

Branching processes

Sophie Hautphenne

University of Melbourne & EPFL

Stochastic Modelling Meets Phylogenetics
16–18 November 2015



Introduction

Branching processes



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**

Introduction

Branching processes

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 ∞ 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

- 1 Introduction
- 2 Galton-Watson branching processes**
- 3 Markovian branching processes
- 4 Markovian binary trees

The Galton-Watson branching process

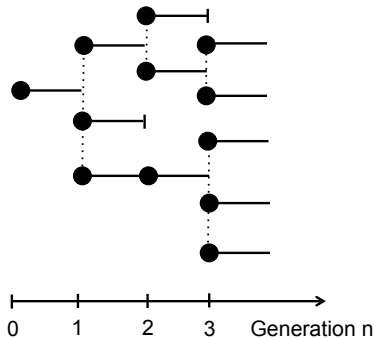
- Time is **discrete** and represents **successive generations**
- Each individual has a **unit lifetime**, at the end of which it might give birth to one or more offsprings simultaneously
- The **offspring distribution** is described by a random variable ξ taking non-negative integer values with corresponding probabilities

$$p_k = \mathbb{P}[\xi = k], \quad k \geq 0.$$

- All individuals behave **independently** of each other

The Galton-Watson branching process

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 ξ_1, ξ_2, \dots are i.i.d. copies of ξ .

- The process $\{Z_n, n \geq 0\}$ is a **discrete-time Markov chain** on the state space $\{0, 1, 2, 3, \dots\}$ where state 0 is **absorbing** and all other states are **transient**.

Population size distribution

The **probability generating function** (p.g.f.) of ξ 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 ,

$$P_n(s) := \sum_{k=0}^{\infty} \mathbb{P}[Z_n = k]s^k.$$

Composition of probability generating functions

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 ξ_1, ξ_2, \dots are i.i.d. copies of ξ .

Proposition

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

$$\begin{aligned} P_n(s) &= P_{n-1}(P(s)) \\ &= \underbrace{P(P(\dots P(s)))}_n \\ &= P(P_{n-1}(s)), \quad n \geq 1, \end{aligned}$$

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

Examples of progeny distributions

- **Binary** case :

$$\xi = \begin{cases} 0 & \text{w.p. } 1/3, \\ 2 & \text{w.p. } 2/3 \end{cases}$$

$$\rightarrow P(s) = \sum_k \mathbb{P}[\xi = k]s^k = \frac{1}{3} + \frac{2}{3}s^2$$

- **Geometric** case : $\xi \sim \text{Geom}(p)$,

$$\mathbb{P}[\xi = k] = (1-p)^k p \quad \rightarrow \quad P(s) = \frac{p}{1 - (1-p)s}$$

- **Poisson** case : $\xi \sim \text{Poisson}(\lambda)$,

$$\mathbb{P}[\xi = k] = e^{-\lambda} \lambda^k / k! \quad \rightarrow \quad P(s) = e^{\lambda(s-1)}$$

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, $M_n \searrow 0$ as $n \rightarrow \infty$
- $m = 1$: **critical** case $M_n = 1$ for all n
- $m > 1$: **supercritical** case, $M_n \nearrow \infty$ as $n \rightarrow \infty$

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

$$P_n(s) = P(P_{n-1}(s)) \quad \Rightarrow \quad q_n = P(q_{n-1}), \quad n \geq 1.$$

- The probability of extinction of the branching process is

$$q = \lim_{n \rightarrow \infty} \mathbb{P}[Z_n = 0] = \mathbb{P}[\lim_{n \rightarrow \infty} Z_n = 0] \quad \Rightarrow \quad q = P(q)$$

Theorem

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

$$s = P(s).$$

Computation of q

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 : $s = \frac{p}{1 - (1-p)s} \rightarrow q = \min(p/(1-p), 1)$,
i.e.

