Generating Functions in Branching Processes and Birth and Death Processes

A Dissertation submitted to the Faculty of Science, University of the Witwatersrand, in fulfillment of the requirements for the degree of Master of Science.

Raeesa Docrat

March 2018, Johannesburg

Abstract

"If I have seen further, it is by standing upon the shoulders of giants"

-Sir Isaac Newton

Birth and death processes and branching processes are areas of stochastic processes that can be applied to a multitude of fields and disciplines, such as biology, economics and engineering. In general, the analysis of distributions and moments of these processes are difficult to obtain in explicit form. The use of generating functions makes computation much easier. The dissertation looks at the theory, application and innovation of using generating functions in the analysis of branching and birth and death processes.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

____

Raeesa Docrat

__________ day of ____________________ 20__________ in ____________________

Acknowledgments

Firstly, I would like to acknowledge God, for blessing me with the ability and the opportunity to complete this dissertation.

I would like to thank my husband and family for all the love and support that they have shown me. Thank you for showing interest and trying to grasp the matter of my research. I know that my family at least knows the meaning of stochastic.

I wish to present special thanks to my supervisor, Professor Frank Beichelt, for all the guidance, patience, motivation, support and knowledge over the past two years. You have been an inspiration and an example of what hard work produces. I am very lucky to have a supervisor with your expertise.

Lastly, I would like to thank the NRF for funding me throughout my Master's dissertation. Thank you for giving me this opportunity.

List of Symbols

$\mathbb{N}$	Set of all positive integers
$\mathbb{Z}$	Set of all integers
$\mathbb{Z}_+$	Set of all non-negative integers
$\mathbb{Q}$	Set of all rational numbers
$\mathbb{Q}'$	Set of all irrational numbers
$\mathbb{R}$	$(-\infty, \infty)$
$\mathbb{R}_+$	$[0, \infty)$
$*$	Convolution $g * h(t) = \int_0^\infty g(t-u)h(du)$
$\square$	Symbol after a proof of a theorem
$f \circ g$	Composition of two functions f and g
$f^*(n)$	n^{th} Convolution power of a function f
$f^{(n)}$	n^{th} derivative of function f
f^+	$f^+ = \begin{cases} f, & f \geq 0 \\ 0, & f < 0 \end{cases}$ is defined for an arbitrary function f
$o(x)$	Landau order symbol
X, Y, Z	Random variables

$\mu_X, E[X\ ]$ Mean or expected value of random variable X

$\sigma_X^2, var(X)$ Variance of random variable X

$P_X(t)$ Probability generating function of a random variable X

$M_X(t)$ Moment generating function of a random variable X

$\phi_X(t)$ Characteristic function of a random variable X

$K_X(t)$ Cumulant generating function of a random variable X

$\{X(t), t \in \mathbf{T}\}$ Stochastic process with parameter space $\mathbf{T}$

$m(t)$ trend function of a stochastic process, $t \in \mathbf{T}$

p_{ij}, $p_{ij}^{(n)}$ one step, n- step transition probabilities of a homogeneous discrete-time Markov chain

$p_{ij}(t)$ transition probabilities of a continuous-time homogeneous Markov chain

π_i stationary state distribution of homogeneous Markov chain, $\{\pi_i, i \in \mathbb{Z}\}$

λ_i, μ_i birth, death rates

L lifetime

Contents

Chapter 1

Introduction

1.1 Background

Birth and death processes and branching processes are stochastic models for quantitatively investigating the development in time of random phenomena. Examples of random phenomena modelled through birth and death processes include biological populations (particularly of threatened species), spread of epidemic disease, mutant gene dynamics, cell kinetics (proliferation of cancer cells) as well as nuclear chain reactions, and for modeling flows of radioactive, cosmic and other particles. These processes have direct applications in actuarial science, in finance and in engineering.

Birth and death processes are an important class of Markov chains where there are only two transitions, “births” and “deaths”. A process with no “deaths” is known as a pure-birth process, and one without “births” is called a pure-death process. Birth and death processes occur in biology, economics, demographics and queuing theory.

Branching processes have a long history dating back to 1845 when a few individuals became interested in studying the survival of noble family names. Today, branching processes form an integral part of population theory and have a variety of uses extending to electron multipliers, neutron chain reactions, population growth and survival of mutant genes.

A key mathematical tool for analysing birth and death processes and branching processes are generating functions of random variables. They allow us to calculate parameters of the probability functions which help to determine the probability distribution. The parameters of the probability function are usually tedious to obtain. Generally, in combinatorics, generating functions are an important tool used for counting combinatorial objects and obtaining asymptotics.

1.2 Generating Functions

1.2.1 Generating Functions used in Mathematics

The following derivations, formulas and definitions follow Drmota (2009).

Ordinary Generating Functions

The ordinary generating function (OGF) of a sequence of real numbers, $(\alpha_n)_{n>0}$, is the formal power series:

$$\alpha(x) = \sum_{n=1}^{\infty} \alpha_n x^n \tag{1.1}$$

The notation $[x^n]\alpha(x) = \alpha_n$ is used to extract the coefficients of x^n.

A generating function $\alpha(x)$ represents an analytic function for $|x| < R$, where

$$R = \lim_{n\to\infty} \left(\sup \left|\alpha_n\right|^{1/n} \right)^{-1} \tag{1.2}$$

denotes the radius of convergence of $\{\alpha_n\}$. Thus if $R > 0$, then we can either use differentiation to obtain the sequence $\{\alpha_n\}$,

$$\alpha_n = \frac{\alpha^{(n)}(0)}{n!}, \quad n > 0, \tag{1.3}$$

or we use Cauchy's Formula:

$$\alpha_n = \frac{1}{2\pi i} \int_\gamma \alpha(x) \frac{dx}{x^{n+1}}, \tag{1.4}$$

where γ is a closed curve inside the region of analyticity of $\alpha(x)$ with winding $+1$ around the origin.

Table 1.1: Combinatorial construction of operations used in counting

Combinatorial construction	**OGF**
$C = A + B$	$\gamma(x) = \alpha(x) + \beta(x)$
$C = A \times B$	$\gamma(x) = \alpha(x)\beta(x)$
$C = A^*$	$\gamma(x) = \frac{1}{1-\alpha(x)}$
$C = A(B)$	$\gamma(x) = \alpha(\beta(x))$

Table 1.1 is a summary of the combinatorial construction of some of the operations used in counting problems. The operations are between two OGFs, A and B. The first row of the table is the sum, the second row is the product, the third row is the inverse and the last row is the composition between A and B. These constructions of operations, are useful as they assist with the computation of complicated generating functions.

Bivariate Generating functions

Although the ordinary generating functions are very useful, bivariate generating functions are the adequate tools when assessing functions where there are two parameters of interest.The ordinary bivariate generating function of a double indexed sequence a_{nk} is defined as the series

$$a(u,z) = \sum_{n\geq 0}\sum_{k\geq 0} a_{nk}u^k z^n \tag{1.5}$$

1.2.2 Generating Functions used in Statistics

Assume X is a discrete random variable with state space $\mathbb{Z}$. Let$f$ denote the probability mass function of X defined as

$$f(j) = Prob(X = j) = p_j, \quad j \in \mathbb{Z} \tag{1.6}$$

where

$$\sum_{j=0}^{\infty} p_j = 1 \tag{1.7}$$

The mean and the variance of X are

$$\mu_X = E[X] = \sum_{j=0}^{\infty} j p_j \tag{1.8}$$

and

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2 \tag{1.9}$$

Probability Generating Function

The probability generating function(PGF), denoted as P_X, of the discrete random variable X is a function defined on a subset of the real numbers defined by:

$$P_X(t) = E[t^X] = \sum_{j=0}^{\infty} p_j t^j \tag{1.10}$$

for some $t \epsilon \mathbb{R}$

The PGF generates probabilities associated with the distribution. In general, the m^{th} derivative of the PGF of X is

$$P_X^{(m)}(0) = m! p_m \tag{1.11}$$

Theorem 1.1

If X and Y have PGFs M_X and M_Y respectively, then

$$M_X(t) = M_Y(t) \quad \text{for all } t \quad (a)$$

$$\text{iff} \quad P(X=k) = P(Y=k) \quad \text{for all } k = 0, 1, 2, ... \quad (b)$$

Proof

We only need to prove that (a) implies (b). The radii of convergence on M_X and M_Y are both greater or equal to one, so they both have a unique power series expansion about the origin:

$$M_X(t) = \sum_{k=0}^{\infty} t^k P(X=k)$$

$$M_Y(t) = \sum_{k=0}^{\infty} t^k P(Y=k)$$

If $M_X = M_Y$, then the two power series have identical coefficients. $\square$

The above theorem shows the uniqueness of probability generating functions, which allows us to get unique probability distributions.

Moment Generating Function

The moment generating function (MGF) of the discrete random variable X with state space $\mathbb{Z}$ and probability mass function $f(j) = p_j, j \in \mathbb{Z}$ denoted $M_X(t)$ is defined as:

$$M_X(t) = E[e^{tX}] = \sum_{j=0}^{\infty} p_j e^{jt} \tag{1.12}$$

for some $t \epsilon \mathbb{R}$

The MGF generates moments $E[X^m]$ of the distribution of random variable X, provided $\sum_{j=0}^{\infty} p_j e^{jt} < \infty$. In general,

$$M_X^{(m)}(0) = E[X^m] \tag{1.13}$$

where $M_X^{(m)}$ is the m^{th} derivative of $M_X(t)$.

Characteristic Function

The characteristic function (CHF) of the random variable X is

$$\varphi_X(t) = E[e^{itX}] = \sum_{j=0}^{\infty} p_j e^{ijt} \tag{1.14}$$

where $i = \sqrt{-1}$

If $f(x)$ is the density function of the random variable X, then the n-fold derivative of $\varphi_X(t)$ with respect to t is

$$\frac{\partial^n \varphi_X(t)}{\partial t^n} = (-1)^n \sum_{x=0}^{\infty} e^{itX} f(x)$$

Hence, the moments of all the orders of X can be obtained from $E[X^0] = E[1] = 1$ and

$$E[X^n] = (-1)^n \frac{\partial^n \varphi_X(t)}{\partial t^n}\bigg|_{t=0}$$

Cumulant Generating Function

The cumulant generating function(CGF), $K_X(t)$, of the discrete random variable X is defined as the natural logarithm of the moment generating function

$$K_X(t) = \ln[M_X(t)] \tag{1.15}$$

The generating functions for continuous random variables are obtained in a similar manner, just integrating instead of taking the sum.

The generating function given in (1.15) is more efficient in calculating the moments and distribution of variables as it is more tractable than using direct methods. Both birth and death processes and branching processes make use of them in analysis.

1.3 Differential Equations

When working with branching and birth and death processes, we often encounter differential equations which need to be solved in order to obtain the desired information. A differential equation for a function f is an equation which contains at least one term involving f and derivatives of f. The order of a differential equation is determined by the highest derivative in that equation. When working with the generating functions of birth and death processes or branching processes, it is usually sufficient to find differential equations of the first and second order.

1.3.1 First Order Differential Equations

A first order differential equation has the following form:

$$\frac{dy}{dx} = f(x, y) \tag{1.16}$$

There are two methods that can be used to solve first order differential equations

1. **Separation of Variables**

   The separation of variables method is used when $f(x, y)$ can be written in the form $f(x, y) = f(y)g(x)$:

   $$\frac{dy}{dx} = f(y)g(x) \tag{1.17}$$

   We solve this by rewriting the equation as

   $$\frac{1}{f(y)}dy = g(x)dx \tag{1.18}$$

   and integrate out both sides of equation (1.18). We use an initial condition to find the constant of integration.

2. **Integrating Factor**

   The integrating factor method is used when the differential equation can be written in the following form:

   $$\frac{dy}{dx} + p(x)y = q(x), \tag{1.19}$$

   where p and q are functions of x only.

   We solve this by rearranging the differential equation into the standard form and finding the integrating factor ($e^{\int p(x)dx}$). Then multiply through by the integrating factor and rewrite the left hand side as the derivative of $y^{\int p(x)dx}$. Integrating both sides gives the general solution.

Riccati Equations

Riccati looked at a differential equation of type

$$\frac{dw}{dz} = A(z) + B(z)w + C(z)w^2 \tag{1.20}$$

There are two transformations that assist with solving the Riccati equations.

1. **Transformation 1**

   The transformation:

   $$w = -\frac{y'}{yC(z)} \tag{1.21}$$

   leads to the second-order linear homogeneous equation:

   $$C(z)y'' - [C'(z) + B(z)C(z)]y' + [C(z)]^2A(z)y = 0 \tag{1.22}$$

   If a particular solution w_1 to equation (1.20) is known, then a more general solution containing a single arbitrary constant can be obtained from:

$$w = w_1(z) + \frac{1}{v(z)} \tag{1.23}$$

where $v(z)$ is a solution to the first-order linear equation.

$$v' = -[B(z) + 2C(z)w_1(z)]v - C(z) \tag{1.24}$$

2. **Transformation 2**

   The transformation

$$y = f + \frac{1}{v} \tag{1.25}$$

reduces the Riccati equation to a linear function, where f is any solution of equation (1.20). Since f is a particular solution,

$$f' = C(z)f^2 + B(z)f + A(z). \tag{1.26}$$

Using equation (1.25), we get

$$y' = f' - \frac{1}{v^2}v' = (C(z)f^2 + B(z)f + A(Z)) - \frac{1}{v^2}v' \tag{1.27}$$

From equation (1.20)

$$y' = C(z)y^2 + B(z)y + A(z) = C(z)(f + \frac{1}{v})^2 + B(z)(f + \frac{1}{v}) + A(z) \tag{1.28}$$

Equating equations (1.27) and (1.28) yields

$$-\frac{1}{v^2}v' = C(z)\frac{1}{v^2} + 2fC(z)\frac{1}{v} + B(z)\frac{1}{v}. \tag{1.29}$$

Simplifying yields:

$$v' = (B(z) + 2C(z)A)v = -A \tag{1.30}$$

which is a linear differential equation in v.

We come across Riccati equations quite often when working with the generating function of birth and death processes.

1.3.2 Second Order Differential Equations

The most general linear second order differential equation is given by

$$A(x)\frac{d^2y}{dx} + B(x)\frac{dy}{dx} + C(x)y = G(x). \tag{1.31}$$

In order to solve equation (1.31), we will need to solve the homogeneous and in-homogeneous problems.

Chapter 2

Birth and Death Processes

2.1 Introduction

The theory of birth and death processes was developed in the 20th century. It was initially used to model the growth of a population with stochastic demographic factors. Over the past few decades, this area has developed into many branches of stochastic analysis. The first models considered by Yule, Feller and Kendal provide a framework for the many applications in biology. Birth and death processes are Markov chains with the property that only transitions to “neighbouring” states can be made. These processes are an important tool for modeling biological data, queuing and economical sciences (Allen (2010), Beichelt (2016), Goel and Richter-Dyn (2013)). The name comes from biology. Beichelt (2016) models the development of the number of individuals in populations. Goel and Richter-Dyn (2013) demonstrates the usefulness of the theory of stochastic processes in understanding biological phenomena. Their monograph, Stochastic models in biology, uses two approaches in making probabilistic models which analyse biological systems. The first approach assumes the model is described by a set if random variables which either have a discrete or continuous state space, while time is continuous. The second approach is used for modelling complex biological systems.

Allen (2010), Beichelt (2016), Goel and Richter-Dyn (2013) provide a detailed foundation on birth and death processes including examples and applications. Pure birth and pure death processes, a subset of birth and death processes are defined as processes where transitions are only possible to the ‘next’ and ‘previous’ state respectively. Restrictions can be set on the first and last state of a birth and death process. This allows for different applications to real-life scenarios such as immigration. These states can either be absorbing or reflecting. In models with absorbing states, the mean time to absorption is of interest. Other tractable quantities one may extract from birth and death processes are the equilibrium of the system, also known as the steady state and the first passage time which is the time it takes to reach a state for the first time.

Birth and death processes with transition probabilities which are time-independent are known as homogeneous Markov chains. For this reason, one usually refers to such processes as time-homogeneous or having stationary transition probabilities. On the other hand, a non-homogeneous birth and death process has birth and death rates which depend on time.

Other important applications of birth and death processes are in queuing theory. The birth and deaths are seen as arrivals and departures at a system. The applications of queuing theory are used in engineering disciplines such as reliability and maintenance. It is also used in economics to model staff fluctuations, number of enterprises (creation and bankruptcies), among others (Beichelt (2016)).

2.2 Discrete-Time Birth and Death Processes

2.2.1 Definition

Let X_n denote the size of a population at time n, where the state space is either finite $\{0, 1, 2, ...N\}$ or infinite $\{0, 1, 2, ...\}$. Assume transition probabilities

$$p_{ji} = P\{X_{n+1} = j | X_n = i\} \tag{2.1}$$

with

$$p_{ji} = \begin{cases} p_i, & \text{if } j = i + 1 \\ q_i & \text{if } j = i - 1 \\ r_i & \text{if } j = i \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

$$\text{where } p_i + q_i + r_i = 1. \tag{2.3}$$

Then $\{X_n\}$ is a homogeneous birth and death process with birth rates p_i and death rates q_i. The transition matrix P for the finite Markov Chain has the following form:

$$P = \begin{pmatrix} 1 & q_1 & 0 & \dots & 0 & 0 \\ 0 & r_1 & q_2 & \dots & 0 & 0 \\ 0 & p_1 & r_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & r_{N-1} & q_N \\ 0 & 0 & 0 & \dots & p_{N-1} & r_N \end{pmatrix} \tag{2.4}$$

In general, so that P is a stochastic matrix, condition (2.3) must hold.

The transitions are illustrated by the following figure:

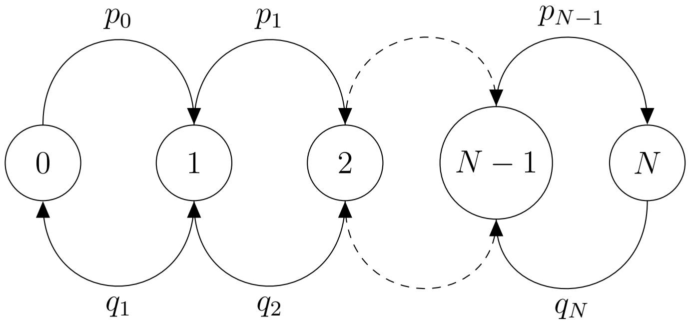

Figure 1.1 Transition graph from state 0 *to* N *of a birth and death process*

Definition: Transient and Recurrent states

A state i is said to be recurrent if $f_{ii} = 1$ and transient if $f_{ii} < 1$, where

$$f_{ij} = \sum_{m=1}^{\infty} f_{ij}^{(m)}$$

and $f_{ij}^{(m)}$ is the probability that a Markov chain starting from state i, first transitions to state j after m steps.

