INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
RICE UNIVERSITY

Multi-type branching process models of cell proliferation

by

David N. Stivers

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Marek Kimmel, Chairman
Professor of Statistics
Rice University

James R. Thompson
Professor of Statistics
Rice University

Richard H. Gomer
Assistant Professor of Biochemistry & Cell Biology
Rice University

Houston, Texas
May, 1995
Abstract

Multi-type branching process models of cell proliferation

by

David N. Stivers

[Axelrod et. al., 1993] carried out experiments in which colonies of mouse fibroblast cells were dispersed and seeded to form secondary colonies. They found that there existed highly significant correlations between sizes for the primary and secondary colonies, which were accompanied by high variances of colony sizes. Simulation experiments by these same authors showed that the data are fitted only by a computer simulation model in which the lifetimes of cells in a secondary clone are, to a large extent, determined by the lifetimes of the founder cell of the clones, selected randomly from a primary colony. To mathematically analyze this unexpected result, we derive, for previously uninvestigated multi-type Bellman-Harris and Galton-Watson branching process models, “sampling formulas” which make it possible to find the mixed second moment of the cell counts in the primary and secondary colonies.

The derived expressions are difficult to evaluate explicitly for all but fairly simple cases; hence, it was necessary to write computer code to do so numerically. Despite
an inherent trade-off between correlation and variance, it was possible to match the Galton-Watson model to the observations. Numerical results for the Bellman-Harris process are still pending; however, some simplified asymptotic equivalents were derived and analyzed. It was found analytically that, in these models, the correlation goes to zero asymptotically. We note that we have added to the number of explicit expressions for branching processes. We present these results and some of their consequences.
Acknowledgments

Thanks to Dr. David E. Axelrod for the use of his data and for constructive comments on parts of this manuscript. Further thanks to Dr. Robert Bartoszynski for reading and for commenting on pieces of the mathematical analysis. Thanks also go to Drs. Tony Gorry and George Phillips of the W. M. Keck Center for Computational Biology, which provided financial support for my final year.

Dr. Frank Jones, Eric Davis, Sue Minkoff, Andrew Stivers and my parents deserve thanks for their support and encouragement through some of the more difficult moments of the past few years. Melinda Oliver is entitled to more than just gratitude for her continued patience and support, as well as for the odd good luck kiss.

Thanks are due to my thesis committee, Drs. Richard Gomer and James Thompson for taking the time to read and comment on this work. Finally, I am deeply indebted to Dr. Marek Kimmel for the patience of Job and the pulsion of Paul.
Contents

Abstract ii
Acknowledgments iv
List of Illustrations ix
List of Tables xi
Preface xii

1 Introduction 1

2 Branching Processes 7
  2.1 The Galton-Watson branching process .................... 7
  2.2 The multi-type Galton-Watson branching process .......... 10
  2.3 The Bellman-Harris branching process .................... 14
  2.4 The multi-type Bellman-Harris branching process .......... 16

3 Branching Process Models of Cell Proliferation 18
  3.1 Biological mechanisms of dependence .................... 18
  3.2 Branching process models in a biological context ........ 20
4 Methylation

4.1 Methylation maintenance and genesis ............................................. 29
4.2 Methylation and mutability .............................................................. 31
4.3 Methylation and cancer ................................................................. 32

5 Description of Experimental Data ...................................................... 34

5.1 Primary and secondary colonies ..................................................... 34
5.2 Sister colonies .............................................................................. 36
5.3 Comment on data ..................................................................... 38

6 Simulation of Data ............................................................................... 40

6.1 Models ......................................................................................... 40
6.2 Results of simulation ................................................................ 43

7 Time-Discrete Generational Inheritance Model: Basic Properties ......... 44

7.1 The sampling distribution and its properties ................................. 44
7.2 Expressions for colony size mean and variance .............................. 45
7.3 The covariance of primary and secondary colonies ...................... 46
7.4 The covariance of sister-sister colonies ........................................ 50
7.5 Correlations ................................................................................ 53
7.6 The distribution of $I_p$ ................................................................ 54
8 Time-Discrete Generational Inheritance Models: Description of Variants Used

8.1 Basic 3-type model (no mutation, no cell death) 55
8.2 3-type model: “mutation” only 57
8.3 3-type model: “mutation” and variable lifetimes 59
8.4 3-type model: “mutation,” variable lifetimes, and cell death 61
8.5 4-type models 62

9 Time-Discrete Generational Inheritance Model: Numerical Investigation

9.1 3-type models: “mutation” only 65
9.2 3-type models: “mutation” and variable lifetimes 66
9.3 3-type models: “mutation,” variable lifetimes, and cell death 67
9.4 4-type models 67

10 Discussion of Numerical Results 69

11 Time-Continuous Clonal Inheritance Model: Single-type Case

11.1 Introduction 73
11.2 The sampling distribution and its properties 74
11.3 Concluding remarks ........................................ 81

12 Time-Continuous Clonal Inheritance Model: Multi-type Case 82

12.1 The sampling distribution and its properties ............... 82
12.2 Asymptotics .................................................. 89
12.3 Remarks and example ........................................ 93
12.4 Proof of Lemma 2 ............................................ 93

13 Conclusions 94

13.1 Mathematical results ........................................ 94
13.2 Biological predictions ....................................... 95
13.3 Future work .................................................. 97

A Algorithms 98

A.1 Recursion algorithms for the p.g.f.'s and their derivatives .... 98

B Code 99

B.1 Calculation of p.g.f.'s and derivatives of p.g.f.'s .......... 99
B.2 Calculation of $E Z_p$ ........................................ 100

Bibliography 102
Illustrations

4.1 Inheritance of methylation. .................................. 29
4.2 Maintenance methylation. .................................. 30
4.3 de novo methylation. ........................................ 31

5.1 Primary and Secondary Colonies. ......................... 35
5.2 Treatment and Control colonies. ......................... 37

6.1 The No-Inheritance Model: $X_m, X_{d_1}, X_{d_2} \sim f(x)$. ........ 41
6.2 The Exact Inheritance Model: $X_{d_1} = X_{d_2} = X_m$. ........ 41
6.3 The Generational Inheritance Model: $X_{d_1}, X_{d_2} \sim f(x; X_m)$. .... 42
6.4 The Clonal Inheritance Model: $X_{d_1}, X_{d_2}, \ldots, X_{d_8} \sim f(x; X_0)$ and $X_{d_1^*}, X_{d_2^*}, \ldots, X_{d_8^*} \sim f(x; X_0^*)$. ........ 42

8.1 Inter-type transitions: basic 3-type model. ............... 57
8.2 Inter-type transitions: 3-type model with "mutation." ........ 58
8.3 Inter-type transitions: 3-type model with mutation and random lifespan. ........................................................... 60

8.4 Inter-type transitions: 3-type model with mutation, random lifespan and cell death. ........................................ 62

8.5 Inter-type transitions: 4 type model. ................................. 64

11.1 Spawning subprocesses: single type. ........................................ 76

12.1 Spawning subprocesses: multi-type. ................................. 84
## Tables

5.1 Correlations Between Primary and Secondary Colonies .................. 36
5.2 Correlations Between Treatment and Control Colonies .................. 38

9.1 Results of 3-type model with “mutation” only. ...................... 66
9.2 Results of 3-type model with “mutation” and variable lifetimes. .... 67
9.3 Results of 3-type model with “mutation,” variable lifetimes, and cell death. ................................................................. 67
9.4 Results of 4-type model. ..................................................... 68
Preface

With respect to the causes of variability, we are in all cases very ignorant; but we can see that in man as in the lower animals, they stand in some relation to the conditions to which each species has been exposed, during several generations.

Charles Darwin, "Descent of Man"

This thesis has two major points: one theoretical and one applied.

The theoretical point of this thesis is a development of a particular type of branching process that has hitherto not been successfully analyzed. We have been able to characterize, exactly and asymptotically, the behavior of both the first and the second moment of the process. Complete descriptions of second moments are relatively rare in the literature. Further, we present expressions for mixed second moments, which are then used to derive correlations.

The applied point of this thesis is analysis of experimental data. The conclusions discussed have implications for primary and secondary tumors; while cell behavior in vitro is frequently quite different than in vivo, we can gain insight into how a metastatic tumor might behave, given the behavior of its parent tumor.
Chapter 1

Introduction

As stated in the Preface, the subject of this work is the investigation of variability and inheritance of growth rates in cell colonies. These two issues are important because these colonies may serve as models for cancer metastasis. From a mathematical point of view, the studies required new tools, the development of which has increased our knowledge of branching processes.

[Axelrod et. al., 1993] carried out experiments with colonies of mouse fibroblast cells, in which colonies of mouse fibroblast cells were dispersed and seeded to form secondary colonies. They found that there exist highly significant correlations between sizes for the primary and secondary colonies, which are accompanied by high variances of colony sizes.

Simulation experiments by these same authors were only able to fit a computer simulation model in which the lifetimes of cells in a secondary clone are to a large extent determined by the lifetimes of the founder cell of the clones; the secondary colonies founder cells were selected randomly from a primary colony. They called this the “clonal inheritance” model.
This model seemed unnatural, since there exist no apparent reasons for this "founder effect." Therefore, we explored another approach, put forward by Dr. Richard Gomer, that there are a number of varieties of cells, differing with respect to their growth rates, but otherwise indistinguishable. Cells may switch from one variety to another, but, most of the time, growth rate is inherited.

In a larger primary colony, faster growing cells predominate; so, a randomly selected founder of the secondary colony tends to belong to the faster growing variety and tends to produce a larger colony. Hence, there is correlation of primary and secondary colony sizes. The question is whether this qualitatively reasonable model is capable of producing the observed values of variability and correlation.

To answer this question mathematically, we derive, for previously uninvestigated multi-type Bellman-Harris and Galton-Watson branching process models, expressions for the joint second moments between the two related but separate processes.

The derived expressions are difficult to evaluate explicitly for all but fairly simple cases; hence, it was necessary to write computer code to do so numerically. With some effort, the code for evaluating the Galton-Watson model was straightforward to implement and is able to efficiently calculate the desired quantities.

One of the more interesting features of these models is the tradeoff between correlation and variability. It is easy to obtain high variability with low correlation, and conversely, low variability with high correlation. However, to elevate both to the observed levels required a fairly complicated model. In fact, it seems that the
observations are located at the boundary of the domain of values possible to generate by these models. We try to relate the models which best fit the data to the dynamics of the epigenetic inheritance of methylation of some genes.

We also made an attempt to mathematically analyze the original hypothesis of [Axelrod et. al., 1993]. We attempted to model their hypothesis using the more complicated Bellman-Harris process. We managed to derive expressions for correlation, which, depending on parameters, may yield positive or negative values. The greater complexity of the Bellman-Harris model made developing satisfactory computer code more difficult. As a result, numerical results for Axelrod's hypothesis of clonal inheritance are still pending.

However, the Bellman-Harris branching process model did yield some to analytical efforts applied in the direction of simplified asymptotic equivalents, which were derived and analyzed. As a result, we found analytically that, in these models, the correlation goes to zero asymptotically.

In conclusion, we have added to the the small collection of explicit expressions for mixed second moments of branching processes. We have developed particular types of branching processes that hitherto have not been successfully analyzed. We have been able to characterize, both exactly and asymptotically, their behavior in both first and second moments; we note that second moment analysis is somewhat rare in general. Also, our analysis of experimental data has implications for primary
and secondary tumors; we have gained insight into how a metastatic tumor might behave, given the behavior of its parent tumor.

The methods developed in this thesis are also applicable to another experiment performed by [Axelrod et. al., 1993]. In this latter experiment, each primary colony was divided into four equally sized sister colonies, which were subsequently grown in an under-nourishing medium. Two of the sister colonies also had an anti-methylation agent added to their culture. Two cell lines were experimented with in this manner, one of which was transfected with the ras oncogene; this gene has been implicated as tumorigenic.

The result of these experiments was that all but the ras transfected genes grown in the culture without the anti-methylation agent showed high correlation between the population sizes of the treated and control colonies. This suggests a strong effect of DNA methylation on the reproduction rate of tumorigenic cell lines. We develop the mathematical framework for analyzing this data, but we have not yet completed the analysis. We include a chapter discussing the effects of methylation and the ras oncogene on cell growth, particularly with respect to carcinogenesis.

The thesis is organized as follows. In Chapter 2, Branching Processes, we present definitions and theorems concerning two basic types of stochastic processes, the Galton-Watson and Bellman-Harris branching processes. In Chapter 3, Branching Process Models of Cell Proliferation, we describe several historical applications of branching process models to the problem of cell proliferation, especially as it
concerns inheritance of cell phenotype. Chapter 4, Methylation, contains a brief overview of DNA methylation and its effect on cell proliferation.

Both sets of experimental data are discussed in Chapter 5, Description of Experimental Data. A discussion of the simulation in [Gusev & Axelrod 1992] is found in Chapter 6, Simulation of Data.

A general mathematical description of the Galton-Watson branching process model is found in Chapter 7, Time-Discrete Generational Inheritance Model: Basic Properties, including expressions for first, second, and joint second moments. A description of several particular models of this type used to model the data in Chapter 5 is found in Chapter 8, Time-Discrete Generational Inheritance Model: Description of Variants Used. These models have been fitted numerically to the data; results are found in Chapter 9, Time-Discrete Generational Inheritance Model: Numerical Investigation, and are discussed and analyzed in Chapter 10, Discussion of Numerical Results.

An analysis of a Bellman-Harris branching process model describing Axelrod’s clonal inheritance hypothesis can be found in Chapter 11, Time-Continuous Clonal Inheritance Model: Single-type Case, and Chapter 12, Time-Continuous Clonal Inheritance Model: Multi-type Case.

Finally, we discuss the mathematical results of the thesis and make some biological predictions based on the results of modeling the data in Chapter 13, Conclusions.
In the appendices, we provide the algorithms used for calculation of numerical results (Appendix A) and the actual computer code (Appendix B).
Chapter 2

Branching Processes

2.1 The Galton-Watson branching process

A brief outline of the Galton-Watson branching process and of relevant formulas is given below. Also, equations for generating functions and first and second moments are given. For a more exhaustive description of the Galton-Watson branching process, see [Athreya & Ney 1972].

**Definition 2.1** A Galton-Watson branching process is a group of proliferating particles satisfying the following:

1. The process starts with a single particle.

2. Each particle has a “lifelength” of 1 time unit.

3. After each time unit, each particle independently produces a random number of progeny distributed as the random variable $Z_1$, which has the distribution $p(k)$.

4. The number of progeny is a random variable independent from time, from the total number of particles in the process, and from the history of the ancestor particle.
The time units can be referred to as “generations.” The distribution of progeny can be characterized by a probability generating function.

**Definition 2.2** A probability generating function (p.g.f.) of the progeny count is given by the following formula:

\[ f(s) = E s^{Z_1} = \sum_{k \geq 0} s^k p(k) \]

where \( p(k) \) is a discrete probability distribution defined exclusively on the non-negative integers. Note that \( f(s) \leq 1 \) for all complex \( s \) such that \( |s| \leq 1 \). Also, \( f(0) = \mathbb{P}(Z_1 = 0) \), and \( f(1) = 1 \).

**Expressions for the probability generating function of \( Z_n \)**

We denote by \( Z_n \) the total number of particles in the process by the \( n \)-th generation.

**Theorem 2.1**

\[ f_n(s) = h(f_{n-1}(s)) \]

where \( f_n(s) \) is the p.g.f. of \( Z_n \), i.e. \( f_n(s) = \sum_{k \geq 0} s^k \mathbb{P}(Z_n = k) \), and \( h(s) = f_1(s) \) is the p.f.g. of \( Z_1 \).

**Expressions for the expectation of \( Z_n \)**

The most important single parameter characterizing the process is the expected progeny count \( \mu \).
Theorem 2.2  
When $E Z_1 < \infty$,

$$
\mu = E Z_1 = f'(1)
$$

where $Z_1$ is the number of progeny produced by a single particle at the time of its death.

As a consequence of Theorems 2.1 and 2.2 we have,

Theorem 2.3  
When $\mu < \infty$,

$$
E Z_n = f'_n(1) = \mu^n
$$

where $\mu = f'(1)$.

Expressions for the second factorial moment of $Z_n$

As we will be analyzing the variance of the population of the branching process, we will need appropriate expressions.

Theorem 2.4  
When $\mu$ and $E Z_1^2$ are both finite,

$$
E \left( Z_1^2 - Z_1 \right) = \left( \frac{d}{ds} \right)^2 f(s)|_{s=1}.
$$

As a consequence of Theorems 2.1 and 2.4 we have,

Theorem 2.5

$$
E \left( Z_n^2 - Z_n \right) = \left( \frac{d}{ds} \right)^2 f_n(s)|_{s=1}.
$$
Finally, combining Theorems 2.3 and 2.5 we have,

\textbf{Theorem 2.6}

\[ \text{Var } Z_n = \frac{(f''(1) + \mu - \mu^2)\mu^n(\mu^n - 1)}{\mu^2 - \mu}. \]

\subsection{The multi-type Galton-Watson branching process}

The single-type Galton-Watson branching process has been modified to include multiple (possibly infinite) distinct particle types. For example, this is a way to model a system in which the particles can have distinguishable behavior. The hypotheses for a multi-type Galton-Watson branching processes are, as might be expected, a direct extension of the single-type Galton-Watson branching process. For an exhaustive description, see [Mode 1971].

\textbf{Definition 2.3} A multi-type Galton-Watson branching process is a group of proliferating particles satisfying the following:

1. The process starts with a single particle, possibly of random type.

2. Each particle has a "lifelong" of 1 time unit.

3. After each time unit, each particle independently produces a random number of progeny of random type. The number of progeny of all types of a particle of type i is a random vector. This random variable is denoted by \( Z_i(n) = (Z_{i1}(n), Z_{i2}(n), \ldots, Z_{im}(n)) \), which has the distribution \( p_i(k_1, k_2, \ldots, k_m) \).
4. The number and type of progeny are random variables which are independent from time, from the total number of particles in the process, and from the history of the ancestor particle.

