APM 541: Stochastic Modelling in Biology
Branching Processes

Jay Taylor
Fall 2013
Branching Processes

Branching processes include a variety of stochastic models of population growth which share the following properties in common:

- Individuals reproduce independently of one another; in particular, the numbers of offspring produced by individuals are independent random variables.
- Usually, the offspring numbers are identically-distributed and do not depend on the population size. Thus population growth is **density-independent**.
- They are Markov processes, but reproduction can either occur at discrete time steps or continuously in time.
- Branching processes are commonly used to model biological invasions: by exotic species, infectious agents, cancer cells or new alleles. These applications are usually justified by arguing that density dependence is weak in newly founded populations.
The first branching processes to be introduced were discrete-time Markov chains now known as **Galton-Watson** processes, after Francis Galton and Henry William Watson.

### Definition

Let \( \left( \xi_i^{(n)}; i, n \geq 0 \right) \) be a collection of independent and identically-distributed random variables with values in the non-negative integers and define a process \( X = (X_n : n \geq 0) \) by setting \( X_0 = k \) and

\[
X_{n+1} = \sum_{i=1}^{X_n} \xi_i^{(n)}.
\]

Then \( X \) is said to be a **Galton-Watson process** with offspring distribution \( \xi \) and initial size \( k \). Also, \( X \) is said to be **subcritical**, **critical**, or **supercritical** depending on whether the mean number of offspring produced by an individual \( \mu = \mathbb{E}[\xi] \) is less than, equal to, or greater than 1, respectively.

**Biological interpretation:** \( X_n \) denotes the population size in generation \( n \) while \( \xi_i^{(n)} \) is the number of **surviving** offspring or lineages contributed by the \( i \)'th individual to the next generation.
Although Galton-Watson processes are discrete-time Markov chains, the key to their analysis is not the transition matrix but the fact that these processes satisfy the branching property. This is described in the next theorem.

**Theorem**

Suppose that $X^{(1)}$ and $X^{(2)}$ are independent Galton-Watson processes with the same offspring distribution and initial sizes equal to $k_1$ and $k_2$, respectively. Then, the process $X = (X_n : n \geq 0)$ defined by setting

$$X_n = X_n^{(1)} + X_n^{(2)}$$

for each $n \geq 0$, is also a Galton-Watson process with offspring distribution $\xi$ and initial size $k_1 + k_2$. This property is called the branching property.

One consequence of the branching property is that a Galton-Watson process with initial size $k$ has the same distribution as the sum of $k$ independent Galton-Watson processes which each have initial size 1. Another consequence is that the expected population size either grows or shrinks geometrically.
Theorem

Let $X$ be a Galton-Watson process with offspring distribution $\xi$ and initial size $X_0 = 1$ and suppose that the average number of offspring born to each individual is finite:

$$\mu = \mathbb{E}[\xi] < \infty.$$ 

Then the expected size of the population in generation $n$ is

$$\mathbb{E}[X_n] = \mu^n$$

Proof: We can prove the result by induction on $n$. First observe that $X_0 = 1 = \mu^0$ holds by assumption. For the general case, assume that $\mathbb{E}[X_n] = \mu^n$. Then, by conditioning on the value of $X_n$ and using the fact that $X_n$ is independent of the variables $\xi^{(n)}_1, \xi^{(n)}_2, \cdots$, we have

$$\mathbb{E}[X_{n+1}] = \mathbb{E}\left[\sum_{i=1}^{X_n} \xi^{(n)}_i\right] = \mathbb{E}\left\{\mathbb{E}\left[\sum_{i=1}^{X_n} \xi^{(n)}_i \mid X_n\right]\right\} = \mathbb{E}[X_n \cdot \mu] = \mu^{n+1}.$$ 

$\square$
Depending on whether \( \mu < 1 \) or \( \mu > 1 \), the expected population size will either tend to 0 or infinity. In fact, a much stronger dichotomy is true. Define the **extinction time** \( T_{\text{ext}} \) of the Galton-Watson process \( X \) to be the first time that the population size is 0:

\[
T_{\text{ext}} = \inf \{ t \geq 0 : X_t = 0 \},
\]

where \( T_{\text{ext}} = \infty \) means that the process never goes extinct.