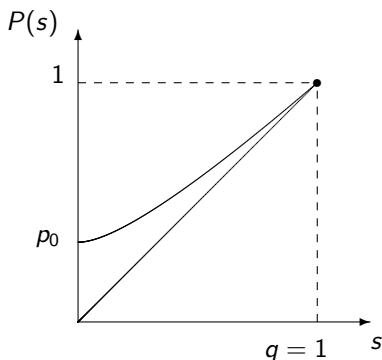
$$q = \begin{cases} 1 & \text{if } p \geq 1/2 \\ p/(1-p) & \text{if } p < 1/2 \end{cases}$$

- Poisson case : $s = e^{\lambda(s-1)} \rightarrow$ the functional iteration is useful here!

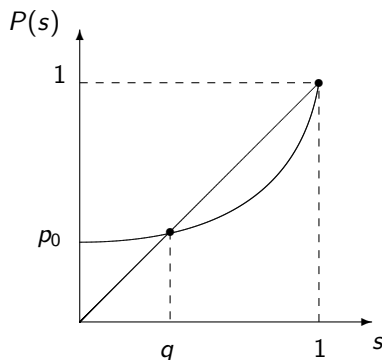
A closer look at the progeny generating function

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

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



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



$$m = P'(1) > 1$$

→ $P(s)$ 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, $m > 1 \Rightarrow q < 1$
- in the **critical** case, $m = 1 \Rightarrow q = 1$
- in the **subcritical** case, $m < 1 \Rightarrow q = 1$

Dichotomy : extinction or explosion !

Theorem

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

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

Moreover,

$$\underbrace{\mathbb{P}[Z_n \rightarrow 0]}_q + \underbrace{\mathbb{P}[Z_n \rightarrow \infty]}_{1-q} = 1.$$

In the supercritical case, conditionally on non-extinction,
 $\lim_{n \rightarrow \infty} Z_n = +\infty$ a.s.

Can we tell more about the growth rate of Z_n ?

Limiting behaviour in the supercritical case $1 < m < \infty$

Let $W_n = Z_n/m^n$. $\{W_n, n \geq 0\}$ is a martingale with $\mathbb{E}[W_n] = 1$ for all n , so it converges a.s. to a nonnegative random variable W ,

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

Theorem (Kesten-Stigum)

Either $\mathbb{P}[W = 0] = q$ or $\mathbb{P}[W = 0] = 1$.

The following are equivalent :

- (i) $\mathbb{E}[\xi \log^+ \xi] < \infty$
- (ii) $\mathbb{P}[W = 0] = q$
- (iii) W_n converges in mean (L^1)
- (iv) $\mathbb{E}[W] = 1$.

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

Theorem (Yaglom)

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

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

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 $\mathbf{b} = (b_1, b_2, \dots)$ 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

$$\lim_{n \rightarrow \infty} n \mathbb{P}[Z_n \neq 0] = \frac{2}{\sigma^2}.$$

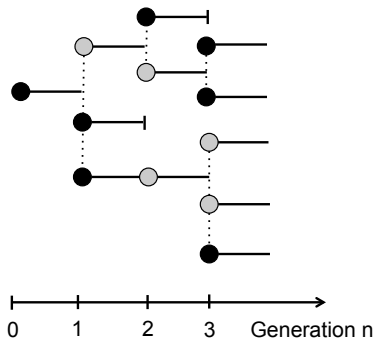
- Yaglom's universal limit law

$$\lim_{n \rightarrow \infty} \mathbb{P}[Z_n/n \geq x \mid Z_n \neq 0] = \exp(-2x/\sigma^2), \quad x > 0.$$

Multitype Galton-Watson branching process

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

Example with $r = 2$:



Multitype Galton-Watson branching process

Population size vector : $\mathbf{Z}_n = (Z_{n1}, Z_{n2}, \dots, Z_{nr})$, $n \in \mathbb{N}$, where

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

$\{\mathbf{Z}_n\}$ is an r -dimensional Markov process with state space \mathbb{N}^r and an absorbing state $\mathbf{0} = (0, 0, \dots, 0)^\top$

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

p_{ij} = probability that a type i gives birth to j_1 children of type 1, j_2 children of type 2, \dots , j_r children of type r .