We denote the number of particles of type \( j \) in the \( n \)-th generation of a process started by a single particle of type \( i \) by \( Z_{ij}(n) \), described by a p.g.f. \( h_i(s_1, \ldots, s_m) \).

**Definition 2.4** The probability generating function of the number of progeny of a type \( i \) particle is defined as follows:

\[
h_i(s_1, s_2, \ldots, s_m) = \sum_{j_1 \geq 0} \sum_{j_2 \geq 0} \cdots \sum_{j_m \geq 0} s_1^{j_1} s_2^{j_2} \cdots s_m^{j_m} \mathbb{P}[Z_{i1}(1) = j_1, Z_{i2}(1) = j_2, \ldots, Z_{im}(1) = j_m]
\]

where \( m \) is the number of types.

We have a recursion equation for the p.g.f.’s similar to that in the single type case:

**Theorem 2.7**

\[
F_i(s, n) = \begin{cases} 
  h_i(F_1(s, n-1), F_2(s, n-1), \ldots, F_m(s, n-1)) & n \geq 1 \\
  s_i & n = 0
\end{cases}
\]

where \( s = (s_1, s_2, \ldots, s_m) \).

Note that the notation for the p.g.f. differs from the usual convention; we merely wish to emphasize the dependence on time.
Theorem 2.8

\[ E Z_{ij}(n) = M_{ij}(n) = \left( \frac{\partial}{\partial s_j} F_i(s_1, s_2, \ldots, s_m, n) \right) \bigg|_{s=1}. \]

Theorem 2.9 The matrix of means \( M_{ij}(n) \) described in Theorem 2.8 is given by

\[ M(n) = M^n \]

where \( M_{ij} = \frac{\partial}{\partial s_j} h_i(s) \big|_{s=1}. \)

Theorem 2.10 Suppose that the process is started by \( u_0 = (u_1, u_2, \ldots, u_m) \), a vector of the number of initial particles of each type. Then,

\[ u(n) = u(n-1)'M = u(0)'M(n) \]

where \( u(n) \) is the vector of the expected number of particles of each type at generation \( n \).

We now derive a formula for the second factorial moment of the total particle count at the \( n \)-th generation derived from a single initial particle of type \( I_p, \phi_{I_p}(n) \).

Theorem 2.11

\[ \phi_{I_p}(n) = \Phi'(n-1)M_{I_p} + \left(1'M^{n-1}\right)' \sum I_p 1'M^{n-1} \]

where \( \Phi(n-1) \) is the vector of second factorial moments of the total particle count at the \( (n - 1) \)-st generation, \( M \) is the matrix of mean
progeny counts $M_{1p}$ is the $I_p$-th row of $M$, $1'$ is a vector of all 1's of length $m$, and the matrix $\Sigma_{I_p}$ is given by

$$
\Sigma_{I_p} = \left[ \frac{\partial^2}{\partial s_i \partial s_j} h_{I_p}(s_1, s_2, \ldots, s_m) \right]_{s=1}.
$$

The proof is included because it is somewhat uncommon, and is not found in the referenced literature.

Proof.

$$
\phi_{I_p}(n) = \left( \frac{d}{ds} \right)^2 F_{I_p}(s, s, \ldots, s, n) \bigg|_{s=1} = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2}{\partial s_i \partial s_j} F_{I_p}(s, s_1, \ldots, s_m, n) \bigg|_{s=1}
$$

$$
= \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2}{\partial s_i \partial s_j} h_{I_p}(F(s, n - 1), F_2(s, n - 1), \ldots, F_m(s, n - 1)) \bigg|_{s=1}
$$

$$
= \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial}{\partial s_i} h_{I_p}(F_1(s, n - 1), F_2(s, n - 1), \ldots, F_m(s, n - 1))
$$

$$
\times \sum_{q=1}^{m} \frac{\partial}{\partial s_j} F_q(s, n - 1) \bigg|_{s=1}
$$

$$
= \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \frac{\partial}{\partial s_i} \frac{\partial}{\partial s_j} h_{I_p}(F_1(s, n - 1), F_2(s, n - 1), \ldots, F_m(s, n - 1))
$$

$$
\times \sum_{q=1}^{m} \frac{\partial}{\partial s_j} F_q(s, n - 1) \sum_{r=1}^{m} \frac{\partial}{\partial s_i} F_r(s, n - 1)
$$

$$
+ \frac{\partial}{\partial s_j} h_{I_p}(F_1(s, n - 1), F_2(s, n - 1), \ldots, F_m(s, n - 1)) \sum_{q=1}^{m} \frac{\partial}{\partial s_j} F_q(s, n - 1) \bigg|_{s=1}
$$

Considering the first term, setting $s = 1$, and recalling that $\frac{\partial}{\partial s_j} F_q(s, n - 1) |_{s=1} = M_{qj}$, we obtain,

$$
\sum_{i=1}^{m} \sum_{j=1}^{m} \left( \Sigma_{I_p} \right)_{ij} \sum_{q=1}^{m} \frac{\partial}{\partial s_j} F_q(s, n - 1) \sum_{r=1}^{m} \frac{\partial}{\partial s_i} F_r(s, n - 1) \bigg|_{s=1}
$$
\[\begin{align*}
&= \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \sum_{q=1}^{m} \frac{\partial}{\partial s_j} F_q(s, n - 1)\left(\Sigma I_{r}\right)_{ij} \sum_{r=1}^{m} \frac{\partial}{\partial s_i} F_r(s, n - 1)\right) \bigg|_{s=1} \\
&= \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \sum_{q=1}^{m} \left(M^{n-1}\right)_{qj} \left(\Sigma I_{r}\right)_{ij} \sum_{r=1}^{m} \left(M^{n-1}\right)_{ri} \right) = \left(1'M^{n-1}\right)' \Sigma I_p 1'M^{n-1}.
\end{align*}\]

\[\square\]

**Theorem 2.12**

\[
\text{Var} Z(n) = \sum_{I_p} \left( \phi_{I_p}(n) + E Z_{I_p}(n) \right) \varphi(I_p) - \left( \sum_{I_p} E Z_{I_p}(n) \varphi(I_p) \right)^2.
\]

### 2.3 The Bellman-Harris branching process

The Galton-Watson branching process is a very powerful tool; however, it does have limitations. Specifically, it is limited to a discrete time axis; additionally, it is Markovian. The Bellman-Harris process was developed to overcome these limitations. A definition is given below. For a more exhaustive description, see [Athreya & Ney 1972].

**Definition 2.5** *The Bellman-Harris branching process is a group of proliferating particles satisfying the following:*

1. *The process starts with a single particle.*

2. *The lifetime of each particle is a non-negative random variable* \(T\) *with probability density function* \(g(t)\).
3. At the end of its lifetime, the particle (independently of its life-
length) produces a random number of progeny which can be described
by the probability generating function $h(s)$.

4. The number of progeny are independent from time, from the to-
tal number of particles in the process, and from the history of the
ancestor particle.

We denote by $Z(t)$ the number of particles at time $t$ in the process started by a
single particle. Let $F(s,t)$ denote the p.g.f. of $Z(t)$.

**Theorem 2.13**  $F(s,t)$ is the unique p.g.f. solution of

$$ F(s,t) = s(1 - G(t)) + \int_0^t h(f(s,t-u))g(u)du $$

where $G(t) = \int_0^t g(u)du$.

Let us denote $E Z(t) = M(t)$.

**Theorem 2.14**  Suppose $\mu = h'(1) < \infty$. Then, $M(t) < \infty$ and $M(t)$
is the unique solution of

$$ M(t) = 1 - G(t) + \mu \int_0^t M(t-u)g(u)du $$

$\mu$ is the mean number of progeny produced by a particle at the end of its
life.
Theorem 2.15  Suppose $\mu, h''(1) < \infty$. Let us denote $E[Z(t)]^2 = M_2(t)$.

$$M_2(t) = \mu \int_0^t M_2(t-u)g(u)du + h''(1) \int_0^t [M(t-u)]^2 g(u)du.$$  

2.4  The multi-type Bellman-Harris branching process

Just as the single-type Galton-Watson branching process can be modified to include multiple (possibly infinite), distinct particle types, so too can the single-type Bellman-Harris branching process. The hypotheses for a multi-type Bellman-Harris branching processes are, as might be expected, very similar to those for the single-type Bellman-Harris branching process. Again, for an exhaustive and slightly more general description, see [Mode 1971].

Discussion of some additional asymptotic properties of the expectation and variance are deferred until Section 12.

Definition 2.6  The multi-type Bellman-Harris branching process is a group of proliferating particles satisfying the following:

1. The process starts with a single particle, possibly of random type.

2. The lifetime of a particle of type $i$ is a non-negative random variable $T$ with probability density function $g_i(t)$.  

3. At the end of its lifetime, the particle (independently of its life-length) produces a random number of progeny of random type which can be described by the probability generating function $h_i(s)$.

4. The number and type of progeny are independent from time, from the total number of particles in the process, and from the history of the ancestor particle.

**Theorem 2.16**

$$F_i(s,t) = s_i(1-G_i(t)) + \int_0^t h_i(F_1(s,t-u), F_2(s,t-u), \ldots, F_m(s,t-u))g_i(u)du$$

where $G_i(t) = \int_0^t g_i(u)du$.

**Theorem 2.17** Let $E Z_{ij}(t) = M_{ij}(t) < \infty$. Then,

$$M_{ij}(t) = \delta_{ij}(1-G_i(t)) + \sum_{\nu=1}^{m} \mu_{ij} \int_0^t M_{\nu j}(t-u)g_i(u)du$$

where $\mu_{ij} = \frac{\partial}{\partial s_j} h_i'(1)$ is the mean number of progeny of type $j$ produced by a particle of type $i$ at the end of its life, and $\delta_{ij}$ is the Kronecker delta.
Chapter 3

Branching Process Models of Cell Proliferation

3.1 Biological mechanisms of dependence

A cell's size is determined by its growth rate and the length of growth (its lifetime); thus, regulation of a cell's size involves one or both of these parameters. Data found in [Hola & Riley 1987] suggest that for their particular cell line (mammalian epithelial (GPK) cells), daughter cells' growth rates were inversely correlated with their mother's, and that sister's lifetimes were strongly correlated. There is considerable amount of other experimental evidence to support the hypothesis of a dependence mechanism for cell lifetimes and growth rates. Most data show a high correlation between the lifetime of sister cells and between the lifetime of mother cells and the mean of its two daughter cells [Staudte et. al., 1984]. [Kimmel & Axelrod 1991] find unequal cell division as a source of negative correlation between sister cell sizes.

A recently published paper [Axelrod et. al., 1993] describes experimental data that shows very high correlations between the number of cells in a primary colony and the number of cells in a secondary colony grown from cells in the primary colony, suggesting a strong dependence of either or both growth rates and lifetimes on ancestor cells.
Other considerations include observed high variability between different clones, and its effect on correlation; see for instance [Kuczek & Axelrod 1986], which describes these effects on statistical analysis) and the effect of drug treatment on interclonal correlations (see [Gusev & Axelrod 1992], which describes clonal size correlations among NIH3T3ras cells, in control colonies and colonies treated with an anti-methylation drug, cytosine arabinoside (Ara-C). They find that growth with the drug almost completely restores correlation between related treated and untreated subclones, while cells not grown in the drug had low correlations between treated and untreated (but related) colonies.

Branching processes, since their introduction, have been useful tools for analyzing biological processes. At present, however, the majority of such tools have failed to take into account the experimentally observed correlations of cells with the behavior of previous generations. Such observations have been made since at least 1955 [Powell 1955]. We are interested particularly in one such set of observations, in [Axelrod et. al., 1993], which shows a very high correlation between the number of cells in a colony, and the number of cells in secondary colonies started by the cells in the first colony (see Chapter 5). Interesting correlation structure has also been found in secondary colonies that were treated by drugs [Gusev & Axelrod 1992].

The colonies in this experiment can serve as model for early phases of cancer growth, during which stochastic phenomena play a major role. It is known that from a very early stage, primary tumors are made up of cells that are heterogeneous in
most characteristics, and metastatic colonies are also quite heterogeneous. However, tumors are believed to be clonal; that is, descendant from a single cell. Thus, a mathematical model of these cell colonies must take into account the observed heterogeneity [Kuczek & Axelrod 1987].

We propose to construct a model with biologically meaningful hypotheses that will describe this heterogeneity. We show that a model without inheritance of lifetimes is incapable of matching the data. We will then develop a method that we believe will be useful for deriving a general model with inheritance. Initially, we use the method to explore the traditional branching process hypothesis of exponentially distributed cell lifetimes.

We believe that our method will allow us to derive a model that is rich enough to adequately explain the experimental observations in [Gusev & Axelrod 1992], and also to be extensible enough to describe other observations. It is hoped that this will lead to a general class of models which may be utilized to describe many different kinds of cell dependence, and thus offer insight into the mechanism of cell growth and reproduction.

3.2 Branching process models in a biological context

Branching processes have historically been used to describe cell proliferation (see e.g. [Harris 1963], [Mode 1971]). Experimental evidence (discussed in Section 3.1) suggests that there is frequently dependence among related cells in a clone. Several
branching process models with differing hypotheses have been studied; we outline several of them below.

Perhaps the original branching process with some sort of dependence is the multitype branching process; this has the advantage that the process may then be considered (with the assumption of exponential lifetimes) Markovian. The difficulty, of course, lies in making the assignment of types in a manner which is both mathematically tractable and biologically plausible.

A branching process may be characterized as follows:

1. A process is started by one or more particles of possibly more than one type.

2. Each particle type is considered to have a density function describing the time to an event for that type.

3. Events may include disappearance, conversion to another type of particle, or production of one or more new particles, possibly of a different type.

Specific examples have been defined in Chapter 1. For the purposes of biologist, the type of a particle may be considered to be (for example) its size, its mother’s lifetime, its genotype, or some other descriptor or combination of descriptors that is deemed to have an effect on its behavior or is of interest.

We briefly describe several applications of branching processes to describing cell populations. The first group will rely on multiple types to describe dependence, while the second will primarily rely on dependence through lifetimes.
Dependence-through-type processes

The first two approaches consider the size of the cell as the dependent parameter.

[Sudbury 1981] makes the following assumptions: the growth rate $\lambda(x)$ and the division rate $\delta(x)$ are deterministic functions of the cell size $x$. Then, under the assumption that the lifetime of a randomly chosen sister cell is distributed uniformly on the interval $[0, x]$, a partial differential equation for the Laplace transform of the quantity $M(x, A, t) = \{\text{expected # of cells at } t \text{ in set } A \text{ produced by a single cell of size } x \text{ at time } t\}$ is found and solved in a few specific cases. It is shown that there exist some conditions which may lead to sub- and super-exponential growth. Asymptotics are then considered. The more biologically realistic assumption that cell sizes are constrained to be in an interval $[a, c]$ is made. Then, again under an assumption of a uniform density for division of the cell (as long as the constraint is met), and what can be characterized as regularity conditions on $\lambda(x)$, a method for finding an asymptotic expression for the first moment is found, which is generally exponential (depending on the growth rate). There are some limitations; the most obvious is assumption of a uniform density of the division size. Secondly, the asymptotics are difficult to find and calculate; several auxiliary equations must be solved. Finally, no expressions are found for higher order moments, so that correlations are impossible to calculate.

[Kimmel et. al., 1984] describe a model of cell division in which the division of DNA between two daughter cells is assumed to be random, and the subsequent
growth of the cell is a deterministic function of the amount of DNA inherited from its mother. An integral equation for the mean number of cells as a function of time is found. It was also found that this was asymptotically exponential. The model was then fitted to experimental data based on Chinese hamster ovary cells. Since only an expression for the mean was found, [Kimmel et al., 1984] had to use numerical calculations to estimate the sister-sister and mother-daughter correlations.

In a more recent paper, [Kimmel & Axelrod 1991] describe a model that is essentially a continuous type Galton-Watson process. Cell lifetimes are all equal, and the event at the end of a lifetime is death, quiescence, or mitosis. Division is assumed to be symmetrically distributed, that is, if $Y$ is the size of the mother cell, then one daughter cell will have size $UY$ and the other will have size $(1 - U)Y$, where $U$ is drawn from a distribution with support on the interval $[0, 1]$, and size at division is assumed to have the functional form $Y = \phi(\Theta) + V$, where $\Theta$ is the size of a cell at birth, $\phi$ is an increasing functions, and $V$ is a non-negative random variable independent of $\Theta$. Recurrence equations for the first two moments of the cell size distribution are found, as well as asymptotics for these moments. The model was successfully fitted to experimental data for normal NIH3T3 mouse fibroblast cells and for ras transformed NIH3T3 mouse fibroblast cells. Limitations include the assumption of constant life-length and the resultant implications for modeling populations with widely varying growth rates. These limitations stem from the use of the Galton-Watson process as a framework.
The models presented heretofore have dealt with the implications of cell size in colony growth. The first attempts to find the implications for colony size, the second for cell size distribution (the mathematics describing the size distribution of a Galton-Watson process are quite well known). Some further work using the cell-type as a means of describing the dependence of a daughter cell on its mother is now outlined.

[Macdonald 1978] performs analysis on a model in which daughters’ lifetimes depend stochastically on their mother’s lifetime. He describes a continuous type process, in which a cell’s type is its mother’s lifetime. This is essentially an extension of a discrete multi-type process (such as that described in [Mode 1971]) to the continuous case. He assumes that two daughters of type \( \tau' \) will have the joint p.d.f. \( p(\tau_1, \tau_2|\tau') \), which is not necessarily symmetric in \( \tau_1, \tau_2 \). [Macdonald 1978] finds asymptotic expressions for both the age distribution and the 1st moment of the number of cells in the process. Also found are expressions for various joint p.d.f.’s of mother cells and daughter cells, depending on the method of selecting the cells (e.g., the p.d.f. of a randomly selected cell and its mother, the p.d.f. of a cell observed from birth and its daughters, and the p.d.f. of a randomly selected cell and its daughters all differ). He asserts that these expressions may be extended to multiple generations. No asymptotic expression for the 2nd moment is found, and hence it is not possible to gauge the effect of the dependence in the long term.
Dependence-through-lifetime processes

Several other workers have attempted to describe the dependence of daughters on their mothers in a somewhat different way. Perhaps the original is outlined briefly in [Harris 1963], where he extends his age-dependent branching process to the case where the lifetimes of two sister cells are positively correlated; they are considered to be the sum of a common and an independent random variable. This analysis is limited to the binary fission model; no cell death or quiescence is allowed.

