Branching Processes

Linda J. S. Allen
Texas Tech University

The study of branching processes began in the 1840s with Irénée-Jules Bienaymé, a probabilist and statistician, and was advanced in the 1870s with the work of Reverend Henry William Watson, a clergyman and mathematician, and Francis Galton, a biometrician. In 1873, Galton sent a problem to the Educational Times regarding the survival of family names. When he did not receive a satisfactory answer, he consulted Watson, who rephrased the problem in terms of generating functions. The simplest and most frequently applied branching process is named after Galton and Watson, a type of discrete-time Markov chain. Branching processes fit under the general heading of stochastic processes. The methods employed in branching processes allow questions about extinction and survival in ecology and evolutionary biology to be addressed. For example, suppose we are interested in family names, as in Galton's original problem, or in the spread of a potentially lethal mutant gene that arises in a population, or in the success of an invasive species. Given information about the number of offspring produced by an individual per generation, branching process theory can address questions about the survival of a family name or a mutant gene or an invasive species. Other questions that can be addressed with branching processes relate to the rate of population growth in a stable versus a highly variable environment.

Some background in probability theory is required to apply branching process theory. We present some of this background in the next section. Then we discuss single-type and multi-type Galton-Watson branching processes, and extensions to random environments that will address questions about population survival and growth.

I. Background

The fundamental tools required for studying branching processes are generating functions. As the name implies, a generating function is a function that "generates" information about the process. For example, a probability generating function is used to calculate the probabilities associated with the process, whereas a moment generating function is used to calculate the moments, such as the mean and variance. We introduce some notation and define some terms in probability theory.

Let X be a discrete random variable taking values in the set $\{0, 1, 2, \dots\}$ with associated probabilities,

$$p_j = \text{Prob}\{X = j\}, \quad j = 0, 1, 2, \dots,$$

that sum to one,

$$p_0 + p_1 + p_2 + \cdots = \sum_{j=0}^{\infty} p_j = 1.$$

Let the **expectation** of $u(X)$ be defined as

$$\mathbb{E}(u(X)) = p_0 u(0) + p_1 u(1) + p_2 u(2) + \cdots = \sum_{j=0}^{\infty} p_j u(j).$$

If $u(X) = X$, then $\mathbb{E}(X)$ is the expectation of X or the **mean** of X,

$$\mathbb{E}(X) = p_1 + 2p_2 + 3p_3 + \cdots = \sum_{j=0}^{\infty} j p_j = \mu.$$

The summation reduces to a finite number of terms if, for example, $p_j = 0$ for $j > n$.

With this notation, three generating functions important in the theory of branching processes are defined. The **probability generating function (pgf)** of X is

$$\mathcal{P}(s) = \mathbb{E}(s^X) = p_0 + p_1 s + p_2 s^2 + \cdots = \sum_{j=0}^{\infty} p_j s^j,$$

where s is a real number. Evaluating $\mathcal{P}$ at $s = 1$ yields $\mathcal{P}(1) = 1$. The **moment generating function (mgf)** of X is

$$M(s) = \mathbb{E}(e^{sX}) = p_0 + p_1 e^s + p_2 e^{2s} + \cdots = \sum_{j=0}^{\infty} p_j e^{js}.$$

Evaluating M at $s = 0$ yields $M(0) = 1$. The **cumulant generating function (cgf)** of X is the natural logarithm of the moment generating function,

$$K(s) = \ln[M(s)].$$

Evaluating K at $s = 0$ yields $K(0) = \ln[M(0)] = \ln 1 = 0$.

As noted previously, the pgf generates the probabilities associated with the random variable X. For example, if X is the random variable for the size of a population, then the probability of population extinction p_0 can be found by evaluating the pgf at $s = 0$:

$$\mathcal{P}(0) = \text{Prob}\{X = 0\} = p_0.$$

Differentiation of $\mathcal{P}$ and evaluation at $s = 0$ equals p_1, the probability that the population size is one.

The mgf generates the moments of the random variable X. Differentiation of M and evaluation at $s = 0$ yields the moments of X about the origin. For example, the mean of X is the first derivative of M evaluated at $s = 0$. That is,

$$\left.\frac{dM(s)}{ds}\right|_{s=0} = M'(0) = \sum_{j=0}^{\infty} j p_j = \mathbb{E}(X).$$

The second moment is

$$\mathbb{E}(X^2) = M''(0).$$

The **variance** of X is $\sigma^2 = \mathbb{E}[(X - \mu)^2]$. Using properties of the expectation, the variance can be expressed as $\sigma^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$. Written in terms of the mgf,

$$\sigma^2 = M''(0) - [M'(0)]^2 = \sum_{j=0}^{\infty} j^2 p_j - \left[\sum_{j=0}^{\infty} j p_j\right]^2.$$

Formulas for the mean and variance of X can be computed from any one of the generating functions by appropriate differentiation and evaluation at either $s = 0$ or $s = 1$. They are defined as follows:

$$\mu = \mathcal{P}'(1) = M'(0) = K'(0)$$

and

$$\sigma^2 = \begin{cases} \mathcal{P}''(1) + \mathcal{P}'(1) - [\mathcal{P}'(1)]^2 \\ M''(0) - [M'(0)]^2 \\ K''(0) \end{cases} .$$

The preceding expressions for the mean and variance will be applied in the following sections.