Multitype Galton-Watson branching processes

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

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

Mean progeny matrix M with elements

$$M_{ij} = \left. \frac{\partial P_i(\mathbf{s})}{\partial s_j} \right|_{\mathbf{s}=\mathbf{1}}$$

= expected number of direct offsprings of type j
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 $\mathbf{q} = (q_1, q_2, \dots, q_r)^\top$, with

$$q_i = \mathbb{P} \left[\lim_{n \rightarrow \infty} |\mathbf{Z}_n| = 0 \mid \varphi_0 = i \right]$$

$$\mathbf{q} = \mathbb{P} \left[\lim_{n \rightarrow \infty} |\mathbf{Z}_n| = 0 \mid \varphi_0 \right],$$

where φ_0 is the type of the first individual in generation 0.

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

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

As before, $\mathbf{q} = \lim_{n \rightarrow \infty} \mathbf{q}_n$ 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

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

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

If $\rho > 1$, then $\mathbf{q} < \mathbf{1}$ (supercritical case)

If $\rho = 1$, then $\mathbf{q} = \mathbf{1}$ (critical case)

If $\rho < 1$, then $\mathbf{q} = \mathbf{1}$ (subcritical case)

Stable type distribution (irreducible case)

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

Then, in the long run, we can show that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{ni}]}{\mathbb{E}[|\mathbf{Z}_n|]} = u_i.$$

The vector \mathbf{u} is called the **stable type distribution**.

Infinitely many types

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. !

- 1 Introduction
- 2 Galton-Watson branching processes
- 3 Markovian branching processes**
- 4 Markovian binary trees

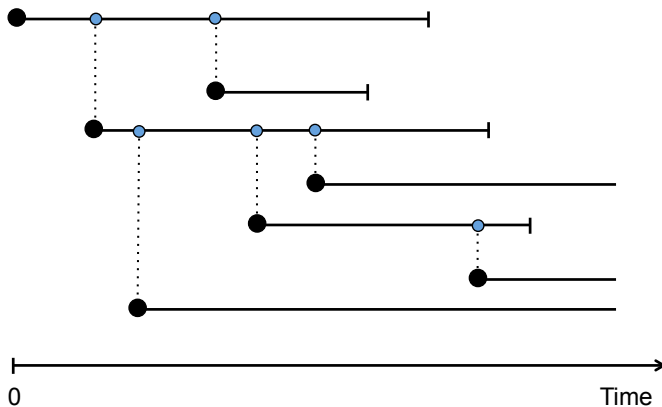
Markovian branching processes

In a Markovian branching process,

- the lifetime of an individual is **exponentially** distributed with parameter μ ;
- during its lifetime, the individual reproduces according to a **Poisson process** with rate λ , 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 $\{0, 1, 2, 3, \dots\}$, where state 0 is **absorbing**, and with generator

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

- 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 $\lim_{t \rightarrow \infty} M(t) = \infty$ (supercritical case),
- if $\lambda = \mu$, then $M(t) = 1$ for all t (critical case),
- if $\lambda < \mu$, then $\lim_{t \rightarrow \infty} M(t) = 0$ (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

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

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

or equivalently

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

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,

$$q = 1 \Leftrightarrow m = P'(1) \leq 1.$$

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 $q = 1 \Leftrightarrow \lambda \leq \mu$.

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 λ_s ;
- the collection of **genera** evolve according to a **pure birth process** with birth rate λ_g .