[Crump & Mode 1969] extend this model and flesh out the details extensively. Their assumptions include the existence of a joint p.d.f. (conditional on the existence of \( k \) sisters) \( g_k(x_1, \ldots, x_k) = g_k(x_{i_1}, \ldots, x_{i_k}) \), where \( i_1, \ldots, i_k \) is any permutation of \( 1, \ldots, k \). The extension includes the possibility of an arbitrary number of sisters cells (so long as the expected number is finite), which, in particular, allows for cell death. However, quiescence is not allowed, because of regularity restrictions on \( g_k \). [Crump & Mode 1969] find asymptotic expressions for the 1st two moments. However, this model does not allow for mother-daughter dependence.

[Green 1980] models yeast cell growth by using a single type branching process, under the biologically motivated assumption that after a cell is born, it must grow to maturity before it can reproduce, thus reflecting the yeasts' budding style of reproduction, where a mother cell produces small buds, which then separate, grow for a time, and then start a cycle of reproduction. He analyzes the model both with and without the assumption of cell death; the no-cell-death assumption eases
analysis to a certain degree. He produces an asymptotic expression for \( X(t)/Z(t) \), where \( X(t) = \sum_x \chi_x(y - \sigma_x) \), \( \sigma_x \) is the birth time of \( x \), \( \chi_x(y) \) is the effect of \( x \) at age \( y \), and \( Z(t) \) is the number of cells in the population at \( t \). For instance, \( \chi(y) \) may be an indicator that \( x \) is in a certain age group, or it may be the number of children of \( x \).

[Staudte et al., 1984] describes a branching process model which allows for several different types of mother-daughter and sister-sister correlation. They consider a binary fission process in which the two sister-lifetimes follow the following model: \( T_1 = U_2 + V_2 + \gamma \), \( T_2 = U_2 + V_3 + \gamma \), where \( U_2 \) is a positive random variable, \( V_2 \) and \( V_3 \) are positive random variables dependent on the mother’s lifetime, and \( \gamma \) is a constant representing the minimum lifetime. The model was fitted to cell lifetime data of EMT6 cells. While exact results were found in several specific instances, these were only of use with extensive pedigree information. No asymptotic results were presented. The correlations found were for only one (mother-daughter) generation. Interestingly, they chose to assume that the cell lifetimes were from a shifted gamma distribution; they offered several arguments for this.

Finally, we discuss some results from [Day 1986]. The model considered may be characterized as a multi-type Bellman-Harris process, with the addition of “killing epochs,” which would correspond (e.g.) to a drug treatment regimen. It is used to describe the reaction of the heterogeneous cell population of a tumor to drug treatment; differing reactions to the drug are allowed, based on the cell type. The analysis is heavily dependent on the assumption of exponentially distributed event
times, and makes a few awkward seeming approximations. The analysis is primarily useful for finding solutions numerically; no closed form solution is found.
Chapter 4

Methylation

Methylation of C and A residues in DNA sequences is a frequently observed phenomena in many prokaryotic and eukaryotic cells. One of a number of postulated functions for methylation in eukaryotic cells is regulation of gene expression [Voet & Voet 1990].

Methylation in mammalian cells occurs predominantly at cytosine residues in CpG sequences. A methyl group is added to the cytosine, making 5-methylcytosine (5-mC). One model for gene regulation by DNA methylation is that hypermethylation in the promoter region of a gene will cause that gene to be inactive. This is based primarily on observations of inverse correlation between gene expression and promoter site methylation. This correlation is not, however, universal; there are genes that show no correlation or positive correlation [Graessmann & Graessman 1993].

The mechanism by which expression would depend on methylation is not known. However, in some genes, e.g. HSV-tk, there appears to be a dependence upon the physical configuration of the chromatin structure, which, combined with methylation, can affect the action of transcriptase [Graessmann & Graessman 1993].
4.1 Methylation maintenance and genesis

In eukaryotes, DNA methylation is self-perpetuating. The current model is that each 5-mC is replicated as an ordinary C residue; then, after mitosis, a maintenance enzyme is active which methylates the unmethylated $\frac{1}{2}$ of the DNA helix according to the template provided by the methylated $\frac{1}{2}$ of the DNA helix inherited from the mother cell [Voet & Voet 1990]. See Figure 4.1.

![Methylation Diagram]

Figure 4.1 Inheritance of methylation.

Of course, the maintenance mechanism is not perfect. There also must be some way of generating de novo methylation. At least two authors have addressed the problem of estimating the efficiency of maintenance and the rate of de novo methylation.

In the first paper discussed here, [Otto & Walbot 1990], the authors model overall levels of 5-mC as a Markov process, and fit the parameters of the model under the
assumption that a steady state had been reached in the genome of the species (maize and mice) studied. [Otto & Walbot 1990] estimate the ratio $E_m:E_d$ to be somewhere between 100:1 and 20:1. Also of interest is the fact that [Otto & Walbot 1990] cite a reference estimating the proportion of potentially methylated sites that are methylated in mouse DNA to be an average of approximately 80%.

[Pfeifer et. al., 1990] describe methylation levels with a simple system of ordinary differential equations involving maintenance and de novo parameters. They fit this model to data obtained from experiments with human-hamster cells; in particular, they concern themselves with the methylation level of the 5’ pre-transcription end of a gene encoding for human phosphoglycerate kinase 1. They estimate the probability of maintenance methylation at 99.9% per site per generation, and the probability of de novo methylation per site per generation at 5%. The mechanism of demethylation is not necessarily through active removal of 5-mC, but might be because of failure to maintain methylation on the duplicated strand of DNA after mitosis. See Figures 4.2 and 4.3.

\[ \begin{array}{c}
\text{CH}_3\text{CG} \\
\text{GC} \\
\end{array} \xrightarrow{\ \text{Maintenance methylation}\ (99.9\% \text{ efficiency per site)}\ } \begin{array}{c}
\text{CH}_3\text{CG} \\
\text{GC} \text{--CH}_3 \\
\end{array} \]

**Figure 4.2** Maintenance methylation.
4.2 Methylation and mutability

Methylation of the C in a CpG dinucleotide palindrome on the 5' end is the only known epigenetic modification in vertebrate DNA. 5-methylcytosine comprises approximately 1-3% of bases in the human genome. However, its effect may be significant, due to the increased mutability of the the 5-mC complex, as well as its ability to influence gene expression. Most CpG islands (clusters) in somatic cells are, however, not methylated. However, the presence of CpG dinucleotides is approximately 4-5 times lower overall, and 3 times lower in coding regions, than would be expected if their appearance were completely random.

5-mC plays a major role in the generation of mutations in tumor suppressor genes in both somatic and germ-line cells. In somatic cells, it may spontaneously deaminate to thymine. This hypermutability has been implicated in inactivating several type of tumor suppressor genes [Spruck et. al., 1993].
4.3 Methylation and cancer

It has been observed that at the tissue specific level, both the amount of 5-mC and the pattern of methylation is altered during tumorigenesis; for example, methylation of CpG islands (clusters) associated with areas of transcriptional regulation is altered in tumorigenic cells relative to non-tumorigenic cells. A causal correlation of CpG methylation with transcriptional inactivity has been observed in experimental systems; methylation of CpG islands may activate oncogenes by repressing tumor suppressor genes. As one example, a high concentration of 5-mC in the 5' regulatory region of the calcitonin gene has been found in several cell lines and tumor types.

Methylation levels also appear to change in non-CpG islands as well. The overall prevalence of 5-mC appears to decrease as tumorigenic cells become more malignant or metastatic. Treatment with a methylase inhibitor such as 5-azacytidine (5-aza-C) significantly alters the metastatic properties of tumor derived cell lines.

The \textit{ras} oncogene

The mechanism by which \textit{ras} oncogene works is as follows; there is a \textit{ras} protein that is critical in the signal pathway for recognizing the reception of growth hormone. Normal \textit{ras} waits for the growth factor to trigger it; mutant \textit{ras} does not require growth factor for it to send a signal to the rest of the cell to grow and divide; hence, the presence of an unsuppressed \textit{ras} oncogene causes uncontrolled growth, and thus cancer.
The c-Ha-*ras* oncogene is extensively methylated in germ-line and somatic tissues where it is expressed. It has also been observed that Ha-*ras*-1 is repressed when its promoter region is heavily methylated.
Chapter 5

Description of Experimental Data

5.1 Primary and secondary colonies

[Axelrod et. al., 1993] carried out the following observations for the 4 cell lines NIH 3T3, NIH 3T3ras, BALB 3T3, and BALB 3T3ras:

- 150 colonies for each cell line were started from individual cells, and were allowed to grow for 4 days.

- Each colony was then counted, resulting in the number $Z_i$, which represents the number of cells in the $i$-th colony at the end of the 4 days. Approximately 40 primary colonies (per cell line) were disbursed to seed new, secondary colonies.

- The secondary colonies were allowed to grow for another 4 days, and were then fixed and counted, resulting in the numbers $Z_{ij}$. This number is the population size of the colony seeded by the $j$-th member of the $i$-th primary colony.

[Axelrod et. al., 1993] then calculated the correlation between the mean number of cells in the secondary colonies and the number of cells in the preceding primary
colony. That is, \( \hat{\rho}(Z_i, \tilde{Z}_i) \) was calculated. In all cases, these numbers were significantly different from 0. We reproduce some of their results in Table 5.1 (compiled from Tables 1 and 2 in [Axelrod et. al., 1993]). Note that the mean and standard deviation provided is for the primary cell colonies that were recloned, not for all primary colonies; these statistics are the relevant ones because they are the for the colonies from which the correlation coefficients were calculated. We will use the simplified notation \( Z_\mu \) for the random variable denoting the colony size of primary colonies and \( Z_s \) for the random variable denoting the colony size of secondary colonies.

Note that although the calculated correlation for primary vs. secondary colonies for the BALBras cell line is only 0.373, it is not statistically significantly different than the calculated correlation for the untransformed BALB line.

*See [Bickel & Doksum 1977] for an approximation for the confidence interval of \( \rho \).
<table>
<thead>
<tr>
<th>Cell Line</th>
<th>$\bar{z}_{\mu}$</th>
<th>95% C.I.</th>
<th>$\bar{\sigma}_{Z_p}$</th>
<th>95% C.I.</th>
<th>$\hat{\rho}_{Z_p,z}$</th>
<th>95% C.I.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH</td>
<td>31.9</td>
<td>(26.0, 37.8)</td>
<td>18.2</td>
<td>(14.9, 23.4)</td>
<td>0.619</td>
<td>(0.381, 0.78)</td>
</tr>
<tr>
<td>NIHras</td>
<td>41.1</td>
<td>(33.4, 48.8)</td>
<td>23.8</td>
<td>(19.5, 30.6)</td>
<td>0.650</td>
<td>(0.424, 0.8)</td>
</tr>
<tr>
<td>BALB</td>
<td>37.2</td>
<td>(32.6, 41.8)</td>
<td>14.1</td>
<td>(11.6, 18.2)</td>
<td>0.550</td>
<td>(0.288, 0.735)</td>
</tr>
<tr>
<td>BALBras</td>
<td>33.8</td>
<td>(28.7, 38.9)</td>
<td>15.9</td>
<td>(13.0, 20.4)</td>
<td>0.373</td>
<td>(0.07, 0.613)</td>
</tr>
</tbody>
</table>

Table 5.1  Correlations Between Primary and Secondary Colonies

5.2  Sister colonies

Similarly high correlation between the sizes of related colonies was noted in another set of unpublished series of experiments by Axelrod which further explore the properties of this inheritance mechanism. In these experiments, the mouse fibroblast cell lines NIH and NIHras were used.

- 33 to 102 colonies were seeded from single cells selected from a large population, and were allowed to grow for 8 days in medium containing 10% serum.

- At the end of the 8 days, each colony was divided into four equal parts: two of which were designated as treatment groups and two as control.

- The treatment groups were placed in a low (1%) serum environment, and the control in 10% serum.

- All groups were then allowed to proliferate for another 7 days, at which point they were fixed and the colonies were stained to infer the number of cells.

- In order to investigate the possible role of DNA methylation in lifetime inheritance, cells were pretreated with 5-azacytidine. A population of cells was
grown for 4 days in 10% serum with 0, 1 or 3 μM 5-azacytidine, then as in Figure 5.2, single colonies were grown without 5-azacytidine in 10% serum for 8 days. The colonies were then divided into 4 parts and grown without 5-azacytidine in 10% or 1% serum for 7 days.

See Figure 5.2.

**Figure 5.2** Treatment and Control colonies.

The correlation between the average sizes of the treatment and control groups was calculated, as in the experiments described in 5.1. The results are shown in Table 5.2.

The effect of the methylase inhibitor is apparently to make ras transfected cells behave like NIH cells. It seems likely that the demethylation agent, 5-azaC, is
<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Conc. 5-azaC</th>
<th>$\hat{\rho}(T,C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH</td>
<td>0 $\mu$M</td>
<td>0.611</td>
</tr>
<tr>
<td>NIHras</td>
<td>0 $\mu$M</td>
<td>0.002</td>
</tr>
<tr>
<td>NIH</td>
<td>1 $\mu$M</td>
<td>0.636</td>
</tr>
<tr>
<td>NIHras</td>
<td>1 $\mu$M</td>
<td>0.467</td>
</tr>
<tr>
<td>NIHras</td>
<td>3 $\mu$M</td>
<td>0.613</td>
</tr>
</tbody>
</table>

Table 5.2  Correlations Between Treatment and Control Colonies

preventing methylation of tumor-suppressor gene, allowing the tumor-suppressor gene to suppress the ras oncogene, and therefore causing the cell to proliferate as a normal cell.

5.3  Comment on data

Primary-Secondary Colony Data

In the first set of experiments described above (Section 5.1), we have observed that larger primary cell colonies tend to produce larger secondary cell colonies, and that smaller primary cell colonies tend to produce secondary cell colonies. The fact that in a given amount of time, a colony becomes larger than another implies that the larger colony is comprised of cells that, on average, proliferate more rapidly than the cells in the smaller colony. Therefore, we can conclude that more quickly growing cells will tend to produce more quickly growing cells, rather than slowly growing cells, and that slowly growing cells will likewise tend to produce cells with growth
rates similar to the parent cells. We note that the observations described in Section 5.2 are consistent with these remarks.

As discussed further in this thesis, we propose to describe the actions of the colonies in Section 5.1 by modeling events at the cellular level.

**Sister-Sister Colony Data**

In the second set of experiment described above (Section 5.2), we can again conclude that there is some mechanism that allows more quickly growing cells to produce more quickly growing progeny; the fact that the progeny of sister cells derived from a uniclonal colony appear to behave similarly is evidence for this.

While the addition of the ras oncogene does not appear to affect the inheritance of cell life time in cell colonies grown in a normal growth medium, it has a profound affect on cells grown under stress situations.

The addition of the ras oncogene has a profound impact on the correlation of colony size between the sister colonies grown in normal growth medium (10% serum) and under stress conditions (1% serum). Further, it appears that the putative suppression of the ras oncogene allows the cell to revert to its normal condition. As discussed in Chapter 4, methylation can have a profound impact on the expression of genes, in particular the ras oncogene. The mechanism is assumed to be that methylation inhibits the transcription of a ras suppressor gene, and demethylation allows this suppressor gene to become transcriptionally active.
Chapter 6

Simulation of Data

The data very clearly indicate that there is a strong inheritance mechanism present. In order to deduce what this mechanism might be, [Axelrod et. al., 1993] attempted to simulate the data using several different models. While one of the models that they used was able to match the data with reasonable precision, it depends on a hypothesis which is very non-standard.

6.1 Models

[Axelrod et. al., 1993] decided upon four possible models to explore.

**No-Inheritance Model:** The lifetimes of each cell are assumed to be independent and random. This can be described using a Bellman-Harris branching process. See Figure 6.1.

**Exact Inheritance Model:** Within a clone, lifetimes are assumed to be inherited exactly. This can be described using a Galton-Watson branching process. See Figure 6.2.

**Generational Inheritance Model:** The lifetime of a daughter cell is assumed to be a random deviation from its mother’s lifetime, i.e. $X_d = g(X_m, \epsilon)$,
where $X_d$ represents the lifetime of a daughter cell, $X_m$ represents the lifetime of its mother, and $\epsilon$ is a random variable. See Figure 6.3.

**Clonal Inheritance Model:** The lifelengths of the cells in a colony are assumed to deviate randomly from the lifetime of the founding member. $X_d = X_0 + \epsilon$, where $X_0$ is lifetime of the founding cell, and $\epsilon$ is a zero mean random variable. $X_d$ represents the lifetime of any cell in a colony other than the founding member. For primary colonies, $X_0$ is assumed to be drawn from a common distribution. For secondary colonies, $X_0 = X_f$, where $X_f$ is the
Figure 6.3  The Generational Inheritance Model: \( X_{d_1}, X_{d_2} \sim f(x; X_m) \).

lifetime of a randomly selected member of the primary colony (i.e. the founder of any particular secondary colony).

Figure 6.4  The Clonal Inheritance Model:
\[
X_{d_1}, X_{d_2}, \ldots, X_{d_6} \sim f(x; X_0) \quad \text{and} \\
X_{d'_1}, X_{d'_2}, \ldots, X_{d'_6} \sim f(x; X_{0'})
\]

In the example shown in Figure 6.4, the primary colony is grown for 2 generations, at which point the fourth daughter is selected to seed a new, secondary colony. Hence, the lifetime of \( d_4 \) is \( X_{d_4} \), which has a distribution depending
directly on $X_6$. The lifetime of (e.g.) $d'_3$ depends directly on $X_{d_4}$, since it is a member of the secondary colony.

