mother, for each 1000 women of the same age class $i$.

→ Female birth rate $\beta_i$ per individual in phase $i$

$$\beta_i = \frac{\text{age-specific fertility rate in } i}{1000 \cdot (\text{sex ratio} + 1)}.$$

$$\beta = \begin{bmatrix} 0_4, & \beta_5, & \ldots & \beta_{11}, & 0_{11} \end{bmatrix}'$$

and

$$B = (e_1 \otimes \text{diag}(\beta)).$$
Age-specific mortality rate \( d_i \) in age class \( i \)

\[
= \frac{\# \text{ deaths during the year of women in age class } i}{\text{population in the age class } i}
\]

\[
d = [d_1, d_2, \ldots, d_{22}]'
\]
$m(t) = (M(t)1)_1$ : Mean total family size generated by a new-born woman after $t$ units of time:
Application of MBTs in demography

$F_1(t)$ : Distribution function of the time until extinction of the family generated by one new-born woman
Application of MBTs\(x\) in demography

\(q\): Extinction probability of the family generated by a single woman as a function of her age class:
Parameter estimation under complete observation

An MBT is represented by the set of parameters $\left( \alpha, D_0, B, d \right)$.

**Complete observation**: All birth and death events are observable but phase transitions of the underlying Markov chain are not observable.
**Aim**: to estimate the parameters $\theta = (\alpha, D_0, B, d)$ from the continuous observation of a population during an interval $[0, T]$.

Remember: phase transitions are not observable.

If individuals are **distinguishable**, we observe the sequences of

1. **intervals of time** between population size changes: $\{T_k\}$,
2. **population sizes**: $\{N_k\}$,
3. **individuals responsible** for each population size change: $\{L_k\}$.

$\rightarrow Y = \{(T_k, N_k, L_k), 1 \leq k \leq K\}$

If individuals are **indistinguishable**, we only observe (1) and (2)

$\rightarrow Y = \{(T_k, N_k), 1 \leq k \leq K\}$. 
Observable variables: \( \mathbf{Y} = \{ (T_k, N_k, L_k), 1 \leq k \leq K \} \).

Hidden variables: \( \mathbf{Z} = \{ (H_i, Z_i, A_{ij}, C_{ij\ell}, D_i), i, j, \ell = 1, \ldots, n \} \)

where

- \( H_i \): number of MBTs starting in phase \( i \);
- \( Z_i \): total cumulated time spent in phase \( i \);
- \( A_{ij} \): total number of jumps from phase \( i \) to phase \( j \) without accompanying observable event, for \( i \neq j \);
- \( C_{ij\ell} \): total number of jumps from phase \( i \) to phase \( \ell \) associated with the birth of a child in phase \( j \);
- \( D_i \): number of terminating jumps from phase \( i \).
Parameter estimation under complete observation
Likelihood of a complete sample

Given the parameters \( \theta = (\alpha, D_0, B, d) \) of the MBT, the likelihood of a complete data sample \( X = (Y, Z) \) can be written in the form

\[
L_c(\theta | X) = \prod_{i=1}^{n} \alpha_i^H_i \prod_{i=1}^{n} \exp(D_{0;ii} Z_i) \prod_{i=1}^{n} \prod_{j=1, j \neq i}^{n} D_{0;ij} A_{ij} \prod_{i=1}^{n} \prod_{j=1}^{n} \prod_{\ell=1}^{n} B_{i,j, \ell} C_{i,j, \ell} \prod_{i=1}^{n} d_i^{D_i}
\]
The maximum likelihood estimator

The MLE is

$$\hat{\theta} = \arg\max_\theta \log L_c(\theta \mid X),$$

leading to the maximum likelihood estimates $\hat{\alpha}, \hat{D}_0, \hat{B}$ and $\hat{d}$:

$$\hat{\alpha}_i = \frac{H_i}{M}, \quad \hat{D}_{0;ij} = \frac{A_{ij}}{Z_i}, \quad \hat{B}_{i,j\ell} = \frac{C_{i,j\ell}}{Z_i}, \quad \hat{d}_i = \frac{D_i}{Z_i},$$

$$\hat{D}_{0;ii} = -\left( \sum_{j=1,j\neq i}^n \hat{D}_{0;ij} + \sum_{j=1}^n \sum_{\ell=1}^n \hat{B}_{i,j\ell} + \hat{d}_i \right)$$

for $1 \leq i, j, k \leq n$.