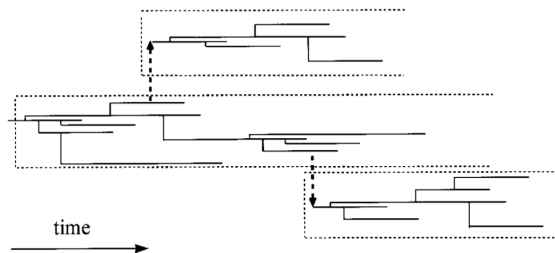
At each ramification, a new genus evolves which has exactly **one species**, and the old genus continues unchanged

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 :



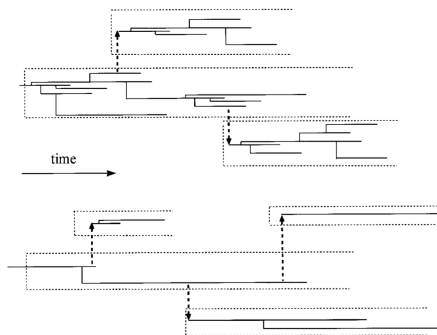
Branching of species is represented by *continuous lines*. Boundaries of genera are represented by *dotted line 'tubes'*, and branching of genera by *arrows*.

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 λ_g/λ_s** ?



Two sample paths of the Yule's process : low and high value of the λ_g/λ_s ratio.

Example of multitype Markovian branching process

Yule's evolutionary process

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).

Example of multitype Markovian branching process

Yule's evolutionary process

Using results on multitype Markovian branching processes, we obtain that the **asymptotic expected proportion of monotypic genera** is equal to

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

→ λ_s/λ_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 !

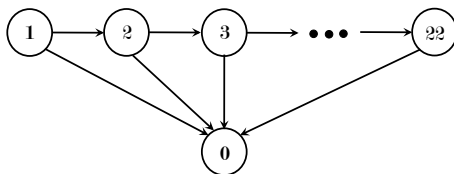
We introduce a type of branching process, called the *Markovian binary tree* (MBT), which is at the same time very **general** and **tractable**.

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,

⋮

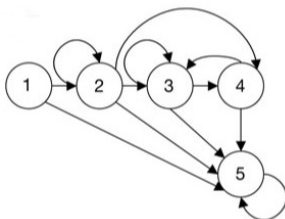
22 = over 100 years,

0 = dead

Phase-structured lifetime

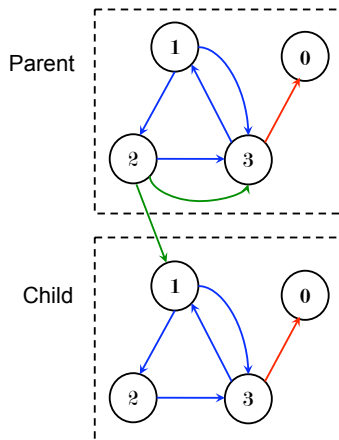
Example 2

The **right whale's** life cycle decomposed into **stages**



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

Phase-structured lifetime

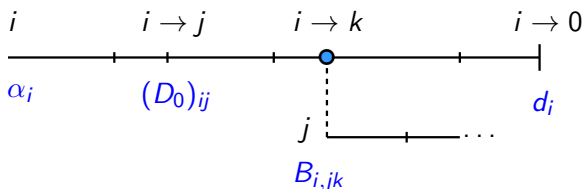


3 types of transitions:

- "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;



- α : **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).

MBT are generalisations of Markovian branching processes

- 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**.

The MBT as a model for the macroevolutionary process

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

- $\mathbf{Z}(t) = (Z_1(t), \dots, Z_n(t))$ the pop. size vector at time t , where $Z_i(t)$: # of individuals in phase i at time t
- $F_i(\mathbf{s}, t)$: the conditional p.g.f. of the pop. size at time t , given that the MBT starts with a first individual in phase i

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

where $\mathbf{s} = (s_1, \dots, s_n)^\top$, $|s_i| \leq 1$ and $\mathbf{s}^{\mathbf{k}} = s_1^{k_1} \cdots s_n^{k_n}$.