Theorem 2.1

Let $N = \infty$. If $p_i > 0$ for $i = 0, 1, 2, ...$ and $q_i > 0$ for $i = 1, 2, ...$, then a birth and death process is transient iff

$$\sum_{i=1}^{\infty} \frac{q_1 q_2 ... q_{i-1} q_i}{p_1 p_2 ... p_{i-1} p_i} < \infty \tag{2.5}$$

Proof

Let α_n denote the probability that a birth and death process starting at state $n \in \{0, 1, 2, ...\}$ ever returns to state 0, Then we have

$$\alpha_n = P\{X_i = 0 \text{ for some } i \geq 1 | X_0 = n\}$$

$$\alpha_n = \sum_k P\{X_i = 0 \text{ for some } i \geq 1 | x_1 = k\} P\{X_1 = k | X_0 = n\}$$

$$\alpha_n = p_n \alpha_{n+1} + q_n \alpha_{n-1} + (1 - p_n - q_n)\alpha_n \tag{2.6}$$

which yields the relation:

$$(p_n + q_n)\alpha_n = p_n\alpha_{n+1} + q_n\alpha_{n-1} \tag{2.7}$$

Then we have the recursive relation:

$$\alpha_n - \alpha_{n+1} = \frac{q_n}{p_n}(\alpha_{n-1} - \alpha_n), \quad n = 1, 2, ..., N \tag{2.8}$$

Iterating equation (2.8) yields:

$$\alpha_n - \alpha_{n+1} = \frac{q_1 ... q_n}{p_1 ... p_k} + 1 \tag{2.9}$$

Finally, we have

$$\alpha_{n+1} = (\alpha_n - 1)\sum_{k=1}^{n} \frac{q_1 ... q_k}{p_1 ... p_k} + 1 \tag{2.10}$$

The following term converges, since it is a finite sum of the product of finite probabilities:

$$\sum_{k=1}^{n} \frac{q_1 ... q_k}{p_1 ... p_k} < \infty \tag{2.11}$$

Then,

$$\alpha_1 = \sum_{k=1}^{\infty} \frac{q_1 ... q_k}{p_1 ... p_k} \Big/ \sum_{k=0}^{\infty} \frac{q_1 ... q_k}{p_1 ... p_k} \tag{2.12}$$

so that

$$\alpha_{n+1} = \frac{1}{\sum_{k=0}^{\infty} \frac{q_1 ... q_k}{p_1 ... p_k}} \sum_{k=n+1}^{\infty} \frac{q_1 ... q_k}{p_1 ... p_k} \to 0 \text{ as } n \to \infty \tag{2.13}$$

This completes the proof $\square$

2.2.2 Expected Time to Extinction

Let there be a population with initial size k. Then τ_k denotes the expected time until extinction for that population with $\tau_0 = 0$. The following relationship holds for τ_k, $k = 1, 2, ...$

$$\tau_k = p_k(1 + \tau_{k+1}) + q_k(1 + \tau_{k-1}) + (1 - (p_k + q_k))(1 + \tau_k) \tag{2.14}$$

If the maximal population size is finite, then for $k = N$,

$$\tau_N = q_N(1 + \tau_{N-1}) + (1 - q_N)(1 + \tau_N) \tag{2.15}$$

The difference equation can be simplified as follows:

$$q_k\tau_{k-1} - (p_k + q_k)\tau_k + p_k\tau_{k+1} = -1 \tag{2.16}$$

for $k = 1, 2, ...$

Theorem 2.2

Suppose $\{X_n, n \in 0, 1, ...\}$ is a birth and death process with $X_0 = m \geq 1$ satisfying $p_0 = 0 = q_0$, $p_i > 0$ for $i = 1, 2, ..., N - 1$ and $q_i > 0$ for $i = 1, 2, ..., N$. The expected time until population extinction is

$$\tau_m = \begin{cases} \dfrac{1}{q_1} + \sum_{i=2}^{N} \dfrac{p_1...p_i}{q_1...q_i} & \text{if } m = 1 \\ \tau_1 + \sum_{s=1}^{m-1} \left[\dfrac{q_1...q_s}{p_1...p_s} \sum_{i=s+1}^{N} \dfrac{p_1...p_{i-1}}{q_1...q_i} \right] & \text{if } m = 2, 3, ..., N \end{cases} \tag{2.17}$$

Proof

For $k = 1, 2, ..., N - 1$, then $q_k\tau_{k-1} - (p_k + q_k)\tau_k + p_k\tau_{k+1} = -1$ can be solved recursively for $\tau_2, ..., \tau_N$ to obtain the formulas

$$\tau_m = \tau_1 + \sum_{k=1}^{m-1} \frac{q_1...q_k}{p_1...p_k} \left[\tau_1 - \frac{1}{q_1} - \sum_{i=2}^{k} \frac{p_1...p_{i-1}}{q_1...q_i} \right] \tag{2.18}$$

for $m = 2, ..., N$. The second summation is zero when $k < 2$, then applying the relation (2.16) when $k = N$:

$$q_N \tau_{N-1} - q_N \tau_N = -1 \tag{2.19}$$

we get

$$\tau_N = \frac{1}{q_N} + \tau_{N-1} \tag{2.20}$$

and equating the two expressions of τ_N, the following formula for τ_1 is obtained:

$$\tau_1 = \frac{1}{q_1} + \sum_{i=2}^{N} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} \tag{2.21}$$

Substituting τ_1 into equation (2.18) above, we get

$$\tau_m = \frac{1}{q_1} + \sum_{i=2}^{N} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} + \sum_{k=1}^{m-1} \frac{q_1 ... q_k}{p_1 ... p_k} \left[\frac{1}{q_1} + \sum_{i=2}^{N} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} - \frac{1}{q_1} - \sum_{i=2}^{k} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} \right]$$

$$\tau_m = \frac{1}{q_1} + \sum_{i=2}^{N} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} + \sum_{k=1}^{m-1} \frac{q_1 ... q_k}{p_1 ... p_k} \left[\sum_{i=2}^{N} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} - \sum_{i=2}^{k} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} \right]$$

The second summation in the bracket cancels out with the first k terms in the first summation, giving:

$$\tau_m = \frac{1}{q_1} + \sum_{i=2}^{N} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} + \sum_{k=1}^{m-1} \frac{q_1 ... q_k}{p_1 ... p_k} \left[\sum_{i=k+1}^{N} \frac{p_1 ... p_{i-1}}{q_1 ... q_i} \right]$$

Simplifying further, we get the formula stated in the theorem. $\square$

Example 1: Gambler's Ruin

Two gamblers initially have stakes k and $m - k$ respectively, with $\{k, z | 0 < k < m, k \in \mathbb{Z}\}$. After each move, a gambler can either win or lose \$1. The game will end if either gambler wins or loses everything.

Let the state space be defined as $\{0, 1, 2, ...\}$ and let $0 < p = 1 - q < 1$.

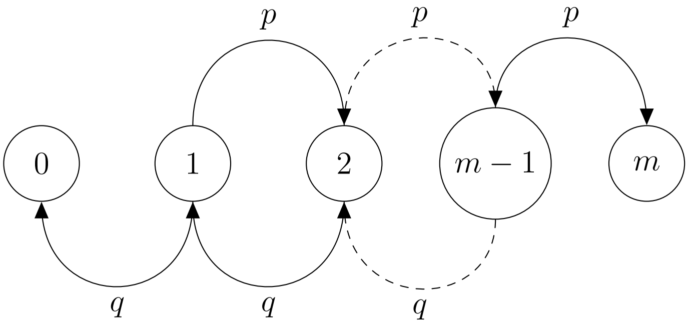

Figure 1.2 Transition graph of Gambler's ruin

The transition probabilities are as follows:

$$p_{00} = 1 \quad p_{mm} = 1$$

$$p_{i,i-1} = q \qquad p_{i,i+1} = p$$

for $i = 1, 2, ...$

We wish to derive the probability, $P_i(0)$, that a gambler is ruined on time i, $0 < i < m$. We have by applying the total law of probability,

$$P_i(0) = pP_{i+1}(0) + qP_{i-1}(0), \quad i = 1, 2, ..., m-1$$

Replacing $P_i(0)$ with $pP_i(0) + qP_i(0)$ yields the following:

$$[P_i(0) - P_{i+1}(0)] = \frac{q}{p}[P_{i-1}(0) - P_i(0)]$$

Using the above relationship recursively gives the following:

$$[P_{m-1}(0) - P_m(0)] = \left(\frac{q}{p}\right)^{m-1}[1 - P_1(0)]$$

Using the boundary conditions, $p_{00} = 1$ and $p_{mm} = 1$, we get

$$\sum_{i=1}^{m} [P_i(0) - P_{i+1}(0)] = 1$$

Solving as a finite geometric series

$$\sum_{i=1}^{m} [P_i(0) - P_{i+1}(0)] = [1 - P_1(0)] \sum_{i=1}^{m} \left(\frac{q}{p}\right)^{i-1} = [1 - P_1(0)] \frac{1 - \left(\frac{q}{p}\right)^m}{1 - \frac{q}{p}} = 1$$

Solving this equation for $P_1(0)$ gives

$$P_1(0) = \frac{\left(\frac{q}{p}\right)^m - \frac{q}{p}}{\left(\frac{q}{p}\right)^m - 1}$$

Starting with $P_0(0) = 1$ and $P_1(0)$, we get

$$P_i(0) = \frac{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^m - 1}, \quad i = 1, 2, ..., m, p \neq q$$

The probability of winning, which is getting the full m when starting at i, is given by $Q_i(m)$, where

$$Q_i(m) = 1 - P_i(0)$$

□

Example 2: Birth and Death process conditional on non-extinction

When the expected time until extinction or absorption is large, the dynamics of the process prior to extinction is examined.

Let $\{X_n, n \in 0, 1, ...\}$ denote a general birth and death process with

$$P_i(n) = P\{X_n = i | i \in 0, 1, ...\} \tag{2.22}$$

Define a conditional probability as

$$Q_i(n) = P\{X_n = i | X_j \neq 0, j \in 0, 1, ..., n\}, \quad i \neq 0 \tag{2.23}$$

Using Bayes' Theorem on conditional probability, we get

$$Q_i(n) = \frac{P_i(n)}{1 - P_0(n)}$$

The distribution $Q(n) = \begin{bmatrix} Q_1(n) \\ Q_2(n) \\ \vdots \\ Q_N(n) \end{bmatrix}$ defines a probability distribution since

$$\sum_{i=1}^{N} Q_i(n) = \frac{\sum_{i=1}^{N} P_i(n)}{1 - P_0(n)} = \frac{1 - P_0(n)}{1 - P_0(n)} = 1$$

2.3 Continuous-Time Birth and Death Processes

Let the increment of a stochastic process, $\{X(t), t \geq 0\}$ in the time interval $[t, t + \Delta t]$, be denoted by $\Delta X(t)$.

$$\Delta X(t) = X(t + \Delta t) - X(t) \tag{2.24}$$

2.3.1 Homogeneous Birth and Death Process

A homogeneous continuous-time birth and death process is a continuous-time Markov chain $\{X(t) : t \in [0, \infty]\}$ with finite space $\{0, 1, 2, ..., N\}$ or infinite space $\{0, 1, 2, ...\}$ and infinitesimal transition probabilities:

$$p_{i+j,i}(\Delta t) = P\{\Delta X(t) = j | X(t) = i\} \tag{2.25}$$

$$p_{i+j,i}(\Delta t) = \begin{cases} \lambda_i \Delta t + o(\Delta t) & \text{if } j = 1 \\ \mu_i \Delta t + o(\Delta t) & \text{if } j = -1 \\ 1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t) & \text{if } j = 0 \\ o(\Delta t) & \text{if } j \neq -1, 0, 1 \end{cases} \tag{2.26}$$

The forward Kolmogorov differential equations for $p_{ji}(t + \Delta t)$ can be derived directly from (2.26). That is

$$p_{ji}(t) = p_{j-1,i}(t)[\lambda_{j-1}\Delta t + o(\Delta t)] + p_{j+1,i}(t)[\mu_{j+1}\Delta t + o(\Delta t)]$$

$$+p_{ji}(t)[1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t)] + \sum_{k \neq -1,0,1}^{\infty} p_{j+k,i}(t)o(\Delta t) \tag{2.27}$$

$$p_{ji}(t+\Delta t) = p_{j-1,i}(t)\lambda_{j-1}\Delta t + p_{j+1,i}(t)\mu_{j+1}\Delta t + p_{ji}(t)[1-(\lambda_j+\mu_j)\Delta t] + o(\Delta t) \tag{2.28}$$

which holds for all i and j in the state space except for $j = 0$ and $j = N$ (if the population size is finite).
If $j = 0$, then

$$p_{0i}(t+\Delta t) = p_{1,i}(t)\mu_1\Delta t + p_{0i}(t)[1-\lambda_0\Delta t] + o(\Delta t) \tag{2.29}$$

If $j = N$ is the maximum population size, then

$$p_{Ni}(t+\Delta t) = p_{N-1,i}(t)\lambda_{N-1}\Delta t + p_{Ni}(t)[1-\mu_N\Delta t] + o(\Delta t) \tag{2.30}$$

where $\lambda_N = 0$ and $p_{kN}(t) = 0$ for all $k > N$. Subtracting $p_{ji}(t)$, $p_{0i}(t)$ and $p_{Ni}(t)$ from the preceding three equations respectively, dividing by Δt and taking the limit as $\Delta t \to 0$, yields the forward Kolmogorov differential equations for the general birth and death process,

$$\begin{aligned} p_0'(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t) \\ p_j'(t) &= \lambda_{j-1}p_{j-1}(t) - (\lambda_j+\mu_j)p_j(t) + \mu_{j+1}p_{j+1}(t),\ j = 1, 2, ... \\ p_N'(t) &= \lambda_{N-1}p_{N-1}(t) - \mu_N p_N(t) \end{aligned} \tag{2.31}$$

To further determine state probabilities, we use their respective probability generating function.

$$M(t,z) = \sum_{i=0}^{\infty} p_i(t) z^i \tag{2.32}$$

using the initial condition

$$p_n(0) = P(X(0) = n) = 1 \tag{2.33}$$

Equation (2.33) is equivalent to

$$M(0, z) = z^n, n = 0, 1, ... \tag{2.34}$$

when written in terms of the probability generating function.

The partial derivatives of the probability generating function are:

$$\frac{\partial M(t, z)}{\partial t} = \sum_{i=0}^{\infty} p_i'(t) z^i \tag{2.35}$$

and

$$\frac{\partial M(t, z)}{\partial z} = \sum_{i=1}^{\infty} i p_i(t) z^{i-1} \tag{2.36}$$

Linear Birth and Death Process

$\{X(t), t \geq 0\}$ is called a linear birth and death process if it has transition rates

$$\lambda_i = i\lambda \qquad\qquad \mu_i = i\mu \tag{2.37}$$

for $i = 0, 1, ...$

This process is analysed on condition that

$$p_1(0) = P(X(0) = 1) = 1 \tag{2.38}$$

The system of differential equations of (2.31) becomes

$$\begin{aligned} &p_0'(t) = \mu p_1(t), \\ &p_j'(t) = (j-1)\lambda p_{j-1}(t) - j(\lambda+\mu)p_j(t) + (j+1)\mu p_{j+1}(t); \quad j = 1, 2, ... \end{aligned} \tag{2.39}$$

Multiplying the j^{th} differential equation by z^j and summing from $j = 0$ to $j = \infty$, taking the partial derivatives of the probability generating function into account yields the following linear homogeneous partial differential equation in $M(t, z)$:

$$\frac{\partial M(t,z)}{\partial t} - (z-1)(\lambda z - \mu)\frac{\partial M(t,z)}{\partial z} = 0 \tag{2.40}$$

The corresponding characteristic differential equation is a Riccati differential equation with constant coefficients:

$$\frac{dz}{dt} = -(z-1)(\lambda z - \mu) = -\lambda z^2 + (\lambda + \mu)z - \mu \tag{2.41}$$

a) $\boldsymbol{\lambda \neq \mu}$

Separating variables in (2.41), we get

$$\frac{dz}{(z-1)(\lambda z - \mu)} = -dt \tag{2.42}$$

Integrating both sides of (2.42) gives:

$$-\frac{1}{\lambda - \mu}\ln\left(\frac{\lambda z - \mu}{z-1}\right) = -t + C \tag{2.43}$$

The general solution $z = z(t)$ of the characteristic differential equation in implicit form is given by

$$c = (\lambda - \mu)t - \ln\left(\frac{\lambda z - \mu}{z - 1}\right) \tag{2.44}$$

where c is an arbitrary constant. The general solution $M(t, z)$ has structure

$$M(t, z) = f\Bigg((\lambda - \mu)t - \ln\left(\frac{z - 1}{\lambda z - \mu}\right)\Bigg) = z, \tag{2.45}$$

where f can be any function with a continuous derivative. To determine f, we make use of the initial condition

$$M(0, z) = f\Bigg(\ln\left(\frac{\lambda z - \mu}{z - 1}\right)\Bigg) = z$$

Thus, f must have structure

$$f(x) = \frac{\mu e^x - 1}{\lambda e^x - 1} \tag{2.46}$$

Thus $M(t, z)$ is

$$M(t, z) = \frac{\mu exp\Bigg\{(\lambda - \mu)t - \ln\left(\frac{\lambda z - \mu}{z - 1}\right)\Bigg\} - 1}{\lambda exp\Bigg\{(\lambda - \mu)t - \ln\left(\frac{\lambda z - \mu}{z - 1}\right)\Bigg\} - 1} \tag{2.47}$$

After simplification, $M(t, z)$ becomes

$$M(t, z) = \frac{\mu[1 - e^{(\lambda-\mu)t}] - [\lambda - \mu e^{(\lambda-\mu)t}]z}{[\mu - \lambda e^{(\lambda-\mu)t}] - \lambda[1 - \mu e^{(\lambda-\mu)t}]z} \tag{2.48}$$

We can expand $M(t, z)$ as a power series in z. To get the state probabilities, $p_j(t)$, we extract the coefficient of z^j:

$$p_0(t) = \frac{1 - e^{(\lambda-\mu)t}}{1 - \left(\frac{\lambda}{\mu}\right) e^{(\lambda-\mu)t}} \tag{2.49}$$

$$p_j(t) = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{j-1} \frac{\left[1 - e^{(\lambda-\mu)t}\right]^{j-1}}{\left[1 - \left(\frac{\lambda}{\mu}\right) e^{(\lambda-\mu)t}\right]^{j+1}} e^{(\lambda-\mu)t}, \quad j = 1, 2, ... \tag{2.50}$$

State 0 is absorbing therefore, $p_0(t)$ is the probability that the population is extinct at time t

$$\lim_{t\to\infty} p_0(t) = \begin{cases} 1 & \text{for } \lambda < \mu \\ \dfrac{\mu}{\lambda} & \text{for } \lambda > \mu \end{cases} \tag{2.51}$$

Thus, in the cases where $\lambda < \mu$, the distribution function of the lifetime L of the population is

$$P(L \leq t) = p_0(t) = \frac{1 - e^{(\lambda-\mu)t}}{1 - \left(\frac{\lambda}{\mu}\right) e^{(\lambda-\mu)t}}, \quad t \geq 0 \tag{2.52}$$

The population will survive in the interval $[0, t]$ with probability

$$P(L > t) = 1 - p_0(t) \tag{2.53}$$

Since

$$E[X] = \int_0^\infty [1 - F(x)]dx$$

for any non-negative random variables X,

$$E[L] = \int_0^\infty 1 - p_0(t)dt \tag{2.54}$$

From this, we get

$$E[L] = \frac{1}{\mu - \lambda} \ln \left(2 - \frac{\lambda}{\mu} \right) \tag{2.55}$$