**Theorem**

Let \( X \) be a Galton-Watson process with offspring distribution \( \xi \) and initial size \( X_0 \geq 1 \), and let \( \mu \) be the mean number of offspring born to each individual.

1. If \( \mu \leq 1 \) and \( \text{Var}(\xi) > 0 \), then \( P( T_{\text{ext}} < \infty ) = 1 \).
2. If \( \mu = 1 \) and \( \text{Var}(\xi) = 0 \), then \( T_{\text{ext}} = \infty \) and \( X_n = X_0 \) for all \( n \geq 1 \).
3. If \( \mu > 1 \), then \( P( T_{\text{ext}} < \infty ) < 1 \). Furthermore, on the event \( \{ T_{\text{ext}} = \infty \} \), the population size grows unboundedly:

\[
P( X_n \to \infty \mid T_{\text{ext}} = \infty ) = 1.
\]
It follows from this theorem that the range of behaviors that can be exhibited by Galton-Watson processes with random offspring numbers is quite limited:

- Every critical and subcritical branching process will suffer extinction in finite time, and the population size will tend to decrease geometrically on the way to extinction.
- Every supercritical branching process will have a positive probability of persisting forever; however, in this case the population size will grow without bound with a geometric rate of increase.
- In either case, no positive state $n$ will be visited more than finitely many times. For this reason, Galton-Watson processes are said to be transient.

The transient behavior of branching processes is a consequence of the assumption that population growth is density-independent. This limits the scope of applications for Galton-Watson processes, since no population can continue growing without eventually exhausting the resources needed to support growth. Nonetheless, Galton-Watson processes and other branching processes can still be used to model the early stages of population growth under conditions in which density-dependence is initially very weak.
Simulations of a Galton-Watson Process with Poisson Offspring Distribution

Dashed black line shows the expected number of offspring in each generation. Of the 20 simulations, 11 suffered extinction in the first 10 generations.
One of the most important questions in studies of biological invasions is whether the invasion will succeed or fizzle out. Branching processes are often used to estimate the probability of invasion by equating invasion with persistence of the branching process. This leads us to study the **extinction probability** of branching processes, which we denote

\[ \rho = \mathbb{P}(T_{\text{ext}} < \infty | X_0 = 1). \]

Of course, if the process is critical or subcritical, then \( \rho = 1 \). Thus interest focuses on the supercritical case, where \( \rho < 1 \). This is an absorption probability which could be studied using the Markov chain tools introduced previously, but we can learn more by exploiting the branching property.
One tool that can be used to study branching processes and other stochastic processes with values in the natural numbers are **probability generating functions**.

**Definition**

Let $X$ be a random variable with values in the natural numbers $\mathbb{N} = \{0, 1, \cdots \}$. The **probability generating function** of $X$ is the function $\psi()$ defined by

$$
\psi_X(t) = \mathbb{E} \left[ t^X \right] = \sum_{n=0}^{\infty} \mathbb{P}(X = n) \cdot t^n
$$

for those values of $t \in \mathbb{R}$ such that the series on the right-hand side converges. The set of all such $t$ is called the **domain** of $\psi_X$.

**Remarks**: The probability generating function is an alternative way of encoding information about the distribution of a random variable. Furthermore, the distribution of a random variable is uniquely determined by the probability generating function of the variable.
Example: If $X \sim \text{Poisson}(\lambda)$, then the probability generating function of $X$ is

\[
\psi_X(t) = \mathbb{E}
\left[
 t^X
\right] =
\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} t^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} = e^{-\lambda} e^{\lambda t} = e^{-\lambda(1-t)},
\]

and the domain of $\psi_X$ is the entire real line.
Example: If $X \sim \text{Geometric}(p)$, then the probability generating function of $X$ is

$$
\psi_X(t) = \mathbb{E}[t^X] \\
= \sum_{n=0}^{\infty} p(1-p)^n t^n \\
= p \sum_{n=0}^{\infty} (t(1-p))^n \\
= \frac{p}{1 - t(1-p)}
$$