- $\mathbf{F}(\mathbf{s}, t) = (F_1(\mathbf{s}, t), \dots, F_n(\mathbf{s}, t))^\top$

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

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

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

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

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

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

→ $M(t) = (M_{ij}(t))$ with

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

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

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

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

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

- ω = P-F eigenvalue of Ω , 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 ,

$$\rightarrow \mathbf{F}(t) = (F_1(t), \dots, F_n(t))^T = \mathbf{F}(\mathbf{0}, t)$$

Theorem

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

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

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

As for $\mathbf{F}(\mathbf{s}, t)$, 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

→ $\mathbf{q} = (q_1, \dots, q_n)^T = \lim_{t \rightarrow \infty} \mathbf{F}(t)$ is the minimal non-negative solution of

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

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

$$\rightarrow \mathbf{s} = \boldsymbol{\theta} + \Psi(\mathbf{s} \otimes \mathbf{s}),$$

or equivalently

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

Extinction probability

$$\begin{aligned}
 \mathbf{s} &= \boldsymbol{\theta} + \Psi(\mathbf{s} \otimes \mathbf{s}) \\
 &\equiv \\
 \mathbf{s} &= [I - \Psi(I \otimes \mathbf{s})]^{-1} \boldsymbol{\theta} \\
 &\equiv \\
 \mathbf{s} &= [I - \Psi(\mathbf{s} \otimes I)]^{-1} \boldsymbol{\theta}
 \end{aligned}$$

- 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

$$\mathbf{s}_0 = \boldsymbol{\theta}$$

$$\mathbf{s}_k = \boldsymbol{\theta} + \Psi(\mathbf{s}_{k-1} \otimes \mathbf{s}_{k-1}), \quad k \geq 1$$

For $k \geq 0$,

- \mathcal{M}_k = the set of MBTs considered at stage k
- $\mathbf{s}_k = \mathbb{P}[\mathcal{M}_k | \varphi_0]$ = the k th approximation of \mathbf{q} .

$$\mathcal{M}_0 = \perp$$

$$\mathcal{M}_k = \perp \cup \begin{array}{c} | \\ \hline \mathcal{M}_{k-1} \quad \mathcal{M}_{k-1} \end{array} \quad \text{for } k \geq 1.$$

Probabilistic interpretation of the Depth algorithm

$$\mathbf{s}_0 = \boldsymbol{\theta}$$

$$\mathbf{s}_k = \boldsymbol{\theta} + \Psi(\mathbf{s}_{k-1} \otimes \mathbf{s}_{k-1}), \quad k \geq 1$$

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)
- $\mathcal{M}_k \subseteq \mathcal{M}_{k+1} \subseteq \dots \subseteq \mathcal{M}$ = the set of all extinct MBTs.
- $\mathbf{s}_k = \mathbb{P}[\mathcal{M}_k \mid \varphi_0] \nearrow \mathbf{q}$ as $k \rightarrow \infty$.

The Order algorithm

$$\mathbf{s}_0 = \boldsymbol{\theta}$$

$$\mathbf{s}_k = [I - \Psi(\mathbf{s}_{k-1} \otimes I)]^{-1} \boldsymbol{\theta}, \quad k \geq 1$$

For $k \geq 0$,

- \mathcal{M}_k = the set of MBTs considered at stage k
- $\mathbf{s}_k = \mathbb{P}[\mathcal{M}_k \mid \varphi_0]$ = the k th approximation of \mathbf{q} .

$$\mathcal{M}_0 = \perp$$

$$\mathcal{M}_k = \begin{array}{c} \text{---} \\ | \\ \text{---} \\ / \quad \backslash \\ \mathcal{M}_{k-1} \quad \text{---} \\ | \quad \backslash \\ \mathcal{M}_{k-1} \quad \text{---} \\ | \quad \backslash \\ \mathcal{M}_{k-1} \quad \dots \end{array} \quad \text{for } k \geq 1.$$

Probabilistic interpretation of the Order algorithm

$$\begin{aligned} \mathbf{s}_0 &= \boldsymbol{\theta} \\ \mathbf{s}_k &= [I - \Psi(\mathbf{s}_{k-1} \otimes I)]^{-1} \boldsymbol{\theta}, \quad k \geq 1 \end{aligned}$$

Order of an MBT = total number of **children generations**

For $k \geq 0$,

- \mathcal{M}_k = the set of **extinct MBTs** with an **order $\leq k$** (**constraint** on the **shape** of the tree)
- $\mathcal{M}_k \subseteq \mathcal{M}_{k+1} \subseteq \dots \subseteq \mathcal{M}$ = the set of all extinct MBTs.
- $\mathbf{s}_k = \mathbb{P}[\mathcal{M}_k \mid \varphi_0] \nearrow \mathbf{q}$ as $k \rightarrow \infty$.