The trend function $m(t)$ can be calculated as follows

$$m(t) = E[X] = \sum_{j=0}^{\infty} j p_j(t) \tag{2.56}$$

We can also derive the mean from the probability generating function:

$$m(t) = \left. \frac{\partial M(t,z)}{\partial z} \right|_{z=1} \tag{2.57}$$

Multiplying the jth differential equation of the process by j and summing from $j = 0$ to ∞ yields the following first-order differential equation

$$m'(t) = (\lambda - \mu) m(t) \tag{2.58}$$

Applying the initial condition $p_1(0) = 1$, its solution is

$$m(t) = e^{(\lambda - \mu)t} \tag{2.59}$$

Multiplying the j^{th} differential equation by j^2 and summing from $j = 0$ to ∞, a second order differential equation in $Var(X(t))$ is obtained. The solution is

$$Var(X(t)) = \frac{\lambda + \mu}{\lambda - \mu} \left[1 - e^{-(\lambda - \mu)t} \right] e^{2(\lambda - \mu)t} \tag{2.60}$$

b) $\boldsymbol{\lambda = \mu}$

In this case the characteristic differential equation simplifies to

$$\frac{dz}{\lambda(z-1)^2} = -dt \tag{2.61}$$

Integrating both sides yields

$$c = \lambda t - \frac{1}{z-1} \tag{2.62}$$

where c is an arbitrary constant. Therefore $M(t,z)$ has structure:

$$M(t,z) = f\left(\lambda t - \frac{1}{z-1}\right) \tag{2.63}$$

where f is a continuously differentiable function. Since $p_1(0) = 1$, f satisfies

$$f\left(-\frac{1}{z-1}\right) = z \tag{2.64}$$

Hence, the function f is given by

$$f(x) = 1 - \frac{1}{x}, x \neq 0 \tag{2.65}$$

The corresponding generating function is

$$M(t,z) = \frac{\lambda t + (1-\lambda t)z}{1 + \lambda t - \lambda t z} \tag{2.66}$$

Expanding $M(t,z)$ as a power series in z yields the absolute state probabilities:

$$p_0(t) = \frac{\lambda t}{1+\lambda t},$$
$$p_j(t) = \frac{(\lambda t)^{j-1}}{(1+\lambda t)^{j+1}}, \tag{2.67}$$

for $j = 1, 2, ..., t \geq 0$

An equivalent form of the state probabilities is

$$p_0(t) = \frac{\lambda t}{1+\lambda t},$$
$$p_j(t) = [1 - p_0(t)]^2 [p_0(t)]^{j-1}, \tag{2.68}$$

for $j = 1, 2, ..., t \geq 0$

Mean value and variance of $X(t)$ are

$$E[X(t)] = 1$$

$$Var(X(t)) = 2\lambda t \tag{2.69}$$

Proof

From (1.8), we have

$$E[X(t)] = \sum_{j=1}^{\infty} j[1 - p_0(t)]^2 [p_0(t)]^{j-1}$$

$$\frac{E[X(t)]}{[1 - p_0(t)]^2} = \sum_{j=1}^{\infty} j[p_0(t)]^{j-1}$$

$$\frac{E[X(t)]}{[1-p_0(t)]^2}[1-p_0(t)] = \frac{1}{[1-p_0(t)]}$$

$$E[X(t)] = 1$$

From (1.9), we have

$$Var(X(t)) = E[X^2(t)] - (E[X(t)])^2$$

$$E[X^2(t)] = \sum_{j=1}^{\infty} j^2[1-p_0(t)]^2[p_0(t)]^{j-1}$$

$$\frac{E[X^2(t)]}{[1-p_0(t)]^2} = \sum_{j=1}^{\infty} j^2[p_0(t)]^{j-1}$$

$$\frac{E[X^2(t)]}{[1-p_0(t)]^2}[1-p_0(t)]^2 + 1 = \sum_{j=1}^{\infty} 2[p_0(t)]^{j-1}$$

$$E[X^2(t)] = \frac{2}{1-p_0(t)} - 1$$

$$E[X^2(t)] = \frac{1+p_0(t)}{1-p_0(t)}$$

Thus

$$Var(X(t)) = \frac{1+p_0(t)}{1-p_0(t)} - 1$$

$$Var(X(t)) = \frac{2p_0(t)}{1-p_0(t)} = 2\lambda t$$

2.3.2 Stationary Distribution

Let $\{\pi_i, i \in \mathbb{Z}\}$ be a stationary distribution of the Markov chain $\{X(t), t \geq 0\}$. The absolute distribution must satisfy (2.31), the Kolmogorov equations.
The left hand side of the equations are equal to 0, since the π_i are constant.
The system of differential equations simplifies to a system of linear algebraic equations in the unknown, π_i.
If the stationary distribution will satisfy the following system of equations, if it exists:

$$\begin{aligned}
&\lambda_0\pi_0 - \mu_1\pi_1 = 0 \\
&\lambda_{i-1}\pi_{i-1} - (\lambda_i + \mu_i)\pi_i + \mu_{i+1}\pi_{i+1} = 0, \text{ for } \quad i = 1, 2, ... \\
&\lambda_{n-1}\pi_{n-1} - \mu_n\pi_n = 0, \text{ for } n \leq \infty \qquad (2.70)
\end{aligned}$$

Theorem 2.3
Let $\{X(t) : t \in [0, \infty)\}$, the continuous-time Markov chain, be a homogeneous birth and death process with infinitesimal transition probabilities given by the equations (2.2).
a) If the state space is infinite, $\{0, 1, 2, ...\}$, a unique positive stationary probability distribution π exists iff

$$\sum_{i=1}^{\infty} \frac{\lambda_0\lambda_1...\lambda_{i-1}}{\mu_1\mu_2...\mu_i} < \infty \qquad (2.71)$$

and

$$\mu_i > 0 \text{ and } \lambda_{i-1} > 0 \text{ for } i = 0, 1, 2... \qquad (2.72)$$

The stationary probability distribution equals

$$\pi_i = \frac{\lambda_0 \lambda_1 ... \lambda_{i-1}}{\mu_1 \mu_2 ... \mu_i} \pi_0, \ i = 1, 2, ..., \tag{2.73}$$

and

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 ... \lambda_{i-1}}{\mu_1 \mu_2 ... \mu_i}} \tag{2.74}$$

b) If the state space is finite, $\{0, 1, 2, ..., N\}$, then a unique positive stationary probability distribution π exist iff

$$\mu_i > 0 \text{ and } \lambda_{i-1} > 0 \text{ for } i = 0, 1, 2, ..., N$$

The stationary distribution is given by (2.73) and (2.74), where the summation goes from $i = 0, 1, 2, ..., N$

Proof

The explicit equations for the stationary distribution are as (2.67). These equations can be recursively solved:

$$\pi_1 = \frac{\lambda_0}{\mu_1} \pi_0. \tag{2.75}$$

Then

$$\begin{aligned} \mu_2 \pi_2 &= (\lambda_1 + \mu_1) \pi_1 - \lambda_0 \pi_0 \\ \mu_2 \pi_2 &= \left[\frac{(\lambda_1 + \mu_1) \lambda_0}{\mu_1} - \lambda_0 \right] \pi_0 \\ \pi_2 &= \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \pi_0 \end{aligned} \tag{2.76}$$

Applying the induction hypothesis, assume π_j is given by

$$\pi_i = \frac{\lambda_1 \lambda_1 .... \lambda_{i-1}}{\mu_1 \mu_2 ... \mu_i} \pi_0$$

for $j = 1, 2, ..., i$

Then

$$\begin{aligned}
\mu_{i+1}\pi_{i+1} &= (\lambda_i + \mu_i)\pi_i - \lambda_{i-1}\pi_{i-1} \\
\mu_{i+1}\pi_{i+1} &= \left[\frac{\lambda_0\lambda_1...\lambda_{i-1}(\lambda_i + \mu_i)}{\mu_1\mu_2...\mu_i} - \frac{\lambda_0\lambda_1...\lambda_{i-1}}{\mu_1\mu_2...\mu_{i-1}}\right]\pi_0 \\
\mu_{i+1}\pi_{i+1} &= \frac{\lambda_0\lambda_1...\lambda_{i-1}}{\mu_1\mu_2...\mu_{i-1}}\left[\frac{\lambda_i + \mu_i}{\mu_i} - 1\right]\pi_0 \\
\pi_{i+1} &= \frac{\lambda_0\lambda_1...\lambda_i}{\mu_1\mu_2....\mu_{i+1}}\pi_0 \qquad (2.77)
\end{aligned}$$

For the infinite case, we apply an additional constraint

$$\sum_{i=0}^{\infty} \pi_i = 1 \text{ or } \pi_0 \left(1 + \sum_{i=1}^{\infty} \frac{\pi_i}{\pi_0}\right) = 1.$$

Then

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda_0\lambda_1...\lambda_{i-1}}{\mu_1\mu_2...\mu_i}}$$

A unique positive stationary distribution exists iff the conditions (2.71) and (2.72) are satisfied. □

2.3.3 Quasi-Stationary Probability Distribution

In birth and death models with an absorbing state, there is no stationary distribution. An approximate stationary distribution, known as the quasi-stationary distribution can be determined prior to extinction. When $\lim_{t\to\infty} p_0(t) = 1$, the probability distribution of $X(t)$ can be approximately stationary for a very long time.

Denote the probability distribution associated with $X(t)$ conditioned on nonextinction as $q_i(t)$. Then

$$q_i(t) = \frac{p_i(t)}{1 - p_0(t)}, \ i = 1, 2, ... \tag{2.78}$$

The quasi-stationary probabilities are solutions of a system of differential equations similar to the forward Kolmogorov differential equations.

$$\begin{aligned} \frac{dq_i}{dt} &= \frac{\frac{dp_i}{dt}}{1 - p_0} + \frac{p_i}{(1 - p_0)} \frac{\frac{dp_0}{dt}}{(1 - p_0)} \\ \frac{dq_i}{dt} &= \lambda_{i-1} q_{i-1} - (\lambda_i + \mu_i) q_i + \mu_{i+1} q_{i+1} + q_i \mu_1 q_1, \end{aligned} \tag{2.79}$$

where it is assumed that $\lambda_0 = 0 = \mu_0$.

The quasi-stationary probability distribution can be approximated by making the assumption $\mu_1 = 0$. Then $\frac{dq}{dt} = \tilde{Q}q$, where

$$\tilde{Q} = \begin{pmatrix} -\lambda_1 & \mu_2 & 0 & \cdots \\ \lambda_1 & -\lambda_2 - \mu_2 & \mu_3 & \cdots \\ 0 & \lambda_2 & -\lambda_3 - \mu_3 & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix} \tag{2.80}$$

A unique positive stationary distribution of the system is given by $\tilde{\pi} = (\tilde{\pi}_1, \tilde{\pi}_2, ...)^{tr}$, if the assumptions of above theorem is satisfied.

The stationary distribution is:

$$\tilde{\pi}_i = \frac{\lambda_1 \lambda_2 ... \lambda_{i-1}}{\mu_2 \mu_3 ... \mu_i} \tilde{\pi}_1 \tag{2.81}$$

and $\sum_{i=1}^{\infty} \tilde{\pi}_i = 1$.

Therefore, a unique solution exists if

$$\sum_{i=0}^{\infty} \frac{\lambda_1 \lambda_2 ... \lambda_{i-1}}{\mu_2 \mu_3 ... \mu_i} < \infty \tag{2.82}$$

The solution approximates the quasi-stationary distribution.

2.3.4 Birth Process

A continuous Markov chain with transitions only from state i to $i+1$, for all $i = 0, 1, 2, ..., n-1$, is called a pure birth process. The state n is absorbing if $n < \infty$. The population can only increase in size.

Then, $q_{i,i+1}$ is the positive transition rates of a birth and death process. They are called the birth rates and have the following notation:

$$\lambda_i = p_{i,i+1}, \quad i = 0, 1, ..., n-1, \tag{2.83}$$

$$\lambda_n = 0 \quad \text{for} \quad n < \infty \tag{2.84}$$

Given the initial distribution

$$p_N(0) = P(X(0) = N) = 1 \tag{2.85}$$

the absolute state probabilities $p_i(t)$ are equal to the transition probabilities p_{Ni}. The $p_i(t)$ are identically equal to 0 for $i < N$ and, the probabilities $p_i(t) = P\{X(t) = i\}$ are the solutions of the forward Kolmogorov differential equations, $\frac{dp}{dt} = Qp$, where Q is the generator matrix

$$Q = \begin{pmatrix} -\lambda_N & \mu_{N+1} & 0 & \cdots \\ \lambda_N & -\lambda_{N+1} - \mu_{N+1} & \mu_{N+2} & \cdots \\ 0 & \lambda_{N+1} & -\lambda_{N+2} - \mu_{N+2} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix} \tag{2.86}$$

and

p'$_i(t) = \lambda(i-1)p_{i-1}(t) - \lambda_i p_i(t), \quad i = N, N+1, ...$

p'$_n(t) = \lambda_{n-1}p_{n-1}(t), \quad n < \infty$

p'$_i(t) = 0, \quad i = 0, 1, 2, ..., N-1$

From the first differential equation in (2.87),

$$p_N(t) = e^{-\lambda_N t}, t \geq 0$$

For $N+1 < i < \infty$ the differential equations in (2.87) are equivalent to

$$e^{\lambda_i t}(p_i'(t) + \lambda_i p_j(t)) = \lambda_{i-1}e^{\lambda_i t}p_{i-1}(t)$$

or

$$\frac{d}{dt}(e^{\lambda_i t}p_i(t)) = \lambda_{i-1}e^{\lambda_i t}p_{i-1}(t)$$

By integration,

$$p_i(t) = \lambda_{i-1}e^{-\lambda_i t}\int_0^t e^{\lambda_i x}p_{i-1}(x)dx$$

These formulas allow the successive calculation of the probabilities $p_i(t)$ for $i = N+1, N+2, ...$ With the conditions $p_0(0) = 1$ and $\lambda_0 \neq \lambda_1$

$$\begin{aligned} p_1(t) &= \lambda_0 e^{-\lambda - 1t}\int_0^t e^{\lambda_1 x}e^{\lambda_0 x}dx \\ &= \lambda_0 e^{-\lambda_1 t}\int_0^t e^{-(\lambda_0 - \lambda_1)x}dx \\ &= \frac{\lambda_0}{\lambda_0 - \lambda_1}(e^{-\lambda_1 t} - e^{-\lambda_0 t}), \quad t \geq 0 \end{aligned}$$

If the birth rates are all different, then by induction:

$$p_j(t) = \sum_{i=0}^{j} C_{ij}\lambda_i e^{-\lambda_i t},\ j = 0, 1, ...,$$

where

$$C_{ij} = \frac{1}{\lambda_j}\prod_{k=0,k\neq i}^{j}\frac{\lambda_k}{\lambda_k - \lambda_i}, 0 \leq i \leq j,\ C_{00} = \frac{1}{\lambda_0}$$

Linear Birth Processes

A pure birth process is called a linear birth process or a Yule-Furry process if the birth rates are given by

$$\lambda_i = i\lambda, i = 0, 1, 2, ...$$

Assuming $p_1 = P(X(0) = 1) = 1$, the system of differential equations (2.75) becomes

$$p_i' = -\lambda[ip_i(t) - (i-1)p_{i-1}(t)],\ \ i = 1, 2, .. \tag{2.87}$$

with

$$p_1(0) = 1,\ \ p_i(0) = 0,\ \ i = 2, 3, ... \tag{2.88}$$

The solution of (2.76) under the initial condition of (2.77) is

$$p_i(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{i-1},\ \ i = 1, 2, ... \tag{2.89}$$

We have that $X(t)$ is distributed geometrically with parameter $p = e^{-\lambda t}$.

2.3.5 Death Processes

A birth and death process where only transitions from i to $i-1$ are possible, for all $i = 1, 2, ...$ is called a pure death process. State 0 is absorbing, i.e $\mu_0 = 0$.

For pure death processes, with the condition

$$p_N(0) = P(X(0) = N) = 1, \tag{2.90}$$

the system of differential equations becomes

$$\begin{aligned} p'_N(t) &= -\mu_N p_N(t) \\ p'_i(t) &= -\mu_i p_i(t) + \mu_{i+1} p_{i+1}(t), \quad i = 0, 1, ..., N-1 \end{aligned}$$

The solution to the first differential equation is

$$p_N(t) = e^{-\mu_N t}, \quad t \geq 0$$

Integrating yields

$$p_i(t) = \mu_{i+1} e^{-\mu_i t} \int_0^t e^{\mu_i x} p_{i+1}(x) dx, \quad i = N-1, ..., 1, 0 \tag{2.91}$$

Starting with $p_N(t)$, the probabilities

$$p_i(t), \quad i = N-1, ..., 0$$

can be recursively determined by assuming $\mu_N \neq \mu_{N-1}$

$$P_{N-1}(t) = \mu_N e^{-\mu_{N-1} t} \int_0^t e^{-(\mu_N - \mu_{N-1})x} dx \quad = \frac{\mu_N}{\mu_N - \mu_{N-1}} (e^{-\mu_{N-1} t} - e^{-\mu_N t})$$

More generally, if all the death rates are different from each other $\mu_k \neq \mu_i$, then

$$p_j(t) = \sum_{i-j}^{N} D_{ij}\mu_i e^{-\mu_i t}, \quad 0 \leq j \leq N, \tag{2.92}$$

where

$$D_{ij} = \frac{1}{\mu_j} \prod_{k=j,k\neq i}^{N} \frac{\mu_k}{\mu_k - \mu_i}, \quad j \leq i \leq N, D_{NN} = \frac{1}{\mu_N}$$

Linear Death processes

A pure death process $\{X(t), t \geq 0\}$ is called a linear death process if it has the death rates

$$\mu_i = i\mu$$

Under the initial distribution (2.79), the process stays in state N and exponential with parameter $n\lambda$ distributed time. Starting with $p_N(t)$, one obtains inductively from (2.80)

$$p_i(t) = \binom{N}{i} e^{-i\lambda t}(1 - e^{-\lambda t})^{N-i}, \;\; i = 0, 1, 2... \tag{2.93}$$

We have that $X(t)$ follows a binomial distribution with parameters N and $p = e^{-\lambda t}$

2.3.6 Birth and Death Process with Immigration

Since birth and death processes have applications in various fields including biology and ecology, another variation is one that accommodates for immigration into the population over time. Due to immigration from the outside the population will increase by one individual in $[t, t + \Delta t]$ with probability $\nu\Delta t + o(\Delta t)$.