II. Galton-Watson branching process

Galton-Watson branching processes are **discrete-time Markov chains**, that is, collections of discrete random variables, $\{X_n\}_{n=0}^{\infty}$, where the time $n = 0, 1, 2 \ldots$ is also discrete. The random variable X_n may represent the population size of animals, plants, cells, or genes at time n or generation n. The term **chain** implies each of the random variables are discrete-valued; their values come from the set of nonnegative integers $\{0, 1, 2, \ldots\}$. The name Markov acknowledges the contributions of probabilist Andrei Markov to the theory of stochastic processes. The **Markov property** means that the population size at time $n+1$ is only dependent on the population size at time n, and is not dependent on the size at earlier times. In this way, the population size X_n at time n predicts the population size in the next generation, X_{n+1}.

The graph in Fig. 1 illustrates what is referred to as a **sample path** or **stochastic realization** of a branching process. Fig. 1 also shows why the name "branching" is appropriate for this type of process. Beginning with an initial size of $X_0 = 1$ such as one mutant gene, in the next two generations, the number of mutant genes are $X_1 = 4$ and $X_2 = 8$, respectively. The parent genes are replaced by progeny genes in subsequent generations. The graph illustrates just one of many possible sample paths for the number of mutant genes in generations 1 and 2. Associated with each parent population is an offspring distribution which specifies the probability of the number of offspring produced in the next generation. The following three assumptions define the Galton-Watson branching process more precisely.

Three important assumptions about the Markov process $\{X_n\}_{n=0}^{\infty}$ define a single-type Galton-Watson branching process (GWbp).

(i) Each individual in the population in generation n gives birth to Y offspring of the same type in the next generation, where Y is a discrete random variable that takes values in $\{0, 1, 2 \ldots\}$. The offspring probabilities of Y are

$$p_j = \text{Prob}\{Y = j\}, \quad j = 0, 1, 2, \ldots.$$

(ii) Each individual in the population gives birth independently of all other individuals.

(iii) The same offspring distribution applies to all generations.

The expression "single-type" refers to the fact that all individuals are of one type such as the same gender, same cell type, or same genotype or phenotype. The pgf of X_n will be denoted as $\mathcal{P}_n$ and the pgf of Y as g.

If, in any generation n, the population size reaches zero, $X_n = 0$, then the process stops and remains at zero, population extinction has occurred. The probability that the population size is zero in generation n is found by setting $s = 0$ in the pgf,

$$\text{Prob}\{X_n = 0\} = \mathcal{P}_n(0).$$

To obtain information about population extinction in generation n, it is necessary to know the pgf in generation n, $\mathcal{P}_n(s)$. It can be shown that if the assumptions (i)-(iii) are satisfied, then the pgf in generation n is just an n-fold composition of the pgf of the offspring distribution g:

$$\mathcal{P}_n(s) = g(g(\cdots(g(s))\cdots)) = g^n(s).$$

To indicate why this is true, note that if the initial population size is one, $X_0 = 1$, then the pgf of X_0 is $\mathcal{P}_0(s) = s$. Then the pgf of X_1 is just the offspring of the one individual from the first generation, that is, $X_1 = Y$ so that $\mathcal{P}_1(s) = g(s)$. In the case of two generations, $\mathcal{P}_2(s) = g(g(s)) = g^2(s)$ so that the probability of population extinction after two generations is simply $\mathcal{P}_2(0) = g^2(0)$. In general, the pgf in generation n is found by taking n compositions of the offspring distribution g, as given above.

The preceding calculation assumed $X_0 = 1$. The probability of extinction can be calculated when the initial population size is greater than one, $X_0 = N > 1$. In this case, the pgf of X_n is just the n-fold composition of g raised to the power N,

$$\mathcal{P}_n(s) = [g^n(s)]^N.$$

Then the probability of population extinction in generation n can be found by evaluating the pgf at $s = 0$:

$$\mathcal{P}_n(0) = [g^n(0)]^N.$$

In most cases, it is difficult to obtain an explicit simple expression for $\mathcal{P}_n(s)$ in generation n. This is due to the fact that each time a composition is taken of another function, a more complicated expression is obtained. Fortunately, it is still possible to obtain information about the probability of population extinction after a long period of time (as $n \to \infty$). Ultimate extinction depends on the mean number of offspring produced by the parents. Recall that the mean number of offspring can be computed from the pgf $g(s)$ by taking its derivative and evaluating at $s = 1$. We will denote the mean number of offspring as m,

$$m = g'(1) = \sum_{j=1}^{\infty} j p_j.$$

The following theorem is one of the most important results in GWbp: if the mean number of offspring is less than or equal to one, $m \leq 1$, then eventually the population dies out, but if $m > 1$, the population has a chance of surviving. The probability of extinction depends on the initial population size N and the offspring distribution g. The theorem also shows how to calculate the probability of extinction if $m > 1$. A fixed point of g is calculated: a point q such that $g(q) = q$ and $0 < q < 1$. Then the probability of extinction is q^N.

Fundamental Theorem I: Let the initial population size of a GWbp be $X_0 = N \geq 1$ and let the mean number of offspring be m. Assume there is a positive probability of zero offspring and a positive probability of more than two offspring.

If $m \leq 1$, then the probability of ultimate extinction is one,

$$\lim_{n\to\infty} \text{Prob}\{X_n = 0\} = \lim_{n\to\infty} \mathcal{P}_n(0) = 1.$$

If $m > 1$, then the probability of ultimate extinction is less than one,

$$\lim_{n\to\infty} \text{Prob}\{X_n = 0\} = \lim_{n\to\infty} \mathcal{P}_n(0) = q^N < 1,$$

where the value of q is the unique fixed point of g: $g(q) = q$ and $0 < q < 1$.