### 6.2 Results of simulation

It is obvious that the No-Inheritance Model cannot possibly describe the data, as there is no way for there to be any correlations. In a similar vein, the Exact Inheritance Model is incapable of allowing imperfect correlations.

[Axelrod et. al., 1993] simulated the remaining 2 models in an attempt to fit them to their data. They found that the Generational Inheritance model was not able to adequately match their data. A reasonable fit was only possible with Clonal Inheritance Model.

This is a very interesting but rather nonintuitive finding, as it suggests that there is an inheritance mechanism at work that “resets” itself periodically; i.e., under this model, cells would have to be “aware” of the dissolution of the primary colony and the establishment of the secondary colonies.

As will be seen in the next chapter, other models, stemming from a modification of the Generation Inheritance model, are also capable of reproducing the observations.
Chapter 7

Time-Discrete Generational Inheritance Model: Basic Properties

7.1 The sampling distribution and its properties

We present a modification of the Generational Inheritance Model described in Chapter 6. Eventually, this model will be able to reproduce the high variances and correlations observed in the data described in Chapter 5. This will be one of the two important results of this thesis.

Of the assumption that we made, there are three which are of great significance: firstly, the assumption that the inherited characteristics of a daughter cell depends directly on the characteristics of the mother (i.e. the Generational Inheritance model is assumed; see Section 6.1 and Figure 6.3), secondly, the use of a discrete time axis, and thirdly, a mechanism of inheritance via cell type rather than cell lifespan. The latter two assumptions make the model easier to analyze and enable making biologically interpretable results.

Hypotheses: The primary colonies are growing according to a version of the Galton-Watson branching process, i.e.

- Each cell evolves independently of all other cells.
At the end of its life, a cell of type $i$ gives birth to a random number of progeny of all types described by the probability generating function

$$h_i(s) = \sum_{(j_1, j_2, \ldots, j_n) \in \mathbb{Z}^+} a_{i_1 j_1, i_2 j_2, \ldots, i_n j_n} s_1^{j_1} s_2^{j_2} \cdots s_n^{j_n}.$$  

Note that the p.g.f. $h_i(s)$ is dependent on the mother cell's type.

The process is started at $t = 0$ by the birth of a single cell of random type.

An additional technical hypothesis will simplify analysis:

We suppose in addition that $h_i(0) = 0$, i.e. each cell leaves at least one progeny.

Note also that a cell may give birth to one cell of similar type, thereby allowing the process to effectively have random lifelengths. The lifelength distribution would be geometric in this case. That is, if the probability of producing exactly 1 cell of the same type is $p$, then the distribution of the waiting time for some event other than the production of exactly 1 cell of the same type is geometric with parameter $1 - p$, and consequentially the expected waiting time is $\frac{1-p}{p}$.

### 7.2 Expressions for colony size mean and variance

Referring to equation (2.8), we have that the mean number of particles produced by a single particle of type $i$ in the time interval $t$ is

$$M(t; i) = \sum_{j=1}^{m} M_{ij}(t). \quad (7.1)$$
Further, the unconditional mean is

\[ M(t) = \sum_{i=1}^{m} M(t; i)p(i) = \sum_{i=1}^{m} \sum_{j=1}^{m} M_{ij}(t)p(i), \tag{7.2} \]

where \( p(i) \) is a discrete density on \( 1, 2, \ldots, m \).

To uncondition the variance, we need to combine the unconditioned moments. Hence,

\[ \text{Var}(t) = \sum_{i=1}^{m} \phi_i(t)p(i) + M(t) - (M(t))^2. \tag{7.3} \]

The expression for the conditional second factorial moment \( \phi_i(t) \) is shown in Lemma 2.11.

### 7.3 The covariance of primary and secondary colonies

We need to develop some tools to characterize the generational inheritance model. Let \( I_p \) be the type of a primary colony's founding particle. \( I_p \) may be random or fixed. Let \( I_s(t) \) be the type of a member of a primary colony randomly selected at time \( t \). Its distribution will be identical to the distribution of the type of the initial particle of a randomly chosen secondary colony.

**Lemma 7.1**

\[ \varphi[I_s(t) = i] = \int_{0}^{1} D_{i}F_{I_{p}}(u, \ldots, u, t)du \tag{7.4} \]

where \( F_{I_{p}}(s_1, \ldots, s_m, t) \) is the joint p.g.f. for the vector of population counts \( (Z_1(t), \ldots, Z_m(t)) \).
Proof.

\[ \varphi[I_s(t) = i] = \sum_{z \geq 1} \varphi[I_s(t) = i, Z(t) = z]. \]

\( Z(t) \) is the total colony population at time \( t \). Note that the summation starts at 1, because of the condition that \( h_i(0) = 0 \), which implies that \( Z(t) \neq 0 \). Now,

\[ \varphi[I_s(t) = i, Z(t) = z] = \sum_{z_1 + \cdots + z_m = z} \varphi[I_s(t) = i, Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \]

\[ = \sum_{z_1 + \cdots + z_m = z} \varphi[I_s(t) = i|Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \varphi[Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \]

\[ = \sum_{z_1 + \cdots + z_m = z} z_i \varphi[Z_1(t) = z_1, \ldots, Z_m(t) = z_m]. \]

Passing to a generating function \( Q(s, t) \),

\[ Q(s, t) = \sum_{z_1 = 0}^{\infty} s_i z_1 \sum_{z_1 + \cdots + z_m = z} \varphi[Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \]

\[ = \sum_{z = 0}^{\infty} s_i \sum_{z_1 + \cdots + z_m = z} z_i \varphi[Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \]

\[ = \int_0^s \sum_{z = 0}^{\infty} u^{z-1} \sum_{z_1 + \cdots + z_m = z} z_i \varphi[Z_1(t) = z_1, \ldots, Z_m(t) = z_m] du \]

\[ = \int_0^s \sum_{z = 0}^{\infty} \sum_{z_1 + \cdots + z_m = z} z_i u^{z_1} \ldots u^{z_{i-1}} u^{z_i-1} u^{z_{i+1}} \ldots u^{z_m} \varphi[Z_1(t) = z_1, \ldots, Z_m(t) = z_m] du \]

\[ = \int_0^s D_i F_p(u, \ldots, u, t_p) du. \]

Finally,

\[ \int_0^1 D_i F_p(u, \ldots, u, t_p) du = Q(1, t) = \sum_{z \geq 1} \varphi[I_s(t), Z(t) = z] = \varphi[I_s(t)]. \]
Corollary 7.1

\[ E(Z_s(t_s)) = \sum_{i=1}^{m} M(t_s; i) \int_0^1 D_t F_t(u, \ldots, u, t_p)du \]

where \( Z_s(t_s) \) is the population count of a randomly selected secondary colony at \( t_s \), \( t_p \) is the time at which the primary colony is dispersed, \( I_p \) is as above, and \( M(t; i) = \sum_{k=1}^{m} D_k F_k(1, \ldots, 1, t) \) is the total progeny count of a type \( i \) cell at time \( t \).

Proof.

\[ E(Z_s(t_s)) = E_{I_s(t_s)} E[Z_s(t_s)|I_s(t_s)] \]

\[ = \sum_{i=1}^{m} E[Z_s(t_s)|I_s(t_s)] \varphi[I_s(t_s) = i]. \]

\[ \square \]

We now present the main theorem of this section.

Theorem 7.1

\[ \text{Cov}(Z_p(t_p), Z_s(t_s)) \]

\[ = \sum_{i=1}^{m} M(t_s; i) \left\{ D_t F_t(1, \ldots, 1, t_p) - M(t_p; I_p) \left[ \int_0^1 D_t F_t(u, \ldots, u, t_p)du \right] \right\} \]

where \( m \) is the number of different cell types, \( I_p \) is the type of the founder cell of the primary colony, \( Z_p(t_p) \) is the total population count in a primary colony, and \( Z_s(t_s) \) is the mean of the total population counts in the secondary colonies derived from the primary colony.
Proof.

\[ \text{Cov}(Z_p(t_p), Z_s(t_s)) = \mathbb{E}(Z_p(t_p)Z_s(t_s)) - \mathbb{E}(Z_p(t_p))\mathbb{E}(Z_s(t_s)) \]

\[ = \mathbb{E}(Z_p(t_p)Z_s(t_s)) - \mathbb{E}(Z_p(t_p))\mathbb{E}(Z_s(t_s)). \]

Let us concentrate on the second moment.

\[ \mathbb{E}(Z_p(t_p)Z_s(t_s)) = \mathbb{E}_{I_s(t_p)}[\mathbb{E}(Z_p(t_p)Z_s(t_s)|I_s(t_p))], \]

where \( I_s(t_p) \) is the type of a cell randomly selected from a primary colony at time \( t_p \) (and is the initial cell type for a randomly selected secondary colony). Continuing,

\[ = \sum_{i=1}^{m} \varphi[I_s(t_p) = i] \mathbb{E}[Z_p(t_p)|I_s(t_p) = i] \mathbb{E}[Z_s(t_s)|I_s(t_p) = i], \]

(since \( Z_p \) and \( Z_s \) are independent when conditioned on \( I_s \))

\[ = \sum_{i=1}^{m} \varphi[I_s(t_p) = i] \frac{\mathbb{E}[Z_p(t_p), I_s(t_p) = i]}{\varphi[I_s(t_p) = i]} \mathbb{E}[Z_s(t_s)|I_s(t_p) = i]. \]

And so

\[ \mathbb{E}(Z_p(t_p)Z_s(t_s)) = \sum_{i=1}^{m} \mathbb{E}[Z_p(t_p), I_s(t_p) = i] \mathbb{E}[Z_s(t_s)|I_s(t_p) = i]. \quad (7.5) \]

Now,

\[ \mathbb{E}[Z_p(t_p), I_s(t_p) = i] = \frac{d}{ds} \int_0^s D_i F_{t_p}(u, \ldots, u, t_p) du \bigg|_{s=1} = D_i F_{t_p}(1, \ldots, 1, t_p). \]

Also,

\[ \mathbb{E}(Z_s(t_s)) = \mathbb{E}_{I_s(t_p)}[\mathbb{E}(Z_s(t_s)|I_s(t_p))], \]

\[ = \sum_{i=1}^{m} \varphi[I_s(t_p) = i] \mathbb{E}[Z_s(t_s)|I_s(t_p) = i]. \]
\[ \sum_{i=1}^{m} \left[ \int_{0}^{1} D_i F_{t_p}(u, \ldots, u, t_p) du \right] M(t_s; i). \]  

Thus, from equations (7.5) and (7.6) we have

\[ \mathbb{E}(Z_p(t_p)Z_s(t_s)) - \mathbb{E}(Z_p(t_p))\mathbb{E}(Z_s(t_s)) \]

\[ = \sum_{i=1}^{m} M(t_s; i) D_i F_{t_p}(1, \ldots, 1, t_p) - M(t_p; I_p) \sum_{i=1}^{m} \left[ \int_{0}^{1} D_i F_{t_p}(u, \ldots, u, t_p) du \right] M(t_s; i) \]

\[ = \sum_{i=1}^{m} M(t_s; i) \left\{ D_i F_{t_p}(1, \ldots, 1, t_p) - M(t_p; I_p) \left[ \int_{0}^{1} D_i F_{t_p}(u, \ldots, u, t_p) du \right] \right\}. \]

\[ \square \]

### 7.4 The covariance of sister-sister colonies

The Theorem 7.1 is not suitable for the data described in Section 5.2, because it does not describe the relationship between sibling colonies.

First, we will prove a lemma.

**Lemma 7.2**

\[ \varphi[I_{c_1}(t) = i, I_{c_2}(t) = j] = \int_{0}^{\nu} \int_{0}^{\nu} D_i D_j F_{t_p}(u, \ldots, u, t) du \, dv \]

where \( I_c(t) \) is the type of the initial cell of a sister colony starting at \( t \),

and \( I_p \) is the type of the founding member of the colony from which the founding members of the sister colonies were drawn.

**Proof.**

\[ \varphi[I_{c_1}(t) = i, I_{c_2}(t) = j] = \sum_{z \geq 2} \varphi[I_{c_1}(t) = i, I_{c_2}(t) = j, Z(t) = z]. \]
Now,
\[
\varphi [I_{c_1}(t) = i, I_{c_2}(t) = j, Z(t) = z] = \sum_{z_1 + \cdots + z_m = z} \varphi [I_{c_1}(t) = i, I_{c_2}(t) = j | Z(t) = z] \varphi [Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \\
= \sum_{z_1 + \cdots + z_m = z} \frac{z_i z_j}{z (z - 1)} \varphi [Z_1(t) = z_1, \ldots, Z_m(t) = z_m]
\]
when \( i \neq j \) and
\[
= \sum_{z_1 + \cdots + z_m = z} \frac{z_i (z_i - 1)}{z (z - 1)} \varphi [Z_1(t) = z_1, \ldots, Z_m(t) = z_m]
\]
when \( i = j \).

Again, let us pass to a generating function. When \( i \neq j \),
\[
Q'(s, t) = \sum_{z_1 \geq 2} \frac{s^z}{z(z - 1)} \sum_{z_1 + \cdots + z_m = z} z_i (z_i - 1) \varphi [Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \\
= \int_0^{\varphi} \int_0^{\nu} \sum_{z_1 \geq 2} \frac{u^z}{z(z - 1)} \sum_{z_1 + \cdots + z_m = z} z_i z_j \varphi [Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \, du \, d\nu \\
= \int_0^{\varphi} \int_0^{\nu} \frac{1}{z_1 + \cdots + z_m} \sum_{z_1 + \cdots + z_m = z} z_i z_j \\
\times \left( u^{z_1} \cdots u^{z_i-1} u^{z_i-1} \cdots u^{z_j-1} u^{z_j-1} \cdots u^{z_m} \right) \\
\times \varphi [Z_1(t) = z_1, \ldots, Z_m(t) = z_m] \, du \, d\nu \\
= \int_0^{\varphi} \int_0^{\nu} D_i D_j F_{I_p}(u, \ldots, u, t) \, du \, d\nu.
\]
The proof for the case \( i = j \) is similar. \( \square \)

We can now state the main theorem for this section.

**Theorem 7.2**

\[
\text{Cov} (Z_{c_1}(t), Z_{c_2}(t))
\]
\[
= \sum_{i=1}^{m} \sum_{j=1}^{m} M(t; i)M(t; j)\varphi [I_{c_1} = i, I_{c_2} = j] - \left\{ \sum_{i=1}^{m} M(t; i)\varphi [I_{c_1} = i] \right\}^2
\]
\[
= \sum_{i=1}^{m} M(t; i) \left\{ \sum_{j=1}^{m} M(t; j) [\varphi [I_{c_1} = i, I_{c_2} = j] - \varphi [I_{c_1} = i] \varphi [I_{c_2} = j]] \right\}
\]
(7.7)

where \( I_{c_i} \) is the type of the founder cell of the \( i \)-th sister colony, and
\( M(t; i) = \sum_{k=1}^{m} D_k F_i (1, \ldots, 1, t) \) is the total number of progeny of a type
\( i \) cell at time \( t \). The probability \( \varphi [I_{c_1} (t) = i] \) is the same as \( \varphi [I_{c_2} (t) = i] \)
as given in Lemma 7.1, and \( \varphi [I_{c_1} (t) = i, I_{c_2} (t) = j] \) is found in Lemma 7.2.

Proof. Again, we use conditioning to divide and conquer.

\[
\text{Cov} (Z_{c_1} (t), Z_{c_2} (t)) = E (Z_{c_1} (t)Z_{c_2} (t)) - E (Z_{c_1} (t)) E (Z_{c_2} (t)).
\]

Since \( Z_{c_i} (t) \) has the same distribution as \( Z_{c_i} (t) \) in Theorem 7.1,

\[
E [Z_{c_1} (t)] E [Z_{c_2} (t)] = [E (Z_{c_1} (t))]^2 = \left\{ \sum_{i=1}^{m} M(t; i)\varphi [I_{c_1} (t) = i] \right\}^2.
\]
(7.8)

We turn to \( E (Z_{c_1} (t)Z_{c_2} (t)) \).

\[
E (Z_{c_1} (t)Z_{c_2} (t)) = E_{I_{c_1}, I_{c_2}} [E (Z_{c_1} (t)|I_{c_1}) E (Z_{c_2} (t)|I_{c_2})]
\]

Now, since \( Z_{c_1} \) is independent of \( Z_{c_2} \) when conditioned on the type of the initial
particle,

\[
E (Z_{c_1} (t)Z_{c_2} (t)) = \sum_{i=1}^{m} \sum_{j=1}^{m} M(t; i)M(t; j)\varphi [I_{c_1} (t) = i, I_{c_2} (t) = j].
\]
(7.9)
Putting equations (7.8) and (7.9) together yields the result. \hfill \Box

Note that equation (7.7) goes to 0 as the size of the primary colony gets very large.

7.5 Correlations

The results in Theorem 7.1 and Theorem 7.2 are conditional on the type of the initial particle, $I_p$. We wish to produce more general results, not dependent on $I_p$. To do so, we need $\varphi(I_p = i)$, $i \geq 0$. The distribution of $I_p$ is discussed below in Section 7.6.

We use Theorem 2.12 to find $\text{Var}_{p}(t_p)$, $\text{Var}_{s}(t_s)$, and $\text{Var}_{c}(t_c)$ using the distributions $p_{I_p}(i)$, $p_{I_s}(i)$, and $p_{I_c}(i)$ respectively. $p_{I_s}(i)$ is given in Lemma 7.1, and the marginal distribution $p_{I_c}(i)$ is identical to $p_{I_s}(i)$.