and the domain of $\psi_X$ is the interval $(-(1-p)^{-1}, (1-p)^{-1})$. 
Theorem

Let $X$ be a Galton-Watson process with offspring distribution $\xi$ and initial population size 1, and let $\psi$ be the probability generating function of $\xi$. Then the probability generating function $\psi_n$ of $X_n$ is the $n$-fold iterate of $\psi$, i.e., $\psi_0(s) = 1$, $\psi_1(s) = \psi(s)$, $\psi_2(s) = \psi(\psi(s))$, and, in general, $\psi_{n+1}(s) = \psi_n(\psi(s))$.

Proof: The result follows by induction on $n$. If $X_0 = 1$, then $X_1 = \xi^{(0)}_1$ has the same distribution as $\xi$ and therefore the two variables have the same probability generating function: $\psi_1(s) = \psi(s)$. For the general case, notice that

$$
\psi_{n+1}(s) = \mathbb{E}\left[s^{X_{n+1}}\right] = \mathbb{E}\left[s^{\sum_{i=1}^{X_n} \xi_i^{(n)}}\right] = \mathbb{E}\left[\prod_{i=1}^{X_n} s^{\xi_i^{(n)}}\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{X_n} s^{\xi_i^{(n)}} \mid X_n\right]\right] = \mathbb{E}\left[\prod_{i=1}^{X_n} \mathbb{E}\left[s^{\xi_i^{(n)}} \mid X_n\right]\right] = \mathbb{E}\left[\psi(s)^{X_n}\right] = \psi_n(\psi(s)).
$$
In particular, since $T_{\text{ext}} \leq n$ holds if and only if $X_n = 0$, it follows that the distribution of the extinction time can also be obtained from $\psi$:

$$
P\{T_{\text{ext}} \leq n\} = P(X_n = 0) = \psi_n(0).
$$

Even more usefully, the extinction probability $\rho$ can be calculated using $\psi$:

$$
\rho = \mathbb{P}\{T_{\text{ext}} < \infty\} = \lim_{n \to \infty} P\{T_{\text{ext}} \leq n\} = \lim_{n \to \infty} \psi_n(0).
$$

Since $\psi$ is a continuous function on $[0, 1]$, it follows that

$$
\psi(\rho) = \psi\left(\lim_{n \to \infty} \psi_n(0)\right)
= \lim_{n \to \infty} \psi(\psi_n(0))
= \lim_{n \to \infty} \psi_{n+1}(0)
= \rho,
$$

which shows that $\rho$ is a fixed point of $\psi$. In fact, it can be shown that $\rho$ is the smallest non-negative fixed-point of $\psi$. 
Example: Let $X$ be a Galton-Watson process with binary offspring distribution $\xi$, where $\mathbb{P}(\xi = 2) = p$ and $\mathbb{P}(\xi = 0) = 1 - p$. Thus, each individual either gives birth to two offspring or dies childless. The probability generating function for $\xi$ is

$$
\psi(s) = (1 - p) + ps^2,
$$

which has two fixed points:

$$
\rho = 1 \quad \text{and} \quad \rho = \frac{1 - p}{p}
$$

When $p \leq 1/2$, $\rho = 1$ is the smaller fixed point and so the process is certain to go extinct in finite time; of course, we could have deduced this directly from the fact that the mean number of offspring $\mu = 2p$ is less than or equal to 1 when $p \leq 1/2$. On the other hand, if $p > 1/2$, then the extinction probability is $\rho = (1 - p)/p$ which is strictly less than 1.
Suppose that \( X \) is a Galton-Watson process with offspring distribution \( \xi \) and let \( \rho \) be the extinction probability of \( X \) when \( X_0 = 1 \). If instead \( X_0 = k \), then we can decompose \( X \) into a sum of \( k \) independent Galton-Watson processes \( X^{(1)}, \cdots, X^{(k)} \) that have the same offspring distribution but which each begin with a single individual,

\[
X_n = X^{(1)}_n + \cdots + X^{(k)}_n.
\]