The GWbp is referred to as **supercritical** if $m > 1$, **critical** if $m = 1$, and **subcritical** if $m < 1$. If the process is subcritical or critical, then the probability of extinction is certain. But if the process is supercritical, then there is a positive probability, $1 - q^N$, that the population will survive. As the initial population size increases, the probability of survival also increases. We apply the Fundamental Theorem 1 to the question of the survival of family names.

Example 1. Survival of family names. In 1931, Alfred Lotka assumed a zero-modified geometric distribution to fit the offspring distribution of the 1920s American male population. The theory of branching processes was used to address questions about survival of family names. In the zero-modified geometric distribution, the probability that a father has j sons is

$$p_j = bp^{j-1}, \quad j = 1, 2, 3, \ldots$$

and the probability that he has no sons is

$$p_0 = 1 - (p_1 + p_2 + \cdots) = 1 - \sum_{j=1}^{\infty} p_j.$$

Lotka assumed $b = 1/5$, $p = 3/5$, and $p_0 = 1/2$. Therefore, the probability of having no sons is $p_0 = 1/2$, the probability of having one son is $p_1 = 1/5$, and so on. Then the offspring pgf has the following form:

$$g(s) = p_0 + p_1 s + p_2 s^2 + \cdots = p_0 + \sum_{j=1}^{\infty} bp^{j-1}s^j = p_0 + \frac{bs}{1-ps} = \frac{1}{2} + \frac{s}{5-3s}.$$

The mean number of offspring is $m = g'(1) = 5/4 > 1$. According to the Fundamental Theorem I, in the supercritical case $m > 1$, there is a positive probability of survival, $1 - q^N$, where N is the initial number of males. Applying the Fundamental Theorem I, the number q is a fixed point of g, that is, a number q such that $g(q) = q$ and $0 < q < 1$. The fixed points of g are found by solving the following equation:

$$\frac{1}{2} + \frac{q}{5 - 3q} = q.$$

There are two solutions, $q = 1$ and $q = 5/6$, but only one solution satisfies $0 < q < 1$, namely $q = 5/6$. It should be noted that $q = 1$ will always be one solution to $g(q) = q$ due to one of the properties of a pgf, $g(1) = 1$.

To address the question about the probability of survival of family names, it follows that one male has a probability of 5/6 that his line of descent becomes extinct and a probability of 1/6 that his descendants will continue forever.

The zero-modified geometric distribution is one of the few offspring distributions where an explicit formula can be computed for the pgf of X_n, that is, the function $\mathcal{P}_n(s) = [g^n(s)]^N$. This particular case is referred to as the linear fractional case because of the form of the generating function. For $X_0 = 1$, the pgf of X_n for the linear fractional case is

$$\mathcal{P}_n(s) = \begin{cases} \dfrac{(m^n q - 1)s + q(1 - m^n)}{(m^n - 1)s + q - m^n}, & \text{if } m \neq 1 \\ \dfrac{[(n+1)p - 1]s - np}{nps - (n-1)p - 1}, & \text{if } m = 1 \end{cases},$$

where the value of q is a fixed point of g, $g(q) = q$. If $m > 1$, then q is chosen so that $0 < q < 1$ and if $m < 1$, q is chosen so that $q > 1$.

Example 2. Probability of extinction. Applying the preceding formula to Lotka's model in Example 1, for $X_0 = 1$, the probability of population extinction at the nth generation or the probability that no sons are produced in generation n is found by evaluating the pgf at $s = 0$:

$$\mathcal{P}_n(0) = \frac{q(m^n - 1)}{m^n - q}, \; m > 1.$$

Substituting $m = 5/4$ and $q = 5/6$ into $\mathcal{P}_n(0)$, the probabilities of population extinction in generation n can be easily calculated. They are graphed in Fig. 2. The probability of extinction approaches $q = 5/6 \approx 0.83$ for a single line of descent, $X_0 = 1$. For three lines of descent $X_0 = 3$, $q^3 \approx 0.58$.

Formulas for the mean and variance of X_n depend on the mean and variance of the offspring distribution, m and σ^2. The expectation of X_n is

$$\mathbb{E}(X_n) = m\mathbb{E}(X_{n-1}) = m^n\mathbb{E}(X_0)$$

and the variance is

$$\mathbb{E}\left[(X_n - \mathbb{E}(X_n))^2\right] = \begin{cases} \dfrac{m^{n-1}(m^n - 1)}{m-1}\sigma^2, & m \neq 1 \\ n\sigma^2, & m = 1. \end{cases}$$

In the subcritical case, $m < 1$, the mean population size of X_n decreases geometrically. In the critical case, $m = 1$, the mean size is constant but the variance of the population size X_n increases linearly, and in the supercritical case, $m > 1$, the mean and variance increase geometrically. This is an interesting result when one considers that in the supercritical case, there is a probability of extinction of q^N. In the next example, branching processes are used to study the survival of a mutant gene.