**Theorem 7.4**

$$
\rho(Z_p(t_p), Z_s(t_s)) = \frac{\sum_{i=1}^{m} \text{Cov}_{I_p}(Z_p(t_p), Z_s(t_s)) \varphi(I_p = i)}{\sqrt{\text{Var}_{p}(t_p)\text{Var}_{s}(t_s)}}
$$

**Theorem 7.4**

$$
\rho(Z_c(t_{c_1}), Z_c(t_{c_2})) = \frac{\sum_{i=1}^{m} \text{Cov}_{I_c}(Z_c(t_{c_1}), Z_c(t_{c_2})) \varphi(I_p = i)}{\sqrt{\text{Var}_{c_1}(t_{c_1})\text{Var}_{c_2}(t_{c_2})}}
$$

where $\text{Var}_{c}(t_c)$ is found using Theorem 2.12.
7.6 The distribution of $I_p$

A reasonable distribution for the type of initial cell should be assumed. We assume that the initial cell is drawn from a population in asynchronous exponential growth. The cell cultures which were used in the experiments described in Chapter 5 have been cultivated for a long time, and therefore can be presumed to have reached some kind of steady state with respect to the component cell types.

That is, if that the cells can be differentiated according to some discrete phenotype, the relative proportions of different types should be in equilibrium. In this case, the probability of selecting a particular type of cell is exactly its proportion. These proportions are easily computed under the assumption that the mean progeny matrix $M$ is irreducible. Unfortunately, this is not generally the case.

However, we can use a normalized eigenvector corresponding to the dominant eigenvalue of $M$. Then, if the cells are in proportion according to the eigenvector, they will be at equilibrium.
Chapter 8

Time-Discrete Generational Inheritance Models: Description of Variants Used

Several multi-type Galton-Watson branching process models were developed in the attempt to fit a model to the experimental data. These models are outlined in this chapter. A description of the results of numerically fitting the models to experimental data can be found in Chapter 9. Chapter 10 discusses the results of Chapter 9. Refer to Section 7.1 for detail on the general model and basic assumptions. For illustration purposes, we start with a description of the simplest model.

These models all have a common theme; they model using a heterogeneous population of cells which are divided into two subpopulations: a quickly growing subpopulation of cells and a more slowly growing subpopulation of cells.

8.1 Basic 3-type model (no mutation, no cell death)

This model has two disjoint irreducible subpopulations of cells: a subpopulation of slowly growing cells and a subpopulation of quickly growing cells. The first subpopulation consists of particle types 1 ("young cells") and 2 ("old cells") and the second subpopulation consists of particles of type 3. The slowly growing cells are growing at 1/2 the rate of the quickly growing cells. The quickly growing cells
reproduce every time unit, and the slowly growing cells reproduce every other time unit.

Description of the model

1. The model has three particle types.

2. The progeny generating functions are given by:

\[ h_1(s_1, s_2, s_3) = s_2 \]
\[ h_2(s_1, s_2, s_3) = s_1^2 \]
\[ h_3(s_1, s_2, s_3) = s_3^2. \]

3. The mean progeny matrix is given by

\[
\begin{bmatrix}
0 & 1 & 0 \\
2 & 0 & 0 \\
0 & 0 & 2
\end{bmatrix}
\]

See Figure 8.1 for a graphical representation of the transitions.

The mean progeny matrix is the matrix \( \mathbf{M} \) as found in Section 2.2; \( M_{ij} \) is the expected number of type \( j \) progeny produced by a single particle of type \( i \) after one time unit.

However, this model is not capable of reproducing the data. It is a deterministic model, and hence (a) the correlation is perfect (i.e. 1) and (b) there is no variation in the size of the colonies produced (i.e. the variance is 0).
Clearly, some kind of stochasticity is required to account for the variation. There are several ways to introduce stochasticity into the model; in the first model described below, we allow transitions between the two subpopulations.

### 8.2 3-type model: “mutation” only

This model differs from the basic model in that it allows for transitions between the two subpopulations; i.e. a cell of the slowly growing subpopulation may (with probability $M_3$) change into a cell of the quickly growing subpopulation, and a cell of the quickly growing subpopulation may (with probability $M_1$) change into a cell of the slowly growing subpopulation. This allows the slowly growing subpopulation to be replenished, as otherwise it would be quickly being outstripped in size by the quickly growing subpopulation. Shunting off some of the cells from the quickly growing subpopulation into the slowly growing subpopulation can help prevent the slowly growing subpopulation from becoming too small to be significant.
Description of the model

1. The model has three particle types.

2. The progeny generating functions are given by:

\[ h_1(s_1, s_2, s_3) = (1 - M_3)s_2 + M_3s_3 \]
\[ h_2(s_1, s_2, s_3) = s_1^2 \]
\[ h_3(s_1, s_2, s_3) = M_1s_1 + (1 - M_1)s_3^2. \]

3. The mean progeny matrix is given by

\[
\begin{bmatrix}
0 & 1 - M_3 & M_3 \\
2 & 0 & 0 \\
M_1 & 0 & 2(1 - M_1)
\end{bmatrix}.
\]

See Figure 8.2 for a graphical representation of the transitions.

![Diagram](image)

**Figure 8.2** Inter-type transitions: 3-type model with "mutation."

The correspondence between the two cell subpopulations allows for variation in the size of the colonies; a colony started by a member of the long-lived (and hence
slowly growing) subpopulation will eventually produce members of the shorter-lived (and hence quickly growing) subpopulation. As this "mutation" is random, it permits the size of the colonies to be random. At the same time, assuming that the probabilities of changing cell subpopulation are low enough, some correlation between the sizes of the primary colony and and secondary colony will be retained; i.e. a larger primary colony is still likely to have a large proportion of particles in the short-lived subpopulation. Hence, a secondary colony would be more likely to be seeded by a member of the short-lived subpopulation, and would be therefore more likely to have a large population itself.

8.3 3-type model: "mutation" and variable lifetimes

In an effort to increase the variance of the population size of the colonies without unduly decreasing correlation, it was decided to make the lifetime of the long-lived subpopulation variable. Note that in the previous models, a particle's reproduction time was exactly 1 or 2 time units, depending upon the subpopulation (i.e. the subpopulation with one particle type (type 3) lived 1 time unit, and the subpopulation with two particle types (types 1 and 2) lived 2 time units. This model allows the cells in the slowly-growing subpopulation to bypass the second type (type 2) with probability $1 - L_1$ and to reproduce in only one time unit.
Description of the model

1. The model has three particle types.

2. The progeny generating functions are given by:

\[
\begin{align*}
    h_1(s_1, s_2, s_3) &= (1 - L_1 - M_3)s_1^2 + L_1 s_2 + M_3 s_3 \\
    h_2(s_1, s_2, s_3) &= s_1^2 \\
    h_3(s_1, s_2, s_3) &= M_1 s_1 + (1 - M_1)s_3^2.
\end{align*}
\]

3. The mean progeny matrix is given by

\[
\begin{bmatrix}
    2(1 - M_3 - L_1) & L_1 & M_3 \\
    2 & 0 & 0 \\
    M_1 & 0 & 2(1 - M_1)
\end{bmatrix}.
\]

See Figure 8.3 for a graphical representation of the transitions.

![Graphical representation of transitions](image)

**Figure 8.3** Inter-type transitions: 3-type model with mutation and random lifespan.

The effect of allowing the reproduction process in the long-lived subpopulation to bypass the type 2 stage is to shorten the expected life-span of the long-lived
subpopulation from \(2 - M_3\) to \(1 - L_1 - M_3 + 2L_1 + M_3 = 1 + L_1\). Note that \(L_1 + M_3 \leq 1\), and hence when \(L_1 = 1\), this model is identical to the “mutation” only 3-type model.

8.4 3-type model: “mutation,” variable lifetimes, and cell death

In this model, we have allowed the quickly growing cells to die; this is represented by the parameter \(D_3\), which is the probability that a type 3 cell will die. Perhaps this is a better way to keep the population of quickly growing subpopulation of cells from outstripping the population of slowly growing cells, instead of allowing quickly growing cells to mutate to the slowly growing cell subpopulation. Because of the extremely low calculated value for \(M_3\) in the models described in Sections 8.2 and 8.3, for computational efficiency, it was decided to assume that \(M_3 \equiv 0\) for this model.

Description of the model

1. The model has three particle types.

2. The progeny generating functions are given by:

\[
\begin{align*}
    h_1(s_1, s_2, s_3) & = L_1s_1^2 + (1 - L_1)s_2 \\
    h_2(s_1, s_2, s_3) & = s_1^2
\end{align*}
\]
\[ h_3(s_1, s_2, s_3) = D_3 + M_1 s_1 + (1 - M_1 - D_3) s_3^2. \]

3. The mean progeny matrix is given by

\[
\begin{bmatrix}
2L_1 & 1 - L_1 & 0 \\
2 & 0 & 0 \\
M_1 & 0 & 2(1 - M_1 - L_3)
\end{bmatrix}.
\]

See Figure 8.4 for a graphical representation of the transitions.

Figure 8.4 Inter-type transitions: 3-type model with mutation, random lifespan and cell death.

8.5 4-type models

As will be noted in the following chapter, the model in Section 8.3 ended up being fitted with a \( L_1 = 1.0 \). Perhaps this was an indication that the slowly growing subpopulation of cells were not growing slowly enough. Hence, a fourth particle type was added, so that the slowly growing cells had to progress through types 1, 2, and 3 before reproducing. The quickly growing subpopulation of cells was moved to particle type 4.
The same basic hypotheses as in Section 8.3 are assumed; refer to Figure 8.5.

Description of the model

1. The model has four particle types.

2. The progeny generating functions are given by:

   \[ h_1(s_1, s_2, s_3, s_4) = (1 - M_4)s_2 + M_4s_4 \]

   \[ h_2(s_1, s_2, s_3, s_4) = s_2 \]

   \[ h_3(s_1, s_2, s_3, s_4) = s_1^2 \]

   \[ h_3(s_1, s_2, s_3, s_4) = M_1s_1 + (1 - M_1)s_4^2. \]

3. The mean progeny matrix is given by

   \[
   \begin{bmatrix}
   0 & 1 - M_4 & 0 & M_4 \\
   0 & 0 & 1 & 0 \\
   2 & 0 & 0 & 0 \\
   M_1 & 0 & 0 & 2(1 - M_1)
   \end{bmatrix}
   \]

   See Figure 8.2 for a graphical representation of the transitions.

Certainly, the same variations (variable lifelength, cell death) can be made on this model as on the 3-type models. Additional types can also be added; there is no mathematical reason to stop at 4 types. Finally, the list of models described in this chapter is not meant to be exhaustive by any means, and is not a complete list
of all models examined; however, these models seemed to be the most useful and likely models.
Chapter 9

Time-Discrete Generational Inheritance Model: Numerical Investigation

Because of the relative complexity of the equations describing the models, it was necessary to generate computer code in order to numerically calculate the means, variances and correlations. See Section 2.2 and Chapter 7 for formulas and derivations. See Appendix A for algorithms and calculation procedures and Appendix B for more detailed information on the code.

9.1 3-type models: "mutation" only

Here, we allow correspondence between the two subpopulations, or "mutation." Refer to Section 8.2 for a complete description of this model. As shown below in Table 9.1, the most suitable values for $M_1$ and $M_3$ are very small. $\mathcal{F}$ is the goodness-of-fit: $\mathcal{F} = \left[\frac{\sigma_1 - \bar{Y}_1}{\bar{Y}_1}\right]^2 + \left[\frac{\sigma_2 - \bar{Y}_2}{\bar{Y}_2}\right]^2$. The values shown were found using the Powell minimum hunting algorithm.

Note that it is possible to exactly match either the variance or the correlation, but not both. However, when $M_1 = 0$ and $M_3 = .095$, both the calculated standard deviation and the calculated correlation are well within the appropriate 95% confidence intervals (see Table 5.1). The calculated mean is not within the confi-
Table 9.1 Results of 3-type model with "mutation" only.

The data is for the NIH line of mouse fibroblast cells discussed in Chapter 5. Refer to Table 5.1. Notes: * Fitted to the coefficient of variation. † Fitted to the standard deviation. ‡ Fitted to the correlation.

<table>
<thead>
<tr>
<th>Models</th>
<th>$M_1$</th>
<th>$M_3$</th>
<th>$\mathcal{F}$</th>
<th>$\hat{\mu}_{Z_p}$</th>
<th>$\hat{\sigma}_{Z_p}$</th>
<th>$\hat{\rho}(Z_p, Z_s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.095</td>
<td>0</td>
<td>.012</td>
<td>32</td>
<td>18</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>.098</td>
<td>0</td>
<td>-*</td>
<td>34</td>
<td>20</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>.10</td>
<td>.20</td>
<td>-*</td>
<td>36</td>
<td>18</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>3.6e-6</td>
<td>0</td>
<td>-*</td>
<td>32</td>
<td>.11</td>
<td>.62</td>
<td></td>
</tr>
</tbody>
</table>

dence interval for the estimated mean; this is due to the use of the discrete-time Galton-Watson branching process.

9.2 3-type models: "mutation" and variable lifetimes

This model allows the slowly-growing subpopulation to have a variable lifetime, in addition to allowing switching between the two subpopulations. Note that the lifetime of the slowly-growing subpopulation of cells will be either 1 or 2 time units; it is 1 time unit with probability $1 - L_1$. A complete description of the model is given in Section 8.3. Refer to Table 9.2 for results. Notice that $M_1$ and $M_3$ are the same as the values in Table 9.1. Also, $L_1$ is 1, meaning that all of the slowly-growing cells will live 2 time units before reproducing.
9.3 3-type models: "mutation," variable lifetimes, and cell death

As we found in the previous two models that $M_3$ was 0, it was decided to assume this from the start in this model in order to reduce numerical complexity. This model allows replenishment of the slower-growing type from the faster growing type, a variable life-time for the slowly growing type of cells and cell death for the rapidly growing cells. A complete description is found in Section 8.4, and numerical results are found below in Table 9.3. Note that it too duplicates the mutation only model.

<table>
<thead>
<tr>
<th>$L_1$</th>
<th>$M_1$</th>
<th>$M_3$</th>
<th>$F$</th>
<th>$\hat{\rho}(Z_p, Z_s)$</th>
<th>$\hat{\mu}_p$</th>
<th>$\hat{\sigma}_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
<td>.095</td>
<td>0</td>
<td>0.012</td>
<td>35</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 9.2 Results of 3-type model with "mutation" and variable lifetimes.

<table>
<thead>
<tr>
<th>$L_1$</th>
<th>$M_1$</th>
<th>$D_3$</th>
<th>$F$</th>
<th>$\hat{\mu}_p$</th>
<th>$\hat{\sigma}_p$</th>
<th>$\hat{\rho}(Z_p, Z_s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
<td>.095</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.3 Results of 3-type model with "mutation," variable lifetimes, and cell death.

9.4 4-type models

It was thought that perhaps the slowly growing cells were not growing slowly enough; after all, the factor in the model discussed above in Section 9.2 that would allow the
lifetime of the slowly growing cell to shorten was pushed to its boundary. The same basic hypotheses as in Section 8.3 are assumed; however, the number of compartments for the slowly growing cells to progress through has been increased to three; hence, the number of time units required for a slowly growing cell to reproduce has been increased to 3. Section 8.5 contains a complete description. Refer to Table 9.4 for results. The fit of the model is slightly worse than the above models.

<table>
<thead>
<tr>
<th></th>
<th>$M_1$</th>
<th>$M_4$</th>
<th>$\mathcal{F}$</th>
<th>$\hat{\mu}_{Z_p}$</th>
<th>$\hat{\sigma}_{Z_p}$</th>
<th>$\hat{\rho}(Z_p, Z_e)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Observed data</em></td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>18</td>
<td>.62</td>
</tr>
<tr>
<td><em>Model</em></td>
<td>.093</td>
<td>0</td>
<td>.014</td>
<td>36</td>
<td>20</td>
<td>.55</td>
</tr>
</tbody>
</table>

**Table 9.4** Results of 4-type model.
Chapter 10

Discussion of Numerical Results

Reviewing the tables in Chapter 9 reveals that the simplest non-trivial model (3-type, “mutation” only, see Section 8.2 for a complete description) best describes the data; no other model has a better fit. Several other models were able to be fitted as well, but were all more complex and/or were indistinguishable from this model.

The model that fits best is a model with two subpopulations: a quickly growing subpopulation and a slowly growing subpopulation; the quickly growing subpopulation grows at a rate of approximately twice the slowly growing subpopulation. Additionally, there is replenishment of the slowly growing subpopulation from the quickly growing subpopulation. Other models considered allowed conversion from the longer-lived subpopulation to the shorter lived subpopulation, increased and/or variable lifetimes of the longer-lived subpopulation, cell death, and additional subpopulations.

The important parameters ($\sigma$ and $\rho$) were able to be fitted to values that were well within the 95% confidence intervals constructed from the observed data. Although an exact fitting of the model to the data was not possible, the mean size
of the cell populations could also be matched to a value within a 95% confidence interval of the experimentally observed values.

We conclude that at least one model for the direct inheritance of cell lifetimes from one generation to another can account for experimental observations of numbers of cells in primary and secondary colonies. That model includes two subpopulations with different cell lifetimes and conversion from the shorter-lived subpopulation to the longer. This conclusion is consistent with some previously proposed models, based on observations of cell lifetimes of individual cells in pedigrees, that have described cell lifetimes as inherited from parent to progeny cells.

The numerical fit provided corresponds to the data for the NIH line of mouse fibroblast cells. The model can easily be fitted to the rest of the observed data described in Table 5.1.

We take note of the biological plausibility of the model; while a significant difference between the growth rates of the two subpopulations is required, the relative size of the slowly growing subpopulation is very small. In other experiments of Axelrod's (unpublished data), time-lapse photographic techniques have revealed a wide range of cell lifetimes, from cells that are growing very slowly or quiescent to cells that are growing rapidly. Observations by Sennerstam [Sennerstam & Strömberg 1984, Sennerstam & Strömberg 1988] on embryonic cells also show two distinct subpopulations, though we do not model this data here. Note also the importance of the $M_1$
parameter; it is necessary that this be non-zero in order that the quickly growing cells not completely take over the population.