Suppose that immigration is included in the simple birth and death process at a constant rate ν. Let $X(0) = N$. The infinitesimal transition probabilities for this process

$$p_{i+j,i}(\Delta t) = P\{\Delta X(t) = j | X(t) = i\}$$

are given by

$$p_{i+j,i}(\Delta t) = \begin{cases} (\nu + \lambda i)\Delta t + o(\Delta t), & \text{if } j = 1 \\ \mu i \Delta t + o(\Delta t), & \text{if } j = -1 \\ 1 - [\nu + (\lambda + \mu) i]\Delta t + o(\Delta t), & \text{if } j = 0 \\ o(\Delta t), & \text{if } j \neq -1, 0, 1 \end{cases}$$

The forward Kolmogorov differential equations are:

$$\begin{aligned} p_0'(t) &= \mu p_1(t) - \nu p_0(t) \\ p_i'(t) &= (\lambda(i-1) + \nu) p_{i-1}(t) + \mu(i+1) p_{i+1}(t) - (\lambda i + \nu + \mu i) p_i(t) \end{aligned}$$

The probability generating function, $M(t, z)$ as defined in (2.32), of the probability distribution $\{p_0(t), p_1(t), ...\}$ satisfies the partial differential equation

$$\frac{\partial M(t,z)}{\partial t} = (\lambda z - \mu)(z - 1)\frac{\partial M(t,z)}{\partial z} + \nu(z-1)M(t,z)$$

The system of the characteristic differential equations belonging to the above is

$$\begin{aligned} \frac{dz}{dt} &= -(\lambda z - \mu)(z - 1), \\ \frac{dM(t,z)}{dt} &= \nu(z-1)M(t,z) \end{aligned}$$

From this, using the initial condition $p_0(0) = 1$, the solution is obtained

$$M(t,z) = \left\{ \frac{(\lambda - \mu)}{\lambda z + \lambda(1-z)e^{(}\lambda - \mu)t - \mu} \right\}^{\frac{\nu}{\lambda}} \quad \text{for} \quad \lambda \neq \mu$$

$$M(t,z) = (1+\lambda t)^{\frac{\nu}{\lambda}} \left\{ 1 - \frac{\lambda t z}{1 + \lambda t} \right\}^{-\frac{\nu}{\lambda}} \quad \text{for} \quad \lambda = \mu$$

The absolute state probabilities $p_i(t)$ can be obtained by differentiation of $M(t,z)$

$$p_i(t) = \left. \frac{\partial^i M(t,z)}{\partial z^i} \right|_{z=0} \quad \text{for } i = 1, 2, ..$$

The trend function can be calculated as follows

$$m(t) = E[X(t)] = \left. \frac{\partial M(t,z)}{\partial z} \right|_{z=1}$$

Therefore, the mean of this process is

$$m(t) = \frac{\nu}{\lambda - \mu}[e^{(\lambda-\mu)t} - 1] \quad \text{for} \quad \lambda \neq \mu$$

$$m(t) = \nu t \quad \text{for} \quad \lambda = \mu$$

If $\lambda < \mu$, the limit as $t \to \infty$ of the probability generating function exists:

$$\lim_{t \to \infty} M(t,z) = \left(1 - \frac{\lambda}{\mu} \right)^{\frac{\nu}{\lambda}} \left(1 - \frac{\lambda}{\mu z} \right)^{-\frac{\nu}{\lambda}}$$

For $\lambda < \mu$, the trend function tends to a positive limit as $t \to \infty$:

$$\lim_{t \to \infty} M(t,z) = \frac{\nu}{\mu - \lambda} \quad \text{for} \quad \lambda < \mu$$

2.3.7 Non-homogeneous Birth and Death Processes

2.3.7.1 Non-homogeneous Birth Processes

1. The most simple non-homogeneous pure birth process is the non-homogeneous Poisson process. Its birth rates are

   $$\lambda_i(t) = \lambda(t), \quad i = 0, 1, ...$$

   Thus, the process makes a transition from state i at time t to state $i+1$ in $[t, t+\Delta t]$ with probability:

   $$\lambda(t)\Delta t + o(\Delta t)$$

2. With certain conditions, mixed Poisson distributions can be considered as a class of non-homogeneous birth processes. Lundberg (1964) proved that a birth process is a mixed Poisson process iff

   $$\lambda_{i+1}(t) = \lambda_i(t) - \frac{d \ln \lambda_i(t)}{dt}, \quad i = 0, 1, ...$$

   A pure birth process $\{X(t) = i, t \geq 0\}$ with transition rates $\lambda_i(t)$ and with absolute state distribution

   $$\{p_i(t) = P(X(t) = i); \quad i = 0, 1, ...\}$$

   is a mixed Poisson process (See *Appendix A*) iff

   $$p_i(t) = \frac{t}{i}\lambda_{i-1}(t)p_{i-1}(t); \quad i = 1, 2, ...$$

2.3.7.2 Non-homogeneous Linear Birth and Death Processes

Consider a birth and death process which has transition rates

$$\lambda_i(t) = \lambda(t)i$$

$$\mu_i(t) = \mu(t)i$$

for $i = 0, 1, ...$ and initial distribution

$$p_1(0) = P(X(0) = 1) = 1$$

The absolute state probabilities $p_i(t)$ satisfy

$$p_0'(t) = \mu(t)p_1(t),$$

$$p_i'(t) = (i-1)\lambda(t)p_{i-1}(t) - i(\lambda(t) + \mu(t))p_i(t) + (i+1)\mu(t)p_{i+1}(t); \quad i = 1, 2, ...$$

The probability generating function, $M(t, z)$, of

$$\{p_i(t) = P(X(t) = i) : \quad i = 0, 1, ...\}$$

is given by the partial differential equation with time-dependent μ and λ:

$$\frac{\partial M(t,z)}{\partial t} - (z-1)[\lambda(t)z - \mu(t)]\frac{\partial M(t,z)}{\partial z} = 0 \tag{2.94}$$

The corresponding characteristic differential equation is a differential equation of the Riccati type with time-dependent coefficients

$$\frac{dz}{dt} = -\lambda(t)z^2 + [\lambda(t) + \mu(t)]z - \mu$$

As mentioned in section (1.3.1), a property of this differential equation is that there exist functions

$$\phi_i(x); \quad i = 1, 2, 3, 4$$

so that the general solution $z = z(t)$ can be implicitly written in the form

$$c = \frac{z\phi_1(t) - \phi_2(t)}{\phi_3(t) - z\phi_4(t)}$$

where c is a constant.

Hence, for all differentiable functions $g(.)$, the general solution has the form

$$M(t, z) = g\left(\frac{z\phi_1(t) - \phi_2(t)}{\phi_3(t) - z\phi_4(t)}\right)$$

From this and the initial condition, $M(0, z) = z$, it follows that there exist two functions $a(t)$ and $b(t)$ so that

$$M(t, z) = \frac{a(t) + [1 - a(t) - b(t)]z}{1 - b(t)z} \tag{2.95}$$

By expanding $M(t, z)$ as a power series in z,

$$\begin{aligned} p_0(t) &= a(t), \\ p_i(t) &= [1 - a(t)][1 - b(t)][b(t)]^{i-1}; \quad i = 1, 2, ... \end{aligned} \tag{2.96}$$

Using (2.96) in (2.97) and comparing the coefficients of z yields a system of differential equations for $a(t)$ and $b(t)$:

$$\begin{aligned} (a'b - ab') + b' &= \lambda(1 - a)(1 - b) \\ a' &= \mu(1 - a)(1 - b) \end{aligned}$$

The transformation $A = 1 - a$ and $B = 1 - b$ simplifies this system to

$$B' = (\mu - \lambda)B - \mu B^2 \tag{2.97}$$

$$A' = -\mu AB \tag{2.98}$$

The differential equation (2.98) is of Bernoulli type. Substituting equation (2.99)

$$y(t) = \frac{1}{B(t)}$$

gives a linear differential equation in y:

$$y' + (\mu - \lambda)y = \mu$$

Since

$$a(0) = b(0) = 0$$

y satisfies $y(0) = 1$. Hence the solution is

$$y(t) = e^{-\omega(t)} \left[\int_0^t e^{\omega(x)} \mu(x) dx + 1 \right],$$

where

$$\omega(t) = \int_0^t \left[\mu(x) - \lambda(x) \right] dx$$

Using (2.99) and (2.100)

$$\frac{A'}{A} = -\mu B = -\frac{\mu}{y} = -\frac{y'}{y} - \omega'$$

Therefore, the functions a and b are:

$$a(t) = 1 - \frac{1}{y(t)} e^{-\omega(t)}$$

$$b(t) = -\frac{1}{y(t)}, \quad t \geq 0$$

Now the process $\{X(t), t \geq 0\}$, with one-dimensional probabilities as in (2.97), is completely characterised by the non-homogeneous birth and death process, we know:

$$p_0(t) = \frac{\int_0^t e^{\omega(x)}\mu(x)dx}{\int_0^t e^{\omega(x)}\mu(x)dx + 1}$$

Hence, the process $\{X(t), t \geq 0\}$ will reach state 0 with probability 1 if the integral

$$\int_0^t e^{\omega(x)\mu(x)}dx$$

diverges as $t \to \infty$.

Mean value and variance of $X(t)$ are

$$E[X(t)] = e^{\omega(t)}$$
$$Var(X(t)) = e^{-2\omega(t)} \int_0^t e^{\omega(x)}[\lambda(x) + \mu(x)]dx$$

2.3.8 Population Extinction

Returning to Section (2.3.1), the homogeneous birth and death process, population extinction will be detailed. In biological applications without immigration, the zero state is absorbing. Eventually, the distribution of the total population size is concentrated at zero. The following theorem gives conditions for the total population extinction in a general birth and death process.

Theorem 2.4

Let $\mu_0 = 0 = \lambda_0$ in a general birth and death chain with $X(0) = m \geq 1$

i Suppose $\mu_1 > 0$ and $\lambda_i > 0$ for $i = 1, 2, ...$

If

$$\sum_{i=1}^{\infty} \frac{\mu_1\mu_2...\mu_i}{\lambda_1\lambda_2...\lambda_i} = \infty$$

then $\lim_{t\to\infty} p_o(t) = 1$

If

$$\sum_{i=1}^{\infty} \frac{\mu_1\mu_2...\mu_i}{\lambda_1\lambda_2...\lambda_i} < \infty$$

and the probability of extinction approaches zero as $m \to \infty$, then for finite m,

$$\lim_{t\to\infty} p_0(t) = \frac{\sum_{i=m}^{\infty} \frac{\mu_1\mu_2...\mu_i}{\lambda_1\lambda_2...\lambda_i}}{1 + \sum_{i=1}^{\infty} \frac{\mu_1\mu_2...\mu_i}{\lambda_1\lambda_2...\lambda_i}}$$

ii Suppose $\mu_i > 0$ for $i - 1, 2, ...$, $\lambda_i > 0$ for $i = 1, 2, ..., N-1$ and $\lambda_i = 0$ for $i = N, N+1, N+2, ...$ Then $\lim_{t\to\infty} p_0(t) = 1$.

2.3.9 Cumulative Population

An important variable that is often used in biology and ecology is the cumulative population. Let the integer-valued time-dependent random variable N_t measure the size of the population at time t and let a variable be defined as follows:

$M_0 = N_0$, while for $t > 0$, M_t counts all the positive jumps of N_t. This is also known as the total count.

To study the relationship of M_t and N_t, we look at the bi-variate generating function:

$$M(t, w, z) = \sum_{i=0}^{\infty}\sum_{M=0}^{\infty} P_{i,M}(t) z^i w^M$$

where $P_{i,M}(t)$ is the joint frequency distribution of N_t and M_t at time t.

We have the boundary condition:

$$N_0 = M_0 = 1$$

In terms of the bivariate probability generating function, the initial condition is equivalent to:

$$M(0, w, z) = zw$$

The partial derivatives of the z, w- transform is

$$\frac{dM(t,w,z)}{dt} = \sum_{i=0}^{\infty}\sum_{M=0}^{\infty} p'_{i,M}(t)z^n w^M$$

$$\frac{dM(t,w,z)}{dz} = \sum_{i=0}^{\infty}\sum_{M=0}^{\infty} p_{i,M}(t)nz^{n-1} w^M$$

$$\frac{dM(t,w,z)}{dw} = \sum_{i=0}^{\infty}\sum_{M=0}^{\infty} p_{i,M}(t)Mz^n w^{M-1}$$

Multiplying adaption of equation (2.29) by z^i and summing from $i = 0$ to $i = \infty$, taking the partial derivatives of the z, w-transform yields:

$$\frac{dM(t,w,z)}{dt} = \lambda z^2 w \frac{dM(t,w,z)}{dz} - (\lambda+\mu)z\frac{dM(t,w,z)}{dz} + \frac{\mu}{w}\frac{dM(t,w,z)}{dz}$$

By separation of variables, we get

$$\frac{dz}{(zw\lambda - \mu)\left(z - \frac{1}{w}\right)} = dt$$

The corresponding characteristic differential is Riccati differential equation:

$$\frac{dz}{dt} = \lambda z^2 w - (\lambda+\mu)z + \frac{\mu}{w} = (zw\lambda - \mu)\left(z - \frac{1}{w}\right)$$

Integrating both sides of the relationship:

$$\ln\left(\frac{wz-1}{w\lambda z - \mu}\right)\frac{1}{\lambda-\mu} + C = t + C$$

giving

$$\ln\left(\frac{wz-1}{w\lambda z - \mu}\right) = t(\lambda - \mu)$$

Thus,

$$c = t(\lambda - \mu) - \ln\left(\frac{wz-1}{w\lambda z - \mu}\right)$$

Where c is some arbitrary constant.

Thus, we have :

$$M(t,w,z) = f\left(t(\lambda - \mu) - \ln\left(\frac{wz-1}{w\lambda z - \mu}\right)\right)$$

Where f is any function with continuous derivative. From the initial condition, this becomes:

$$M(0,w,z) = f\left(-\ln\left(\frac{wz-1}{w\lambda z - \mu}\right)\right) = zw$$

f must have structure

$$f(x) = \frac{\mu e^x - 1}{\lambda e^x - 1}$$

Thus $M(t,w,z)$ is

$$\begin{aligned}
M(t,w,z) &= \frac{\mu exp\left\{t(\lambda - \mu) - \ln\left(\frac{wz-1}{w\lambda z - \mu}\right)\right\} - 1}{\lambda exp\left\{t(\lambda - \mu) - \ln\left(\frac{wz-1}{w\lambda z - \mu}\right)\right\} - 1} \\
M(t,w,z) &= \frac{\mu\left\{e^{(\lambda-\mu)t}\left(\frac{wz-1}{\lambda z - \mu}\right)\right\} - 1}{\lambda\left\{e^{(\lambda-\mu)t}\left(\frac{wz-1}{\lambda z - \mu}\right)\right\} - 1} \\
M(t,w,z) &= \frac{\mu(1 - e^{(\lambda-\mu)t}) - w(\lambda - \mu e^{(\lambda-\mu)t})z}{(\mu - \lambda e^{(\lambda-\mu)t}) - \lambda w(1 - e^{(\lambda-\mu)t})z}
\end{aligned}$$

Using the generating function, we can obtain the mean and the variance of the cumulative population.

2.4 Applications

2.4.1 Epidemiology, Evolution, and Future of the HIV/AIDS Pandemic

A mathematical model was developed by Levin et al. (2001) to model the dynamics of the HIV/AIDS population. The model was used to study the evolution and epidemiology of HIV/AIDS. The model that they used was an adaption of the model usually used in demography. A birth and death process was used to represent the changes in the number of people infected. New infections were seen as the "births" and a host leaving the population was the "deaths".

Levin et al. (2001) characterised the HIV infections by the following four stages, given in Table (2.1)

Table 2.1: The stages of AoI Model

Stages in Age of Infection Model	
Stage	Characteristics
0	Establishment- Short time after infection of host.
1	Primary infection- Initial presence of HIV in blood. At this time, the circulation of HIV increases to high levels.
2	Asymptotic- The long "asymptotic" period when circulating virus remains at low levels.
3	AIDS- The increase in density HIV associated with AIDS moving around. (This stage terminates the death of infected individuals)

Levin et al. (1996) developed a mathematical model with the following notation:

x- The age of infection (AoI), the time since infection.

$b(x)$- Rate of new infections of host infected by another host with AoI x.

$I(x)$- The probability that infection persists at least to AoI x.

$N_t(x)$- The AoI density of infected host at time t.

The following assumptions are made by Levin et al. (2001)

1. A deterministic model can be adopted due to the large population at risk.
2. There are no deaths in the first 3 stages.
3. The weekly rate of transaction varies between stages but remains constant within each stage.

Levin et al. (2001) has found that "$N_t(x)$ grows approximately exponentially with t, and as a function of x, approaches a limiting shape at which the relative densities of different AoIs remain constant, known as the stable age of infection distribution."

The Model

The following model was defined as in the Levin et al. (1996) paper. Assume there is a single profile. Let $U(t)$ be the number of new infections expected per unit time, then

$$U(t) = \int_0^D U(t-x)I(x)b(x)dx$$

Where D is an arbitrary number big enough, that a negligible amount of infections are greater than AoI of D. Assume the profiles are not inheritable. The rate in which new host acquire infectious of the j^{th} profile is $f_j \int_0^D U(t-x)I(x)b(x)dx$. The sum of $U(t)$ taken over different profiles satisfies:

$$U(t) = \int_0^D U(t-x)v(x)dx$$

where $v(x) = \sum_{j=1}^{J} f_j I_j(x) b_j(x) = \sum f_j \bar{v}_{k(j,x)}$ where $K(j,x)$ is the stage reached by the j^{th} profile at AoI x, and the second sum is taken over those profiles where hosts are still alive at AoI x.

From demography, the "intrinsic rate" of increase of an epidemic, or its "Malthusian parameter" is used to determine the potential for epidemic spread. For a single $I(x)$ and $b(x)$, the intrinsic rate of increase, r_0, satisfies

$$\int_0^D e^{r_0 x} I(x) b(x) dx = 1$$

Taking various rates at which HIV infections develop:

$$\int_0^D e^{r_0 x} v(x) dx = 1$$

We have

$$R_0 = \int_0^D v(x) dx$$

This can be interpreted as follows:

- $R_0 = 1 \leftrightarrow r_0 = 0$- The disease will not increase or decrease.
- $R_0 > 1 \leftrightarrow r_0 > 0$- The disease will become epidemic
- $R_0 < 1 \leftrightarrow r_0 < 0$- The disease will die out.

The research provides the following conclusions from their analysis of the models:

- When HIV enters a human sub-population, the epidemic is driven by early transmissions.
- New infections in a subgroup may decrease since all susceptible host already have been infected and not necessarily due to other reasons such as intervention or education.
- Human resistance to the virus will require thousands of years.
- The severity of HIV is not likely increased by evolution.

- If transmission of HIV is reduced by chemotherapy, then the treatment of patients can decrease the amount of HIV infections and deaths in populations.

Modelling HIV needs more data, which not much interest is vested in, without it, the modelling will remain strictly mathematical.

2.4.2 Stochastic Evolution Dynamic of the Rock-Paper-Scissors Game Based on a Quasi-Birth and Death Process

In this application, Yu et al. (2016) modelled the "Rock-Scissors-Paper (RSP) game as a finite , state dependent Quasi-Birth and Death (QBD) process". It is based on the assumption that players can change their strategies based on the previous game. The application of the RSP game extends to areas such as biological systems and species interaction as well as to economic and social systems. "Previous studies of the RSP game identify various factors influencing the cyclical behaviour of populations such as noise, alliance specific heterogeneous rate, mutations and group interactions"(Yu et al. (2016)).

Yu et al. (2016) has shown that "the long-run equilibrium of the RSP game played by bounded individuals in a finite population can be explained by the limiting distribution of the QBD process of the game evolutionary dynamic. With the limiting distribution, the probability of a stable state of RSP game evolution dynamic can be more profoundly predicted and interpreted".

The individuals are bounded by 3 hypotheses. Inertia, where not all individuals react instantaneously to the environment, myopia, where not all individuals take into account long-term implications and mutation, where individuals will alter their strategy randomly with a small probability.

The Model

Denote the 3 strategies R, S and paper as 1, 2 and 3 and define the stochastic process as follows:

Let the strategy at time t be represented by a two-dimensional stochastic variable, $\omega(t) = (\omega_1(t), \omega_2(t))$. Let $\omega_1(t)$ and $\omega_2(t)$ be the number of players that choose strategy R and S respectively. The number of people who choose strategy P is the remainder, $N - \omega_1(t) - \omega_2(t)$.