Example 3. Survival of a mutant gene. Suppose the population size is very large. A new mutant gene appears in N individuals of the population; the remaining individuals in the population do not carry the mutant gene. Individuals reproduce according to a branching process and those individuals with a mutant gene have offspring that carry the mutant gene. Suppose the mean number of offspring from parents with a mutant gene is m. If $m \leq 1$, then the line of descendants from the individuals with a mutant gene will eventually become extinct with probability one. But suppose the mutant gene has a mean that is slightly greater than 1,

$$m = 1 + \epsilon, \quad \epsilon > 0,$$

there is a probability $1 - q^N$ that the subpopulation with the mutant gene will survive. The value of q can be approximated from the cgf $K(s) = \ln M(s)$ using the fact that $K(0) = 0$,

$K'(0) = m$, and $K''(0) = \sigma^2$. Let $q = e^\theta$, where θ is small and negative, so that q will be less than but close to one. Then $e^\theta = g(e^\theta) = M(\theta)$ or equivalently, $\theta = \ln M(\theta) = K(\theta)$. Expanding $K(\theta)$ in a Maclaurin series about zero leads to

$$\theta = K(\theta) = 0 + m\theta + \sigma^2\frac{\theta^2}{2!} + \cdots .$$

Truncating the preceding series, gives an approximation to θ:

$$\theta \approx m\theta + \sigma^2\frac{\theta^2}{2}.$$

Solving for θ yields $\theta \approx -2\epsilon/\sigma^2$ and an approximation to $q = e^\theta$ is

$$q \approx e^{-\frac{2}{\sigma^2}\epsilon}.$$

The probability that the mutant gene survives in the population is $1 - q^N \approx 1 - e^{-\frac{2N}{\sigma^2}\epsilon}$. Suppose the offspring distribution has the form of a Poisson distribution, where the mean and variance are equal, $m = \sigma^2 = 1 + \epsilon$, and $\epsilon = 0.01$. Then if $N = 1$, the probability of survival of the mutant gene is $1 - q \approx 0.02$ but if $N = 100$, the probability of survival is much greater, $1 - q^{100} \approx 0.86$.

III. Random environment

Plant and animal populations are subject to their surrounding environment which is constantly changing. Their survival depends on environmental conditions such as food and water availability and temperature. Under favorable environmental conditions, the number of offspring increases, but under unfavorable conditions the number of offspring declines. With environmental variation, assumption (iii) in the GWbp no longer holds. Suppose the other two assumptions still hold for the branching process. That is,

(i) Each individual in the population in generation k gives birth to Y_k offspring of the same type in the next generation, where Y_k is a discrete random variable.

(ii) Each individual in the population gives birth independently of all other individuals.

In some cases, the random environment can be treated like a GWbp. For example, suppose the environment varies periodically, a good season followed by a bad season, so that the offspring random variables follow sequentially as Y_1, Y_2, Y_1, Y_2, and so on. As another example, suppose the environment varies randomly between good and bad seasons so that a good season occurs with probability p and a bad one with probability $1-p$.

In general, if m_n is the mean number of offspring produced in generation $n-1$, the expectation of X_n depends on the previous generation in the following way:

$$\mathbb{E}(X_n) = m_n \mathbb{E}(X_{n-1}).$$

Repeated application of this identity leads to an expression for the mean population size in generation n,

$$\mathbb{E}(X_n) = m_n \cdots m_1 \mathbb{E}(X_0).$$

If the mean number of offspring $m_i = m$ is the same from generation to generation, then this expression is the same as in a constant environment, $\mathbb{E}(X_n) = m^n \mathbb{E}(X_0)$.

Suppose the environment varies periodically over a time period of length T. The offspring distribution is $Y_1, Y_2, \ldots, Y_T$, then repeats. The preceding formulas can be used to calculate the mean size of the population after nT generations. Using the fact that the expectation $\mathbb{E}(Y_i) = m_i$, and $\mathbb{E}(X_T) = m_T \cdots m_2 m_1 \mathbb{E}(X_0)$ it follows that after nT generations,

$$\begin{aligned} \mathbb{E}(X_{nT}) &= (m_T \cdots m_2 m_1)^n \mathbb{E}(X_0). \\ &= \left[(m_T \cdots m_2 m_1)^{1/T}\right]^{nT} \mathbb{E}(X_0). \end{aligned}$$

The mean population growth rate each generation in an environment that varies periodically is $\mu_r = (m_T \cdots m_2 m_1)^{1/T}$, which is the geometric mean of $m_1, m_2, \ldots, m_T$. It is well-known that the geometric mean is less than the arithmetic mean (average),

$$\mu_r = (m_1 m_2 \cdots m_T)^{1/T} \leq \frac{1}{T}(m_1 + m_2 + \cdots + m_T) = \mu.$$

The mean population growth rate in a periodic environment is less than the average of the growth rates, $\mu_r \leq \mu$. If $\mu_r < 1$, then the population will not survive.

Suppose the environment varies randomly and the random variable for the offspring distribution is Y_k. Let $m_k = \mathbb{E}(Y_k)$ be the mean number of offspring in generation $k-1$. For example, suppose there are two random variables for the offspring distribution that occur with probability p and $1-p$, and their respective means are equal to μ_1 and μ_2. The order in which these offspring distributions occur each generation is random, that is, m_k could be either μ_1 or μ_2 with probability p or $1-p$, respectively. Thus, m_k is a random variable with expectation $\mathbb{E}(m_k) = p\mu_1 + (1-p)\mu_2 = \mu$. The set $\{m_k\}_{k=1}^{\infty}$ is a sequence of **independent and identically distributed (iid)** random variables, that is, random variables that are independent and have the same probability distribution. To compute the mean population growth rate, the same technique is used as in the previous example, but the limit is taken as $n \to \infty$. Rewriting the geometric mean using the properties of an exponential function leads to

$$\begin{aligned}(m_n \cdots m_2 m_1)^{1/n} &= e^{(1/n)\ln[m_n \cdots m_2 m_1]} \\ &= e^{(1/n)[\ln m_n + \cdots + \ln m_2 + \ln m_1]}.\end{aligned}$$