One possible explanation for the difference in growth rates is methylation or lack of methylation of sites upstream of a key gene. Our $M_1$ would then, in some sense, represent the probability of loss of methylation. Also, our probability of regaining methylation, $M_3$, is 0 (cf. Chapter 4).

These findings might appear to be at odds with the estimates for the efficiency of maintaining methylation and de novo methylation mentioned in Chapter 4*. However, two important points can be made.

First, as noted in [Pfeifer et. al., 1990], many sites must be demethylated to affect transcription. [Otto & Walbot 1990] cite a reference estimating that the proportion of potentially methylated sites that are methylated in mouse DNA to be approximately 80%. The number of potentially methylated sites upstream of a gene varies widely, but can be of within an order of magnitude of $10^2$. Hence, $E_d$ and $E_m$ (which are defined to be the per site per generation probability of new methylation and maintaining methylation respectively) can not be directly compared to the probability of dropping below a specific level of methylation and hence losing the transcription suppression effect.

---

*The reader may wish to refer to Figure 4.2; recall that methylation maintenance refers to methylation of an unmethylated strand of DNA that is the copy generated in mitosis of a methylated strand of DNA.
Second, the estimates in [Otto & Walbot 1990] and [Pfeifer et al., 1990] are calculated under the assumption that all cells grow at the same rate, regardless of methylation level. However, we specifically assuming that methylation level has an effect upon the grow rate of cell. Hence, any estimate of the relative probability of de novo methylation and maintenance of methylation would have to take this into account. While beyond the scope of this work, it should be a strait-forward, if non-trivial exercise to derive estimates of our $M_1$ and $M_3$ from estimates of $E_d$ and $E_m$. This is a topic for future work.

It should be noted that the two subpopulations of cells growing at different (fixed) rates is an artificial simplification necessary for the analysis of this model. The point is not that we believe that there are slowly growing cells growing at exactly 1/2 the rate of the quickly growing subpopulation of cells, but that the data is able to fit such a straitjacket with reasonable ease. It seems reasonable to assume that there is a slowly growing subpopulation, but because of its small size and because of the fairly broad and continuous distribution of lifelengths, it is not directly observable as a separate group in a plot of lifetime distributions; we merely observe its effects on the behavior of the population as a whole.

We are able to conjecture a biologically feasible explanation for this switching of lifelength, namely, methylation level. That methylation can have an effect on lifetime inheritance is indicated by the effect of a demethylation drug on the NIHras cell line as given in Table 5.2.
Chapter 11

Time-Continuous Clonal Inheritance Model:
Single-type Case

11.1 Introduction

In Chapters 11 and 12, we will analyze two variants of the Clonal Inheritance Model described in Section 6.1. This model was defined as follows:

Clonal Inheritance Model: The lifelengths of the cells in a colony are assumed to deviate randomly from the lifetime of the founding member. \( X_d = X_0 + \epsilon \), where \( X_0 \) is lifetime of the founding cell, and \( \epsilon \) is a zero mean random variable. \( X_d \) represents the lifetime of any cell in a colony other than the founding member. For primary colonies, \( X_0 \) is assumed to be drawn from a common distribution. For secondary colonies, \( X_0 = X_f \), where \( X_f \) is the lifetime of a randomly selected member of the primary colony (i.e. the founder of any particular secondary colony).

Here, we present a key equation enabling analysis of second moment properties of one possible mathematical model describing the experiments of [Axelrod et. al., 1993]. This formula, of independent interest, characterizes the joint distribution of the lifetime of a particle randomly selected from a Bellman-Harris branching process and the number of particles present in the process, at the same time. Numerical re-
results have proven somewhat more difficult to produce, due to the complexity of the derived formulas.

11.2 The sampling distribution and its properties

We analyze the clonal inheritance model. In order to relate the sizes of the primary and the secondary colonies in this model it is necessary to characterize the distribution of the lifelength $X_f$ of the founder cell of a secondary colony (being a randomly selected descendant of the primary colony) joint with the number of cells $Z_p(t_p)$ present in the primary colony at time $t_p$. In other words, we would like to find

$$f_{Z_p(t),X_f}(s,x)$$

which is a probability generating function (p.g.f.) in $s$ and a probability density function in $x$. This formula is difficult to analyze but in fact, we will need only

$$\frac{d}{ds}f_{Z_p(t),X_f}(s,x)|_{s=1},$$

which can be put into a compact form.

Indeed, let us begin by deriving an expression for the covariance between $Z_p(t_p)$ and $\tilde{Z}_s(t_s)$,

$$\begin{align*}
E\left[Z_p(t_p)\tilde{Z}_s(t_s)\right] &= E\left[E\left[Z_p(t_p)\tilde{Z}_s(t_s)\right|Z_p(t_p)\right]\right]
= E\left[Z_p(t_p)\left[E\left[Z_s(t_s)\right|X_f,Z_p(t_p)\right]\right]Z_p(t_p)\right]
= E\left[Z_p(t_p)E\left[\mu_{X_f}(t_s)\right|Z_p(t_p)\right]\right] = E\left[Z_p(t_p)\mu_{X_f}(t_s)\right]
\end{align*}$$

where $X_f$ denotes the lifetime of a randomly selected member of the colony alive at $t_p$, and $\mu_X(t)$ is the expected number of members in a colony at $t$ in a process
started by an founding cell with lifetime \( X \). From the properties of the probability generating functions, we obtain that the last expression in (1) is equal to

\[
\int_0^\infty \mu_x(t_x)E(x,t_p)dx
\]

where \( E(x,t_p) = \frac{d}{ds} f_{z_p(t_x),X_f(s,x)}|_{s=1} \).

**Hypotheses:** It is supposed that the primary colonies are growing according to an age-dependent branching process of the Bellman-Harris type, i.e.

- Each cell evolves independently of all other cells.

- The lifetime of each cell is a non-negative random variable \( T \) with probability density function \( g(\cdot) \).

- At the end of its lifetime, the cell (independently of its lifelength) gives birth to a random number of progeny described by the probability generating function \( h(s) \). We suppose in addition that \( h(0) = 0 \), i.e. each cell leaves at least one progeny. Then, \( h(s)/s \) is also a p.g.f.

- The process is started at \( t = 0 \) by the birth of a single cell.

The basic idea underlying our derivations is to trace the ancestry of the founder cell, noting the time of birth of each of its ancestors. Let us assume that there are \( l \) cell divisions between \( X_0 \) and \( X_f \); then, there are \( l \) times \( (t_0,t_1,\ldots,t_{l-1}) \) which gave rise to new process. Note that \( t_i - t_{i-1} \) is the lifetime of a cell that was born at time \( t_{i-1} \), and \( t_0 = x_0 \) is the lifetime of the founding cell. See Figure 11.1.
If we assume that the subprocesses spawned at each time \( t_i \) are independent of each other, then the probability generating function (p.g.f.) for the number of new processes spawned at \( t_i \) is given by \( h(s)/s \). So, \( \frac{h(F(s, t - t_k))}{F(s, t - t_k)} \) is the p.g.f. of the number of cells at \( t \) in the subprocess started by all the cells born at \( t_k \) less the ancestor of our final cell, where \( F(s, t) \) is the p.g.f. of the number of cells at time \( t \) in a process started at 0 by a single cell. Then the p.g.f. of the total number of cells at time \( t \) conditional on there being divisions at \( t_0, t_1, \ldots, t_{l-1} \) is given by

\[
\frac{s^{l-1} \prod_{k=0}^{l-1} h(F(s, t - t_k))}{F(s, t - t_k)}.
\]

Let \( t_k = \sum_{j=0}^{k} x_j \), where \( x_j \) is the lifetime of the \( j \)-th generation in the lineage of the final cell. Then

\[
\int_{Z(t), X_j} f(s, x) = s g(x) \mathbb{I}[t < x]
\]

\[
+ \sum_{l \geq 2} \int_{t_{l-1}}^{t} \cdots \int_{t_0}^{t} s^{l-1} \prod_{k=0}^{l-1} \frac{h(F(s, t - t_k))}{F(s, t - t_k)} \prod_{k=0}^{l} g(x_k) \mathbb{I}[t_k \leq t < t_{k+1}] dt_l \cdots dt_0
\]

---
**Figure 11.1** Spawning subprocesses: single type.
where \( t_{-1} \equiv 0 \) and \( x = x_{t+1} \). The first term is for the case where the initial cell does not reproduce before \( t \). The above equation is quite complex. We will show how a much simpler formula can be derived for \( E(x, t) \). Differentiating (4) with respect to \( s \) and setting \( s = 1 \), we obtain

\[
E(x, t) = g(x)I[t < x]
\]

\[+ \sum_{i \geq 0} \int_0^t \int_0^t \ldots \int_0^t [1 + \sum_{k=0}^i (m - 1)M(t - t_k)] g(x_0) \ldots g(x_{i+1})
\]

\[I[t_i \leq t < t_{i+1}] dt_i \ldots dt_0,
\]

where \( x = x_{t+1} \), \( m = h'(1) \) is the expected progeny count per cell, and \( M(\tau) = F''(1, \tau) \) is the expected cell count in the Bellman-Harris process at time \( \tau \).

**Theorem 11.1** Suppose \( h(0) = 0 \). Then the following holds:

\[
E(x, t) = g(x) \left( [\Gamma(t) - \Gamma_x(t)] + (m - 1)\gamma(t) * \{M(t) [\Gamma(t) - \Gamma_x(t)]\} \right)
\]

\[x, t \geq 0\]

where

\[
\gamma(t) = \sum_{n \geq 1} g^*n(t)
\]

\[
\Gamma(t) = \sum_{n \geq 0} G^*n(t)
\]

The convolutions \( * \) and \( \star \) are defined by

\[
a(t) * b(t) = \int_0^t a(t - \tau)b(\tau)d\tau
\]

\[
A(t) \star B(t) = \int_0^t A(t - \tau)dB(\tau)
\]

and \( A_x(t) = A(t - x), \ t \geq x, \ A_x(t) = 0, \ t < x \), for \( t, x \geq 0 \).
Proof. We have

\[ E(x,t) = g(x)I[t < x] + \sum_{l \geq 0} A_l + (m - 1) \sum_{l \geq 0} \sum_{k=0}^{l} A_{lk} \]

where

\[ A_l = g(x) \int_{0}^{t} \int_{0}^{u_0} \cdots \int_{0}^{u_{l-2}} g(u_{l-1} - u_l) \cdots g(u_0 - u_1)g(t - u_0)I[u_l < x] \ du_l \cdots du_0, \]

\[ A_{lk} = g(x) \int_{0}^{t} g(t - u_0) \int_{0}^{u_0} g(u_0 - u_1) \cdots \int_{0}^{u_{k-1}} M(u_k)g(u_{k-1} - u_k) \]

\[ \int_{0}^{u_k} g(u_k - u_{k+1}) \cdots \int_{0}^{u_{l-1}} I[u_l < x]g(u_{l-1} - u_l) \ du_l \cdots du_{k+1} du_k \cdots du_1 du_0 \]

following a change of variables \( u_j = t - t_j, \ j = 0, \ldots, l + 1 \) (\( u_{-1} = t \)).

The innermost integral in \( A_l \) (and in \( A_{lk} \)) is equal to

\[ \int_{0}^{\min(u_{l-1},x)} g(u_{l-1} - u_l) \ du_l = \int_{\max(0,u_{l-1} - x)}^{u_{l-1}} g(\nu) d\nu = G(u_{l-1}) - G_x(u_{l-1}) \]

Carrying out the integration in the outward direction, we obtain

\[ A_l = g(x) \left[ G^{*(l+1)}(t) - G_x^{*(l+1)}(t) \right], \]

and

\[ A_{lk} = g(x)g^{*(k+1)}(t) \ast \left\{ M(t) \left[ G^{*(l-k)}(t) - G_x^{*(l-k)}(t) \right] \right\}. \]

Summing \( A_l \) over \( l \geq 1 \) and noticing that \( I[t < x] = G^{*0} - G_x^{*0} \) we obtain the first term of our assertion.

The summation of the \( A_{lk} \) is carried out by changing the summation indices, \( j = l - k, i = k \) so that the double sum is written

\[ \sum_{l \geq 0} \sum_{k=0}^{l} g^{*(i+1)} \ast \left[ M \left( G^{*j} - G_x^{*j} \right) \right] \]
The proof is now complete. □

This result is illustrated by the following application.

**Corollary 11.1** Suppose that \( g(x) = \lambda e^{-\lambda x} \) and \( h(s) = s^2 \). Then

\[
E(x, t) = \lambda e^{-\lambda x} \left[ I_{[t < x]} + \lambda \min (x, t) \left( 1 + e^{\lambda t} \right) \right].
\]

*Proof.* Follows by noticing that, in this case, \( m - 1 = 1 \), \( \gamma = \lambda \), \( \Gamma(t) = (1 + \lambda t) \), \( \Gamma_x(t) = I[t < x] + \lambda \min(x, t) \), and \( M(t) = e^{\lambda t} \). □

Let us note that if \( X_f \) and \( Z_p \) were independent, then \( E(x, t) \) would be equal to the product

\[
g(x) E[Z(t)] = g(x) M(t),
\]

which is not the case.

However, let us notice that if our calculations are correct, then \( \int_0^\infty E(x, t) dx = E[Z(t)] = M(t) \).

On the other hand,

\[
\int_0^\infty g(x) G_x^*(t) dx = G^{*(i+1)}(t),
\]

which implies \( \int_0^\infty g(x) \Gamma_x(t) dx = \Gamma(t) - 1 \), and \( \int E(x, t) dx = 1 + (m - 1) \gamma(t) * M(t) \).

Therefore, we have the following

**Corollary 11.2** \( M(t) \) satisfies the following Volterra-type linear integral equation:

\[
M(t) = (m - 1) \gamma(t) * M(t) + 1, \quad t \geq 0.
\]
This result does not seem to be commonly found in the literature. As is well known [Athreya & Ney 1972], $M(t)$ satisfies the following Volterra-type equation

$$M(t) = mg(t) \ast M(t) + 1 - G(t), \quad t \geq 0.$$  

Equivalence of (11.2) and (11.2) can be heuristically checked by the use of the Laplace transform. We denote by $\hat{a}(s)$ the transform of the function $a(t)$. (ii) gives

$$\hat{M}(s) = \frac{1}{s} \frac{1 - \hat{g}(s)}{1 - m\hat{g}(s)},$$

while (ii) implies

$$\hat{M}(s) = \frac{1}{s} \frac{1}{1 - (m - 1)\hat{g}(s)}.$$  

However, $\hat{y}(s) = \sum_{i \geq 1} [\hat{g}(s)]^i = \frac{\hat{g}(s)}{1 - \hat{g}(s)}$, which shows that both expressions are equivalent.

We conclude this section by a remark concerning the case $g(x) = \lambda e^{-\lambda x}$. Let us assume that the mean lifetime of cells in a secondary colony, started by a founder cell with lifetime $x_f$ is equal to $x_f$, according to the Clonal Inheritance Model. Consequently, $\lambda_s = \frac{1}{x_f}$ and

$$\mu_{x_f}(t) = \mathbb{E}(Z_s(t \mid x_f) = e^{\lambda s x_f} = e^{\frac{t}{x_f}}.$$  

We now attempt to evaluate $EZ(t)\mu_{x_f}(t)$. Evaluating the appropriate integral, we obtain a term of the form

$$\lambda^2 \left(1 + e^{\lambda t}\right) \int_0^t x_f e^{\frac{t}{x_f}} e^{-\lambda x_f} dx_f.$$
This integral is divergent and therefore Cov\((Z_p, \tilde{Z}_s)\) is undefined. This indicates that as far as clonal inheritance is concerned, the hypothesis of an exponential cell lifetime distribution is unjustifiable.

11.3 Concluding remarks

The sampling formula of Theorem 11.1 enables calculations of mixed second moments of primary and secondary colony counts in the clonal inheritance models. Its applications are twofold. First, it will serve to evaluate explicit forms of \(E(x, t)\) for those variants of the Bellman-Harris process in which explicit forms of \(M(t)\) are known. Second, the formula for \(E(x, t)\) depends on the first moment \(M(t)\) of the Bellman-Harris process and on the renewal density \(\gamma(t)\) and renewal function \(\Gamma(t)\). Since asymptotics of these three function are known, so is that of \(E(x, t)\) (as \(t \to \infty\)).
Chapter 12

Time-Continuous Clonal Inheritance Model: Multi-type Case

12.1 The sampling distribution and its properties

We analyze the clonal inheritance model, and propose adding different cell "types" to reflect the difference in behavior of e.g. cells with methylated vs. non-methylated cell DNA. In order to relate the sizes of the primary and the secondary colonies in this model it is necessary to characterize the distribution of the lifelength $X_f$ of the founder cell of a secondary colony (being a randomly selected descendant alive at the disbursement of the primary colony) joint with the vector of the numbers of cells of all types $Z_p(t_p)$ present in the primary colony at time $t_p$. In other words, we would like to find

$$f_{Z_p(t_p), X_f(s, x)}$$

which is a probability generating function (p.g.f.) in $s = \{s_1, s_2, \ldots, s_n\}$ and a probability density function in $x$. This formula is difficult to analyze, but in fact, we will need only

$$\sum_{\nu=1}^{n} \frac{d}{ds_\nu} f_{Z_p(t_p), X_f(s, x)}|_{s=1},$$

which can be put into a compact form.
Let us begin by deriving an expression for the mixed second moment of $Z_p(t_p)$ and $\tilde{Z}_s(t_s)$, where $Z_p(t_p)$ is the total number of cells of all types in the primary colony at time $t_p$, and $Z_s(t_s)$ is the total number of cells of all types in the primary colony at time $t_s$.

$$E \left[ Z_p(t_p) \tilde{Z}_s(t_s) \right] = E \left[ E \left[ Z_p(t_p) \tilde{Z}_s(t_s) \mid Z_p(t_p) \right] \right]$$

$$= E [Z_p(t_p) \mid E [Z_s(t_s) \mid X_f, Z_p(t_p)] ] Z_p(t_p)]$$

$$= E [Z_p(t_p)E [M_s(t_s \mid X_f) \mid Z_p(t_p)] ] = E [Z_p(t_p)M_s(t_s \mid X_f)] = \int_0^\infty \mu_x(t_s)E(x, t_p) dx.$$  

(12.1)

where $X_f$ is the lifetime of a randomly selected cell in the primary colony that is alive at $t_s$, $M_s(t \mid X)$ is the expected number of cells in a colony at $t$ in a process started by an founding cell with lifetime $X$, and $E(x, t) = \sum_{r=1}^{\infty} \frac{d}{ds} \int f_{Z_p(t),X_f}(s, x) |_{s=1}$.