The pay-off matrix of the RSP game is given by

$$G = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \tag{2.99}$$

The pay-off matrix of the generalised RSP game can be normalised as follows:

$$G = \begin{pmatrix} 0 & -a_2 & b_3 \\ b_1 & 0 & -a_3 \\ -a_1 & b_2 & 0 \end{pmatrix} \tag{2.100}$$

The average payoff will be as follows:

- If R is chosen,

$$\pi_1(\omega) = \frac{0.\omega_1 + (-a_2).\omega_2 + (b_3)(N - \omega_1 - \omega_2)}{N}$$

- If S is chosen,

$$\pi_2(\omega) = \frac{b_1.\omega_1 + (0).\omega_2 + (-a_3)(N - \omega_1 - \omega_2)}{N}$$

- If P is chosen,

$$\pi_3(\omega) = \frac{-a_1.\omega_1 + (b_2).\omega_2 + (0)(N - \omega_1 - \omega_2)}{N}$$

At $t+1$, players switch strategies at a rate which depends on the payoff at time t.

The inertia and myopia hypotheses are satisfied since players can only move to adjacent states, taking the payoff matrix into consideration when moving from one strategy to the next. Let the mutation be modelled by a perturbation factor $\epsilon \geq 0$, which allows a player to deviate with a small probability.

Assume $\omega_1(t) = i$ and $\omega_2(t) = j$. Individuals can change from strategy $l - k$ at rate

$$\tau_{kl}(i,j) = \epsilon + [\pi_k(i,j) - \pi_l(i,j)]^+$$

where $(i,j) \neq \{(0,N),(N,0),(0,0)\}$

$$\tau_{kl}(0,N) = \tau_{kl}(N,0) = \tau_{kl}(0,0)$$

"The stochastic evolutionary process of the strategy becomes a Markov-stochastic process that can be described as a QBD process" (Yu et al. (2016)). Given $\lambda_{kl}(i,j) = \tau_{kl}(i,j)$ for $k < l$ and $\mu_{kl}(i,j) = \tau_{kl}(i,j)$ for $k > l$.

In economics or game theory, the stable state of a system involving interaction of different participants, in which no participant can gain by a unilateral change of strategy if the strategy of the others remains unchanged, is known as a Nash equilibrium. There is only one Nash equilibrium in mixed strategies, when both players randomise uniformly.

$$\begin{pmatrix} x_1^* \\ x_2^* \\ x_3^* \end{pmatrix} = \begin{pmatrix} y_1^* \\ y_2^* \\ y_3^* \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} \tag{2.101}$$

There exists a Nash Equilibrium that is,

$$P = \begin{pmatrix} x_1^* \\ x_2^* \\ x_3^* \end{pmatrix} = \begin{pmatrix} y_1^* \\ y_2^* \\ y_3^* \end{pmatrix} = \frac{1}{\Gamma} \begin{pmatrix} a_2a_3 + a_3b_2 + b_2b_3 \\ a_1a_3 + a_1b_3 + b_1b_3 \\ a_1a_2 + a_2b_1 + b_1b_2 \end{pmatrix}$$

where Γ is the normalising constant.
The Nash equilibrium P is asymptotically stable when $a_1a_2a_3 < b_1b_2b_3$, neutrally stable when $a_1a_2a_3 = b_1b_2b_3$, and unstable when $a_1a_2a_3 > b_1b_2b_3$.

2.4.3 Physiological and Pathological Population Dynamics of Circulating Human red blood cells

Systems controlling the number, size, hemoglobin concentrations of populations of human blood cells are poorly understood. Higgins and Mahadevan (2010) develop a master equation model for red blood cells maturation and clearance. Their model has accurately identified patients with anemia and distinguishes talassemia-trait anemia from ion deficiency anemia. It also , can predict pre-anemia patients before anemia is clinically diagnosed. Since the volume and hemoglobin of red blood cells of individual is extremely complex, they have worked with the average behaviour since it is more tractable. The model that they have developed is applicable to the typical behaviour of red blood cells in healthy or mild cases of anemia.

The Model

Notation:

v – volume

h – height

t – time

f – deterministic reductions

ζ – random fluctuations

β – deterministic fast change

α – deterministic slow change

The random fluctuation is modelled as a Gaussian random variable, with mean zero and variance given by:

$$\begin{bmatrix} \frac{dv}{dt} \\ \frac{dh}{dt} \end{bmatrix} = f + \zeta, \tag{2.102}$$

where

$$f = \begin{cases} \alpha . e^{\beta_v (v-h)} \\ \alpha . e^{\beta_h (h-v)} \end{cases}$$

$$\zeta = \begin{cases} N(0, 2D_v) \\ N(0, 2D_h) \end{cases}$$

According to Higgins and Mahadevan (2010), " The dynamics of the entire circulating population of the red blood cells may then be described by a master equation for the time-dependent joint volume-hemoglobin probability distribution $P(v, h, t)$ which can be approximated by the Fokker-Planck equation. Equation (2.105) describes the drift (f), diffusion (D), birth (b), and death (d) of probability density for this joint volume-hemoglobin distribution.”

$$\frac{\partial P}{\partial t} = -\Delta \cdot (Pf) + \Delta \cdot (D \cdot \Delta P) + b(v,h,t) - d(v,h,t)P \tag{2.103}$$

$$D = \begin{bmatrix} D_v & 0 \\ 0 & D_h \end{bmatrix}$$

The birth and death process account for the red blood cells that are constantly added to and removed from the population.

In the case of individuals who are healthy or who have mild-anemia, the total number of cells added is equal to the total number of cells removed.

$$\int\int d(v,h)Pdvdh = \int\int b(v,h)dvdh$$

Complete blood counts(CBC) reach a stable state in healthy individuals since they do not vary significantly. Let us denote the stable state by P_∞, where

$$\lim_{t\to\infty} P(v,h,t) \to P_\infty(v,h) \leftrightarrow \frac{\partial P}{\partial t} = 0$$

For each patient sample, Higgins and Mahadevan (2010) identifies an optimal parameter set (α, β, D, v_c) that reproduces the steady state for that patient. They then used least-square fit between the measured CBC and simulated steady-state distribution to find the best fit.

The steady-state distribution for P_∞ was determined analytically since the clearance terms (w) are linear operators and that the integral scaling the birth process is a constant equal to the reciprocal of twice the mean age $(\frac{1}{2\tau})$.

That is,

$$
\begin{aligned}
\frac{\partial P_\infty}{\partial t} &= 0 = -JP_\infty + LP_\infty + wP_\infty + P_0 \int_h \int_v d(v,h) \\
\frac{\partial P_\infty}{\partial t} &= (-J + L + w)P_\infty + P_0 \frac{1}{2\tau} \leftrightarrow P_\infty = -(-J + L + w)^{-1} P_0 \frac{1}{2\tau}
\end{aligned}
$$

where w is the probability of clearance and J and L are the first and second order of numerically solved solutions of (2.105) given by,

$$J = \Delta_k \frac{[f - P](v)}{k} + \Delta_k \frac{[f - P](h)}{k}$$

and

$$L = \frac{D_v(\delta_k^2[p](v)}{k^2} + \frac{D_h(\delta_k^2[p](h)}{k^2}$$

2.4.4 Speciation Rates Decline through Time in Individual-based Models of Speciation and Extinction

Wang et al. (2013) noted that one pattern found in fossil record data is the long-term decline in origination rate of new taxa following the rebound of diversity after mass extinction. The taxon selection hypothesis explains the decline in speciation rates by the replacement of early dominant high-speciation rate taxa by those with lower rates. Wang et al. (2013) have found that the assumption of correlation between rates of speciation and extinction, though supported by empirical data, lacks a clear mechanism. Wang et al. (2013) investigate the macro-evolutionary prediction of an individual-based birth-death model, where speciation and extinction rates emerge from the population dynamics. They start with the simplest model, Hubbel's Neutral theory of biodiversity model. With this model, the predicted decline per species speciation rates are too fast to explain the long-term trend in fossil data. This is the reason that they consider models with a variation among species. This shows that a variation in speciation rate can induce differences among species to resist extinction. A model that predicts slow temporal decline in speciation rates, provides a mechanistic explanation for fossil patterns.

The Model

Let q_{uv} represent the transition probability which is the distribution of the per capita speciation rate v for a new species, given that the per capita rate of its ancestral species is u.

Let $0 < v_{min} < v_{max} < 1$ and let v_0, the initial rate be in the middle of the interval.

Define the models as follows:

M_T Non-heritable random model

M_t Partially heritable model

Under M_T, the speciation rate of new species is randomly drawn from uniform probability density of the interval $[v_{min}, v_{max}]$, thus

$$q_{uv}dv = \frac{dv}{v_{max} - v_{min}}$$

Under M_t, the speciation rate of new species is randomly drawn from a uniform probability density around that of its parental species (u). Thus,

$$v \sim U[u - \frac{L}{2}, u + \frac{L}{2}]$$

where $\frac{L}{2}$ is a measure of the heritability.

For any v satisfying $|u-v| \leq \frac{L}{2}$, the transition probability is

$$q_{uv}dv = \begin{cases} \dfrac{dv}{\frac{L}{2}+u-v_{min}}, & u < v_{min}+\frac{1}{2} \\ \dfrac{dv}{L}, & u \in [v_{min+\frac{L}{2}, v_{max}-\frac{L}{2}}] \\ \dfrac{dv}{\frac{L}{2}+v_{max}-u}, & u > v_{max}-\frac{L}{2} \end{cases}$$

Master Equation of the Frequency distribution of the speciation rate

Let M denote the total number of individuals in the system. Let $K(v_i,t)$ denote the expected value of the number of individuals with speciation rate v_i at time t.

The change of $K(v_i,t)$ is determined by the probability of increasing or decreasing by one with speciation rate v_i. Since birth and deaths are completely random,

$$K(v,t+1)-K(v,t) = P_v^+ - P_v^-$$

where:

P_v^+ is the probability that a new species with speciation rate v_i is produced by a species with rate not equal to v_i.

P_v^- is the probability that a species with rate v_i produces an offspring with rate not equal to v_i.

$$K(v,t+1)-K(v,t) = \sum_{j\neq 1} \frac{K(v,t)}{M} v_i q_{ij} - (\frac{K(v,t)}{M} v_i.(1-q_{ii})$$
$$K(v,t+1)-K(v,t) = \sum_{j} \frac{K(v,t)}{M} v_i q_{ij} - \frac{K(v,t)}{M} v_i$$

2.4.5 Mathematical Modelling for Human Immunodeficiency Virus (HIV) Transmission using Generating Functions Approach

Branching and birth and death processes are also used to model populations. Thus, the application of generating functions has been extended to the studies of populations. Until recently, researchers have previously been focusing on deterministic models to model the epidemiology of infectious disease. Waema and Olowofeso (2015) study deterministic models, and they develop stochastic functions. They make use of generating functions to solve the differential equations. They specifically model the transmission rates between mother and child and heterosexual transmission. They also try to develop a combined model of the two mentioned.

The authors motivation for their use of stochastic models is as follows:

- Biological factors are subject to random variation, which in essence would make them stochastic by nature.
- Stochastic models are more informative than deterministic models.
- Under certain conditions, both approaches are equivalent.

Waema and Olowofeso (2015) state that each model has three components:

- The Susceptible Model (S)- These are individuals which do not have the virus, but can contract it either from breast-feeding from HIV- positive mothers or having sexual intercourse with HIV-positive partners.
- The Infection Model (I)- Those who contracted the disease from infected mothers or sexual partners.
- The AIDS Model (A)- The former infected which develop full-blown AIDS.

The Model

Notation:

k – age groups defined in model

$$k = \begin{cases} 1, & \text{Children between the age of 0-5 years} \\ 2, & \text{Children aged 5-15 years} \\ 3, & \text{Young adults.} \end{cases}$$

μ_k – death rate unrelated to HIV/AIDS or immigration, $k = 1, 2, 3$

λ – birth rate for a sexually mature person

α – immigration for a sexually mature person

t – present time

x_i – time n years

$S(t)$ – number of people in S at t

$I(t)$ – number of people in I at t

$A(t)$ – number of people in A at t

$N(t)$ – total number of population

ω – sexual contact

δ – probability I transmits to S

β – rate at which infected mother does not infect newborn

γ – transition rate from I-S

The Combined Model as defined by Waema and Olowofeso (2015):

$$P(X(t + \Delta t) = I | X(t) = n) =$$

$$
\begin{cases}
\alpha\Delta t + nS_3\lambda\Delta t + nI_3\beta\lambda\Delta t + o(\Delta t), \quad i = n+1 \\
nS_k\mu_k\Delta t + nI_3\omega\delta\Delta t + o(\Delta t), \quad i = n-1 \\
1 - nS_3\lambda\Delta t + \alpha\Delta t - nI_3\beta\lambda\Delta t - nS_k\mu_k\Delta t - nI_3\omega\delta\Delta t - o(\Delta t), \quad i = n-1 \\
o(\Delta t), \quad i < n-1, \quad i > n+1
\end{cases}
\tag{2.104}
$$

Let $S_n(t) = PS(t) = n|S(0) = m$ for $m < n, m = 0, 1, ...$ From (2.104), we get:

$$
\lambda_n(t) = nS_3\lambda + \alpha + nI_3\beta\lambda
$$
$$
\mu_n(t) = nS_k\mu_k + nI_3\omega\delta
$$

The Kolmogorov forward differential equations:

$$
S_0'(t) = -\alpha S_0(t) + [S_k\mu_k + I_3\omega\delta]S_1(t), \quad n = 0 \tag{2.105}
$$

$$
\begin{aligned}
S_n'(t) &= -[nS_3\lambda + \alpha + nS_k\mu_k + nI_3\beta\lambda + nI_3\omega\delta]S_n(t) \\
&\quad + [\alpha + (n-1)S_3\lambda + (n-1)I_3\beta\lambda]S_{n-1}(t) + [(n+1)S_k\mu_k + (n+1)I_3\omega\delta]S_{n+1}(t), \quad n \geq 1
\end{aligned}
\tag{2.106}
$$

To solve, we get the generating function:

$$
M(t, z) = \sum_{j=0}^{\infty} S_j(t)z^j
$$

We will be using the initial condition $S_n(0) = P(X(0) = n) = 1$ which gives,

$$
M(0, z) = z^j, \quad j = 1, 2, ..
$$

Multiplying (2.105) by z^j and summing from $j = 0$ to $j = \infty$, we get

$$\frac{\partial M(t,z)}{\partial t} = (az - b)(z-1)\frac{\partial M(t,z)}{\partial z} + \alpha(z-1)M(t,z)$$

where $a = S_k\mu_k + I_3\omega\delta$ and $b = S_3\lambda + I_3\beta\lambda$.

We can write the above as follows,

$$\frac{\partial z}{\partial t} = (az - b)(z-1) \tag{2.107}$$

$$\frac{\partial M(t,z)}{\partial t} = \alpha(z-1)M(t,z) \tag{2.108}$$

Applying the initial conditions:

$$p_0(0) = 1$$

$$M(0,z) = 1$$

.

We get a solution analogous to that in section (2.3.6)

$$M(t,z) = \left\{\frac{a-b}{az + z(1-z)e^{(a-b)t} - b}\right\}^{\frac{\alpha}{a}} \quad \text{for } a \neq b$$

and

$$M(t,z) = (1+\alpha t)^{\frac{\alpha}{a}}\left\{1 - \frac{atz}{1+at}\right\}^{-\frac{\alpha}{a}} \quad \text{for } a = b$$

A power series expansion is not possible in this case, thus, to get $p_i(t)$, the absolute state probabilities, we differentiate the probability generating function:

$$p_i(t) = \left.\frac{\partial^i M(t,z)}{\partial z^i}\right|_{z=0} \quad \text{for } i = 1, 2, ...$$

The trend function, which can be calculated by evaluating the first derivative of the generating function wrt to z, and evaluating at $z = 1$ is

$$m(t) = \frac{\alpha}{a-b}\left[e^{(a-b)t} - 1\right] \qquad \text{for } a \neq b$$

and

$$m(t) = \alpha t \qquad \text{for } a = b$$

2.4.6 Consolidating Birth-Death and Death-Birth Processes in Structured Populations

Zukewich et al. (2013) look at the application of birth and death processes in the Gaming theory. Traditionally network models are used, however, they are sensitive to model architecture. Zukewich et al. (2013) investigate " two biologically motivated models of evolution in finite populations". They look at Death-Birth and Birth-Death processes. Both cases have reproduction being directly proportional to the fitness of the populations and death being random. The selection is based on where there is cooperation under the mixed rule. They derive conditions for all social scenarios.
This is an application of discrete birth and death processes as they are used to model the evolution on networks. There are two updating rules:

- Birth-Date Update

  In this case, the probability that an individual will reproduce is a function of how fit that individual is. The offspring of that individual, will replace a neighbour at random.

- Death-Birth Update

  In this case, the death of an individual is chosen at random. It will then be replaced with a neighbour's offspring. The offspring is chosen with a probability proportional to fitness of a parent.

2.4.7 Stochastic Processes in Science, Engineering and Finance

An important application of birth and death processes, which occurs in the modelling of service facilities is given in Beichelt (2006). Customers' arrival at a service system can be seen as a random point process. An available server services the customer. If there are no available servers, the customer will either wait or leave without being served.

A queuing sytem can be classified as follows:

- A loss system- This system has no waiting capacity and customers who are not served will leave.
- A waiting system- In this system, customers will wait until served. The system has a waiting capacity that is unlimited or infinite.
- A waiting-loss system- In this system, only a certain number of customers can wait due to waiting capacity which is limited.

A multi-server queuing system has more than one server. A system with only one server is known as a single-server system.

The tools neccesary for the design and analysis of service systems are provided in queuing theory.

Beichelt (2006) give the criteria to consider for making an efficient system:

1. Probability that on arrival, a customer will be served.
2. How long on average will a customer wait for service.

Notation

The structure of a queuing system is characterised by Kendall's notation $A/B/s/m$, where

A The input

B The service

s The number of servers

m The number of customers

In this dissertation, we are interested in those queuing systems which make use of birth and death processes.

The $M/M/\infty/-$ Model

In this model, arrival and service model is a Markovian. We have that $\{X(t), t \geq 0\}$ is a homogeneous birth and death process with state space $\mathbb{Z}$ and transition rates

$$\lambda_i = \lambda, \quad \mu_i = i\mu$$

with initial distribution $p_0(0) = P(X(0) = 0) = 1$

Thus, the system of differential equations that describe the system is

$$\begin{aligned} p_0'(t) &= \mu p_1(t) - \lambda p_0(t) \\ p_j'(t) &= \lambda p_{j-1}(t) - (\lambda + \mu j)p_j(t) + (j+1)\mu p_{j+1}(t), \quad j = 1, 2, ... \end{aligned}$$

Multiplying the j^{th} equation by z^j and summing from $j = 0$ to ∞ yields a homogeneous linear partial differential equation for the moment generating function.

$$\frac{\partial M(t,z)}{\partial t} + \mu(z-1)\frac{\partial M(t,z)}{\partial z} = \lambda(z-1)M(t,z)$$

The corresponding system of differential equations is

$$\frac{\partial z}{\partial} = \mu(z-1)$$

$$\frac{\partial M(t,z)}{\partial t} = \lambda(z-1)M(t,z)$$

Separating the variables and integrating the first equation yields

$$c_1 = \ln(z-1) - \mu t$$

where c_1 is an arbitrary constant of integration.

Combining both differential equations, we get:

$$\frac{\partial M(t,z)}{M(t,z)} = \frac{\lambda}{\mu} dz.$$

Integrating yield:

$$c_2 = lnM(t,z) - \frac{\lambda}{\mu} z$$

where c_2 is an arbitrary constant of integration.