Clearly, \( X \) will go extinct in finite time if and only if each of the \( k \) independent processes \( X^{(1)}, \cdots, X^{(k)} \) goes extinct in finite time. Since these processes are independent and since the extinction probability of each of these processes is just \( \rho \), it follows that the extinction probability of \( X \) is just

\[
P \left( T_{\text{ext}} < \infty \mid X_0 = k \right) = \rho^k.
\]

Thus the extinction probability of a supercritical Galton-Watson process is a geometrically decreasing function of the initial population size.
Example: Branching processes have also been used to estimate the fixation probability of selectively favorable alleles in finite populations. We begin by formulating a version of the Wright-Fisher model with selection. Let us assume the following:

- The population size is constant, with \( N \) haploid individuals.
- Generations are non-overlapping.
- The population contains two alleles, \( A_1 \) and \( A_2 \), with relative fitnesses \( 1 + s \) and 1.
- The parents of the \( N \) individuals alive in generation \( n + 1 \) are chosen at random and with replacement, but each \( A \)-type individual is \((1 + s)\)-times more likely to be chosen than an \( a \)-type individual.

The parameter \( s \) is called the selection coefficient of \( A_1 \). If \( s = 0 \), then the two alleles are selectively neutral and this model reduces to the Wright-Fisher model that we studied previously. However, if \( s > 0 \), then \( A_1 \) has a selective advantage over \( A_2 \), which is incorporated into this model as a form of fecundity selection, i.e., \( A_1 \) individuals are likely to have greater numbers of surviving offspring than \( A_2 \) individuals.
If $X_n$ denotes the number of copies of $A$ present in the $n$'th generation, then $X = (X_n : n \geq 0)$ is again a discrete-time Markov chain with transition probabilities

$$p_{ij} = \mathbb{P}(X_{n+1} = j|X_n = i) = \binom{n}{j} \tilde{p}_i (1 - \tilde{p}_j)^{N-j},$$

where

$$\tilde{p}_i = \frac{p_i \cdot (1 + s)}{p_i \cdot (1 + s) + (1 - p_i)} = p_i \left( \frac{1 + s}{1 + p_i s} \right) = p_i \left( \frac{w_1}{\bar{w}} \right)$$

and $p_i = i/N$. The quantity $\tilde{p}_i$ is the probability that an $A$-type individual is chosen as a parent, while $\bar{w}$ is the mean fitness of the population.

If $p_t = X_t/N$ denotes the frequency of $A_1$ at time $t$, then the process $(p_t : t \geq 0)$ is also a Markov chain, with the same transition matrix (but relabeled states) as $X$. Since we have not incorporated mutation into this model, it is clear that $p = 0$ and $p = 1$ are absorbing states for this process. Furthermore, fixation of the allele $A_1$ corresponds to absorption by the state $p = 1$. 
Our goal is to calculate the fixation probability of the beneficial allele as a function of its initial frequency:

\[ u(p) = \mathbb{P}\{p_t = 1 \text{ for some } t \geq 0 | p_0 = p \}. \]

We previously showed that if \( s = 0 \), then \( u(p) = p \), i.e., the fixation probability of a neutral allele is equal to its initial frequency. This calculation relied on the observation that the expected frequency of \( A_1 \) is constant in a neutral model. However, this is no longer true when selection is incorporated into the model since

\[ \mathbb{E}[p_1 | p_0 = p] = p \left( \frac{1 + s}{1 + ps} \right) > p \]

whenever \( s > 0 \). Although the Markov property can still be used to derive a system of linear equations satisfied by the fixation probabilities, the complicated structure of the transition matrix makes it difficult to solve these equations. Instead, we will use an alternative approach proposed by J. B. S. Haldane in 1927 which relies on a branching process approximation.
Haldane’s approach relies on the following observations:

- If \( \eta \) denotes the number of surviving offspring contributed by a single \( A_1 \)-type individual, then \( \eta \) is binomially distributed with parameters \( N \) and 
  \[
  \frac{1}{N} \left( \frac{1 + s}{1 + ps} \right).
  \]
- Provided that \( s \) and \( p \) are small and \( N \) is large, \( \eta \) is approximately Poisson distributed with mean \((1 + s)/(1 + ps) \approx 1 + s\).
- As long as the frequency of \( A_1 \) is low, the numbers of offspring contributed by different \( A_1 \)-type individuals are approximately independent of one another.
- The fate of a beneficial allele is primarily determined by its dynamics when rare. Although a beneficial allele can be lost through demographic stochasticity when rare, loss is exceedingly unlikely once the allele becomes common and selection dominates genetic drift.
These observations suggest that we can approximate the initial dynamics of the $A_1$ allele by a Galton-Watson process $Y = (Y_n : n \geq 0)$ in which the offspring number $\xi$ is Poisson distributed with mean $1 + s$ and the initial population size is $Y_0 = 1$. In other words,

$$Y_{n+1} = \sum_{i=1}^{Y_n} \xi_i^{(n)},$$

where $\left(\xi_i^{(n)} : i, n \geq 0\right)$ is a collection of i.i.d. Poisson-distributed random variables with parameter $1 + s$. In this case, $\mu = 1 + s > 1$ if $s > 0$, while the probability generating function of $\xi$ is

$$\psi(t) = \sum_{k=0}^{\infty} e^{-(1+s)} \frac{(1+s)^k}{k!} t^k = e^{(1+s)(t-1)},$$

and the extinction probability $\rho$ of $Y$ starting from a single individual is the smallest non-negative fixed-point of $\psi$:

$$\rho = e^{(1+s)(\rho-1)}.$$
Although $\rho = 1$ is always a solution to this equation, it is the smallest non-negative solution only when $s \leq 0$. To find other solutions, let us assume that we can expand $1 - \rho$ in a power series in $s$: $1 - \rho = c_1 s + c_2 s^2 + \cdots$. Substituting this series into the fixed-point equation and expanding the exponential function on the right-hand side in a power series gives:

$$1 - c_1 s - c_2 s^2 + O(s^3) = \exp \left\{ - (1 + s) \left( c_1 s + c_2 s^2 + O(s^3) \right) \right\}$$

$$= \exp \left\{ - c_1 s - (c_1 + c_2) s^2 + O(s^3) \right\}$$

$$= 1 - c_1 s - \left( c_1 - \frac{1}{2} c_1^2 + c_2 \right) s^2 + O(s^3).$$

Equating the coefficients of the monomials $s^2$ on both sides leads to the identity

$$2c_1 - c_1^2 = 0$$

which has two solutions, $c_1 = 0$ and $c_1 = 2$. 
It can be shown that if we take \( c_1 = 0 \), then we must have \( c_n = 0 \) for all \( n \geq 1 \), which gives the solution \( \rho = 1 \). Thus any non-unit solution must have \( c_1 = 2 \) and so the extinction probability for the branching process \( Y \) is \( \rho = 1 - 2s \). Since fixation of \( A_1 \) corresponds to non-extinction of the supercritical branching process, it follows that the fixation probability of a single beneficial allele is approximately

\[
u(1/N) \approx 2s
\]

for \( s \in [1/N, 0.1] \). The restrictions on \( s \) are necessary for the following reasons:

- **When** \( s < 1/N \), the allele is said to be **nearly neutral** and demographic stochasticity and selection are of comparable strength at all frequencies. In this case, fixation also depends on the dynamics of the allele when it is common and not adequately modeled by branching processes.

- **When** \( s \) is large, a second-order Taylor series approximation may not be sufficiently accurate.
Comparison of the branching process and diffusion approximations for the fixation probability of a beneficial mutation in a population of size $N = 10^4$.

- The branching process approximation underestimates the fixation probability when $s$ is close to $1/N$ and overestimates it when $s$ is greater than 0.05.
- However, between these values, the approximation is quite accurate.
- Unless a beneficial mutation has a very strong selection advantage, i.e., $s > 0.2$, it is more likely to be lost than fixed in a population.