It follows from probability theory that the average of a sequence of iid random variables $(1/n)[\ln m_n + \cdots + \ln m_2 + \ln m_1]$ approaches its mean value $\mathbb{E}(\ln m_k) = \ln \mu_r$. Thus, in the limit, the mean population growth rate in a random environment is

$$\lim_{n\to\infty} (m_n m_{n-2} \cdots m_1)^{1/n} = e^{\mathbb{E}(\ln m_k)} = e^{\ln \mu_r} = \mu_r.$$

As in the periodic case, the mean population growth rate in the random environment is less than the average of the growth rates: $\mu_r \leq \mu$:

$$\mu_r = e^{\ln \mu_r} = e^{\mathbb{E}(\ln m_k)} \leq e^{\ln \mathbb{E}(m_k)} = e^{\ln \mu} = \mu.$$

Interestingly, a population subject to a random environment may not survive, $\mu_r < 1$, even though the mean number of offspring for any generation is greater than one, $\mu > 1$, that is, $\mu_r < 1 < \mu$. A random environment is a more hostile environment for a population than a constant environment.

IV. Multi-type Galton-Watson branching process

In a single-type GWbp, the offspring are of the same type as the parent. In a multi-type GWbp, each parent may have offspring of different types. For example, a population may be divided according to age, size, or developmental stage, representing different types, and in each generation, individuals may age or grow to another type, another age, size, or stage. In genetics, genes can be classified as wild or mutant types and mutations change a wild type into a mutant type.

A multi-type GWbp $\{\vec{X}(n)\}_{n=0}^{\infty}$ is a collection of vector random variables $\vec{X}(n)$, where each vector consists of k different types, $\vec{X}(n) = (X_1(n), X_2(n), \ldots, X_k(n))$. Each random variable $X_i(n)$ has k associated offspring random variables for the number of offspring of type $j = 1, 2, \ldots, k$ from a parent of type i.

As in a single-type GWbp, it is necessary to define a pgf for each of the k random variables $X_i(n)$, $i = 1, 2, \ldots, k$. One of the differences between single-type and multi-type GWbp is that the pgf for X_i depends on the k different types. The pgf for X_i is defined assuming initially $X_i(0) = 1$ and all other types are zero, $X_l(0) = 0$. The offspring pgf corresponding to X_i is denoted as $g_i(s_1, s_2, \ldots, s_k)$. The multi-type pgfs have properties similar to a single-type pgfs. For example, $g_i(1, 1, \ldots, 1) = 1$, $g_i(0, 0, \ldots, 0)$ is the probability of extinction for X_i given $X_i(0) = 1$ and $X_l(0) = 0$ for all other types, and differentiation and evaluation when the s variables are set equal to one leads to an expression for the mean number of offspring. But there are different types of offspring, so there are different types of means. The mean number of j-type offspring by an i-type parent is denoted as m_{ji}. A value for m_{ji} can be calculated from the pgf g_i by differentiation with respect to s_j and evaluating all of the s variables at one:

$$m_{ji} = \left. \frac{\partial g_i(s_1, s_2, \ldots, s_k)}{\partial s_j} \right|_{s_1=1, s_2=1, \ldots, s_k=1}.$$

There are k^2 different means m_{ji} because of the k different offspring and k different parent types. When the means are put in an ordered array, the $k \times k$ matrix is called the expectation matrix:

$$\mathcal{M} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1k} \\ m_{21} & m_{22} & \cdots & m_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ m_{k1} & m_{k2} & \cdots & m_{kk} \end{pmatrix}.$$

Population extinction for a multi-type GWbp depends on the properties of the expectation matrix $\mathcal{M}$. Therefore, some matrix theory is reviewed. Also see the glossary. An **eigenvalue** of a square matrix $\mathcal{M}$ is a number λ satisfying $\mathcal{M}\vec{V} = \lambda\vec{V}$, for some nonzero vector $\vec{V}$, known as an **eigenvector**. If all of the entries m_{ij} in matrix $\mathcal{M}$ are positive or if all of the entries are nonnegative and $\mathcal{M}^2$ or $\mathcal{M}^3$ or some power of $\mathcal{M}$, $\mathcal{M}^n$, has all positive entries, then $\mathcal{M}$ is referred to as a **regular** matrix. It follows from matrix theory that a regular matrix $\mathcal{M}$ has a positive eigenvalue λ that is larger than any other eigenvalue. This eigenvalue λ is referred to as the dominant eigenvalue. It is this eigenvalue that determines whether the population grows or declines. The dominant eigenvalue λ plays the same role as the mean number of offspring m in the single-type GWbp.

The second fundamental theorem for GWbp extends the Fundamental Theorem I to multi-type GWbp. If $\lambda \leq 1$, the probability of extinction is one as $n \to \infty$, but if $\lambda > 1$, then there is a positive probability that the population survives. This latter probability can be computed by finding fixed points of the k generating functions.

Fundamental Theorem II: Let the initial sizes for each type be $X_i(0) = N_i$, $i = 1, 2, \ldots, k$. Suppose the generating functions g_i for each of the k types are nonlinear functions of s_j with some $g_i(0, 0, \ldots, 0) > 0$, the expectation matrix $\mathcal{M}$ is regular, and λ is the dominant eigenvalue of matrix $\mathcal{M}$.