$M(t,z)$ satisfies

$$\begin{aligned} c_2 &= f(c_1) \\ \ln M(t,z) - \frac{\lambda}{\mu} z &= f(\ln(z-1) - \mu t) \\ M(t,z) &= e^{f(\ln(z-1)-\mu t) + \frac{\lambda z}{\mu}} \end{aligned}$$

Since $p_0(0) = 1 \leftrightarrow M(0,z) = 1$, we have

$$f(\ln(z-1)) = \frac{\lambda z}{\mu}$$

Thus, f can be represented as

$$f(x) = \frac{-\lambda}{\mu}(e^x + 1)$$

and then the probability generating function is

$$M(t,z) = exp\left\{ \frac{-\lambda}{\mu}(e^{\ln(z-1)-\mu t}+1) + \frac{\lambda z}{\mu} \right\}$$
$$M(t,z) = e^{\frac{-\lambda}{\mu}(1-e^{-\mu t})} e^{\frac{\lambda}{\mu}(1-e^{-\mu t})z}$$

To get the absolute state probabilities, we need to expand the probability generating function and extract the coefficient of z^j:

$$p_j(t) = \frac{\frac{\lambda}{\mu}(1-e^{-\mu t}))^j}{j!} e^{\frac{-\lambda}{\mu}(1-e^{-\mu t})}, \quad j = 0, 1, 2...$$

This is a Poisson distribution with intensity $\frac{\lambda}{\mu}(1-e^{-\mu t})$. The birth and death process trend function is

$$m(t) = \frac{\lambda}{\mu}(1-e^{-\mu t})$$

For $t \to \infty$, the absolute state probabilities $p_j(t)$ converge to stationary state probabilities

$$\pi_j = \lim_{t\to\infty} p_j(t) = \frac{\left(\frac{\lambda}{\mu}\right)^j}{j!} e^{-\frac{\lambda}{\mu}}, \quad j = 0, 1, ...$$

In the steady state, the mean of the busy servers is equal to the traffic intensity of the system

$$E[X] = \frac{\lambda}{\mu}$$

Chapter 3

Branching Processes

3.1 Introduction

Branching Processes date back to around 1845 when scientists such as L.F Benoiston de Chateanuel (1776-1816) and Sir Francis Galton (1822-1911) were interested in studying the extinction of English nobles (Jagers (1975); Kimmel and Axelrod (2015)). Malthus said that human populations grow exponentially, however, this did not refer to the problem of extinction of noble families. Bienayme (1796-1878) treated the problem mathematically. Galton posed questions about the extinction of noble families and how long this would take to occur. It was Watson who formulated the problem with the use of generating functions. The Danish actuary J. F Steffenson was the first to publish a complete solution to the questions posed by Galton. Today, population theory is built on the foundation of branching processes which do have limitations in measuring the time of the different generations (Jagers (1975)).

Branching processes as defined in Axelrod and Kimmel (2002) is an area of mathematics that a situation is described in which an object exists for a unit of time and then is replaced by one or more offspring, independent of all other individuals. Most of the theoretical background in this chapter follows Goel and Richter-Dyn (2013)

Illustration of a branching process with one common ancestor is given in Figure 3.1.

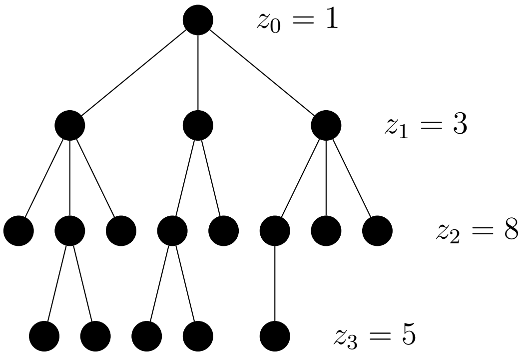

Figure 3.1 Branching process with one ancestor $z_0 = 1$*, and*

$z_1 = 3, z_2 = 8, z_3 = 5.$

3.2 Notation

Let the whole population stemming from ancestor a be represented by the vector

$$(a, j_1, j_2, ..., j_{n-1}, j_n) \tag{3.1}$$

where j_i is the j_i^{th} child of the j_{i-1}^{th} individual.

Let I_a be the set of all possible labels of the descendants of a, including a.

$$I_a = \{a\} \cup \bigcup_{n=1}^{\infty} \{(a; x) : x \in \mathbb{N}^n\} \tag{3.2}$$

The n^{th} generation $I_a(n)$ is the set of individuals (a, x) such that $x \in \mathbb{N}^n$.

The ancestor belongs to the zeroth generation $I_a(0)$.

For a given ancestor a and $x \in \mathbb{N}^k$, for some k:

$$I_{a,x)} = \{(a, x)\} \cup \bigcup_{n=1}^{\infty} \{(a, x, y), y \in \mathbb{N}^n\} \tag{3.3}$$

where $(a, x, y) = (a, j_1, j_2, ..., j_k, i_1, i_2, ..., i_n)$ if $x = (j_1, j_2, ..., j_k)$ and $y = (i_1, i_2, ..., i_n)$.

This defines the family stemming from any individual in I_a

$$I_a = \bigcup_{n=0}^{\infty} I_a(n) = I_a(0) \cup \bigcup_{n=1}^{\infty} I_a(n) \tag{3.4}$$

Each $x \in I$ is associated with one $\mathbb{Z}_+$- valued random variable, ξ_x, the number of children of x. The ξ_x are assumed to be iid with distribution $\{p_k; k \in \mathbb{Z}_+\}$ called the reproduction law.

3.3 Galton-Watson Process

Let

$$r_x = \begin{cases} 1, \text{ if } x \text{ is realised} \\ 0, \text{ otherwise} \end{cases}$$

Then a Galton Watson process is defined as

$$z_n = \sum_{x \in N^n} r_x = \sum_{x \in N^{n-1}} r_x \xi_x, \tag{3.5}$$

for $n \in \mathbb{N}$

A Galton-Watson process is then the number of individuals realised in different generations.

If $\{X_{nj}, n \in \mathbb{N}, j \in \mathbb{N}\}$ is a double array of random variables distributed according to the reproduction law $\{p_k; k \in \mathbb{Z}_+\}$, then we can write

$$z_0 = 1$$

$$z_{n+1} = \sum_{j=1}^{z_n} X_{nj} \tag{3.6}$$

Let $\mathcal{B}_n = \sigma(z_0, z_1, ..., z_n)$ be the sigma algebra generated by $z_0, z_1, ..., z_n$, then

$$P[z_{n+1} = k|\mathcal{B}_n] = P\left[\sum_{j=1}^{z_n} X_{nj} = k|z_n\right] \tag{3.7}$$

Hence $\{z_n\}$ is a homogeneous Markov chain with transition probabilities:

$$p_{jk} = P[z_{n+1} = j|z_n = k] = \sum_{i_1+i_2+...+i_j=k} p_{i_1}...p_{i_j} = p_j^{*j} \tag{3.8}$$

where p_j^{*j} is the conventional convolution notation.

Lemma 3.1

Let $A \in \mathcal{B}(\mathbb{Z}_+^\infty), k \in \mathbb{N}$ and $\{z_n^{(1)}\}, \{z_n^{(2)}\}, ...$, be independent Galton-Watson processes with the reproduction law of $\{z_n\}$. Then for any $r \in \mathbb{N}$

$$P[\{z_n; n > r\} \in A | z_r = k] = P\left[\left\{\sum_{j=1}^{k} \{z_n^{(j)}\right\}; n \geq 1\} \in A\right] \tag{3.9}$$

3.3.1 Moments and generating functions

The lemma (3.1) is used to calculate the generating function of z_n,

$$f_n(s) = E[s^{z_n}] = \sum_{k=0}^{\infty} P[z_n = k] s^k \tag{3.10}$$

for $0 \leq s \leq 1$.

Lemma 3.2

Assume that $Y, X_1, X_2, ...$ is a sequence of independent $\mathbb{Z}_+$-valued random variables. Let all X_i have the same generating function g, and Y the generating function h. Then, with

$$S = \sum_{j=1}^{Y} X_j,$$

$$E[s^S] = h \circ g(s), 0 \leq s \leq 1$$

$$E[S] = E[X_1][Y]$$

$$Var(S) = Var(X_1)E[Y] + E^2[X_1]Var(Y)$$

Using the lemma (3.2) and generating function we have the following:

$$f_n = f_{n-1} \circ f = ... = f \circ f_{n-1}$$

$$E[z_n] = mE[z_{n-1}]$$

$$Var(z_n) = \sigma^2 E[z_{n-1}] + m^2 Var(z_{n-1})$$

where

$$f(s) = \sum_{k=0}^{\infty} p_k s^k = E[s^{\xi_x}] = E[s^{z_1}] = f_1(s)$$

is the reproduction generating function.

$$m = \sum_{k=1}^{\infty} kp_k = E[\xi_x] = E[z_1] = f'(1)$$

is the mean number of offspring per individual, and

$$\sigma^2 = \sum_{k=1}^{\infty} k^2 p_k - m^2 = Var(\xi_x) = Var(z_1) = f''(1) + f'(1) - (f'(1))^2$$

is the reproduction variance.

Theorem 3.3

The generating function of z_n is the composition of $f \circ f \circ ... \circ f$, its expectation is m^n and the variance is

$$\frac{\sigma^2 m^{n-1}(m^n - 1)}{(m-1)}, \text{ if } m \neq 1,$$

$$n\sigma^2, \text{ if } m = 1.$$

3.3.2 The Extinction probability

The event

$$Q = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{z_k = 0\} = \{z_n \to 0\} \tag{3.11}$$

is called the extinction of the process.

$$P(Q) = \lim_{n\to\infty} P\left[\bigcup_{k=1}^{n} \{z_k = 0\}\right] = P \lim_{n\to\infty} [z_n = 0] = \lim_{n\to\infty} f_n(0) \tag{3.12}$$

We call $P(Q)$ the extinction probability.

Theorem 3.4

The equation $f(s) = s$ has exactly one root in $[0, 1)$ if $m > 1$ and none if $m \leq 1$ and $p_1 \neq 1$. The extinction probability, $q = P(Q)$, is the smallest non-negative root of the equation, that is

$$\begin{aligned} & m > 1 \Rightarrow q < 1, \\ & p_1 = 1 \Rightarrow q = 0 \\ m = 1, \quad & p_1 < 1 \Rightarrow q = 1 \\ & m < 1 \Rightarrow q = 1 \end{aligned}$$

Proof

We discard the trivial case where $p_1 = 1$.

Since $f_n(0) \uparrow q$, and f is continuous on $[0, 1]$.

$$f(q) \leftarrow f \circ f_{n+1}(0) \rightarrow q.$$

Hence $f(q) = q$. Assume that a is any number in $[0, 1]$ such that $f(a)$. Then

$$f_1(0) = f(0) \leq f(a)$$

and

$$f_n(0) \leq a \Rightarrow f_{n+1}(0) = f \circ f_n(0) \leq f(a) = a$$

and so, for all n, $f_n(0) \leq a$ proving that $q \leq a$. So q is the smallest root.

If $m \leq 1$, $p_1 \neq 1$, $s < 1$,

$$(f(s) - s)' = f'(s) - 1 < f'(1) - 1 \leq 0$$

and $f(s) - s$ must decrease strictly. As $f(1) = 1$, it follows that $f(s) > s$ for

$0 \leq s < 1$.

If $m > 1$, $f(s)$ increases quicker towards $f(1) = 1$ than does s.

Hence $f(s) < s$ for s in some left nieghbourhood of 1.

But since $f(0) \geq 0$ there must be at least one $0 \leq s < 1$ such that $f(s) = s$

Now assume that there are two:

$$f(s_1) = s_1, \quad f(s_2) = s_2, \quad 0 \leq s_1 \leq s_2 < 1$$

With $\phi(s) = f(s) - s$, in this case:

$$\phi(s_1) = \phi(s_2) = \phi(1) = 0$$

and hence $f'(a) = f'(b)$ contradicting the fact that f' is strictly increasing- as it must be if $m > 1$. $\square$

The above results motivate the following classification of Galton-Watson process:

- Super-critical if $m > 1$
- Critical if $m = 1$
- Sub-critical if $m < 1$

3.3.3 Critical Process

Lemma 3.5

Assume that $m = 1$ and $\sigma^2 < \infty$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \left\{ \frac{1}{1 - f_n(s)} - \frac{1}{1 - s} \right\} = \frac{\sigma^2}{2}$$

uniformly in $0 \leq s < 1$.

Theorem 3.6

If $m = 1$ and $\sigma^2 < \infty$, then

$$\begin{aligned}
&(a) \quad \lim_{n\to\infty} nP[z_n > 0] = \frac{2}{\sigma^2} \\
&(b) \quad \lim_{n\to\infty} E\left[\frac{z_n}{n}|z_n > 0\right] = \frac{\sigma^2}{2} \\
&(c) \quad \lim_{n\to\infty} P\left[\frac{z_n}{n} \le u|z_n > 0\right] = 1 - exp\left(-\frac{2u}{\sigma^2}\right), u \ge 0
\end{aligned}$$

Proof

(a) $\quad nP[z_n > 0] = n\{1 - f_n(0)\} = \left\{\frac{1}{n}\left(\frac{1}{1-f_n(0)} - 1\right) + \frac{1}{n}\right\}^{-1} \to \frac{2}{\sigma^2}$

(b) $\quad E\left[\frac{z_n}{n}|z_n > 0\right] = E[z_n]/n\{1 - f_n(0)\} = \frac{1}{n}\{1 - f_n(0)\} \to \frac{\sigma^2}{2}$

(c) $\quad$ Let $u > 0$. $\quad E\left[exp\left(\frac{-uz_n}{n}\right)|z_n > 0\right] = 1 - \frac{1 - f_n(exp(-u/n))}{1 - f_n(0)}$

$$= 1 - \frac{1}{n\{1-f_n(0)\}}\left\{\frac{1}{n}\left[\frac{1}{1-f_n(exp(-u/n)} - \frac{1}{1-exp(-u/n)}\right] + \frac{1}{n(1-exp(-u/n))}\right\}^{-1}$$

$$\to 1 - (\sigma^2/2)(\sigma^2/2 + 1/u)^{-1} = 1/(1 + u\sigma^2/2)$$

3.3.4 The Total Progeny of a Branching Process

The equation

$$y_\infty = \sum_{n=0}^{\infty} z_n \tag{3.13}$$

defines y_∞ as an integer random variable if $y_\infty < \infty$. Also

$$P[y_\infty < \infty] = q$$

and

$$y_n = \sum_{k=0}^{n} z_k \uparrow y_\infty.$$

The generating function of h_n of y_n are recursively related:

It follows that h_∞,

$$h_\infty(s) = E[s^{y_\infty}],$$

satisfies

$$h_\infty(s) = sf \circ h_\infty(s)$$

3.4 General Branching Process

Consider a population with one ancestor. With each individual $x \in I$, associate one non-negative random variable λ_x, the life length of x, and one point process ξ_x, the reproduction of x. Assume (λ_x, ξ_x) are iid with probability distribution Q. This is a measure on product space $\mathbb{R}_+ \times \mathbb{N}$ with corresponding product σ-algebra. Its margin on $\mathbb{R}_+$, L

$$L(u) = P[\lambda_x \leq u] \tag{3.14}$$

is the life length distribution and its margin on $\mathbb{N}$ the reproduction law.
We shall assume

$$P[\xi_x(\lambda_x, \infty) = 0] = 1 \tag{3.15}$$

A realised individual (x, k) is born, when its mother x is aged

$$\tau_x(k) = inf\{t : \xi_x(t) \geq k\}$$

and this occurs at time $\sigma_{(x,k)}$, the birth time of (x, k).
If $x = (j_1, j_2, ..., j_n) \in I$, then

$$\sigma_x = \tau_0(j_1) + ... + \tau_{(j_1,j_2,...,j_{n-1})}(j_n)$$

Define $\sigma_0 = 0$. The individual x is alive at time $t \geq 0$ if it has been born and has not yet died.

$$\sigma_x < t < \sigma_x + \lambda_x$$

Its age is $t - \sigma_x$, and so, $t \geq 0, a \geq 0$

$$z_t^a(x) = \sum_{x \in I} z_t^a(x) < \infty \tag{3.16}$$

The stochastic process $\{z_t^a; t, a > 0\}$ is the general branching process.

Let $I_n = \{(n, x); x \in I\}$ be the set of all decendants from the ancestors n^{th} child and define

$$z_t^{(n)a} = \sum_{x \in I_n} z_t^a(x).$$

Then

$$z_t^a = z_t^a(0) + \textstyle\sum_{n=1}^{\infty} z_t^{(na)a} = z_t^a(0) + \textstyle\sum_{n=1}^{\xi_0(t)} z_t^{(n)a}, \tag{3.17}$$

Since

$$\xi_0(t) < n \Rightarrow z_t^{(n)} = 0$$

It is clear that $\{z_t^{(n)a}; t, z \geq 0\}$ is a new branching process, initiated at the birth of individual n

Proof

The random variables $z_t^a(x), x \in I$ and therefore also z_t^a and $z_t^{(n)a}$ are defined as functions of $\{(\lambda_y, \xi_y); y \in I\}$.

Define for $x \in I$ an operator S_x on the range space Ω of all (λ_y, ξ_y) by

$$S_x(\{(\lambda_y, \xi_y); y \in I\}) = \{(\lambda_{(x,y)}, \xi_{(x,y)}), y \in I\}$$

Then, since the pairs $(\lambda_x, \xi_x), x \in I$ are iid, $S_x(\{(\lambda_y, \xi_y); y \in I\})$ has the same distribution as its argument $\{(\lambda_y, \xi_y); y \in I\}$.

This implies that any composed function $f \circ S_x$ as a random variable on Ω follows the same probability law as f itself. $\square$

Lemma 3.7

For any a, t

$$z_t^{(n)a} = z_{t-\tau_0(n)}^a \circ S_n \tag{3.18}$$

Define $z_t^a(x) = 0$ for negative t

Proof

If $\tau_0(n) > t$, then for $x \in I_n$

$$\sigma_x \geq \tau_0(n) > t$$

implying that $z_t^a(x) = 0$ for $x \in I_n$, and so $z^{(n)a} = 0$

Assume that $\tau_0(n) \leq n$. Then

$$z_{t-\tau_0(n)}^a(x) \circ S_n = 1 \leftrightarrow t - \tau_0(n) - 1 < \sigma_x \circ S_n$$

$$= \sigma_{(n,x)} - \tau_0(n) \leq t - \tau_0(n) < \sigma_{(n,x)} - \tau_0(n) + \lambda_{(n,x)}$$

$$\leftrightarrow t - a < \sigma_{(n,x)} \leq t < \sigma_{(n,x)} + \lambda_{(n,x)} \leftrightarrow z_t^a(n,x) = 1.$$

The summation over all $x \in I$ completes the proof. $\square$

Theorem 3.8

For $t, a \geq 0$, z_t^a can be written as follows:

$$z_t^a = z_t^a(0) + \textstyle\sum_{n=1}^{\xi_0(t)} z_{t-\tau_0(n)}^a \circ S_n \tag{3.19}$$

If the reproduction has no multiple points, $P[\forall u : \xi\{u\} \leq 1] = 1$, this can be given in convolution form,

$$z_t^a = z_t^a(0) + \int_0^t z_{t-u}^a \circ S_{\xi_0(u)} \xi_0(du)$$

The process $\{z_t^a\}$ is generally not Markovian.