Application of MBTs in demography

Branching processes have numerous applications in **population biology** and in **demography**.

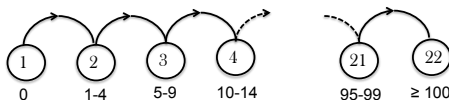
Here, we use the MBT to model the dynamics of **female families** in several countries.

Since our model is **asexual**, we look at **women** only.

Application of MBTs in demography

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 \rightarrow **22** age classes or **phases**.

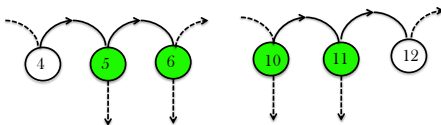


We assume **exponential** sojourn time in each phase (time unit = **1 year**).

$$D_0 = \begin{bmatrix} * & 1 & & & & \\ & * & 1/4 & & & \\ & & * & 1/5 & & \\ & & & \ddots & & \\ & & & & * & 1/5 \\ & & & & & * \end{bmatrix}$$

Application of MBTs in demography

Birth rates



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 .

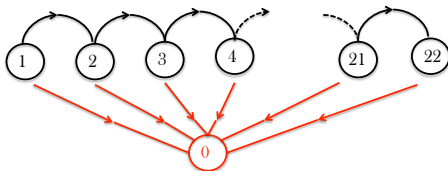
→ **Female birth** rate β_i per individual in phase i

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

$$\beta = [\mathbf{0}_4, \beta_5, \dots, \beta_{11}, \mathbf{0}_{11}]', \quad \text{and} \quad B = (\mathbf{e}_1 \otimes \text{diag}(\beta))$$

Application of MBTs in demography

Death rates



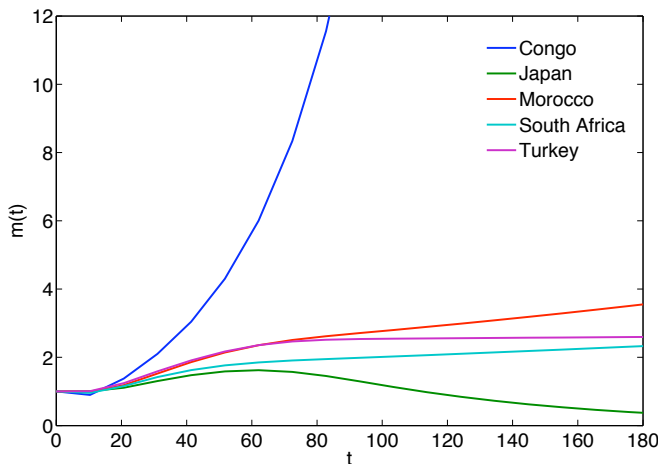
Age-specific **mortality** rate d_i in age class i

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

$$\mathbf{d} = [d_1, d_2, \dots, d_{22}]'$$

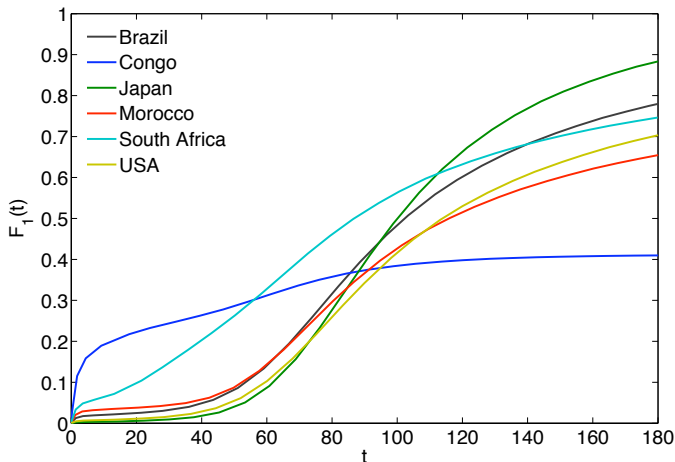
Application of MBTs in demography

$m(t) = (M(t)\mathbf{1})_1$: Mean total family size generated by a new-born woman after t units of time :