If $\lambda \leq 1$, then the probability of ultimate extinction is one,

$$\lim_{n\to\infty} \text{Prob}\{\vec{X}(n) = \vec{0}\} = 1.$$

If $\lambda > 1$, then the probability of ultimate extinction is less than one,

$$\lim_{n\to\infty} \text{Prob}\{\vec{X}(n) = \vec{0}\} = q_1^{N_1} q_2^{N_2} \cdots q_k^{N_k},$$

where $(q_1, q_2, \ldots, q_k)$ is the unique fixed point of the k generating functions $g_i(q_1, \ldots, q_k) = q_i$ and $0 < q_i < 1$, $i = 1, 2 \ldots, k$. In addition, the expectation of $\vec{X}(n)$ is

$$\mathbb{E}(\vec{X}(n)) = \mathcal{M}\mathbb{E}(\vec{X}(n-1)) = \mathcal{M}^n\mathbb{E}(\vec{X}(0)).$$

The mean population growth rate is λ.

The following example is an application of a multi-type GWbp and the Fundamental Theorem II to an age-structured population.

Example 4. Age-structured population. Suppose there are k different age classes. The number of females of age k are followed from generation to generation. The first age, type 1, represents newborn females. A female of age i gives birth to r females of type 1 with probability $b_{i,r}$, then survives, with probability $p_{i+1,i}$ to the next age $i+1$. There may be no births with probability $b_{i,0}$ or one birth with probability $b_{i,1}$, etc. The mean number of female offspring by a female of age i equals $b_i = b_{i,1} + 2b_{i,2} + 3b_{i,3} + \cdots$. Age k is the oldest age; females do not survive past age k. The probability generating functions are

$$\begin{aligned} g_i(s_1, s_2, \ldots, s_k) &= [p_{i+1,i}s_{i+1} + (1 - p_{i+1,i})] \sum_{r=0}^{\infty} b_{i,r}s_1^r, \quad i = 1, \ldots, k-1 \\ g_k(s_1, s_2, \ldots, s_k) &= b_{k,0} + b_{k,1}s_1 + b_{k,2}s_1^2 + \cdots = \sum_{r=0}^{\infty} b_{k,r}s_1^r. \end{aligned}$$

Note that $g_i(1, 1, \ldots, 1) = 1$ and $g_i(0, 0, \ldots, 0) = (1 - p_{i+1,i})b_{i,0}$ except for the oldest age, where $g_k(0, 0, \ldots, 0) = b_{k,0}$. The expression $(1 - p_{i+1,i})b_{i,0}$ is the probability of extinction given $X_i(0) = 1$ and $X_l(0) = 0$. The term $1 - p_{i+1,i}$ is the probability an individual of age i does not survive to the next age and $b_{i,0}$ is the probability an individual of age i has no offspring.

Taking the derivative of the generating functions and evaluating the s variables at one gives the the following expectation matrix $\mathcal{M}$:

$$\mathcal{M} = \begin{pmatrix} b_1 & b_2 & \cdots & b_{k-1} & b_k \\ p_{21} & 0 & \cdots & 0 & 0 \\ 0 & p_{32} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & p_{k,k-1} & 0 \end{pmatrix},$$

where the mean birth probabilities are on the first row and the survival probabilities are on the subdiagonal. In demography, matrix $\mathcal{M}$ is known as a **Leslie matrix** in honor of the contributions of Patrick Holt Leslie. The mean population growth rate is the dominant eigenvalue of $\mathcal{M}$.

As a specific example, consider two age classes with generating functions

$$\begin{aligned} g_1(s_1, s_2) &= [(1/2)s_2 + 1/2][1/2 + (1/6)s_1 + (1/6)s_1^2 + (1/6)s_1^3] \\ g_2(s_1, s_2) &= 1/4 + (1/4)s_1 + (1/4)s_1^2 + (1/4)s_1^3. \end{aligned}$$

The mean number of offspring for ages 1 and 2 is

$$b_1 = b_{1,1} + 2b_{1,2} + 3b_{1,3} = (1/6) + 2(1/6) + 3(1/6) = 1$$

and

$$b_2 = b_{2,1} + 2b_{2,2} + 3b_{2,3} = (1/4) + 2(1/4) + 3(1/4) = 3/2.$$

The expectation matrix is regular,

$$\mathcal{M} = \begin{pmatrix} 1 & 3/2 \\ 1/2 & 0 \end{pmatrix},$$

with a dominant eigenvalue equal to $\lambda = 3/2$, which is the mean population growth rate. According to the Fundamental Theorem II, there is a unique fixed point (q_1, q_2), one q_i for each type satisfying $g_1(q_1, q_2) = q_1$ and $g_2(q_1, q_2) = q_2$. This fixed point can be calculated and shown to be equal to $(q_1, q_2) \approx (0.446, 0.443)$. For example, if there are initially two females of age 1 and three females of age 2, $X_1(0) = 2$ and $X_2(0) = 3$, then according to the Fundamental Theorem II, the probability of ultimate extinction of the total population is approximately

$$(0.446)^2(0.433)^3 \approx 0.016.$$

On the other hand, if $X_1(0) = 1$ and $X_2(0) = 0$, then the probability of population extinction is 0.446.