**Hypotheses:** The primary colonies are growing according to a version of the age-dependent branching process of the Bellman-Harris type, i.e.

- Each cell evolves independently of all other cells.
- The lifetime of each cell is a non-negative random variable with probability density function $g_i(\cdot)$, where $i$ is the type of the cell.
- At the end of its lifetime, the cell (independently of its lifelength) gives birth to a random number of random type progeny described by the probability gen-
erating function
\[ h(s) = \sum_{(j_1, j_2, \ldots, j_n) \in \{\mathbb{Z}^+\}^n} a_{j_1, j_2, \ldots, j_n} s_1^{j_1} s_2^{j_2} \cdots s_n^{j_n}. \]

Note that \( h(s) \) is independent of the mother cell's type. We suppose in addition that \( h(0) = 0 \), i.e. each cell leaves at least one progeny.

- The process is started at \( t = 0 \) by the birth of a single cell of random type.

The basic idea underlying our derivations is to trace the ancestry of the founder cell, noting the time of birth of each of its ancestors. Let us assume that there are \( l \) cell divisions between \( X_0 \) and \( X_f \). Then, there are \( l \) points in time \((t_0, t_1, \ldots, t_{l-1})\) at which new processes are started. Note that \( x_\nu = t_\nu - t_{\nu-1} \) is the lifetime of a cell that was born at time \( t_{\nu-1} \), and \( t_0 = x_0 \) is the lifetime of the founding cell. See Figure 12.1.

![Figure 12.1 Spawning subprocesses: multi-type.](image)

We have \( t_k = \sum_{\nu=0}^{k} x_\nu \), where \( x_\nu \) is the lifetime of the \( \nu \)-th generation in the lineage of the final cell. Let us denote by \( i_j \) the type of the cell which is born at \( t_{j-1} \)
and dies at $t_j$, and $p_{i_0} = \Pr\{1st \ cell \ is \ type \ i_0\}$, and $x = t_{i+1} - t_i$ the lifetime of the “final” cell in our lineage (the cell selected randomly from the primary colony at $t$).

The joint p.g.f.-density of $Z_f(t)$ and $X_f$ conditional on the vector $\{t_0, t_1, \ldots, t_l\}$ is given by

$$F(x, s, t|t_0, t_1, \ldots, t_l) = \sum_{(i_0, i_1, \ldots, i_l) \in \{1, 2, \ldots, n\}^{l+1}} p_{i_0} \left[ \prod_{k=0}^{l} \sum_{(j_1, j_2, \ldots, j_n) \in (\mathbb{Z}^+)^n} \Pr\{j_1 \text{ type } 1, j_2 \text{ type } 2, \ldots, j_n \text{ type } n\} \times \Pr\{\text{pick type } i_k|\{j_1, j_2, \ldots, j_n\}\} \prod_{m=1}^{n} F_m^{j_m}(s, t - t_k) \frac{1}{F_{i_{k+1}}(s, t - t_k)} \right] s_{i_{l+1}}$$

$$= g_{i_{l+1}}(x) \sum_{(i_0, i_1, \ldots, i_l) \in \{1, 2, \ldots, n\}^{l+1}} p_{i_0} \left[ \prod_{k=0}^{l} \sum_{(j_1, j_2, \ldots, j_n) \in (\mathbb{Z}^+)^n} a_{j_1, j_2, \ldots, j_n} \left( \frac{j_{i_{k+1}}}{\sum_{\nu=1}^{n} j_{\nu}} \right) \right] \times \left( \prod_{m=1}^{n} F_m^{j_m}(s, t - t_k) \frac{1}{F_{i_{k+1}}(s, t - t_k)} \right) s_{i_{l+1}},$$

where $F_m(s, u)$ is the p.g.f. of the process started by a single cell of type $m$ at $u = 0$, and $g_i(t)$ the lifetime distribution of a cell of type $i$.

We let

$$\hat{h}_{i_{k+1}}(\tau) = \sum_{(j_1, j_2, \ldots, j_n) \in (\mathbb{Z}^+)^n} a_{j_1, j_2, \ldots, j_n} \left( \frac{j_{i_{k+1}}}{\sum_{\nu=1}^{n} j_{\nu}} \right) \left( \prod_{m=1}^{n} F_m^{j_m}(s, \tau) \frac{1}{F_{i_{k+1}}(s, \tau)} \right).$$

(12.2)

Next, we remove conditioning on $t_0, \ldots, t_l$, but retain the $l + 1$ division times before $t$, denoting $t_{-1} \equiv 0$.

$$F(x, s, t, l) =$$

$$g_{i_{l+1}}(x) \int_0^t \int_0^t \cdots \int_0^t \sum_{(i_0, i_1, \ldots, i_l) \in \{1, 2, \ldots, n\}^{l+1}} p_{i_0} \prod_{k=1}^{l+1} \hat{h}_{i_k}(t - t_{k-1}) s_{i_{l+1}}.$$
\[
\times \prod_{k=0}^{l} g_k(t_k - t_{k-1})I_{[t_l \leq t_{l+1}]} dt_l \ldots dt_0.
\]

We rewrite this as:

\[
\int_0^t \int_0^t \ldots \int_0^t \sum_{k=1}^{n} p_k g_k(t_0) \prod_{j=1}^{l} \left[ \sum_{k=1}^{n} \hat{h}_k(t - t_{j-1}) g_k(t_j - t_{j-1}) \right] \times \sum_{k=1}^{n} \hat{h}_k(t - t_i) s_k g_k(x) I_{[0 \leq t_i - t_l < x]} dt_l \ldots dt_0.
\]

(12.3)

Let

\[
\hat{M}_i(t) = \left. \sum_{\nu=1}^{n} \frac{\partial}{\partial s_\nu} \right|_{s=1} \hat{h}_i(t)
\]

\[
= \sum_{(j_1, j_2, \ldots, j_n) \in (\mathbb{Z}^+)^n} a_{j_1, j_2, \ldots, j_n} \left( \frac{j_i}{\sum_{\nu=1}^{n} j_\nu} \right) \left( \sum_{\nu=1}^{n} j_\nu M_\nu(t) - M_i(t) \right)
\]

(12.4)

where \(M_i(t) = \left. \sum_{\nu=1}^{n} \frac{\partial}{\partial s_\nu} \right|_{s=1} F_i(s, t)\) is the expected number of cells of all types at \(t\) in a process started at 0 by a single cell of type \(i\).

Applying the differential operator and denoting \(p_i = \hat{h}_i(1), i = 1, \ldots, n\), (where \(\sum_{i=1}^{n} p_i = 1\)), and \(g(t) = \sum_{\nu=1}^{n} p_\nu g_\nu(t)\), we get

\[
\int_0^t \int_0^t \ldots \int_0^t g(t_0)
\]

\[
\left\{ g(x) \sum_{j=1}^{n} \sum_{\nu=1}^{n} \hat{M}_\nu(t - t_{j-1}) g_\nu(t_j - t_{j-1}) \prod_{k=1}^{l} g(t_k - t_{k-1}) + g(x) \prod_{k=1}^{l} g(t_k - t_{k-1})
\right.
\]

\[
+ \sum_{\nu=1}^{n} \hat{M}_\nu(t - t_i) g_\nu(x) \prod_{k=1}^{l} g(t_k - t_{k-1}) \right\} I_{[t_l < t_x]} dt_l \ldots dt_0.
\]

Finally, we sum over \(l \geq 0\), obtaining

\[
E(x, t) =
\]

\[
g(x) I_{[t < x]} + \sum_{l \geq 0} \left[ g(x) \int_0^t \int_0^t \ldots \int_0^t g(t_0) \prod_{k=1}^{l} g(t_k - t_{k-1}) I_{[t_l < t_x]} dt_l \ldots dt_0
\]

(12.5)
\[ + g(x) \sum_{j=1}^{l} \int_{t_0}^{t} \cdots \int_{t_{l-1}}^{t} g(t_0) \sum_{\nu=1}^{n} \hat{M}_\nu(t - t_{j-1}) g_\nu(t_j - t_{j-1}) \times \prod_{k=1, k \neq j}^{l} g(t_k - t_{k-1}) I_{[t_{k-1} < x]} dt_1 \cdots dt_0 \]

\[ + \int_{t_0}^{t} \cdots \int_{t_{l-1}}^{t} \sum_{\nu=1}^{n} \hat{M}_\nu(t - t_l) g_\nu(x) \prod_{k=1}^{l} g(t_k - t_{k-1}) I_{[t_{k-1} < x]} dt_1 \cdots dt_0. \]

We are ready to state the following result:

**Theorem 12.1** Suppose that the hypotheses spelled out at the beginning of Section 12.1 are satisfied. Then the following holds:

\[
E(x, t) = g(x) \left\{ \left[ \Gamma(t) - \Gamma_{(x)}(t) + I_{[t < x]} \right] + \gamma(t) \ast \left[ \sum_{\nu=1}^{n} \hat{M}_\nu(t) \left( \Gamma_\nu(t) - \Gamma_{(x)\nu}(t) \right) \right] \right. \\
+ \left. \gamma(t) \ast \left[ \sum_{\nu=1}^{n} g_\nu(x) \hat{M}_\nu(t) \right] I_{[t < x]} \right\}
\]

for \( x, t \geq 0 \) where \( \gamma(t) = \sum_{n \geq 1} g^n(t), \Gamma_1(t) = \sum_{n \geq 1} G_i^n(t), \) and \( \hat{M}_i(t) \) is defined in equation (12.4). The convolutions \( \ast \) and \( \ast \) are defined by

\[ a(t) \ast b(t) = \int_{0}^{t} a(t - \tau) b(\tau) d\tau \]

\[ A(t) \ast B(t) = \int_{0}^{t} A(t - \tau) dB(\tau) \]

and

\[ A_{(x)}(t) = \begin{cases} 
A(t - x), & t \geq x \\
0, & t < x 
\end{cases} \]

for \( t, x \geq 0 \).

**Proof.** We have

\[ E(x, t) = g(x) I_{[t < x]} + \sum_{l \geq 0} \sum_{k=0}^{l} \sum_{\nu=1}^{n} A_{l, k, \nu} \]
where

\[ A_l = g(x) \int_0^t \int_0^{u_0} \cdots \int_0^{u_{l-1}} g(u_{l-1} - u_l) \cdots g(u_0 - u_1) g(t - u_0) I[u_l < x] \, du_l \cdots du_0, \]

and when \( k \neq l, \)

\[ A_{lk,\nu} = \]

\[ g(x) \int_0^t g(t - u_0) \int_0^{u_0} g(u_0 - u_1) \cdots \int_0^{u_{k-1}} \dot{M}_\nu(u_k) g(u_{k-1} - u_k) \int_0^{u_k} g_\nu(u_k - u_{k+1}) \]

\[ \times \int_0^{u_{k+1}} g(u_{k+1} - u_{k+2}) \cdots \int_0^{u_l-1} I[u_l < x] g(u_{l-1} - u_l) du_l \cdots du_{k+1} du_k \cdots du_1 du_0 \]

following a change of variables \( u_j = t - t_j, \) \( j = 0, \ldots, l + 1 \) \((u_{-1} = t).\) The innermost integral in \( A_l \) and \( A_{lk,\nu} \) is equal to

\[ \int_{0}^{\min(u_{l-1}, x)} g(u_{l-1} - u_l) du_l = \int_{\max(0, u_{l-1} - x)}^{u_{l-1}} g(\nu) d\nu = G(u_{l-1}) - G(x)(u_{l-1}). \]

Performing the integration in the outward direction, we obtain

\[ A_l = g(x) \left[ G^{*l+1}(t) - G^{*l+1}(x) \right], \]

and

\[ A_{lk,\nu} = g(x) g^{*(k+1)}(t) * \left\{ \dot{M}_\nu(t) \left[ \dot{G}_\nu^{*(l-k)}(t) - G^{*(l-k)}(t) \right] \right\}. \]

where

\[ \dot{G}_\nu^{*(l-k)}(t) = G_\nu(t) * G^{*(l-k-1)}(t). \]

For the special case \( k = l, \)

\[ A_{ll,\nu} = \]

\[ \int_0^t g(t - u_0) \int_0^{u_0} g(u_0 - u_1) \cdots \int_0^{u_{l-1}} \dot{M}_\nu(u_l) g(u_{l-1} - u_l) g_\nu(x) I[u_l < x] du_l \cdots du_0 \]
\[ = g(t)^{l+1} \ast M(t)I_{[t < x]}g(x), \]

so that
\[
E(x, t) = g(x)I_{[t < x]} + \sum_{l \geq 0} \left( A_l + \sum_{\nu=1}^{n} A_{l, \nu} \right) + \sum_{l \geq 0} \sum_{k=0}^{l-1} \left( \sum_{\nu=1}^{n} A_{l, k, \nu} \right)
\]
\[
= g(x)I_{[t < x]} + \sum_{l \geq 0} \left\{ g(x) \left[ G^{*\ast(l+1)}(t) - G^{*\ast(l+1)}_{\ast x}(t) \right] + \sum_{\nu=1}^{n} g_{\nu}(x)g^{*\ast(l+1)}(t) \ast \hat{M}_{\nu}(t) \right\}
\]
\[
+ g(x) \sum_{l \geq 0} \sum_{k=0}^{l-1} \left\{ g^{*\ast(l+1)}(t) \ast \sum_{\nu=1}^{n} \left[ \hat{M}_{\nu}(t) \left( G^{*\ast(l-k)}_\nu(t) - G^{*\ast(l-k)}_{\nu x}(t) \right) \right] \right\}.
\]

Summing \( A_l \) over \( l \geq 1 \) and noticing that \( I_{[t < x]} = G^{*\ast 0} - G^{*\ast \ast 0}_{(x)} \), we obtain the first term of our assertion.

Summation of the \( A_{l, k, \nu} \)'s is carried out by changing the summation indices, \( j = l - k, i = k \). Thus, the double sum is written
\[
g(x) \sum_{l \geq 0} \sum_{j \geq 1} g^{*\ast(j+1)} \ast \sum_{\nu=1}^{n} \left[ \hat{M}_{\nu}(t) \left( \hat{G}^{*\ast i}_{\nu}(t) - \hat{G}^{*\ast i}_{\nu x}(t) \right) \right]
\]
\[
= \gamma(t) \ast \left[ \sum_{\nu=1}^{n} \hat{M}_{\nu}(t) \left( \Gamma_{\nu}(t) - \Gamma_{\nu x}(t) \right) \right]
\]
where \( \gamma(t) = \sum_{n \geq 1} g^{*\ast n}(t) \) and \( \Gamma_{i}(t) = \sum_{n \geq 1} G^{*\ast n}_{i}(t) \).

12.2 Asymptotics

**Theorem 12.2** Suppose that all hypotheses of Theorem 12.1 are satisfied and that \( \sum_{i=1}^{n} \mu_i > 1 \), where \( \mu_i = \left. \frac{\partial}{\partial s_i} h(s) \right|_{s=1} \), i.e. that the primary colony process is supercritical. Then
\[
\text{Cov} \left( Z_p(t_p), Z_s(t_s) \right) e^{-\alpha_j t_p} \to 0 \text{ as } t_p \to \infty.
\]
The Malthusian parameter $\alpha_f$ is the only real root of the characteristic equation $\mu \mathcal{L}_{\alpha_f}(j) = 1$ where the probability density $j(t) = \sum_{i=1}^{n} \frac{\mu_i q_i(t)}{\mu}$, $\mu = \sum_{i=1}^{n} \mu_i$ and $\mathcal{L}_{\phi}(f)$ is the Laplace transform of the function $f$ evaluated at $\phi$, i.e. $\mathcal{L}_{\phi}(f) = \int_{0}^{\infty} e^{-\phi t} f(t) dt$.

**Proof.** We first derive the asymptotics of $E(x, t)$. We need only be concerned with the term

$$g(x) \left\{ \gamma(t) \ast \left[ \sum_{\nu=1}^{n} \hat{M}_{\nu}(t) (\Gamma_{\nu}(t) - \Gamma_{\nu x}(t)) \right] \right\}$$

where $\hat{M}_{\nu}(t)$ is defined in equation (12.4). From the definition of our process, $M = A \ast M + \tilde{G}$, where $M = (M_1, \ldots, M_n)'$, $G = (G_1, \ldots, G_n)'$, and $A = G(\mu_1, \ldots, \mu_n)$ (for clarity of notation we suppress $t$). Then, $M = \sum_{i \geq 0} A^{*i} \ast \tilde{G}$. Let us observe that $\tilde{G}(t) = \sum_{\nu=1}^{n} \mu_{\nu} G_{\nu}$, $A^{*i} = \tilde{G}^{*i-1} \ast A$ where $i \geq 1$, and $A^{*0} = I$, so that

$$M = \tilde{G} + \sum_{i \geq 0} \tilde{G}^{*i} \ast A \ast \tilde{G} = \tilde{G} + \sum_{i \geq 0} \tilde{G}^{*i} \ast G \ast \left( \sum_{\nu=1}^{n} \mu_{\nu} \tilde{G}_{\nu} \right). \quad (12.5)$$

Let us define the distribution function $J(t) = \frac{\hat{G}(t)}{\mu}$, where $\mu = \sum_{\nu=1}^{n} \mu_{\nu}$. Equation (12.5) implies

$$M_{k}(t) = \tilde{G}_{k}(t) + \mu \sum_{i \geq 0} [\mu J(t)]^{*i} \ast \tilde{J}(t) \ast G_{k}(t).$$

Since $\sum_{i \geq 0} [\mu J(t)]^{*i} \ast \tilde{J}(t) = E(Z(t))$ for a single type process $Z(t)$ with $h'(1) = \mu > 1$ and lifetime distribution $J(t)$ with density existing, we have that

$$\sum_{i \geq 0} [\mu J(t)]^{*i} \ast \tilde{J}(t) = C_{d} e^{\gamma t} + O(e^{(\alpha_{J} - \gamma)t}) \text{ as } t \to \infty$$

where $C_{d} = \frac{1}{\alpha_{J} \mu^{2}} \int_{0}^{\infty} e^{-\alpha_{J} t} dJ(t)$ [Harris 1963].
We need a lemma.