3.4.1 The Finiteness of the Process

In the general case, the process is not always finite. The finiteness is determined by the reproductive function

$$\mu(t) = E[\xi(t)] \tag{3.20}$$

and its value at the origin $\mu(0)$

Define y_t as the total number of individuals born up to time t

$$y_t(x) = \begin{cases} 1, \text{ when } \sigma_x \leq t \\ 0, \text{ otherwise} \end{cases}$$

$$y_t = \textstyle\sum_x y_t(x) \tag{3.21}$$

The criteria for the finiteness of $\{y_t\}$: $z_t \leq y_t$

Theorem 3.9

If $\mu(0) > 1$, then for all $t \geq 0$

$$P[y_t = \infty] > 0 \tag{3.22}$$

Proof

Let η_n be the number of individuals $x \in N^n$ with $\sigma_x = 0$.

The sequence $\{\eta_n\}$ is a Galton-Watson process with reproduction mean $\mu(0) > 1$. Hence there is a positive chance that $\eta_n \to \infty$. But $\eta_n \leq y_t$ for $t \geq 0$. $\square$

Theorem 3.10

If $\mu(0) < 1$ and $\mu(t)$ is finite for some $t > 0$, then

$$P[\forall t; y_t < \infty] = 1 \tag{3.23}$$

Hence $P[\forall t : z_t < \infty] = 1$.

3.4.2 Moments and Generating Function

Let the process generating function be defined by φ_t^a, where

$$\varphi_t^a(s) = E[s^{z_t^a}], \quad 0 \leq s \leq 1 \tag{3.24}$$

$$\varphi_t = \varphi_t^a \quad for \quad a > t \tag{3.25}$$

For age-dependent branching processes, let f_u be the generating function $\xi(u)$, given that $\lambda = u$ and generally, let the reproduction generating function f be

$$f(s) = E[s^{\xi(\infty)}] \tag{3.26}$$

This f is the reproduction generating function of the embedded Galton-Watson process.

It follows that

$$\varphi_t^a(s) = E\left[E\left[s^{z_t^a(0)} + \sum_{1}^{\xi_0(t)} z_{t-\tau_0(n)}^a \circ S_n | \lambda_0, \xi_0\right]\right]$$

$$= E\left[s^{z_t^a(0)} \prod_{n=1}^{\xi_0(t)} \varphi_{t-\tau_0(n)}^a(s)\right]$$

$$= E\left[s^{z_t^a(0)} exp \int log \varphi_{t-u}^a(s)\xi_0(du)\right]$$

Theorem 3.11

For each $0 \leq s \leq 1$ and $0 \leq a \leq \infty$, there is just one measurable function, φ_t, $\mathbb{R}_+ \to [0,1]$ satisfying

$$\varphi_t^a(s) = E\left[s^{z_t^a(0)} exp \int log \varphi_{t-u}^a(s)\xi_0(du)\right] \tag{3.27}$$

Proof

Let $u > 0$ be such that $\mu(u) < 1$, and suppose that there are two functions g and h from R_+ into the unit interval both satisfy the equation for some s. Then for $0 \leq t \leq u$

$$|g(t) - h(t)| \leq E\left[\left|\prod_{n=1}^{\xi_0(t)} g\{t - \tau_0(n)\} - \prod_{n=1}^{\xi_0(t)} h\{t - \tau_0(n)\}\right|\right]$$

$$\leq E\left[\sum_{n=1}^{\xi_0(t)} \left|g\{t - \tau_0(n)\} - h\{t - \tau_0(n)\}\right|\right]$$

$$\leq \sup_{0 \leq t \leq u} \left|g(t) - h(t)\right| \mu(u),$$

and

$$\sup_{0 \leq t \leq u} \left|g(t) - h(t)\right| \leq \mu(u) \sup_{0 \leq t \leq u} \left|g(t) - h(t)\right|.$$

It follows that $h(t) = g(t)$ for $0 \leq t \leq u$.

Next assume that the two functions have been shown to coincide on some interval $[0, t_0]$. We conclude the proof of showing that they are the same on $[t_0, t_0 + u]$.

Indeed, then for $t_0 \leq t \leq t_0 + u$

$$\left|g(t) - h(t)\right| \leq E\left[\sum_{0 \leq \tau_0(n) \leq u} \left|g\{t - \tau_0(n)\} - h\{t - \tau_0(n)\}\right|\right]$$

$$\leq \sup_{0\leq t\leq t_0+u} |g(t)-h(t)|\mu(u)$$

and the argument used, applies again. $\square$

Theorem 3.12

If the reproduction function is finite, then so is $E[y_t]$ and therefore also $m_t = E[z_t]$ for all t.

$$m_t^a = E[z_t^a] \tag{3.28}$$

satisfies

$$m_=^a 1_{[0,a)}(t)\{1-L(t)\} + \int_0^t m_{t-u}^a \mu(du) \tag{3.29}$$

If $m = \mu(\infty) < 1$, the subcritical case, then as $t \to \infty$

$$m_t \to 0$$

If $m = 1$, the critical case, and μ is non-lattice, then for $0 \leq a < \infty$

$$m_t^a \to \int_0^a \{1-L(u)\}du / \int_0^\infty u\mu(du) \tag{3.30}$$

If further

$$\int_0^\infty tL(dt) < \infty,$$

then also

$$m_t \to \int_0^\infty uL(du) / \int_0^\infty u\mu(du) \tag{3.31}$$

When $m > 1$, the supercritical case, μ is not lattice, and $\alpha > 0$ is the

Malthusian parameter defined by $\tilde{\mu}(\alpha) = 1$, then for $0 \leq a \leq \infty$

$$m_t^a \sim e^{\alpha t} \int_0^a e^{-\alpha u}\{1 - L(u)\} / \int_0^\infty u e^{-\alpha u} \mu(du) \tag{3.32}$$

In the lattice, the corresponding assertions hold.

3.4.3 The Extinction Probability

For a process to reach extinction, state zero should be absorbing, that is that

$$z_t = 0 \Rightarrow z_{t+u} = 0$$

for all $u \geq 0$

Certainly, $z_{t+u}(0) = 1 \Rightarrow z_t(0) = 1$.

If $z_{t+u}(j_1, ..., j_n) = 1$, it means that

$$\sigma_{(j_1,...,j_k)} \leq t + u < \sigma_{(j_1,...,j_n)} + \lambda_{(j1,...,j_n)}$$

Obviously, either $\tau_0(j_1) > t$, implying that $z_t(0) = 1$, $\sigma_{(j_1,...j_n)} \leq t$, implying that $z_t(j_1, ...j_n) = 1$, or for some $1 \leq k < n$

$$\sigma_{(j_1,...,j_k)} \leq t < \sigma_{(j_1,..j_{k+1})}$$

By the assumption of no births after a death implies that

$$\sigma_{(j_1,...,j_k)} \leq t < \sigma_{(j_1,...,j_k)} + \lambda_{(j_1,...,j_k)}$$

that is that $z_t(j_1, ..., j_k) = 1$. Thus, if $z_{t+u}(x) = 1$ for some $x, z_t(x') = 1$ for some x'. In other words, $z_{t+u} > 0 \Rightarrow z_t > 0$.

The extinction Q of the process can now be defined as

$$Q = \{\exists t \in R_+; z_t = 0\} = \{\exists n \in Z_+; z_n = 0\}$$

and

$$q = P(Q) = \lim_{t \to 0} P[z_t = 0] = \lim_{t \to \infty} \varphi_t(0).$$

The limits are monotone. q is independent of the time structure of the process.

Theorem 3.13

The extinction probability q is the smallest non-negative root of the equation $f(s) = s$. The size of q is determined by $m = f'(1) = \mu(\infty)$

Proof

Let q^* be any non-negative root of the equation

$$f(s) = s$$

f is non-decreasing on the interval $[0, 1]$, since it has a non-negative coefficient, and thus

$$s_1 = f(s_0) = f(0) \leq s(q^*) = q^*$$

From this, we have

$$s_2 = f(s_1) \leq s(q^*) = q^*$$

By induction $s_n \leq q^*$, for $n = 1, 2, ...$

so $s = \lim_{n \to \infty} s_n \leq q^*$

It follows that s is the smallest non-negative root. □

Theorem 3.14

Except in the degenerate case $P[\xi_(\infty) = 1] = 1$,

$$P[z_t \to 0 \text{ or } \infty] = 1$$

3.5 Applications

3.5.1 The Impact of Gene-tree / Species-tree discordance on diversification Rate

One of the uses of evolutionary trees in many areas of biology, is that they render a foundation for understanding the process. An important requirement of the evolutionary trees, is to provide credible estimates of the underlying species-tree. Many tree based statistics are sensitive to bias and error in phylogenic inference. The use of gene-trees rather than species-trees might cause bias in molecular phylogenies. This is due to the fact that gene-trees have disproportionate lengthening of terminal branches. Another source of bias is discordance between gene-trees and the underlying species-tree.

The Pybus and Harvey's γ, is a statistic which uses ordered inter-mode distances on dated phylogeny to determine when diversification occurs. This is commonly used to detect deviations from pure birth models. The γ-statistic has the quality that under a pure-birth process, the distribution follows a standard normal distribution.

The statistic is defined as follows:

Let $g_2, g_3, ..., g_n$ be inter-mode distances of a reconstructed phylogeny with n taxa.

$$\gamma = \frac{(\frac{1}{n-2}\sum_{i=2}^{n-1}(\sum_{k=2}^{i} kg_k)) - \frac{\tau}{2}}{\tau\sqrt{\frac{1}{12(n-2)}}}$$

where $\tau = (\sum_{j=2}^{n} jg_j)$

The above was modified from Cox and Lewis (1966) by Pybus and Harvey. Burbrink and Pyron (2011) justify their claims by simulating gene-trees within species trees. They use a Yule process and a birth and death process. They make use of a selection of values for the population parameter, and then observe how gene-trees and species-tree estimates diverge under known conditions. They then calculate the statistic error and the tree-depth error, and then test for correlation. They then use the Pybus and Harvey's γ to test their predictions. Once they have completed the simulations, they look if their predictions are true for an empirical data set, namely that of a lizard, genus Sceloporus.

Burbrink and Pyron (2011) have found that their simulations support their predictions that earlier divergences are formed by gene-trees. Estimates of the diversification rates which are produced by high values of the population parameter differ significantly. Low values of the population parameter have less of an effect on the bias.

3.5.2 Evolutionary Rescue in Structured Populations

Environmental changes such as global warming or an introduction of a competing species, if severe, can cause the extinction of a population unless it adapts to the new conditions. According to Uecker et al. (2013), a population has one of the following three options:

- Disperse and find an alternate habitat.
- Change due to phenotypic plasticity without genetic adaption.
- Evolution by genetic adaption.

There are generally two approaches to modelling evolutionary rescue. The first type is a quantitative approach in genetics. Small effects to fitness are contributed by a lot of loci. Additive genetic variance plays an important part. The second class, which is used in this article, is adaption relying on a single mutation. Uecker et al. (2013) examine how ecological factors affect the probability of evolutionary rescue. Genetic variation and selection strength are factors that affect the rate of evolution. They use a generalised island model where islands experience environmental deterioration one after the other, to describe the population structure.“The analysis is based on the mathematical theory of time-inhomogeneous branching processes and is complemented by computer simulations”(Uecker et al. (2013)). They have found that evolutionary rescue occurs more readily in harsher environmental changes, due to the fact that mutant individuals experience reduced competition.

Uecker et al. (2013) assumes the life cycle to be as follows:

1. Reproduction and mutation- A large number of offspring, X, is produced by each individual of the parent generation.
2. Migration- A fraction of m migrates and of those, $1/d$ settles that deme.
3. Selection and density- A maximum of k can be supported in each deme.

From the Island model, the number of mutants is distributed as follows:

$$bin\left(p = \frac{\alpha(N_m^{(i)'} + uN_w(i)')}{(1-u)N_w^{(i)'} + \alpha(N_m^{(i)'} + uN_w^{(i)'})}\right)$$

where

$$N_{w/m}^{(i)'} = (1-m)N_{w/m}^{(i)} + \frac{m}{d}\sum_{k=1}^{d} N_{w/m}^{(k)}$$

The limit as $X \to \infty$, the number of wild type and mutant individuals after selection prior to density regulation is distributed as follows

$$N_w^{(i)} \sim Poisson\left((1-u)(1-r)N_w^{(i)'}\right)$$

$$N_m^{(i)} \sim Poisson\left((1+S_i)(N_m^{(i)'} + uN_w^{(i)'}\right)$$

Uecker et al. (2013) model the dynamics of wildtype population deterministically, failing to take in to account demographic stochasticity. To fix this, Uecker et al. (2013) calculate establishment probabilities using a time inhomgeneous branching process. Since mutation does not occur often, the mutant offspring experience independent fates and therefore can be described by a branching process. Uecker et al. (2013) focus on limiting cases in a lesser dimension and they use single difference equations to model the wildtype's deterministic dynamics. The early phase of mutant growth is modelled by a single-type branching process with a time-dependent effective growth parameter $s_{eff}(t)$.

Uecker et al. (2013) approximate the discrete-time branching process by a continuous time branching process. For the continuous branching process they used the following per capita birth and death rates:

$$\lambda(t) = 0.5 + 0.5 sign[\ln(1 + s_{eff}(t))] min[|\ln(1 + s_{eff}(t))|, 1]$$

$$\mu(t) = 0.5 - 0.5 sign[\ln(1 + s_{eff}(t))] min[|\ln(1 + s_{eff}(t))|, 1]$$

where

$$\hat{s}_{eff}(t) := sign[\ln(1 + s_{eff}(t))] min[|\ln(1 + s_{eff}(t))|, 1]$$

3.5.3 Potential of Branching Processes as a Modeling Tool for Conservation Biology

Gosselin and Lebreton (2000) note that there is a need for models with enough biological relevance and enough mathematical tractability to solve problems.

Extinction models need the following structural characteristics:

- The model must be able to consider different types of variability in population processes.

- It should emphasis the demographic parameters prior to genetic one.
- It should be written in terms of demographic parameters.

In this article, discrete-time branching processes are discussed as Gosselin and Lebreton (2000) meet the above characteristics. Branching processes also have the following characteristics relevant to population variability analysis:

- Population sizes in models only take non-negative integers.
- Branching processes are fundamentally stochastic.
- Individuals are considered explicitly
- Branching processes can be generalised to account for several types of individuals.

The Bienayme'-Galton-Watson Branching process, is an example of a density-independent branching process that is used. This is said to be the simplest due to the fact that the individuals performance is constant over time and is identical over time. There is no age structure and the expected individual performance is constant.

$$E[x_{nj}] = m$$

each individual, on average is replaced by m individuals.

$$E[z_n] = mE[z_{n-1}]$$

$$E[z_n] = m^t E[z_0]$$

If $m > 1$, we have the supercritical case, the population diverges over time. When $m \leq 1$, ultimate extinction is certain, this is the subcritical case.

Gosselin and Lebreton (2000) study the behaviour of the subcritical case before extinction. “In a subcritical BGW BP, the probability distribution at time t of a population size conditioned on non-extinction, converges, when t tends to infinity. It converges to a probability distribution $(b_k)_{k\in N}$, called the quasi-stationary distribution” (Gosselin and Lebreton (2000)). It is stationary in the sense that probability distribution converges irrespective of initial value. Also, it only converges conditional on non-extinction.

Under mild conditions, the expectation of $\sum_{k=0}^{\infty} kb_k$, towards which $E[z_t|z_t > 0]$ converges when $t \to \infty$, and higher moments of the quasi-stationary distribution are finite.

A case study was done on the Alsace White Stalk. A more general model was used for reaching some realism and applicability.

It is concluded that the structure of branching processes allow us to handle population variability analysis models with a variety of demographic features.

3.5.4 Genealogy for Supercritical Branching Processes

Lagerås and Martin-Löf (2006) study a specific kind of branching processes, the supercritical branching process. Lagerås and Martin-Löf (2006) has defined a supercritical branching process as “a process where each particle gives birth to at least two new particles”. Since the process can never die out, it is also known as an immortal branching process. The first part of the article is dedicated to recalling some well-known results mentioned in Chapter 3 above. This is then followed by basic results on the compound geometric distribution. The main result that Lagerås and Martin-Löf (2006) has proved is that “ the number of individuals in an immortal branching process has a compound geometric distribution”. A description of the Yule process, which is said to be the simplest immortal branching process is given.

The Model

Lagerås and Martin-Löf (2006) have used the following notation for their branching process.

Let $Z = \{Z_t\}_{t \geq 0}$, be the number of new particles at each birth with distribution $\{p_k\}_{k \geq 0}$. The process starts with $Z_0 = 1$.

"If a process is in state i at any time, then it remains there for an amount of time which is exponentially distributed with parameter $i\mu$, where μ is the intensity of the process. It jumps to state $j > i - 1$ with probability p_{j-i+1} and the process repeats in this manner"(Lagerås and Martin-Löf (2006)).

Let the generating function of $\{p_k\}$ be given by

$$f(s) = \sum_{k=0}^{\infty} p_k s^k$$

Then $m = f'(1)$.

A condition for the process not to explode $(P(z_t < \infty) = 1$

$$\int_{1-\epsilon}^{1} \frac{ds}{f(s) - s}$$

diverges for all ϵ, $0 < \epsilon < 1$.

This holds for $m < \infty$.

The Kolmogorov forward equation for the generating function $F(s,t) = E[s^{Z_t}]$ is

$$\frac{\partial}{\partial t} F(s,t) = \mu(f(s) - s) \frac{\partial}{\partial s} F(s,t)$$

From the above equation, $p_1 = 0$, since Z is a branching process with $p_1 > 0$ and intensity μ, then it is distributed in the same way as Z^* with intensity $\mu^* = \dfrac{\mu}{(1-p_1)}$. The generating function for the offspring is

$$f^*(s) = \frac{f(s) - p_1 s}{1 - p_1} = \sum_{k=0}^{\infty} p_k^* s^k$$

where $p_1^* =$ and $p_k^* = \dfrac{p_k}{(1-p_1)}$ for $k = 0, 2, 3...$

The Kolmogorov backward equation for the generating function is

$$\frac{\partial}{\partial t}F(s,t) = \mu(f(F(s,t)) - F(s,t))$$
$$\frac{\partial}{\partial t}F(s,t) = \mu F(s,t)(g(F(s,t)) - 1)$$
$$\frac{\partial}{\partial t}F(s,t) = \mu F(s,t)(F(s,t)k(F(s,t)) - 1)$$

where $f(s) = sg(s) = s^2k(s)$, where $g(s)$ and $k(s)$ are generating functions for distributions on $\mathbb{N}$ and $\mathbb{N}_0$, respectively.

An interesting connection between a supercritical process and immortal branching process, found by Lagerås and Martin-Löf (2006), is given by the following:

Proposition 3.5.1

The proportion $\dfrac{\tilde{Z}_t}{Z_t}$ will, conditional on Z exploding, converge almost surely to $1-q$ as $t \to \infty$, where q is the extinction probability of the supercritical branching process, and

Z is a supercritical branching process,

$\tilde{Z}_t$ is the number of individuals before t.

"Since individuals behave independently, conditional on $Z_t > n$, we have that $\tilde{Z}_t \sim Bin(n, 1-q)$. Furthermore, $\tilde{Z} = \{\tilde{Z}_t\}_{t \geq 0}$ itself is an immortal branching process if it is positive"(Lagerås and Martin-Löf (2006)).

Other results proved by Lagerås and Martin-Löf (2006) are:

Proposition 3.5.2

All finitely divisible random variables X with $P(X = 0) > 0$ are compound Poisson, and all infinitely divisible distributions can be determined as the weak limit of a compound Poisson distribution. Furthermore, all weak limits of infinitely divisible distributions are infinitely divisible.

Proposition 3.5.3

All compound exponential random variables X with $P(X = 0) > 0$ are compound geometric, and all compound exponential distributions can be obtained as the weak limit of compound geometric distributions. Furthermore, all weak limits of compound exponential distributions are compound exponential.