Suppose a plant or animal population which has several developmental stages and is subject to a random environment is modeled by a branching process. Environmental variation causes the expectation matrix $\mathcal{M}$ to change from generation to generation, $\mathcal{M}_1$, $\mathcal{M}_2$, etc. Matrix $\mathcal{M}$ may change in a periodic fashion such as seasonally or randomly according to some distribution. As shown in the case of a single-type GWbp, if the environment varies periodically, a good season followed by a bad season, where $\mathcal{M}_1$ is the expectation matrix in a good season and $\mathcal{M}_2$ in a bad season, then after $2n$ seasons, the mean population size for each stage is

$$\mathbb{E}(\vec{X}(2n)) = (\mathcal{M}_2\mathcal{M}_1)^n\mathbb{E}(\vec{X}(0)).$$

The mean population growth rate is λ_r, where $(\lambda_r)^2$ is the dominant eigenvalue of the product of the two matrices $\mathcal{M}_2\mathcal{M}_1$. Note that $(\lambda_r)^2$ does not necessarily equal $\lambda_2\lambda_1$, where λ_1 and λ_2 are the dominant eigenvalues of $\mathcal{M}_1$ and $\mathcal{M}_2$. In general,

$$\lambda_r \leq \frac{\lambda_1 + \lambda_2}{2} = \lambda.$$

The mean population growth rate in a random environment is less than the average of the growth rates. If the environment varies randomly each generation, and if the expectation matrices $\{\mathcal{M}_i\}_{i=1}^{\infty}$ are drawn from a set of iid regular matrices, then a similar relationship holds. In this case, if the expectation in each generation $\mathbb{E}(\mathcal{M}_k)$ has a dominant eigenvalue of λ, then there exists a mean population growth rate λ_r, where $\lambda_r \leq \lambda$. The population growth rate in the random environment is generally less than the growth rate in a constant environment.

The previous discussion was confined to discrete-time processes, where the lifetime is a fixed length $n \to n+1$, which for convenience was denoted as one unit of time. At the end of that time interval, the parent is replaced by the progeny. In continuous-time processes, an individual's lifetime is not fixed but may have an arbitrary probability distribution. In the case of an exponentially distributed lifetime, the branching process is said to have the Markov property because the exponential distribution has the memoryless property; the process at time t only depends on the present and not the past. If the lifetime distribution is not exponential, then it is referred to as an **age-dependent** branching process, also known as a Bellman-Harris branching process, where the name is in honor of the contributions of Richard Bellman and Theodore Harris. Similar results as the Fundamental Theorem I and II can be derived for single-type and multi-type branching processes of the continuous type. Additional biological applications of branching processes to discrete-time processes and applications to continuous-time processes can be found in the references.

Glossary

discrete-time Markov chain is a collection of discrete random variables $\{X_n\}_{n=0}^{\infty}$, where time is discrete $n = 0, 1, 2, \ldots$. The term **Markov** refers to the memoryless property of the process. That is, the conditional probability at time n only depends on the state of the system at time $n-1$ and not on the earlier times,

$$\text{Prob}\{X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0\} = \text{Prob}\{X_n = x_n | X_{n-1} = x_{n-1}\}.$$

eigenvalue of a square matrix $\mathcal{M}$ is a real or complex number λ satisfying $\mathcal{M}\vec{V} = \lambda\vec{V}$ or $(\mathcal{M} - \lambda I)\vec{V} = \vec{0}$, where the nonzero vector $\vec{V}$ is known as an **eigenvector** and I is the identity matrix. A square matrix of size $k \times k$ has k eigenvalues. The eigenvalues of $\mathcal{M}$ can be calculated by finding the roots of the **characteristic equation** of $\mathcal{M}$ which is the determinant of the following matrix $(\mathcal{M} - \lambda I)$.

exponential distribution is a continuous distribution $X \sim E(\lambda)$ with probability density function $f(t) = \lambda e^{-\lambda t}$, $t \in [0, \infty)$, and mean and variance $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$.

Poisson distribution is a discrete distribution $X \sim Po(\lambda)$ with probabilities

$$p_j = \frac{\lambda^j e^{-\lambda}}{j!}, \; j = 0, 1, 2, \ldots,$$

and mean and variance $\mu = \lambda = \sigma^2$.

sample path or stochastic realization is one solution over time of a stochastic process. For a discrete-time Markov chain, a sample path is $x_0, x_1, x_2, \ldots$, where $X_n = x_n$ are particular values of the random variable at times $n = 0, 1, 2, \ldots$.

stochastic process is a collection of random variables $\{X_n(\omega),\ n \in \mathcal{T}, \omega \in \Omega\}$, where $\mathcal{T}$ is an index set and Ω is a common sample space. In the case of continuous random variables, the notation $X(t, \omega)$ is used in place of $X_n(\omega)$, where $t \in \mathcal{T}$. The index set $\mathcal{T}$ is generally the time domain, either discrete $n \in \{0, 1, 2, \ldots\}$ or continuous $t \in [0, \infty)$. The sample space Ω for a discrete random variable with values in the set $A = \{0, 1, 2, 3, \ldots\}$ are infinite sequences, where sequence elements are taken from the set A. For example, $\omega = 1, 2, 3, 2, 1, 0, 0, 0, \ldots \in \Omega$ is one element in the sample space, where $X_0 = 1$ and $X_5 = 0$; extinction occurs after five generations. Often the sample space notation is omitted, so that $X_n(\omega) = X_n$.

Further Reading

Allen, L. J. S. 2003. An Introduction to Stochastic Processes with Applications to Biology. Upper Saddle River, NJ: Prentice Hall.