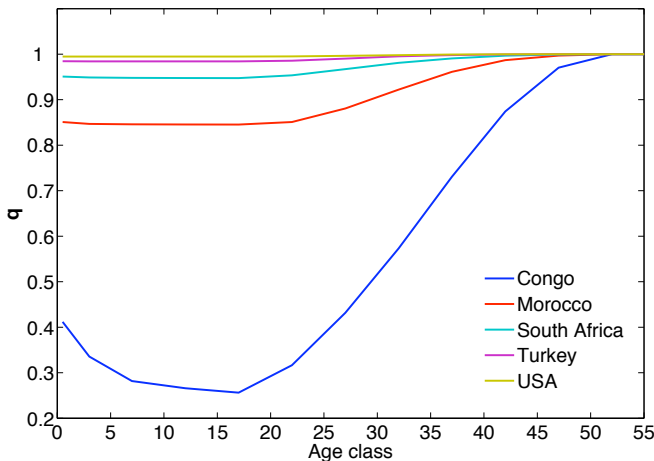
Application of MBTs in demography

$F_1(t)$: Distribution function of the time until extinction of the family generated by one new-born woman



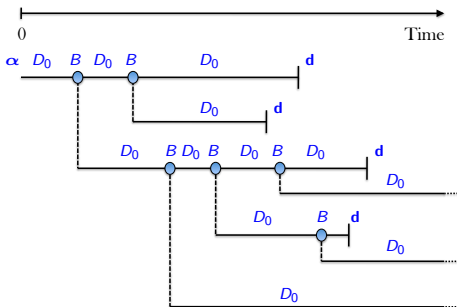
Application of MBT_{st} in demography

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 $(\alpha, D_0, B, \mathbf{d})$.



Complete observation : All birth and death events are observable

BUT phase transitions of the underlying Markov chain are **not** observable.

Parameter estimation under complete observation

Observed variables

Aim : to estimate the parameters $\theta = (\alpha, D_0, B, \mathbf{d})$ 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 : $\{L_k\}$.

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

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

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

Parameter estimation under complete observation

Hidden variables

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

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

- H_i : # MBTs starting in phase i ;
- Z_i : total cumulated time spent in phase i ;
- A_{ij} : total # jumps from phase i to phase j without accompanying observable event, for $i \neq j$;
- $C_{i,j\ell}$: total # jumps from phase i to phase ℓ associated with the birth of a child in phase j ;
- D_i : # terminating jumps from phase i

Parameter estimation under complete observation

Likelihood of a complete sample

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

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

Complete observation

The maximum likelihood estimator

The MLE is

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \log L_c(\theta | \mathbf{X}),$$

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

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

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

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

They are expressed in terms of the **hidden** variables \mathbf{Z} .

Complete observation

The EM algorithm applied to MBTs (Hautphenne and Fackrell, 2014).

Since $\mathbf{Z} = \{(H_i, Z_i, A_{ij}, C_{i,jl}, D_i)\}$ are **hidden** random variables, instead of computing

$$\hat{\theta} = \operatorname{argmax}_{\theta} \log L_c(\theta | \mathbf{X} = (\mathbf{Y}, \mathbf{Z})),$$

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

$$\hat{\theta}^{(n)} = \operatorname{argmax}_{\theta} \mathbb{E}_{\mathbf{Z}} \left[\log L_c(\theta | \mathbf{X}) | \mathbf{Y} = \mathbf{y}, \hat{\theta}^{(n-1)} \right],$$

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

→ 2 steps :