**Lemma 12.1** Suppose \( a(t) = e^{ct} + O\left(e^{(c-t)t}\right) \) and \( 0 < b(t) \sim B < \infty \), as \( t \to \infty \). Then \( a(t) \ast b(t) \sim b(t) \ast e^{ct} \sim \mathcal{L}_c(b)e^{ct} \), as \( t \to \infty \).

**Proof.** The proof of Lemma 12.1 is elementary.

We recall that

\[
M_k(t) = \bar{G}_k(t) + \mu \sum_{i \geq 0} \left[ \mu \mathcal{J}(t) \right]^{i} \ast \bar{J}(t) \ast G_k(t) = \bar{G}_k(t) + \mu \mathcal{E}(Z(t)) \ast g_k(t).
\]

Using Lemma 12.1, and the fact that \( \bar{G}_k(t) \to 0 \) as \( t \to 0 \), we have

\[
M_k(t) \sim \mu C \alpha^t \mathcal{L}_{\alpha}(g_k).
\] (12.6)

By the Renewal Theorem, [Feller 1966] p.347, \( \Gamma_\nu(t) - \Gamma_{\nu,x}(t) \to \frac{x}{E(T)} \) as \( t \to \infty \).

Therefore,

\[
\sum_{\nu=1}^{n} \hat{M}_\nu \left( \Gamma_\nu - \Gamma_{\nu,x} \right) \sim
\]

\[
\mu C \alpha^t \frac{x}{E(T)} \sum_{\nu=1}^{n} \left[ \sum_{(j_1,j_2,\ldots,j_n) \in \{Z^+ \}^n} a_{j_1,j_2,\ldots,j_n} \left( \frac{j_\nu}{\sum_{\eta=1}^{n} j_\eta} \right) \left( \sum_{\eta=1}^{n} j_\eta \mathcal{L}_\eta - \mathcal{L}_\nu \right) \right],
\]

where \( \mathcal{L}_\nu = \mathcal{L}_{\alpha,j}(g_\nu) \).

Again applying Lemma 12.1,

\[
E(x,t) \sim g(x) \left\{ \gamma(t) \ast \left[ \sum_{\nu=1}^{n} \hat{M}_\nu \left( \Gamma_\nu - \Gamma_{\nu,x} \right) \right] \right\}
\]

\[
\sim \mathcal{L}_{\alpha,j}(\gamma) \mu C \alpha^t \frac{g(x)x}{E(T)} \sum_{\nu=1}^{n} \left[ \sum_{(j_1,j_2,\ldots,j_n) \in \{Z^+ \}^n} a_{j_1,j_2,\ldots,j_n} \left( \frac{j_\nu}{\sum_{\eta=1}^{n} j_\eta} \right) \left( \sum_{\eta=1}^{n} j_\eta \mathcal{L}_\eta - \mathcal{L}_\nu \right) \right].
\] (12.7)
We will simplify this by first concentrating on
\[
\sum_{\nu=1}^{n} \sum_{(j_1,j_2,\ldots,j_n) \in \{2^+\}^{\times n}} a_{j_1,j_2,\ldots,j_n} \left( \frac{j_\nu}{\sum_{\eta=1}^{n} j_\eta} \right) \left( \sum_{\eta=1}^{n} j_\eta \mathcal{L}_\eta - \mathcal{L}_\nu \right) 
= \sum_{\nu=1}^{n} \sum_{(j_1,j_2,\ldots,j_n) \in \{2^+\}^{\times n}} a_{j_1,j_2,\ldots,j_n} j_\eta \mathcal{L}_\eta - \sum_{\nu=1}^{n} p_\nu \mathcal{L}_\nu 
= \sum_{\nu=1}^{n} (\mu_\nu - p_\nu) \mathcal{L}_\nu = \mu \mathcal{L}_{\alpha,j}(j) - \mathcal{L}_{\alpha,j}(g) = 1 - \mathcal{L}_{\alpha,j}(g)
\]
Thus, the summation term in equation (12.7) reduces to
\[
\mathcal{L}_{\alpha,j}(\gamma) \left[ 1 - \mathcal{L}_{\alpha,j}(g) \right] = \frac{\mathcal{L}_{\alpha,j}(g)}{1 - \mathcal{L}_{\alpha,j}(g)} \mathcal{L}_{\alpha,j}(g) 
\]
and so \( E(x,t) \sim \mathcal{L}_{\alpha,j}(g) \mu C e^{\alpha_j g(x)} \frac{g(x)}{E(T)} \). This implies that (12.1), \( E(Z_p(t_p) \mathcal{Z}_s(t_s)) \sim \mathcal{L}_{\alpha,j}(g) \mu C e^{\alpha_j \int_X M_s(t_s|x) \frac{g(x)}{E(T)} dx} \) as \( t_p \to \infty \). To conclude the proof of Theorem 12.2, we need the asymptotics of \( E(Z_p(t_p)) \) and \( E(Z_s(t_s)) \) as \( t_p \to \infty \).

**Lemma 12.2** The lifetime distribution of a randomly selected cell in the primary colony alive at time \( t \) is
\[
f(x;t) = g(x) \left( \Gamma(t) - \Gamma(x)(t) + I_{t<\infty} \right). \quad (12.8)
\]

**Proof.** Lemma 12.2 is proved in Section 12.4.

Based on equation (12.8) and the Renewal Theorem [Feller 1966], we conclude that \( f(x;t) \to \frac{g(x)}{E(T)} \) as \( t \to \infty \). Therefore, \( M_s(t_s) \to \int_0^\infty M_s(t_s|x) \frac{g(x)}{E(T)} dx \) as \( t_p \to \infty \). Now, \( E(Z_p(t_p)) = \sum_{i=1}^{n} p_i M_i(t_p) \) which by equation (12.6) is asymptotically equal to \( \mu C e^{\alpha_j \int_p} \mathcal{L}_{\alpha,j}(g) \).

Collecting these results, we obtain \( \text{Cov}(Z_p(t_p), Z_s(t_s)) e^{-\alpha_j \int_p} \to 0 \) as \( t_p \to \infty \). \( \square \)
12.3 Remarks and example

The sampling formula of Theorem 12.1 enables calculations of mixed second moments of primary and secondary colony counts in the clonal inheritance models. It will serve to evaluate explicit forms of $E(x, t)$ for those variants of the Bellman-Harris process in which explicit forms of $M(t)$ are known.

We also note that the covariance $\text{Cov}(Z_p(t_p), Z_s(t_s))$ is, in general, nonzero.

12.4 Proof of Lemma 2

In equation (12.3), we let $s = 1$. Note that $\hat{h}_i$ becomes $p_i$ (see equation (12.2)).

Thus, denoting $F(x, 1, t, l)$ by $f(x, t, l)$, we get

\[
\int_0^t \int_0^t \cdots \int_{t_{l-1}}^t \sum_{k=1}^n p_k g_k(t_0) \prod_{j=1}^l \left[ \sum_{k=1}^n p_k g_k(t_j - t_{j-1}) \right] \sum_{k=1}^n p_k g_k(x) I_{[0 \leq t_{l-i} < x]} dt_l \cdots dt_0.
\]

\[
= \int_0^t \int_0^t \cdots \int_{t_{l-1}}^t G(t_0, x) \prod_{j=1}^l g(t_j - t_{j-1}) g(x) I_{[0 \leq t_{l-i} < x]} dt_l \cdots dt_0.
\]

Changing variables $u_k = t - t_k$, we obtain

\[
\int_0^t g(u_0) \int_0^{u_0} g(u_1) \cdots \int_0^{u_{l-1}} g(u_l) I_{[u_l < x]} du_l \cdots du_0 = g(x) \left[ G^{(l+1)}(t) - G^{(l+1)}(t) \right]
\]

as in the proof of the theorem. It remains only to sum over $l$, obtaining $f(x; t) = g(x) \left[ \Gamma(t) - \Gamma(t) I_{[t < x]} \right]$. \qed
Chapter 13

Conclusions

13.1 Mathematical results

Generational inheritance model

In Chapter 7, we analyzed a general Galton-Watson branching process model that includes generational inheritance. For this process, we derived an equation for the mixed second moment that describes the relationship between distinct but related colonies. In addition, equations for the standard second moments were derived. The relationship between the colonies may be in one of two ways: “vertically”; that is, one colony is comprised of the offspring of a member of the other, or “horizontally”; that is, each colony is comprised of the offspring of members of a third colony, and hence the two colonies are considered sister colonies.

Knowledge of this second moment allowed us to derive the correlation coefficient describing the relationship between the population sizes of the two colonies at arbitrary times. We note the relative infrequency of second moment analysis in the literature for all but the most straightforward cases. This second moment analysis is absolutely essential to match a model to the data.
Clonal inheritance model

In Chapters 11 and 12, we described Bellman-Harris branching processes with "clonal" inheritance; i.e. the lifetime inheritance mechanism works from mother colony to daughter colony, rather than directly from mother cell to daughter cell.

For these processes, we also derived expressions for the second moments and mixed (mother colony/daughter colony or primary colony/secondary colony) second moments of the colony population size.

Again, we were able to derive expressions for correlation coefficients describing the relationship of colony population size between primary and secondary colonies. Also, we demonstrated that asymptotically, the correlation tends to 0 as \( t \to \infty \).

Further, the example described in Corollary 11.1 shows conclusively that under this model, an exponential cell lifetime distribution can *not* be assumed. This is of some consequence, as the exponential cell lifetime distribution is frequently assumed in analysis of other models due to its greater mathematical tractability.

Chapters 11 and 12 are of more serious mathematical interest, although they have not yet yielded numerical predictions.

13.2 Biological predictions

Based upon a simple two cell-type model, we make some experimentally testable biological observations and predictions.
Firstly, the calculated parameters of our model indicate that in order to retain a significant amount of correlation of cell size from mother to daughter when observing large populations in combination with a very large observed variation in progeny population, a relatively rigid inheritance mechanism is sufficient, if combined with a small subpopulation with significantly different behavior, and some correspondence between the two populations.

Again, it should be pointed out that experimental observations reveal a broad and continuous range of cell lifelengths; a heuristic interpretation of the two life-length classes of our model could be that the mean of the lifelength of one class of cells is 1/2 the mean of the lifelength of the other class of cells. Experimental observations show a continuous distribution of lifelengths with no apparent grouping, because in reality variation of the cell lifelengths blurs the distinctions between the groups, especially since one group is very small.

Secondly, if it is supposed that the disparity of growth rates is due to methylation, or rather methylation-inhibition of some gene affecting growth, then our calculation of $M_3$ to be 0 indicates that at least in the case of genes regulating growth, there is no noticeable de novo methylation of CpG sites affecting the expression of that gene, once the methylation proportion has dropped below a certain level*. We point to the experiments described in Section 5.2 as experimental evidence that methylation can have a severe impact upon the inheritance of growth rates of cells.

*We note that if there were no de novo methylation, eventually all methylation would be lost, as methylation levels could only decrease.
We also note that investigation of the hypothesis that methylation levels affect growth rates may require new estimates of $E_m$ and $E_d$ (the per site per generation probabilities of maintaining methylation levels and de novo methylation respectively).

13.3 Future work

There are several problems that provide opportunities for future research.

Firstly, an attempt at fitting the multi-type Bellman-Harris branching process described in Chapter 12 to the experimental data described in Chapter 5 should be made. Secondly, as mentioned in Chapter 10, a Bellman-Harris process similar to the Galton-Watson process described in Chapter 7 should be analyzed and fitted to the primary-secondary colony data described in Section 5.1; it is expected that with an appropriate choice of lifetime distribution functions, an even better fit of the model can be made. There should be an attempt to model the fluctuating methylation levels of gene promotion regions by using a Markov chain approach similar to [Otto & Walbot 1990]. However, the probability of the process dropping below a specific level of methylation would have to be analyzed, rather than a simple mean methylation level analysis. Finally, these approaches can also be applied to the sister-sister data described in Section 5.2.
Appendix A

Algorithms

A.1 Recursion algorithms for the p.g.f.'s and their derivatives

From Theorem 2.7,

\[ F_1(s, n) = h_i(F_1(s, n - 1), F_2(s, n - 1), \ldots, F_m(s, n - 1)) \]

Hence, the p.g.f. may be calculated recursively. Further,

\[ \frac{\partial}{\partial s_j} F_i(s, n) = \frac{\partial}{\partial s_j} h_i(F_1(s, n - 1), F_2(s, n - 1), \ldots, F_m(s, n - 1)) \]

\[ = \frac{\partial}{\partial s_j} \left\{ \sum_{j_1 \geq 0, j_2 \geq 0} \cdots \sum_{j_m \geq 0} [F_1(s, n - 1)]^{j_1} [F_2(s, n - 1)]^{j_2} \cdots [F_m(s, n - 1)]^{j_m} \varphi[Z_i(1) = j] \right\} \]

\[ = \sum_{j_1 \geq 0, j_2 \geq 0} \cdots \sum_{j_m \geq 0} \frac{\partial}{\partial s_j} [F_1(s, n - 1)]^{j_1} \frac{\partial}{\partial s_j} [F_2(s, n - 1)]^{j_2} \]

\[ \times \cdots \times \frac{\partial}{\partial s_j} [F_m(s, n - 1)]^{j_m} \varphi[Z_i(1) = j] \]

\[ = \sum_{j_1 \geq 0, j_2 \geq 0} \cdots \sum_{j_m \geq 0} j_1 [F_1(s, n - 1)]^{j_1-1} \frac{\partial}{\partial s_j} F_1(s, n-1) j_2 [F_2(s, n - 1)]^{j_2-1} \frac{\partial}{\partial s_j} F_2(s, n-1) \]

\[ \times \cdots \times j_m [F_m(s, n - 1)]^{j_m-1} \frac{\partial}{\partial s_j} F_m(s, n-1) \varphi[Z_i(1) = j] \]
Appendix B

Code

Below are selected segments of a computer code which are implementations of algorithms discussed in Appendix A. The are written in ANSI C programming language, compiled with the Free Software Foundation’s gcc C compiler, and were run on a Sun SPARC 10. A significant difference between the notation used in the code and that used in the discussion is that the indices start at 0 in the code rather than at 1 as in the rest of the thesis.

B.1 Calculation of p.g.f.’s and derivatives of p.g.f.’s

The specific example below calculates $f_0(s,t), f_1(s,t), f_2(s,t), \frac{\partial}{\partial s_0} f_0(s,t), \frac{\partial}{\partial s_0} f_1(s,t)$ and $\frac{\partial^2}{\partial s_0^2} f_2(s,t)$ for the case described in Section 8.2. The algorithm used is described in Section A.1.

```c
struct Evals {
    double F0;
    double F1;
    double F2;
    double DF0;
    double DF1;
    double DF2;
};

extern double M0;          /* probability of switching types */
extern double M2;            /* probability of switching types */

void G1om0(PGF,s0,s1,s2,t)
struct Evals *PGF;
double s0, s1, s2;
int t;
{
struct Evals pgf;

if(t>0) {
Glom0(&pgf, s0, s1, s2, t-1);
(*PGF).F0 = M2 * pgf.F2 + (1.0 - M2) * pgf.F1;
(*PGF).F1 = intpow(pgf.F0,2);
(*PGF).F2 = MO * pgf.F0 + (1.0 - MO) * intpow(pgf.F2,2);
(*PGF).DF0 = M2 * pgf.DF2 + (1.0 - M2) * pgf.DF1;
(*PGF).DF1 = 2 * pgf.F0 * pgf.DF0;
(*PGF).DF2 = MO * pgf.DF0 + 2 * (1.0 - MO) * pgf.F2 * pgf.DF2;
}
else if (t == 0){
(*PGF).F0 = s0;
(*PGF).F1 = s1;
(*PGF).F2 = s2;
(*PGF).DF0 = 1.0;
(*PGF).DF1 = 0.0;
(*PGF).DF2 = 0.0;
}
else {printf("serious error in Glom0: t = %i\n", t); exit(-999);} return;
}

B.2 Calculation of E Z_p

Calculation of E Z_s is identical with the exception that p0s, p1s, p2s are used in
place of p0p, p1p, p2p.
extern double p0p, p1p, p2p;

double EZp(T)
    int T;
{
    struct Evals Diff0, Diff1, Diff2;

    /* N is mean matrix */
    double N[9];

    /* probability transition matrix */
    /* included purely for labeling reasons; i.e. to keep track of matrix */
    double m00,m01,m02,m10,m11,m12,m20,m21,m22;
Glom0(&Diff0,1.0,1.0,1.0,T);
Glom1(&Diff1,1.0,1.0,1.0,T);
Glom2(&Diff2,1.0,1.0,1.0,T);

m00 = M[0] = Diff0.DF0;
m01 = M[1] = Diff1.DF0;
m02 = M[2] = Diff2.DF0;

m10 = M[3] = Diff0.DF1;
m11 = M[4] = Diff1.DF1;
m12 = M[5] = Diff2.DF1;

m20 = M[6] = Diff0.DF2;
m21 = M[7] = Diff1.DF2;
m22 = M[8] = Diff2.DF2;

if(T == 0) return 1.0;

return
   p0p*(M[0] + M[1] + M[2])
Bibliography