Lemma 3.5.4

If X, given that $X > 0$, is compound exponential, then X is compound geometric.

Theorem 3.5.5

All immortal branching processes are finitely divisible.

Theorem 3.5.6

All immortal branching processes have compound geometric distributions.

3.5.5 Applied Probability and Stochastic Processes

The following example is extracted from Beichelt (2016). Let the number of offspring Y have a mixed Poisson distribution with continuous structure parameter L with density function $f_L(\lambda)$. Then Y has the following generating function:

$$M(z) = \int_0^\infty e^{\lambda(z-1)} f_L(\lambda) d\lambda$$

The structure parameter L is supposed to have a gamma distribution with density given by

$$f_L(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0, \beta > 0$$

Then $M(z)$ becomes

$$M(z) = \int_0^\infty e^{\lambda(z-1)} f_L(\lambda) \lambda$$
$$M(z) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-(\beta+1-z)} \lambda^{\alpha-1} d\lambda$$

Substituting $x = (\beta + 1 - z)\lambda$ gives the final form of $M(z)$

$$M(z) = \left(\frac{\beta}{\beta + 1 - z} \right)^\alpha$$

which is the generating function of a negative binomial distribution with parameters α and β. The first derivative is

$$M'(z) = \frac{\alpha \beta^\alpha}{(\beta + 1 - z)^{\alpha+1}}$$

Thus, the average number of offspring is

$$E[Y] = M'(1) = \frac{\alpha}{\beta}$$

A general solution of equation $M(z) = z$ has a complicated structure. Two special cases are considered.

1. $\alpha = 1$

   In this case, L has an exponential distribution with parameter β. The equation $M(z) = z$ becomes

   $$z^2 - (\beta + 1)z + \beta = 0$$

and the solutions are

$$z_1 = 1 \quad \text{and} \quad z_2 = \beta$$

Hence the probability of extinction will be

$$\pi_0 = \begin{cases} 1 & \text{for} \quad \beta \geq 1 \\ \beta & \text{for} \quad \beta < 1 \end{cases}$$

2. $\alpha = 2, \beta = 1, 2$

   In this case $M(z) = z$ becomes

$$z^3 - 4.4z^2 + 4.84z - 1.44 = 0$$

   The solutions are $z_1 = 1, z_2 = 0.496$.

   The probability of extinction is $\pi_0 = 0.496$

3.5.6 Assessing local population vulnerability with branching process models: An application to wind energy development

Erickson et al. (2015) look at branching processes to quantify the impact of environmental development on local populations. Erickson et al. (2015) considered using branching processes since they are easily calculated and that they account for demographic stochasticity.

The Model

Let X_t denote the size of a population at time t. Erickson et al. (2015) are initially interested in the probability of extinction of a population at time t.

$$e_t = P(X_t = 0 | X_0 = 1)$$

If the probability of survival , growth and reproduction of the individuals are independent, then

$$e_t = p_0 + p_1 e_{t-1} + p_2 e_{t-11}^2 + p_3 e_{t-1}^3 + ... = f(e_{t-1}) \tag{3.33}$$

where

f- the generating function

p_i- probability an individual produces i offspring including itself.

Since $e_0 = 0$, by iterating (3.35), one can compute e_t

$$e_t = f^t(0)$$

Since

$$0 \leq e_1 \leq e_2 \leq ... \leq 1$$

we know that e_t has a limit e^*, known as the asymptotic probability of extinction. The limit is found by solving

$$e^* = f(e^*)$$

If

- $\lambda \leq q$, then $e^* = 1$, and the population will definitely go extinct.
- $\lambda > 1$, then $e^* < 1$, and there is a probability that the population will survive.

$\lambda = p_1 + 2p_2 + 3p_3 + ...$, is the expected growth of the population.

For modeling species with different phases, a multi-type branching process should be used. The generating function for a multi-type branching process, follows the same process as above, but with an increase in dimensions.

In a multi-stage case, with n stages, the extinction probability is modeled with an n-dimensional vector, $\bar{e}_t$. The i^{th} element is the probability the entire population going extinct by time t, given that it started with one individual at time i.

Assume growth is followed by reproduction and recruitment into newborn stage. Then we have

$$\bar{e}_t = (1 - S)I + (STE^d_{t-1}PE^r_{t-1}J)$$

where

I- $n \times n$ identity matrix

S- $n \times n$ survival matrix

$$\begin{bmatrix} s_1 & 0 & 0 & \dots & 0 \\ 0 & s_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & s_n \end{bmatrix}$$

where s_i is the probability of survival of stage i in a given time step.

T- $n \times n$ transition matrix

$$\begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ t_{21} & t_{22} & t_{23} & \dots & t_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & t_{n3} & \dots & t_{nn} \end{bmatrix}$$

where t_{ij} is the probability of transitioning from stage i to j.

E^d_{t-1}- $n \times n$ time-varying matrix, denoting the distribution of individuals after they transition.

$$\begin{bmatrix} e_{1,t-1} & 0 & 0 & \dots & 0 \\ 0 & e_{2,t-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & e_{n,t-1} \end{bmatrix}$$

P- $n \times N$ reproduction matrix

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1N} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nN} \end{bmatrix}$$

where p_{ij} is the probability of i producing j newborns in a given time step. E^r_{t-1}- $n \times n$ time-varying matrix, tracking the extinction probability of population due to reproductive events.

$$\begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ e_{1,t-1} & e_{2,t-1} & e_{3,t-1} & \cdots & p_{n,t-1} \\ e^2_{1,t-1} & e_{2,t-1} & e^2_{3,t-1} & \cdots & e^2_{n,t-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^N_{1,t-1} & e^N_{2,t-1} & e^N_{3,t-1} & \cdots & e^N_{n,t-1} \end{bmatrix}$$

J- Juvenile distribution vector,

$$J = [j_1, j_2, ..., j_n]^T$$

where j_i is the probability that newborns end up at stage i during their first time step.

3.5.7 Estimating the Survival of the Blue Crane population

In previous research (Docrat (2014)), capture-recapture data was used to estimate the survival probabilities of the Blue Crane. The Cormack-Jolly-Seber (CJS) model was used to estimate the survival probabilities of the Blue Crane, South Africa's national bird. Taking the biology of the bird into account, several models were used under the Bayesian framework. The deviance information criterion (DIC) was used to select the best model.

The Blue Crane (Antropoides paradise) is South Africa's national bird. Blue Cranes are tall and small in comparison to other cranes. The crane is a pale blue-grey colour becoming darker at the head. Blue Crane live in dry grassy uplands and nest both uplands and in wetlands. Blue Cranes are found mainly in the Eastern Cape, Lesotho, Swaziland, Free State, Limpopo, Gauteng and Mpumalanga.

The real data set on the Blue Cranes consisted of 894 individuals that were captured and marked over a 20 year period (from1993-2012). The data were obtained from the Endangered Wildlife Trust. When captured, an individual was classified as an adult, juvenile, chick or unknown. Two age groups were considered. The first age group was chicks and juveniles less than a year old and the second age group was that of adults and unknown, which consisted of individuals older than a year. Based on the best model, the mean juvenile and adult survival probabilities were 0.3886 and 0.8085 respectively.

Using the above information we can create a two-stage model to derive an expression for the probability of extinction using methods described in Erickson et al. (2015).

The Model

The model will consist of two stages. The first stage is juveniles and second stage is adults as defined above.

Let us define the following:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

I be the 2×2 identity matrix

Using the above quoted mean survival rates for juveniles and adults, we get the survival matrix to be

$$S = \begin{bmatrix} 0.3886 & 0 \\ 0 & 0.8085 \end{bmatrix}$$

The transition matrix, conditional on survival is as follows:

$$T = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

The 2×2 time varying matrix is as follows:

$$E^d_{t-1} = \begin{bmatrix} e_{1,t-1} & 0 \\ 0 & e_{2,t-1} \end{bmatrix}$$

According to literature, Blue cranes usually produce at most 2 eggs (Harrison and Cherry (1997)). Assuming that males and females are equally likely, we will calculate the reproduction probability with a binomial distribution.

$$p_i \sim bin(2, 0.5)$$

Also, assuming juveniles do not reproduce, we get the following reproduction matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

The varying matrix tracking the extinction probability due to reproduction is

$$E^r_{t-1} = \begin{bmatrix} 1 & 1 \\ e_{1,t-1} & e_{2,t-1} \\ e^2_{1,t-1} & e^2_{2,t-1} \end{bmatrix}$$

Lastly, we have the juvenile vector as $J = [1, 0]^T$

The probability of the entire population in stage i going extinct at time t, given that it started with on individual, as per Erickson et al. (2015), is:

$$\bar{e}_t = (1 - S)I + (STE^d_{t-1}PE^r_{t-1}J)$$

Using the values above we get

$$e_{1,t} = 0.6114 + 0.09715e_{2,t-1} + 0.1943e_{2,t-1}e_{1,t-1} + 0.09715e_{2,t-1}e^2_{1,t-1}$$

$$e_{2,t} = 0.1915 + 0.20213e_{2,t-1} + 0.40425e_{2,t-1}e_{1,t-1} + 0.20213e_{2,t-1}e^2_{1,t-1}$$

Chapter 4

Summary and Conclusion

Stochastic models play an integral role in modeling the random influences that occur in a wide variety of areas and fields. These stochastic processes could be of various types, but the two we are interested in throughout this research are branching and birth and death processes. These processes account for the stochasticity which we find present in random events, which makes a favourable tool in areas like engineering, finance, ecology and genetics to name a few.

Branching and birth and death processes are easy to calculate and require little computation. This results from the fact that one can use generating functions to calculate probabilities and other moments. Generating functions, which are useful tools in mathematics, mainly in combinatorics, makes the calculations and models that we are faced with more tractable.

We start off with the introduction of generating functions, first defined in mathematics. We then look at the generating functions applied to statistics along with all of its variations. Generating functions in connection with differential equations allow us to understand and derive different kinds and aspects of branching and birth and death processes.

Birth and death processes are Markov chains where transitions to neighbouring states are possible. They can either be discrete or continuous depending on the time. Discrete-time birth and death processes are used in any situation when the time intervals are discrete. A well-studied example of such a discrete birth and death process is Gambler's Ruin, which is examined in section(2.2.2). Another use of discrete birth and death processes is to serve as an approximation for continuous birth and death processes (Beichelt (2016)).

Continuous birth and death processes tend to be more important than discrete processes as they are applied to a wider variety of areas. Continuous birth and death processes are presented in detail. We start with the homogeneous birth and death process before considering variations such as pure birth processes, pure death processes, processes with immigration and non homogeneous birth and death processes. The quasi-stationary birth and death process along with stationary distributions are also covered in this research. An important section is that of population extinction as it has direct applications in areas such as biology and ecology. We finish the theory of birth and death processes with the cumulative population, which too has many uses in biology, ecology and demography.

After the theoretical background of birth and death processes, we look at the applications. Levin et al. (2001) looks at the HIVvirus with the use of a mathematical model to understand the population dynamics of the virus. Birth and death processes were used to model the changes in numbers of infected people. New infections were taken as the "births" and when a host was removed from the system, it was taken to be the "deaths". Yu et al. (2016) applied birth and death processes to game theory. They modeled the Rock-Paper-Scissors game as a finite, state-dependent quasi-stationary birth and death process. Yu et al. (2016) used the limiting distribution of the quasi-stationary birth and death process to explain the long-run equilibrium of the game. Higgins and Mahadevan (2010) developed a master equation to model red blood cells. This model was then used to detect early cases of anemia. In their model, birth and death processes account for the constant flow of red blood cells. Wang et al. (2013) look at fossil data and investigate the macro-evolutionary prediction of an individual based on birth and death model, where speciation and extinction rates emerge from population dynamics. Waema and Olowofeso (2015) look at birth and death processes applied to the modeling of the HIV transmission rates. They use birth and death properties for the properties mentioned above, namely that they account for stochasticity and that they are tractable. With the use of generating functions, they derive the moments of the transmission rates of different cases. Another application of birth and death processes in game theory was done by Zukewich et al. (2013), who investigate biologically motivated models of evolution in finite populations. Beichelt (2006) has provided many applications of branching processes in engineering, science and finance. We look at the application in queuing theory discussed in Beichelt (2006). We look at the $M/M/\infty/-$ model which is a homogeneous birth and death process.

Branching processes are Markov chains where each individual produces a number of offsprings. The theoretical background of branching processes in this dissertation focuses on two main types, namely Galton-Watson branching processes and general branching processes. The use of generating functions to obtain moments is discussed. Other sections covered are the extinction probability and the total progeny of Galton-Watson processes. The Galton-Watson process is the simplest of the branching processes. It is generalised by the general branching process. An important classification of branching processes, is that of critical processes. To ascertain whether a process is critical, sub-critical or super-critical, helps us better understand the extinction probabilities, if they exist.

The application of branching process also span over many areas. Burbrink and Pyron (2011) looks at bias formed from tree-based statistics. Uecker et al. (2013) studies the effect of environmental changes on populations. They use island models to describe population structure, but their analysis is based on the mathematical modeling of time in-homogeneous branching processes. Gosselin and Lebreton (2000) look at using branching processes as a modeling tool in conservational biology. They begin by motivating the use of branching process in biology and then go on to the use of Galton-Watson branching processes. They specifically look at the behaviour of sub-critical populations, before extinction. They look at the distribution it tends to before extinction, known as the quasi-stationary distribution. Lagerås and Martin-Löf (2006) takes a very theoretical approach of branching processes. They study "immortal" branching processes which they then prove to be equivalent to supercritical branching processes. They derive many results of supercritical branching processes. Erickson et al. (2015) looks at branching processes applied to conservational ecology. They calculate the extinction probabilities of 4 populations which consist of different stages. In order to do this, Erickson et al. (2015) use a multi-type branching process. Beichelt (2016) provides some examples of how the generating functions in branching processes can be used to obtain the moments in a concise manner. An expression for the extinction probability of the Blue Crane is derived from work that was done in previous research. The survival probabilities of the Blue Crane used, were obtained from a previous study by Docrat (2014).

It is clear that branching and birth and death processes have many uses due to their properties. They have applications in many fields and there is a possibility of using them in more unconventional fields. Such theoretical approach to branching and birth and death processes has some limitations.

One of the problems, is that, though most of the fundamentals were covered, there are other sections which exist.
Future research, will focus on derivation of more applications of branching and birth and death processes including focusing in detail in a particular field.

Bibliography

Allen, L. J. (2010). *An introduction to stochastic processes with applications to biology*. CRC Press.

Axelrod, D. E. and Kimmel, M. (2002). Branching processes in biology.

Beichelt, F. (2006). *Stochastic Processes in Science, Engineering and Finance*. CRC Press.

Beichelt, F. (2016). *Applied Probability and Stochastic Processes*. CRC Press.

Burbrink, F. T. and Pyron, R. A. (2011). The impact of gene-tree/species-tree discordance on diversification-rate estimation. *Evolution*, 65(7):1851–1861.

Docrat, R. (2014). Estimating the survival of the blue crane population. Unpublished research report.

Drmota, M. (2009). *Random trees: an interplay between combinatorics and probability*. Springer Science & Business Media.

Erickson, R. A., Eager, E. A., Stanton, J. C., Beston, J. A., Diffendorfer, J. E., and Thogmartin, W. E. (2015). Assessing local population vulnerability with branching process models: an application to wind energy development. *Ecosphere*, 6(12):1–14.

Goel, N. S. and Richter-Dyn, N. (2013). *Stochastic models in biology*. Elsevier.

Gosselin, F. and Lebreton, J.-D. (2000). Potential of branching processes as a modeling tool for conservation biology. In *Quantitative methods for conservation biology*, pages 199–225. Springer.

Harrison, J. and Cherry, M. (1997). *The atlas of southern African birds*, volume 1. BirdLife South Africa Johannesburg.

Higgins, J. M. and Mahadevan, L. (2010). Physiological and pathological population dynamics of circulating human red blood cells. *Proceedings of the National Academy of Sciences*, 107(47):20587–20592.

Jagers, P. (1975). *Branching processes with biological applications.* Wiley.

Kimmel, M. and Axelrod, D. E. (2015). Biological background. In *Branching Processes in Biology*, pages 19–36. Springer.

Lagerås, A. N. and Martin-Löf, A. (2006). Genealogy for supercritical branching processes. *Journal of applied probability*, 43(04):1066–1076.

Levin, B., Bull, J., and Stewart, F. (1996). The intrinsic rate of increase of hiv/aids: epidemiological and evolutionary implications. *Mathematical biosciences*, 132(1):69–96.

Levin, B. R., Bull, J., and Stewart, F. M. (2001). Epidemiology, evolution, and future of the hiv/aids pandemic. *Emerging infectious diseases*, 7(3 Suppl):505.

Lundberg, O. (1964). *On Random Processes and Their Application to Sickness and Accident Statistics.* Almqvist & Wiksells.

Uecker, H., Otto, S. P., and Hermisson, J. (2013). Evolutionary rescue in structured populations. *The American Naturalist*, 183(1):E17–E35.

Waema, R. and Olowofeso, O. E. (2015). Mathematical modeling for human immunodeficiency virus (hiv) transmission using generating function approach.

Wang, S., Chen, A., Fang, J., and Pacala, S. W. (2013). Speciation rates decline through time in individual-based models of speciation and extinction. *The American Naturalist*, 182(3):E83–E93.

Yu, Q., Fang, D., Zhang, X., Jin, C., and Ren, Q. (2016). Stochastic evolution dynamic of the rock–scissors–paper game based on a quasi birth and death process. *Scientific Reports*, 6.

Zukewich, J., Kurella, V., Doebeli, M., and Hauert, C. (2013). Consolidating birth-death and death-birth processes in structured populations. *PLoS One*, 8(1):e54639.

Appendix

Mixtures of Random Variables

Let P_X denote the probability distribution of any random variable X. This distribution can depend on one or more parameters. The notation $P_{X,\theta}$ will be used to indicate the dependency on a special parameter θ. We will also use the following notation:

$$F_X(x) = F_X(x, \theta)$$

for the cumulative distribution of X and

$$f_X(x) = f_X(x, \theta)$$

for the probability distribution of X.

Mixtures of random variables or their probability distributions arise from the assumption that the parameter θ is a realisation of a random parameter θ, and all the probability distributions being elements of the set $\{P_{X,\theta}, \theta \in R_\theta\}$ are mixed.

Mixed Poisson Process

Let X have a Poisson distribution with parameter λ:

$$P_{X,\lambda} = \{P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, i = 0, 1, 2, ...; \lambda > 0\}$$

A random variable Y with range $\{0, 1, ...\}$ is said to have a mixed Poisson distribution if its probability is a mixture of Poisson distributions $P_{X,\lambda}$ with regard to any structure distribution.
If the structure distribution is given by the density $f_L(\lambda)$ of a positive random variable L, the distribution of Y is given by

$$P(Y = i) = \int_0^\infty \frac{\lambda^i}{i!} e^{-\lambda} f_L(\lambda) d\lambda$$

$i = 0, 1, ...$

Properties of mixed Poisson distributed random variable Y:

1. $E[Y] = E[L]$
2. $Var(Y) = E[L] + Var(L)$
3. $P(Y > n) = \int_0^\infty \frac{\lambda^n}{n!} e^{-\lambda} \bar{F}_L(\lambda) d\lambda$

where $F_L(\lambda) = P(L \leq \lambda)$ is the distribution function of L and
$\bar{F}_L(\lambda) = 1 - F_L(\lambda)$.