1. E-step : $\mathbb{E}_{\mathbf{Z}} \left[\log L_c(\theta | \mathbf{X}) | \mathbf{y}, \hat{\theta}^{(n-1)} \right] \propto \mathbb{E} \left[\mathbf{Z} | \mathbf{y}, \hat{\theta}^{(n-1)} \right],$
2. M-step : $\hat{\theta}^{(n)} = \operatorname{argmax}_{\theta} \mathbb{E}_{\mathbf{Z}} \left[\log L_c(\theta | \mathbf{X}) | \mathbf{y}, \hat{\theta}^{(n-1)} \right].$

Application in phylogenetics

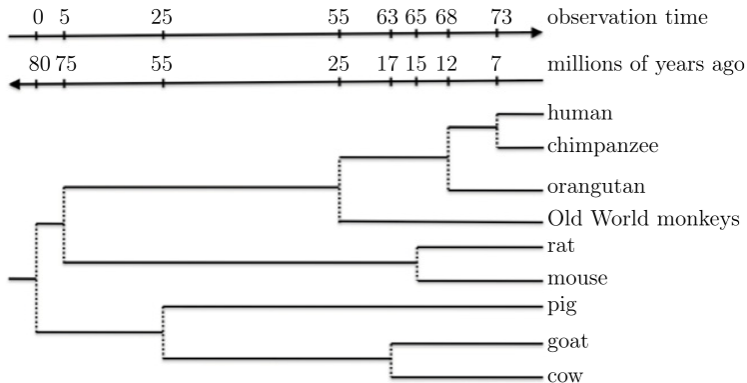


Fig. 9. A phylogenetic tree showing the *DNA* mutations among a group of mammals.

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 :

n	$\hat{\omega}$	Loglikelihood
1	0.0256	-32.64
2	0.0211	-30.66
3	0.0268	-29.12
4	0.0281	-28.14

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.

Springer, Berlin, 1963.



C.J. Mode.

Multitype Branching Processes.

Elsevier, New York, 1971.



K.B. Athreya and P.E. Ney.

Branching Processes.

Springer, Berlin, 1972.



P. Jagers.

Branching Processes with Biological Applications.

Wiley, London, 1975.

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.
Springer, New York, 1997.
-  M. Kimmel and D.E. Axelrod.
Branching Processes in Biology.
Springer, New York, 2002.
-  P. Haccou, P. Jagers, V.A. Vatutin (eds.).
Branching Processes : Variation, Growth and Extinction of Populations.
Cambridge University Press, Cambridge, 2005.

... on phylogenetics...



J.H. Gillespie and C.H. Langley.

Are evolutionary rates really variable?

Journal of Molecular Evolution, 13(1) :27–34, 1979.






J.H. Gillespie.

Natural Selection and the Molecular Clock.

Molecular Biology and Evolution, 3(2) :138–155, 1986.

... and on Markovian binary trees I

-  N. Bean, N. Kontoleon, and P. Taylor.
Markovian trees : Properties and algorithms.
Annals of Operations Research, 160(1) :31–50, 2008.
-  S. Hautphenne, G. Latouche, and M.-A. Remiche.
Newton's iteration for the extinction probability of a
Markovian Binary Tree.
Linear Algebra and its Applications, 428(11-12) :2791–2804,
2008.
-  S. Hautphenne, G. Latouche, and M.-A. Remiche.
Transient features for Markovian binary trees.
*Proceedings of the Fourth International ICST Conference on
Performance Evaluation Methodologies and Tools*, Article No.
18, 2009.

... and on Markovian binary trees II



S. Hautphenne, G. Latouche, and M.-A. Remiche.

Algorithmic approach to the extinction probability of branching processes.

Methodology and Computing in Applied Probability,
13(1) :171–192, 2011.



S. Hautphenne and G. Latouche.

The Markovian binary tree applied to demography.

Journal of Mathematical Biology, 64(7) :1109–35, 2012.



S. Hautphenne and M. Fackrell.

An EM algorithm for the model fitting of Markovian binary trees.

Computational Statistics and Data Analysis, 70 :19–34, 2014.