They are expressed in terms of the hidden variables $Z$. 

Complete observation

The maximum likelihood estimator
Complete observation
The EM algorithm applied to MBTs (Hautphenne and Fackrell, 2014).

Since \( Z = \{(H_i, Z_i, A_{ij}, C_{i,j,\ell}, D_i)\} \) are hidden random variables, instead of computing

\[
\hat{\theta} = \arg\max_{\theta} \log L_c(\theta \mid X = (Y, Z)),
\]

we iteratively compute a sequence \( \{\hat{\theta}^{(n)}\}_{n \geq 1} \):

\[
\hat{\theta}^{(n)} = \arg\max_{\theta} \mathbb{E}_Z \left[ \log L_c(\theta \mid X) \mid Y = y, \hat{\theta}^{(n-1)} \right],
\]

with an arbitrary starting point \( \hat{\theta}^{(0)} \).

\( \rightarrow \) 2 steps:

1. E-step : \( \mathbb{E}_Z \left[ \log L_c(\theta \mid X) \mid y, \hat{\theta}^{(n-1)} \right] \propto \mathbb{E} \left[ Z \mid y, \hat{\theta}^{(n-1)} \right] \),

2. M-step : \( \hat{\theta}^{(n)} = \arg\max_{\theta} \mathbb{E}_Z \left[ \log L_c(\theta \mid X) \mid y, \hat{\theta}^{(n-1)} \right] \).
Application in phylogenetics

Fig. 9. A phylogenetic tree showing the DNA mutations among a group of mammals.

Figure taken from Hautphenne and Fackrell, 2014
Application in phylogenetics
Estimation of speciation rate

We compare the estimated mean growth rate $\hat{\omega}$ of the species population for different numbers of phases $n$:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\hat{\omega}$</th>
<th>Loglikelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0256</td>
<td>-32.64</td>
</tr>
<tr>
<td>2</td>
<td>0.0211</td>
<td>-30.66</td>
</tr>
<tr>
<td>3</td>
<td>0.0268</td>
<td>-29.12</td>
</tr>
<tr>
<td>4</td>
<td>0.0281</td>
<td>-28.14</td>
</tr>
</tbody>
</table>
Complete observation
Future work

Use alternative methods for the parameter estimation, such as

- Nonlinear regression methods
- Metropolis-Hasting (MH)-algorithm as part of a Monte Carlo EM-algorithm
- Bayesian Markov chain Monte Carlo approach
Partial observation

This happens for instance

- when a population is observed only at discrete time points. E.g. when the data consist of annual population counts
- when only a fraction of the individuals in a population are sampled. E.g. in the binomial sampling case, individuals are sampled with some probability $0 < p < 1$, independently of each other.

Future work: Adapt the previous methods and develop new methods to account for partial observation
Further reading on branching processes... I

T.E. Harris.  
*The Theory of Branching Processes.*  

C.J. Mode.  
*Multitype Branching Processes.*  

K.B. Athreya and P.E. Ney.  
*Branching Processes.*  

P. Jagers.  
*Branching Processes with Biological Applications.*  
Further reading on branching processes... II

S. Asmussen and H. Hering.
*Branching Processes.*
Birkhauser, Boston, 1983.

K.B. Athreya and P. Jagers (eds).
*Classical and Modern Branching Processes.*

M. Kimmel and D.E. Axelrod.
*Branching Processes in Biology.*

P. Haccou, P. Jagers, V.A. Vatutin (eds.).
*Branching Processes : Variation, Growth and Extinction of Populations.*
... on phylogenetics...

Are evolutionary rates really variable?  

J.H. Gillespie.  
Natural Selection and the Molecular Clock.  
... and on Markovian binary trees I

N. Bean, N. Kontoleon, and P. Taylor.
Markovian trees: Properties and algorithms.

Newton’s iteration for the extinction probability of a Markovian Binary Tree.

Transient features for Markovian binary trees.
... and on Markovian binary trees II