Athreya, K. B. and P. E. Ney. 1972. Branching Processes. Berlin, Germany: Springer-Verlag.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis and Interpretation. 2nd ed. Sunderland, MA: Sinauer Assoc. Inc.

Haccou, P., P. Jagers, and V. A. Vatutin. 2005. Branching Processes Variation, Growth, and Extinction of Populations. Cambridge Studies in Adaptive Dynamics. Cambridge, UK: Cambridge Univ. Press.

Harris, T. E. 1963. The Theory of Branching Processes. Berlin, Germany: Springer-Verlag.

Jagers, P. 1975. Branching Processes with Biological Applications. London: John Wiley & Sons.

Kimmel, M. and D. Axelrod. 2002. Branching Processes in Biology. New York, NY: Springer-Verlag.

Mode, C. J. 1971. Multitype Branching Processes Theory and Applications. Oxford, UK: Elsevier.

Tuljapurkar, S. 1990. Population Dynamics in Variable Environments. Berlin and New York: Springer-Verlag.

Figure Captions

Figure 1. One sample path or stochastic realization of a branching process, beginning with $X_0 = 1$. The parent population produces four offspring in generation 1, $X_1 = 4$. These four parents produce three, zero, four, and one progeny in generation 2, respectively, making the total population size eight in generation 2, $X_2 = 8$. The sample path is $1, 4, 8, \ldots$.

Figure 2. Probability of extinction of family names with one line of descent, $X_0 = 1$, and three lines of descent, $X_0 = 3$, based on Lotka's model.

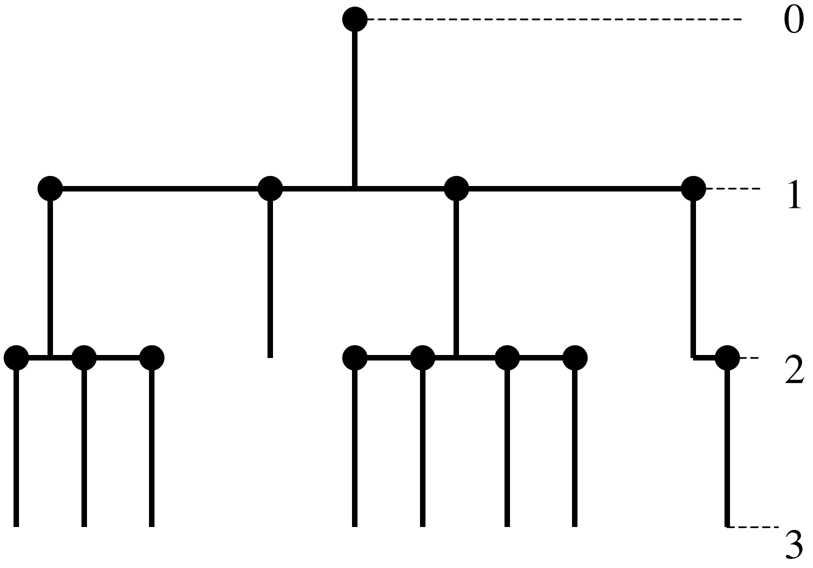

Figure 1:

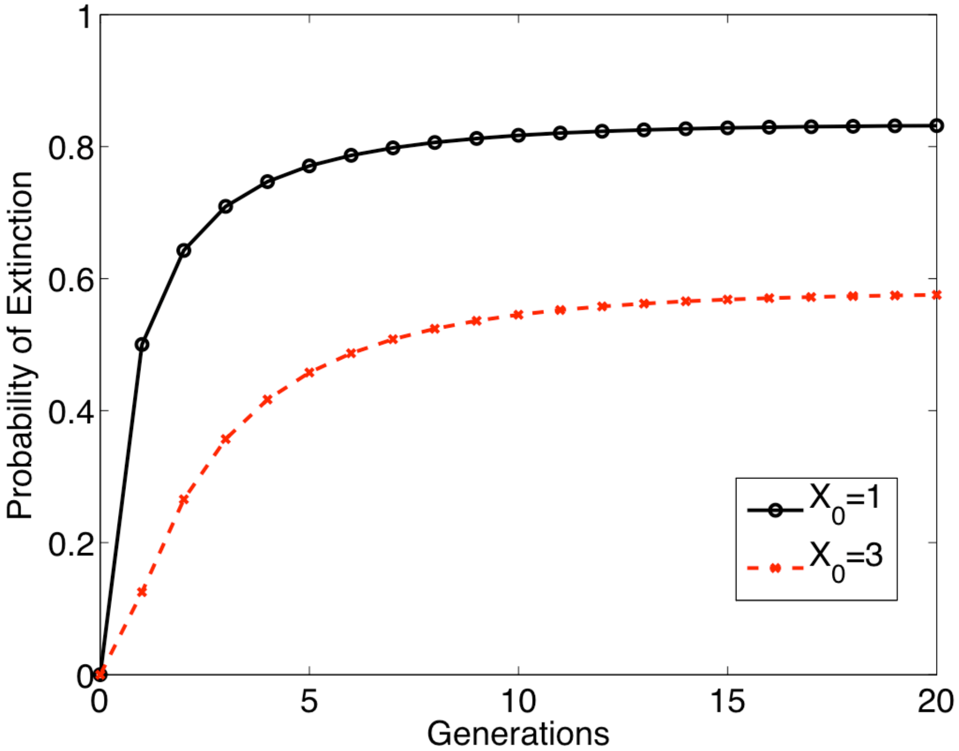
Probability of Extinction
Generations
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
X_0=1
X_0=3

Figure 2: