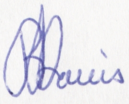
A CHAIN BINOMIAL MODEL WITH IMMIGRATION

A thesis submitted for the degree of
MASTER OF SCIENCE
of The Australian National University

Brett Andrew DAVIS
August 1996

DECLARATION OF AUTHENTICITY

This thesis is my own work and all sources used have been acknowledged.



Brett Andrew Davis

August 1996

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Professor C. R. Heathcote, for his role in the preparation of this thesis. Not only did he offer numerous helpful suggestions and constructive criticism but also a great deal of encouragement.

I also wish to thank Dr T. J. O'Neill for his invaluable advice and ideas on the use of Monte Carlo data augmentation techniques.

My thanks also go to Michael Hiron and Borek Puza for their kind assistance in the computer laboratory.

CONTENTS

SUMMARY

Chain binomial models are a commonly used discrete time model for the spread of an infectious disease in a closed population. The best known chain binomial models are those due to Greenwood and to Reed and Frost. In the Greenwood model it is assumed that the infection rate is the same at all time points, while in the Reed-Frost model, the infection rate at time t depends on the number of individuals infected at time $t-1$.

The aim of this thesis is to extend the Greenwood model (for a closed population) to a model which allows for the possibility of migration into the population at risk of infection from the disease. In fact, we modify the Greenwood model to a model which assumes individuals enter the susceptible population as a Poisson process. In particular, the number of migrants entering during a unit time interval $(t-1, t]$ is a Poisson random variable.

The model we use assumes that migration counts during disjoint unit time intervals are independent and identically distributed Poisson random variables. That is, we assume the migration rate (and infection rate) to be constant.

Chapter One discusses some estimators for the infection rate in the Greenwood model for a closed population. In particular, we review a technique used by Saunders (1980b) to derive a strongly consistent estimator. Our modified Greenwood model, which allows for the possibility of (constant rate) migration, is also developed in this chapter. All subsequent analysis is based on this model.

Let X_t denote the number of newly infected individuals at time t. In Chapter Two we establish some limit theorems for the sequence $\{X_t : t = 1, 2, 3, \ldots\}$. Specifically, we show that a strong law of large numbers holds for this sequence of random variables. Also the cumulative infected count,

$$\sum X_t ,$$

obeys a central limit theorem.

The sequence $\{Y_t: t = 0, 1, 2, \ldots\}$, where Y_t is the number of susceptible individuals who escape infection at time t, forms a Markov chain. In Chapter Three we show that this sequence of random variables can be formulated as the convolution of a Galton-Watson branching process and an independent Poisson process. That is, it is an example of what is known in the literature as a *branching process with immigration.* Using this interpretation we show that, in the limit as time goes to infinity, the count of susceptible individuals escaping infection converges in distribution to a Poisson random variable.

It is also shown that, the bivariate Markov chain $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ has a proper limiting distribution. From this limiting distribution we deduce that, as t tends to infinity, X_t converges in distribution to a Poisson random variable X_∞. Furthermore, the parameter of the distribution of X_∞ is equal to the rate at which individuals migrate into the susceptible population.

Since $\{Y_t: t = 0, 1, 2, \ldots\}$ is a *branching process with immigration* then we can use the estimation theory for this type of process to derive estimators of the infection and migration rates. We give formulae for estimators derived in this fashion which are jointly asymptotically normal.

These estimators assume we can observe the count of susceptible individuals escaping infection at regular time intervals. However, for many infectious diseases this data is not available. For example, in the case of cholera in South East Asia, the available data only consists of information about the number of infected individuals. In Chapter Four we demonstrate that the Gibbs sampler (a Monte Carlo data augmentation algorithm) can be used to obtain estimates of the infection and migration rates when the data consists of just counts of infected individuals (and the initial size of the susceptible population).

CHAPTER ONE

DISCRETE TIME MODELS FOR INFECTIOUS DISEASES

1.0 Introduction

In this chapter we introduce the class of so-called *chain binomial models*, a flexible type of discrete time model used extensively in the modelling of the spread of an infectious disease.

As pointed out in Section 14.1 of Bailey (1975) there are diseases where it is not appropriate to assume that infection occurs as a continuous time process. For diseases where the incubation and latent periods are of low variability and the infectious period is short, discrete time models should be employed.

Compared to continuous time models the amount of theory developed for discrete time models is quite small. However, chain binomial models have proven to be a useful tool in modelling epidemics since their introduction by Reed and Frost in about 1928 and Greenwood in 1931.

Gani and Jerwood (1971) relate chain binomial models to the general theory of stochastic processes with particular emphasis on Markov chains. This approach has several advantages, in particular, it leads to the formulation of a relatively simple method of calculating the probability distributions of the duration time and total infected count for an epidemic. Note that an epidemic terminates when either all susceptibles become infected or when, at a particular time point, there are no new infecteds. Prior to the work of Gani and Jerwood these distributions had been obtained by calculating the probabilities associated with all possible paths of infection.

In Section 1.1 we discuss the structure of chain binomial models and look at the special cases due to Greenwood and to Reed and Frost. In Section 1.2 we investigate determining the probability distributions of the duration time and total infected count. We first use a small example to illustrate the amount of work involved in enumerating all patterns of infection to solve this problem and then show that the work of Gani and Jerwood leads to great simplification.

In Section 1.3 we look at the problem of estimation in chain binomial models and describe an approximate maximum likelihood estimator for the infection rate. The asymptotic properties of this estimator will also be discussed. In Section 1.4 we show how chain binomial models can be extended to include the possibility of individuals migrating into the population at risk of infection. Finally, in Section 1.5 we compare the methods suggested in Section 1.4 with the capture-recapture methods used to estimate mortality rate and population size in animal populations.

1.1 Chain Binomial Models for Closed Populations

Chain binomial models are commonly used to model the spread of an infectious disease through a population. The structure of these models can be described as follows.

In an initial population of size N_0 a number of individuals, X_0, become infected at time $t=0$. Prior to time $t=1$ there will be $Y_0 = N_0 - X_0$ susceptible individuals. We assume that each susceptible individual has the same probability $p(1)$ of being infected at time $t=1$ and that infection of an individual occurs independently of infection to any other individual. If we let X_1 be the number of newly infected individuals at time $t=1$ then in a chain binomial model we assume X_1, conditional on $N_0 - X_0$, is binomial with parameters $N_1 = N_0 - X_0$ and $p(1)$. In general, let X_t be the number of newly infected individuals at time t then, conditional on $\{N_0, X_0, X_1, ..., X_{t-1}\}$, X_t is binomial with parameters N_t and $p(t)$ where

$$N_t = N_0 - \sum_{j=0}^{t-1} X_j$$

and $p(t)$ is the probability that an individual susceptible to the disease immediately prior to time t becomes infected at time t. Of course, we make this assumption provided N_t is strictly greater than zero. As soon as N_t reaches zero for some t we conclude the epidemic has ended.

Chain binomial models are best applied to diseases where the incubation period and latent period (where the disease is present but the infected individual can not transmit the disease) are of low variability and the infectious period is relatively short. Measles, mumps and chickenpox are examples of diseases with these characteristics. For such diseases the epidemic may be thought of as commencing with the infection of X_0 individuals simultaneously. The epidemic then spreads as a series of generations where the time between successive generations is equal to the sum of the incubation and latent periods.

The probability of infection at time t, $p(t)$, is usually assumed to be a function of N_t, I_t and λ where N_t is the size of the susceptible population,

$$I_t = \sum_{j=0}^{t-1} X_j$$

is the total number of infected individuals in the population prior to time t and λ is a constant. Typically, the values of N_t and I_t are readily calculated from the data and our aim is to make inference about λ.

The two best known examples of chain binomial models are those developed by Reed and Frost in about 1928 and by Greenwood in 1931. The simpler model is the Greenwood model where the chance of infection depends only on whether there are infected individuals already present. That is, it does not depend on the number of such individuals. For this model we have

$$p(t) = \lambda \qquad \text{if} \quad I_t > 0$$
$$0 \qquad \text{otherwise}$$

for some $0 < \lambda \leq 1$. In the Reed-Frost model it is assumed that $p(t)$ depends on the number of individuals infected at time $t-1$. We suppose that any of the X_{t-1} individuals can independently infect a susceptible with probability λ. Thus, for a susceptible, the probability of avoiding infection is

$$(1-\lambda)^{X_{t-1}}.$$

Thus,

$$p(t) = 1-(1-\lambda)^{X_{t-1}}.$$

If we let Y_t be the number of individuals escaping infection at time t then, from Gani and Jerwood (1971), in the Greenwood model $\{Y_t : t = 0, 1, 2, \ldots\}$ forms a Markov chain. The probability transition matrix $M = \{ m_{ij} \}$ is given by

$$m_{ij} = \binom{i}{j} \lambda^{i-j} (1-\lambda)^j \qquad \text{for} \quad 0 \leq j \leq i$$

$$0 \qquad \text{otherwise.}$$

In the Reed-Frost model a bivariate Markov chain is formed by

$$\{(X_t, Y_t) : t = 0, 1, 2, \ldots\}$$

for which the probability transition matrix takes the form $\{(m_{ij})_{rs}\}$, where

$$\left(m_{ij}\right)_{rs} = \mathrm{P}(X_{t+1} = j, Y_{t+1} = s \mid X_t = i, Y_t = r)$$

$$= \binom{r}{j}\left[1-(1-\lambda)^i\right]^j (1-\lambda)^{i(r-j)} \qquad \text{for } r = s + j$$

$$0 \qquad \text{otherwise.}$$

In the Reed-Frost model, we can approximate $p(t)$ by its first order Taylor polynomial when λ is small. That is, $p(t)$ can be approximated by

$$1-(1-\lambda X_{t-1}) = \lambda X_{t-1}. \qquad (1.1.1)$$

This form of $p(t)$ is particularly appropriate for diseases with very short infectious periods because we are assuming the chance of infection does not depend on the number of individuals infected prior to time $t-1$. The model is also appropriate for modelling diseases which often effect animal populations where infected individuals inevitably die from the disease. That is, 'old' infecteds are removed from the population by death.

Saunders (1980a) uses a chain binomial model to study an epidemic of myxomatosis in a rabbit population. In this case $p(t)$ is of the form

$$p(t) = \lambda a_t, \qquad (1.1.2)$$

where a_t depends on the total population size, the size of the susceptible population and the number of active infecteds prior to time t. This approach may be regarded as an extension of the approximation to the Reed-Frost model given in equation (1.1.1).

Chain binomial models where the probability of infection is of the form given in equation (1.1.2) have been extensively studied by Saunders (1980b) and Huggins (1993). The work of these two authors will be discussed in Section 1.3.

As mentioned earlier, we aim to extend chain binomial models to a model which includes immigration. The application of the theory of Markov chains to such a model will be discussed in Chapter Three. Gani and Jerwood (1971) show that, for the Greenwood model, the susceptible counts at times $t = 0, 1, 2, \ldots$ may be regarded as a Markov chain embedded in a (continuous time) pure death process. Similarly, for the Reed-Frost model, over a unit time interval $(t, t+1]$, the susceptible count at time $t+1$ (given the susceptible count at time t) may also be approximated by a continuous time process.

1.2 The Probability Distributions of the Duration Time and Total Infected Count

We now consider the problem of determining the joint probability distribution of the duration time and total infected count in a chain binomial model. In Subsection 1.2.1 we use an example to show that enumerating all possible paths of infection is tedious even when the population size is small. In Subsection 1.2.2 we discuss the more sophisticated Markov chain method of Gani and Jerwood which leads to a convenient formula for the joint probability generating function for these two variables.

1.2.1 An Example of Enumerating Infection Patterns

Consider an epidemic in a household of size four in which there are initially two infected individuals. There are four possible paths of infection.

1) No new infecteds at time $t = 1$. The epidemic ends with the total number of infecteds equal to two.

2) One new infected at time $t = 1$ and no new infecteds at time $t = 2$. The epidemic ends with the total number of infecteds equal to three.

3) One new infected at time $t = 1$ and one new infected at time $t = 2$. The epidemic ends with all susceptibles becoming infected.

4) Two new infecteds at time $t = 1$. Again the epidemic ends with all susceptibles becoming infected.

Since infection at different time points are independent processes then the probabilities of these four possible patterns are easily calculated. As shown on page 79 of Bailey (1957) the probabilities are

Pattern	Probability in the Reed-Frost Model	Probability in the Greenwood Model
1	$(1-\lambda)^4$	$(1-\lambda)^2$
2	$2\lambda(1-\lambda)^3(2-\lambda)$	$2\lambda(1-\lambda)^2$
3	$2\lambda^2(1-\lambda)^2(2-\lambda)$	$2\lambda^2(1-\lambda)$
4	$\lambda^2(2-\lambda)^2$	λ^2

Of course, this table is essentially the joint distribution of the duration time and total infected count for both models. These distributions can be presented explicitly as follows.

Total infecteds	Duration	Probability in the Reed-Frost Model	Probability in the Greenwood Model
2	1	$(1-\lambda)^4$	$(1-\lambda)^2$
3	2	$2\lambda(1-\lambda)^3(2-\lambda)$	$2\lambda(1-\lambda)^2$
4	2	$2\lambda^2(1-\lambda)^2(2-\lambda)$	$2\lambda^2(1-\lambda)$
4	1	$\lambda^2(2-\lambda)^2$	λ^2

From this example it is clear that enumerating all possible paths of infection will quickly become tedious as the population size increases. However, this approach makes little use of the Markovian structure of these models. Gani and Jerwood exploit this aspect and show that it leads to a simple method for calculating the probability generating function of the joint distribution of the duration time and total infected count.

1.2.2 Applying Markov Chain Methods to the Greenwood and Reed-Frost Models

Gani and Jerwood view the Greenwood and Reed-Frost models as special cases of the Markov chain $\{z_t : t = 0, 1, 2, \ldots\}$ with finite state space $\{0, 1, 2, \ldots, k\}$ and probability transition matrix M whose diagonal entries are strictly positive. In this subsection we give a discussion of their method and its advantages. We consider the variable T defined by

$$T = \min\{t > 0 : z_t = z_{t-1}\}.$$

In the application to the Greenwood and Reed-Frost models $\{z_t : t = 0, 1, 2, \ldots\}$ is defined in such a way that T is the duration time of the epidemic.

For t > 0 and j = 0, 1, . . ., k,

$$P(T = t,\ z_t = j \mid z_0 = i) = P(T = t,\ z_t = z_{t-1} = j \mid z_0 = i)$$

$$= (P^{t-1})_{ij} m_{jj}.$$

Therefore, for $0 \le \theta, \phi \le 1$,

$$E[\theta^T \phi^{z_T} \mid z_0 = i] = \sum_{t=1}^{\infty} \sum_{j=0}^{k} \theta^t \phi^j A_i P^{t-1} R,$$

where

A_i is the $1 \times (k+1)$ row vector with one in the $(i+1)$th entry and zeroes elsewhere,

$P = \{p_{ij}\}$ is the $(k+1) \times (k+1)$ matrix whose entries are given by

$$p_{ij} = m_{ij} \qquad \text{if } i \neq j$$

$$0 \qquad \text{otherwise,}$$

and R is the transpose of $[m_{00}, m_{11}, \ldots, m_{kk}]$.

That is,

$$E[\theta^T \phi^{z_T} | z_0 = i] = \sum_{t=1}^{\infty} \theta^t A_i(\phi)(P(\phi))^{t-1} R,$$

where

$A_i(\phi)$ is the $1 \times (k+1)$ row vector with ϕ^i in the $(i+1)$th entry and zeroes elsewhere,

$P(\phi)$ is the $(k+1) \times (k+1)$ matrix whose (i, j) th entry is given by

$$p_{ij}\phi^{j-i}.$$

That is,

$$E[\theta^T \phi^{z_T} | z_0 = i] = A_i(\phi)(I - \theta P(\phi))^{-1} \theta R. \qquad (1.2.1)$$

In the Greenwood model the sequence $\{Y_t : t = 0, 1, 2, \ldots\}$ is a special case of the above model in which the state space is equal to $\{0, 1, 2, \ldots, Y_0 = N_0 - X_0\}$ and the probability transition matrix $M = \{m_{ij}\}$ is given by

$$m_{ij} = \binom{i}{j} \lambda^{i-j} (1-\lambda)^j \qquad \text{for } 0 \leq j \leq i$$

$$0 \qquad \text{otherwise.}$$

The total infected count is equal to

$$Y_0 - Y_T = (Y_0 - Y_1) + (Y_1 - Y_2) + \ldots + (Y_{T-1} - Y_T).$$

Therefore, by analogy with the above discussion for the general case, the joint probability generating function for the duration time and the total infected count is given by the last entry ($(Y_0 + 1)$ th entry) of

$$\left(I - \theta P(\phi)\right)^{-1} \theta R,$$

where $P(\phi)$ is the $(Y_0 + 1) \times (Y_0 + 1)$ matrix whose (i, j) th entry is equal to

$$p_{ij} \phi^{i-j}.$$

For the Reed-Frost model we have observed that $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ forms a bivariate Markov chain. We can apply the above theory to this Markov chain with only slight modifications. The method involved in this application is outlined below.

We take as state space the Cartesian product of $\{0, 1, 2, \ldots, k\}$ with itself. The probability transition matrix M of size $(k+1)^2 \times (k+1)^2$ is defined as follows. M consists of $(k+1)^2$ submatrices $(m_{ij})_{rs}$ each of size $(k+1) \times (k+1)$, where for $r, s \in \{0, 1, \ldots, k\}$, the submatrix $(m_{ij})_{rs}$ is defined by

$$(m_{ij})_{rs} = \begin{cases} \binom{s+j}{j} [1 - (1-\lambda)^i]^j (1-\lambda)^{is} & \text{for } r = s + j \\ 0 & \text{otherwise.} \end{cases}$$

Define R to be the $k(k+1) \times 1$ column vector formed by concatenating the diagonals of $(m_{ij})_{10}$, $(m_{ij})_{20}$, . . . , $(m_{ij})_{k0}$. Also define $P(\phi)$ as the $k(k+1) \times k(k+1)$ matrix

consisting of k^2 submatrices $(P(\phi)_{ij})_{rs}$ where, for $r, s \in \{1, 2, \ldots, k\}$, the submatrix $(P(\phi)_{ij})_{rs}$, of size $(k+1)\times(k+1)$, is defined by

$$(P(\phi)_{ij})_{rs} = (m_{ij})_{rs}\phi^{r-s}.$$

We note that this approach not only alleviates the need to enumerate all paths of infection but provides a method which is applicable to a much wider class of models than just chain binomial models. In fact, this method can be applied to any Markov chain with finite state space provided the probability transition matrix has strictly positive diagonal entries.

1.3 An Approximate Maximum Likelihood Estimator for Chain Binomial Models

In this section we review work on an approximate maximum likelihood estimator for the infection rate in a chain binomial model developed by Saunders (1980b) and Huggins (1993). The estimator is developed using the Poisson approximation to the binomial distribution.

In Subsection 1.3.1 we derive the formula for the estimator by replacing the binomial probabilities in the likelihood with their Poisson approximations. Subsection 1.3.2 deals with some statistical properties of this estimator as we increase the time period over which the epidemic is observed. In Subsection 1.3.3 we assume the period of observation is fixed and we investigate the asymptotic behaviour as the size of the initial susceptible population increases.

1.3.1 Definition of the Approximate Maximum Likelihood Estimator

Before discussing the work of Saunders (1980b) and Huggins (1993) we switch to the notation they use which is appropriate for chain binomial models where the probability of infection is of the form given in equation (1.1.2).

Let $\{X_t : t = 1, 2, \ldots\}$ be a sequence of random variables and let $\mathcal{J}_t$ be the σ-field generated by $\{N_0, X_0, X_1, \ldots, X_t\}$. Suppose that, for t = 1, 2, . . . , X_t, conditional on $\mathcal{J}_{t-1}$, is binomially distributed with parameters N_t and λa_t where N_t and a_t are $\mathcal{J}_{t-1}$-measurable random variables. That is, N_t and a_t are determined by $N_0, X_0, \ldots, X_{t-1}$. The quantity λ is a constant which we aim to estimate.

Define

$$A_T = \max\{a_t : t = 1, 2, \ldots, T\}$$

then the likelihood function for $0 < \lambda < A_T^{-1}$, given $N_0, X_0, X_1, \ldots, X_{T-1}$, is defined by

$$L(\lambda) = \prod_{t=1}^{T} \binom{N_t}{X_t} (\lambda a_t)^{X_t} (1 - \lambda a_t)^{N_t - X_t} . \qquad (1.3.1)$$

From the log-likelihood

$$l(\lambda) = \sum_{t=1}^{T} \left[\ln \binom{N_t}{X_t} + X_t \ln(\lambda a_t) + (N_t - X_t) \ln(1 - \lambda a_t) \right]$$

it is readily seen that the maximum likelihood estimator $\hat{\lambda}$ is the solution of the equation

$$\sum_{t=1}^{T} \frac{X_t - \lambda a_t N_t}{1 - \lambda a_t} = 0 .$$

Following Saunders (1980b) we now replace each term in equation (1.3.1) with its Poisson approximation. That is, we assume X_t, conditional on $\mathcal{J}_{t-1}$, is approximately Poisson with parameter $\lambda a_t N_t$. This gives an approximate likelihood function L^* defined, for $0 < \lambda < A_T^{-1}$, by

$$L^*(\lambda) = \prod_{t=1}^{T} \frac{e^{-\lambda a_t N_t} (\lambda a_t N_t)^{X_t}}{X_t!}.$$

We define an approximate maximum likelihood estimator $\tilde{\lambda}$ as the value of λ which maximises L^*.

$$l^*(\lambda) \equiv \ln L^*(\lambda) = \sum_{t=1}^{T} [-\lambda a_t N_t + X_t \ln(\lambda a_t N_t) - \ln(X_t!)]$$

$$\frac{dl^*}{d\lambda} = -\sum_{t=1}^{T} a_t N_t + \frac{1}{\lambda} \sum_{t=1}^{T} X_t$$

Therefore,

$$\tilde{\lambda} = \frac{\sum_{t=1}^{T} X_t}{\sum_{t=1}^{T} a_t N_t}.$$

1.3.2 Some Statistical Properties of this Approximate MLE

We now derive some statistical properties of $\tilde{\lambda}$ which depend both on the analytic form of the estimator and the assumption that, for t = 1, 2, . . . , X_t (conditional on $\mathcal{I}_{t-1}$) is approximately Poisson.

In this subsection we outline the work of Saunders (1980b) which deals with the asymptotic behaviour of $\tilde{\lambda}$ and also provides a bound for its mean square error under the assumption that the conditional expectation of X_t, given $\mathcal{I}_{t-1}$, is bounded.

In order to derive these properties Saunders introduces the following conditions.

There exists a strictly positive constant m such that

$\lambda a_t N_t \geq m$ almost surely. (A)

There exists a strictly positive constant M such that

$\lambda a_t N_t \leq M$ almost surely. (B)

The first theorem of Saunders (1980b) states that, for a chain binomial model satisfying condition (A), $\tilde{\lambda}$ is strongly consistent. That is, as T tends to infinity, $\tilde{\lambda}$ converges to λ almost surely.

This result is obtained by noting that

$$\tilde{\lambda} - \lambda = \frac{\sum_{t=1}^{T} (X_t - \lambda a_t N_t)}{B_T}, \qquad (1.3.2)$$

where $B_T = \sum_{t=1}^{T} a_t N_t$.

To see that this quantity converges to zero almost surely we consider the zero mean martingale $\{U_t, \mathcal{I}_t : t = 1, 2, \ldots\}$ where,

$$U_t = \sum_{j=1}^{t} \frac{X_j - \lambda a_j N_j}{B_j}$$

and $\mathcal{I}_t$ is the σ-field generated by $\{N_0, X_0, X_1, \ldots, X_t\}$. That this is a zero mean martingale can be shown, using the fact that B_t is $\mathcal{I}_{t-1}$-measurable, as follows

$$E[U_t | \mathcal{I}_{t-1}] = U_{t-1} + \frac{1}{B_t} E[X_t - \lambda a_t N_t | \mathcal{I}_{t-1}]$$

$$= U_{t-1} .$$

Furthermore,

$$E[U_t^2] = \sum_{j=1}^{t} E[E[(X_j - \lambda a_j N_j)^2 / B_j^2 \mid \mathcal{I}_{t-1}]]$$

$$= E\left(\sum_{j=1}^{t} \frac{\lambda a_j N_j (1 - \lambda a_j)}{B_j^2} \right)$$

$$< E\left(\sum_{j=1}^{t} \frac{\lambda a_j N_j}{B_j^2} \right)$$

$$= \lambda E\left(\sum_{j=1}^{t} \frac{B_j - B_{j-1}}{B_j^2} \right)$$

$$< \lambda E\left[\sum_{j=2}^{t} \left(\frac{1}{B_{j-1}} - \frac{1}{B_j} \right) + \frac{1}{B_1} \right]$$

$$\leq \frac{2\lambda}{a_1 N_1} < \infty.$$

Hence, we can apply the Martingale Convergence Theorem (see page 242 of Feller (1971)) to $\{U_t, \mathcal{I}_t : t = 1, 2, \ldots\}$. This implies that $\{U_t : t = 1, 2, \ldots\}$ converges almost surely. The result now follows from Kronecker's Lemma (see page 239 of Feller (1971)).

We remark that, in a closed population, N_t is montonically decreasing. It is intuitively clear that N_t approaches zero as t tends to infinity. For a rigorous proof of this we note that, the least value of t for which N_t is zero, is the duration time of the epidemic. The result now follows from page 593 of Gani and Jerwood (1971) where it is shown that the probability distribution of the duration time sums to one. Since N_t converges to

zero then it follows from the definition of $\tilde{\lambda}$ that asymptotic normality does not hold for this estimator as T tends to infinity.

Saunders uses equation (1.3.2) to show that if both conditions (A) and (B) hold in a chain binomial model, the mean square error of $\tilde{\lambda}$ is bounded by

$$\frac{\lambda^2 M}{t^2 m^2}.$$

That is, the relative mean square error of $\tilde{\lambda}$ is bounded by

$$\frac{M}{t^2 m^2}.$$

Conditions (A) and (B) are likely to hold in most practical situations. However, in (B) we are putting a condition on $E[X_t | \mathcal{I}_{t-1}]$ which is not entirely natural. We would expect the performance of $\tilde{\lambda}$ to improve as the expected number of infecteds increases. So it seems a little inappropriate to put an upper bound on this quantity.

The upper bound on $MSE(\tilde{\lambda})$ involves the unnatural bound M. However, an upper bound for $MSE(\tilde{\lambda})$ can be derived from condition (A) alone if we assume the coefficient of variation of B_T is low and the dependence between B_T and $X_t - \lambda a_t N_t$ is also low for all $t = 1, 2, 3, \ldots, T$. Again using equation (1.3.2), Saunders shows the mean square error is bounded above by

$$\frac{\lambda^2}{tm}.$$

Note that this agrees with the bound derived above when m and M are approximately the same. We also point out that this gives a bound for the relative mean square error of

$$\frac{1}{tm}.$$

1.3.3 Asymptotic Inference When the Susceptible Count is Large

In Huggins (1993) asymptotic properties of $\tilde{\lambda}$ are given when the time period over which the epidemic is observed is fixed and the size of the initial susceptible population tends to infinity.

We now introduce notation analogous to that used in the previous subsection. For a strictly positive integer n, let $\{X_{nt}: t = 1, 2, \ldots\}$ be a sequence of random variables and let $\mathcal{F}_{nt}$ be the σ–field generated by $\{X_{n1}, X_{n2}, \ldots, X_{nt}\}$. Conditional on $\mathcal{F}_{nt-1}$, X_{nt} has a binomial distribution with parameters N_{nt} and λa_{nt}, where N_{nt} and a_{nt} are $\mathcal{F}_{nt-1}$-measurable random variables. We consider values of λ in the interval $(0, A_{nT}^{-1})$, where

$$A_{nT} = \max\{a_{nt}: t = 1, 2, \ldots, T\}.$$

Define

$$\tilde{\tilde{\lambda}} = \frac{\sum_{t=1}^{T} w_{nt} X_{nt}}{\sum_{t=1}^{T} w_{nt} a_{nt} N_{nt}}.$$

where $\{w_{nt}: t = 1, 2, \ldots, T\}$ is a sequence of non-negative $\mathcal{F}_{nt-1}$-measurable random variables which do not depend on λ and have the property that, for each n, $w_{nt} > 0$ for at least one $t \in \{1, 2, \ldots, T\}$. We note that $\tilde{\lambda}$ is a special case of $\tilde{\tilde{\lambda}}$ when $w_{nt} = 1$ for all $t = 1, 2, \ldots, T$.

Huggins shows that, under the following regularity conditions, asymptotic normality holds for $\tilde{\tilde{\lambda}}$ as n tends to infinity.

$$N_{nt} \to \infty \qquad \text{in probability.} \tag{1.3.3}$$

For $t = 1, 2, \ldots, T$, there exist constants $a_t \in (0, \lambda^{-1})$ and $b_t \in (0, 1)$ such that

$$a_{nt} \to a_t \qquad \text{in probability,} \tag{1.3.4}$$

$$b_{nt} \to b_t \qquad \text{in probability,} \tag{1.3.5}$$

and there exist constants w_t, where $w_t > 0$ for at least one value of t, such that

$$w_{nt} \to w_t \qquad \text{in probability.} \tag{1.3.6}$$

Under these conditions

$$\left(\sum_{t=1}^{T} w_{nt} a_{nt} N_{nt}\right)^{1/2} (\tilde{\tilde{\lambda}} - \lambda) = \left(\sum_{t=1}^{T} w_{nt} a_{nt} N_{nt}\right)^{-1/2} \sum_{t=1}^{T} w_{nt} (X_{nt} - \lambda a_{nt} N_{nt})$$

converges in distribution to a normally distributed random variable with mean zero and variance

$$\frac{\sum_{t=1}^{T} w_t b_t \lambda a_t (1 - \lambda a_t)}{\sum_{t=1}^{T} w_t b_t a_t}.$$

We comment that there is little point in including the possibility of migration into the susceptible population if the initial size of the susceptible population is increasing to infinity. It is intuitive that as the initial size of the susceptible population increases the effect of migration decreases. Heuristically, we expect the possibility of migration will have no effect on the asymptotic results derived from this approach.

Hence, although Huggins' approach is very innovative, this line of investigation will not be discussed further in this thesis. That is, in our analysis the initial size of the susceptible population will be a known constant.

1.4 A Model Incorporating Migration

So far we have discussed a model for the spread of an infectious disease in a closed population. That is, we have not considered the possibility of migration into or out of the population at risk of infection. In such a model, change in the size of the susceptible population is solely due to individuals becoming infected.

When we use a closed population model we are assuming that changes in the susceptible population due to factors other than the disease are negligible. This may be appropriate, for example, when the susceptible population is very large. However, for most infectious diseases this is not an appropriate assumption because the susceptible population may be changing due to factors such as individuals changing location, and perhaps more importantly, changing their behaviour.

The models we have considered also assume that once an individual has become infected they are permanently removed from the susceptible population. This is true for diseases such as measles , mumps and chickenpox where infection confers protective immunity. That is, upon recovery infected individuals are no longer susceptible to the disease. Another situation where this assumption is valid is where infected individuals inevitably die from the disease. As mentioned earlier, this is often the case for diseases effecting animal populations such as myxomatosis in rabbits. However, many diseases do not have this characteristic, for example, influenza and many sexually transmitted diseases such as gonorrhea.

The model we now consider is subject to the same constraints as a chain binomial model for a closed population except that it allows for the possibility of individuals migrating into the population at risk of infection. A model which allows for migration can be applied to diseases in which infected individuals are not removed from the susceptible population. This can be done by regarding an individual recovering from infection as a migrant coming into the susceptible population if they continue to exhibit behaviour which puts them at risk of infection.

In order to develop a model which allows for migration into the susceptible population we proceed as follows. Suppose migrants of susceptibles arrive as a Poisson stream with

parameter $\mu(t)$. That is, the probability of M_t new susceptibles arriving between time $t-1$ and time t is given by

$$P(M_t = m_t) = \frac{e^{-\mu(t)}\mu(t)^{m_t}}{m_t!} \qquad \text{for } m_t = 0, 1, 2, \ldots .$$

We also assume that the number of migrants entering the susceptible population during the time interval $(t-1, t]$ is independent of the number of individuals infected at time $t-1$. Thus we have, for $n_t \geq n_{t-1} - x_{t-1}$ and $0 \leq x_t \leq n_t$,

$$P(N_t = n_t, X_t = x_t \mid N_{t-1} = n_{t-1}, X_{t-1} = x_{t-1})$$

$$= P(X_t = x_t \mid N_t = n_t, N_{t-1} = n_{t-1}, X_{t-1} = x_{t-1})$$

$$\times P(N_t = n_t \mid N_{t-1} = n_{t-1}, X_{t-1} = x_{t-1})$$

$$= \binom{n_t}{x_t} p(t)^{x_t} (1-p(t))^{n_t - x_t} \frac{e^{-\mu(t)}\mu(t)^{[n_t - (n_{t-1} - x_{t-1})]}}{[n_t - (n_{t-1} - x_{t-1})]!}. \qquad (1.4.1)$$

Furthermore, since migration counts in disjoint time intervals are assumed independent and infection counts at distinct time points are also assumed independent (in chain binomial models) then

$$P(N_1 = n_1, X_1 = x_1, N_2 = n_2, \ldots, N_T = n_T, X_T = x_T \mid \mathcal{I}_0)$$

$$= \prod_{t=1}^{T} \binom{n_t}{x_t} p(t)^{x_t} (1-p(t))^{n_t - x_t} \frac{e^{-\mu(t)}\mu(t)^{n_t - n_{t-1} + z_t}}{(n_t - n_{t-1} + x_{t-1})!}$$

$$= e^{-\sum_1^T \mu(t)} \prod_{t=1}^{T} \frac{\mu(t)^{n_t - n_{t-1} + x_{t-1}}}{(n_t - n_{t-1} + x_{t-1})!} \binom{n_t}{x_t} p(t)^{x_t} (1-p(t))^{n_t - x_t}$$

Note that the natural logarithm of this expression is equal to

$$\sum_{t=1}^{T} x_t \ln(p(t)) + \sum_{t=1}^{T} (n_t - x_t) \ln(1 - p(t)) - \sum_{t=1}^{T} \mu(t)$$

$$+ \sum_{t=1}^{T} (n_t - n_{t-1} + x_{t-1}) \ln(\mu(t)) + \sum_{t=1}^{T} \ln\left(\frac{\binom{n_t}{x_t}}{(n_t - n_{t-1} + x_{t-1})!} \right).$$

Consider the case where $p(t)$ and $\mu(t)$ are of the form described in equation (1.1.2). That is,

$$p(t) = \lambda a_t$$

and

$$\mu(t) = \mu b_t,$$

where λ and μ are unknown constants and a_t and b_t are known functions of the data. Then the log-likelihood for λ and μ is given by

$$l(\lambda, \mu \mid N_0 = n_0, X_0 = x_0, \ldots, N_T = n_T, X_T = x_T)$$

$$= \sum_{t=1}^{T} x_t (\ln \lambda + \ln a_t) + \sum_{t=1}^{T} (n_t - x_t) \ln(1 - \lambda a_t) - \mu \sum_{t=1}^{T} b_t$$

$$+ \sum_{t=1}^{T} (n_t - n_{t-1} + x_{t-1})(\ln \mu + \ln b_t) + C,$$

where

$$C = \sum_{t=1}^{T} \ln \left(\frac{\binom{n_t}{x_t}}{(n_t - n_{t-1} + x_{t-1})!} \right).$$

Maximising with respect to λ we have

$$\frac{\partial l}{\partial \lambda} = \sum_{t=1}^{T} \frac{x_t - \lambda a_t n_t}{\lambda(1 - \lambda a_t)}.$$

Hence, the maximum likelihood estimator $\hat{\lambda}$ of λ satisfies

$$\sum_{t=1}^{T} \frac{x_t - \lambda a_t n_t}{1 - \lambda a_t} = 0.$$

Maximising with respect to μ we have

$$\frac{\partial l}{\partial \mu} = -\sum_{t=1}^{T} b_t + \frac{1}{\mu} \sum_{t=1}^{T} (n_t - n_{t-1} + x_{t-1}).$$

Thus, the maximum likelihood estimator $\hat{\mu}$ of μ satisfies

$$\hat{\mu} = \frac{n_T - n_0 + \sum_{t=0}^{T-1} x_t}{\sum_{t=1}^{T} b_t}.$$

In the Greenwood model we put $a_t = 1$ for all t. That is, we assume the probability of infection only depends on whether there are infected individuals already present, but not on the number of such individuals. We now make the analogous assumption regarding the migration rate. That is, we put $b_t = 1$ for all t. Of course, in making this

assumption we are assuming the migration process does not depend on any aspect of the history of the epidemic such as the cumulative number of infected individuals.

The likelihood is then given by

$$L(\lambda,\mu \mid N_0 = n_0, X_0 = x_0, \ldots, N_T = n_T, X_T = x_T)$$

$$= C' e^{-\mu T} \mu^{n_T - n_0 + \sum_{t=0}^{T-1} x_t} \prod_{t=1}^{T} \lambda^{x_t} (1-\lambda)^{n_t - x_t}, \qquad (1.4.2)$$

where

$$C' = \prod_{t=1}^{T} \left(\frac{\binom{n_t}{x_t}}{(n_t - n_{t-1} + x_{t-1})!} \right).$$

The log-likelihood is equal to

$$\sum_{t=1}^{T} x_t \ln\lambda + \sum_{t=1}^{T} (n_t - x_t) \ln(1-\lambda) - \mu T + \left(n_T - n_o + \sum_{t=0}^{T-1} x_t \right) \ln\mu + C.$$

The maximum likelihood estimators become

$$\hat{\lambda} = \frac{\sum_{t=1}^{T} x_t}{\sum_{t=1}^{T} n_t}$$

and

$$\hat{\mu} = \frac{n_T - n_0 + \sum_{0}^{T-1} x_t}{T}.$$

At this stage we wish to make it clear that all analysis in this thesis is based on the assumption that, for all t, $p(t) = \lambda$ and $\mu(t) = \mu$.

The data on the spread of an infectious disease usually consists of counts of diagnosed infecteds. Rarely will information on the size of the susceptible population at regular time intervals be available. Note that this is certainly the case for the AIDS epidemic and also for diseases, such as cholera, which are still endemic in parts of South East Asia. In Chapter Four we will consider the problem of estimating λ and μ where the data consists of a realisation of $X_1, X_2, \ldots, X_T$ (where T is fixed) and the initial conditions $N_0 = n_0$ and $X_0 = x_0$.

1.5 Capture-Recapture Methods

There is a close analogy between estimating infection and migration rates in epidemiology and the *capture-recapture* methods of estimating mortality rate and population size in animal populations. Capture-recapture methods are discussed in detail in Seber (1982).

The capture-recapture method closest to discrete time epidemiological models is where information is gathered from a single batch of tagged animals. See Chapter Six of Seber (1982). In this type of experiment a batch of tagged animals is released at time $t = 0$ and the whole population (tagged and untagged animals) is subsequently sampled at discrete time points $t = 1, 2, \ldots, T$.

To illustrate how estimation is performed we introduce the following notation which is used by Seber. For $t = 1, 2, \ldots, T$, let

N_t = size of the whole population immediately before the t-th sample is drawn,

M_t = size of the marked (tagged) population immediately before the t-th sample is drawn,

n_t = size of the t-th sample,

m_t = number of marked animals in the t -th sample.

Expressions for M_t and N_t are derived in a series of deterministic equations involving the initial population size N_0, the natural mortality rate ϕ and the rate at which animals enter the population τ. For $t = 1, 2, \ldots, T$, N_t, M_t and n_t are known constants and estimation of N_0, ϕ and τ is performed as follows.

In general, M_t will be small compared to N_t so it is assumed that m_t has a Poisson distribution with parameter α_t, where

$$\alpha_t = \frac{M_t}{N_t} n_t .$$

This follows from the assumption that marked and unmarked animals have the same chance of being selected in any particular sample. Thus the likelihood function is

$$L(N_0, \phi, \tau \mid m_1, m_2, \ldots, m_T) = \prod_{t=1}^{T} \frac{e^{-\alpha_t} \alpha_t^{m_t}}{m_t !} .$$

Capture-recapture methods rely on having information on marked animals whose behaviour is assumed to be representative of the whole population with respect to the parameters of interest. The analogous information when studying the spread of an infectious disease would be counts of infected individuals for an identifiable subpopulation. If this type of information is not available then capture-recapture methods do not seem applicable to modelling the spread of an infectious disease.

We also note that in capture-recapture models there are a large number of unknown parameters to be imputed, namely, $\alpha_1, \alpha_2, \ldots, \alpha_T$. This is also the case in the model of the previous section when the migration counts can not be observed. In this case the susceptible counts can be thought of as unknown parameters. In Chapter Four we show that the Gibbs sampler can be used to estimate the unknown parameters in this situation.

Hence, it is quite possible that a Markov chain Monte Carlo method could be applied to estimate the parameters of a capture-recapture model. However, we will not attempt to deal with this problem.

CHAPTER TWO

MOMENTS AND CENTRAL LIMIT THEOREMS FOR THE GREENWOOD MODEL

2.0 Introduction

In Section 1.4 we discussed a possible extension of chain binomial models which would allow for the possibility of migration into the population at risk of infection. In particular, we showed that the Greenwood model could be extended to allow for immigration (see equation (1.4.2)). As mentioned in Section 1.4, the results of this thesis are restricted to this particular model. That is, we assume, for $t = 1, 2, 3, \ldots,$

$$p(t) = \lambda$$

and

$$\mu(t) = \mu .$$

We also assume that N_0 and X_0 are known constants.

The aim of this chapter is to establish a central limit theorem for the cumulative infected count $\sum X_t$. Central limit theorems are also given for $\sum N_t$ and $\sum Y_t$, where Y_t is the number of susceptible individuals escaping infection at time t. That is,

$$Y_t = N_t - X_t .$$

However, when interpreting the results for $\sum N_t$ and $\sum Y_t$ we need to keep in mind that a single individual can contribute to N_t or Y_t for more than one value of t.

In Section 2.1 we calculate the moments of X_t, N_t and Y_t and look at the properties of these moments as t tends to infinity. In Section 2.2 we obtain the limiting values of the expectation and variance of the averages:

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t, \qquad \overline{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t, \qquad \overline{N}_n = \frac{1}{n}\sum_{t=1}^{n} N_t.$$

We also show, using the limit theory of martingales, that a strong law of large numbers holds for each of the sequences $\{X_t : t = 1, 2, 3, \ldots\}$, $\{Y_t : t = 1, 2, 3, \ldots\}$ and $\{N_t : t = 1, 2, 3, \ldots\}$.

The limit theory of martingales is again used in Section 2.3 to establish central limit theorems for each of $\sum X_t$, $\sum N_t$ and $\sum Y_t$.

The results of this chapter can be formally stated as follows.

Theorem 2.1 For the model described in equation (1.4.2), we have the following formulae for the moments associated with X_t and N_t.

$$E(X_t | \mathcal{J}_0) = \mu - \mu(1-\lambda)^t + \lambda(1-\lambda)^{t-1}(N_0 - X_0) \tag{2.1.1}$$

$$E(N_t | \mathcal{J}_0) = \frac{\mu}{\lambda} - \frac{\mu}{\lambda}(1-\lambda)^t + (1-\lambda)^{t-1}(N_0 - X_0) \tag{2.1.2}$$

$$E(Y_t | \mathcal{J}_0) = \frac{\mu}{\lambda}(1-\lambda) - \frac{\mu}{\lambda}(1-\lambda)^{t+1} + (1-\lambda)^t(N_0 - X_0) \tag{2.1.3}$$

$$Var(X_t | \mathcal{J}_0) = \mu\left[1-(1-\lambda)^t\right] + \lambda(1-\lambda)^{t-1}\left[1-\lambda(1-\lambda)^{t-1}\right](N_0 - X_0) \tag{2.1.4}$$

$$Var(N_t|\mathcal{J}_0) = \mu + (1-\lambda)\left[1-(1-\lambda)^{t-1}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{t-2}(N_0 - X_0)\right] \quad (2.1.5)$$

$$Var(Y_t|\mathcal{J}_0) = (1-\lambda)\left[1-(1-\lambda)^{t}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{t-1}(N_0 - X_0)\right] \quad (2.1.6)$$

$$Cov(X_t, Y_t|\mathcal{J}_0) = -\lambda(1-\lambda)^{2t-1}(N_0 - X_0) \quad (2.1.7)$$

For $j > i \geq 1$,

$$Cov(X_i, X_j|\mathcal{J}_0) = -\lambda^2(1-\lambda)^{i+j-2}(N_0 - X_0) \quad (2.1.8)$$

$$Cov(N_i, N_j|\mathcal{J}_0) = (1-\lambda)^{j-i}\left\{\mu + (1-\lambda)\left[1-(1-\lambda)^{i-1}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{i-2}(N_0 - X_0)\right]\right\} \quad (2.1.9)$$

$$Cov(Y_i, Y_j|\mathcal{J}_0) = (1-\lambda)^{j-i+1}\left[1-(1-\lambda)^{i}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{i-1}(N_0 - X_0)\right] \quad (2.1.10)$$

The proof of this theorem is given in Section 2.1.

Theorem 2.2 For the model described in equation (1.4.2):

(i) As n tends to infinity

$$E(\overline{X}_n) \to \mu,$$

$$E(\overline{Y}_n) \to \frac{\mu}{\lambda}(1-\lambda),$$

$$E(\overline{N}_n) \to \frac{\mu}{\lambda}.$$

(ii) $Var(\overline{X}_n)$, $Var(\overline{Y}_n)$ and $Var(\overline{N}_n)$ all converge to zero as n tends to infinity.

(iii) A weak law of large numbers holds for each of the sequences $\{X_t: t = 1, 2, 3, \ldots\}$, $\{Y_t: t = 1, 2, 3, \ldots\}$ and $\{N_t: t = 1, 2, 3, \ldots\}$. That is, as n tends to infinity, $\overline{X}_n$, $\overline{Y}_n$ and $\overline{N}_n$ converge in probability to μ, $\frac{\mu}{\lambda}(1-\lambda)$ and $\frac{\mu}{\lambda}$, respectively.

(iv) A strong law of large numbers holds for each of the sequences $\{X_t: t = 1, 2, 3, \ldots\}$, $\{Y_t: t = 1, 2, 3, \ldots\}$ and $\{N_t: t = 1, 2, 3, \ldots\}$. That is, as n tends to infinity, $\overline{X}_n$, $\overline{Y}_n$ and $\overline{N}_n$ converge almost surely to μ, $\frac{\mu}{\lambda}(1-\lambda)$ and $\frac{\mu}{\lambda}$, respectively.

The proof of this theorem is given in Section 2.2. Parts (i) and (ii) are proved in Subsection 2.2.1, part (iii) is proved in Subsection 2.2.2 and part (iv) is proved in Subsection 2.2.3. Since almost sure convergence implies convergence in probability then part (iii) follows immediately from part (iv). However, a separate proof of part (iii) is given to show that this result is a simple consequence of Chebychev's inequality. Part (iv) is established using a result from the limit theory of martingales. This mirrors the situation for a sequence of independent and identically distributed random variables in that the weak law of large numbers follows from Chebychev's inequality while the strong law requires a far more complex argument.

Theorem 2.3 If $\mathcal{F}_t^*$ is the σ-field generated by $\{N_0, X_0, N_1, X_1, \ldots, N_t, X_t\}$ then, for the model described in equation (1.4.2), we have the following central limit theorems.

(i) $$\frac{1}{\sqrt{n}}\left[\sum_{t=1}^{n} X_t - \sum_{t=1}^{n} E\left(X_t | \mathcal{F}_{t-1}^*\right)\right]$$

converges in distribution to a normal random variable with mean zero and variance $\mu(1-\lambda+\lambda^2)$.

(ii) $$\frac{1}{\sqrt{n}}\left[\sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} E\left(Y_t | \mathcal{F}_{t-1}^*\right)\right]$$

converges in distribution to a normal random variable with mean zero and variance $\mu(1-\lambda)(2-\lambda)$.

(iii) $$\frac{1}{\sqrt{n}}\left[\sum_{t=1}^{n} N_t - \sum_{t=1}^{n} E\left(N_t | \mathcal{F}_{t-1}^{*}\right)\right]$$

converges in distribution to a normal random variable with mean zero and variance μ.

The proof of this theorem is given in Section 2.3. Parts (i), (ii) and (iii) are proved in Subsections 2.3.1, 2.3.2 and 2.3.3, respectively.

2.1 Moments of X_t, Y_t and N_t

In this section we derive formulae for the means, variances and covariances of X_t, Y_t and N_t. We will also discuss the asymptotic behaviour of these quantities as t tends to infinity. In particular, we note that for these three variables, the limit of the expectation is equal to the limit of the variance.

In Subsection 2.1.1 we derive a recurrence relation for the probability generating function of $\{Y_t : t = 0, 1, 2, \ldots\}$ from which we obtain our formulae. This recurrence relation will be used again in Chapter Three where we investigate the existence of a limiting distribution for $\{Y_t : t = 0, 1, 2, \ldots\}$. Means, variances and covariances are derived in Subsections 2.1.2, 2.1.3 and 2.1.4, respectively.

2.1.1 The Probability Generating Function of Y_t

Let $Y_t(z)$ denote the probability generating function of Y_t. That is,

$$Y_t(z) = E\left(z^{Y_t}\right).$$

In order to evaluate this expectation we begin with the conditional expectation $E\left(z^{Y_t} \mid Y_{t-1}\right)$, which is equal to

$$E_{N_t}\left(E\left(z^{Y_t} \mid Y_{t-1}, N_{t-1}\right)\right).$$

In the Greenwood model Y_t is binomially distributed with parameters N_t and $1-\lambda$. Thus,

$$E\left(z^{Y_t} \mid Y_{t-1} = y_{t-1}\right) = E_{N_t}\left((\lambda + (1-\lambda)z)^{N_t}\right)$$

$$= \sum_{n_t = y_{t-1}}^{\infty} \frac{(\lambda + (1-\lambda)z)^{n_t} e^{-\mu} \mu^{n_t - y_{t-1}}}{(n_t - y_{t-1})!}$$

$$= e^{-\mu}(\lambda + (1-\lambda)z)^{y_{t-1}} \sum_{n_t = y_{t-1}}^{\infty} \frac{(\lambda + (1-\lambda)z)^{n_t - y_{t-1}} \mu^{n_t - y_{t-1}}}{(n_t - y_{t-1})!}$$

$$= e^{-\mu(1-\lambda)(1-z)}[\lambda + (1-\lambda)z]^{y_{t-1}}.$$

Taking expectations,

$$Y_t(z) = e^{-\mu(1-\lambda)(1-z)} Y_{t-1}(\lambda + (1-\lambda)z). \qquad (2.1.11)$$

2.1.2 Expectations

Differentiating both sides of equation (2.1.11) gives

$$Y_t'(z) = (1-\lambda)e^{-\mu(1-\lambda)(1-z)}\left[\mu Y_{t-1}(\lambda + (1-\lambda)z) + Y_{t-1}'(\lambda + (1-\lambda)z)\right]$$

Putting $z = 1$ we have

$$E(Y_t) = (1-\lambda)(\mu + E(Y_{t-1})).$$

This implies

$$E(Y_t) = \mu\left[(1-\lambda) + (1-\lambda)^2\right] + (1-\lambda)^2 E(Y_{t-2}).$$

Continuing this procedure we obtain

$$E(Y_t) = \mu\left[(1-\lambda) + (1-\lambda)^2 + \ldots + (1-\lambda)^t\right] + (1-\lambda)^t E(Y_0).$$

Since we assume N_0 and X_0 are known constants then

$$E(Y_t) = \frac{\mu}{\lambda}(1-\lambda)\left[1-(1-\lambda)^t\right] + (1-\lambda)^t (N_0 - X_0).$$

Note that

$$\lim_{t\to\infty} E(Y_t) = \frac{\mu}{\lambda}(1-\lambda).$$

The expectation of N_t can now be calculated using

$$N_t = Y_{t-1} + M_t,$$

where M_t is the number of susceptible migrants arriving during the time interval $(t-1, t]$. Since M_t is a Poisson random variable with parameter μ then

$$E(N_t) = \mu + E(Y_{t-1})$$

$$= \frac{\mu}{\lambda}\left[1-(1-\lambda)^t\right] + (1-\lambda)^t (N_0 - X_0).$$

We observe that

$$\lim_{t\to\infty} E(N_t) = \frac{\mu}{\lambda}.$$

The limiting expectation of N_t is an increasing function of μ and a decreasing function of λ. Both of these results are clearly intuitive.

Finally,

$$E(X_t) = E_{N_t}\left(E(X_t | N_t)\right)$$

$$= E(\lambda N_t)$$

$$= \mu\left[1-(1-\lambda)^t\right] + \lambda(1-\lambda)^t(N_0 - X_0).$$

Hence,

$$\lim_{t\to\infty} E(X_t) = \mu.$$

That is, X_t is a consistent estimator of the migration rate.

The formulae given in this subsection are expectations conditional on the initial values N_0 and X_0. However, the results may be stated more generally as follows. For $0 < r \le t$,

$$E\left(X_t | \mathcal{F}^*_{t-r}\right) = \mu - \mu(1-\lambda)^r + \lambda(1-\lambda)^{r-1}(N_{t-r} - X_{t-r}), \qquad (2.1.1')$$

$$E\left(N_t | \mathcal{F}^*_{t-r}\right) = \frac{\mu}{\lambda} - \frac{\mu}{\lambda}(1-\lambda)^r + (1-\lambda)^{r-1}(N_{t-r} - X_{t-r}), \qquad (2.1.2')$$

$$E\left(Y_t | \mathcal{F}^*_{t-r}\right) = \frac{\mu}{\lambda}(1-\lambda) - \frac{\mu}{\lambda}(1-\lambda)^{r+1} + (1-\lambda)^r(N_{t-r} - X_{t-r}), \qquad (2.1.3')$$

where $\mathcal{F}^*_{t-r}$ is the σ-field generated by $\{N_0, X_0, N_1, X_1, \ldots, N_{t-r}, X_{t-r}\}$.

2.1.3 Variances

Differentiating equation (2.1.11) a second time gives

$$Y_t''(z) = (1-\lambda)^2 e^{-\mu(1-\lambda)(1-z)}\left[\mu^2 Y_{t-1}(\lambda+(1-\lambda)z)+\mu Y_{t-1}'(\lambda+(1-\lambda)z)\right]$$

$$+(1-\lambda)^2 e^{-\mu(1-\lambda)(1-z)}\left[\mu Y_{t-1}'(\lambda+(1-\lambda)z)+Y_{t-1}''(\lambda+(1-\lambda)z)\right].$$

Putting $z=1$ we have

$$Y_t''(1) = (1-\lambda)^2\left[\mu^2 Y_{t-1}(1)+2\mu Y_{t-1}'(1)+Y_{t-1}''(1)\right].$$

Now

$$Var(Y_t) = Y_t''(1)+Y_t'(1)-\left[Y_t'(1)\right]^2$$

$$=(1-\lambda)^2\left[\mu^2+2\mu Y_{t-1}'(1)+Y_{t-1}''(1)\right]+(1-\lambda)\left[\mu+Y_{t-1}'(1)\right]$$

$$-(1-\lambda)^2\left[\mu^2+2\mu Y_{t-1}'(1)+\left(Y_{t-1}'(1)\right)^2\right]$$

$$=(1-\lambda)\mu+(1-\lambda)^2 E[Y_{t-1}(Y_{t-1}-1)]+(1-\lambda)E(Y_{t-1})-(1-\lambda)^2[E(Y_{t-1})]^2$$

$$=(1-\lambda)\mu+\lambda(1-\lambda)E(Y_{t-1})+(1-\lambda)^2 Var(Y_{t-1}).$$

That is,

$$Var(Y_t) = \mu(1-\lambda)+\mu(1-\lambda)^2-\mu(1-\lambda)^{t+1}+\lambda(1-\lambda)^t(N_0-X_0)+(1-\lambda)^2 Var(Y_{t-1}).$$

We now demonstrate that the solution of this recurrence relation is

$$Var(Y_t|\mathcal{I}_0) = (1-\lambda)\left[1-(1-\lambda)^t\right]\left[\frac{\mu}{\lambda}+(1-\lambda)^{t-1}(N_0-X_0)\right].$$

Using this formula to obtain $Var(Y_{t-1}|\mathcal{I}_0)$, the right hand side of the recurrence relation becomes

$$\mu(1-\lambda)+\mu(1-\lambda)^2-\mu(1-\lambda)^{t+1}+\lambda(1-\lambda)^t(N_0-X_0)$$

$$+(1-\lambda)^3\left\{\left[1-(1-\lambda)^{t-1}\right]\left[\frac{\mu}{\lambda}+(1-\lambda)^{t-2}(N_0-X_0)\right]\right\}.$$

This is equal to

$$(1-\lambda)\mu\left[1+(1-\lambda)-(1-\lambda)^t+\frac{(1-\lambda)^2}{\lambda}\left(1-(1-\lambda)^{t-1}\right)\right]$$

$$+(N_0-X_0)(1-\lambda)^t\left[\lambda+(1-\lambda)\left(1-(1-\lambda)^{t-1}\right)\right].$$

That is,

$$\frac{(1-\lambda)\mu}{\lambda}\left(1-(1-\lambda)^t\right)+(1-\lambda)^t\left(1-(1-\lambda)^t\right)(N_0-X_0),$$

which is the proposed value of $Var(Y_t|\mathcal{I}_0)$.

We observe that

$$\lim_{t\to\infty} Var(Y_t) = \frac{\mu}{\lambda}(1-\lambda) = \lim_{t\to\infty} E(Y_t).$$

The variance of N_t is now readily obtained using

$$Var(N_t) = Var(M_t + Y_{t-1}).$$

Since we assume the migration and infection processes are independent then

$$Var(N_t|\mathcal{I}_0)=\mu+(1-\lambda)\left[1-(1-\lambda)^{t-1}\right]\left[\frac{\mu}{\lambda}+(1-\lambda)^{t-2}(N_0-X_0)\right].$$

Hence,

$$\lim_{t\to\infty} Var(N_t)=\frac{\mu}{\lambda}=\lim_{t\to\infty} E(N_t).$$

We now use the formulae for $E(N_t)$ and $Var(N_t)$ to derive $Var(X_t)$.

$$Var(X_t)=Var_{N_t}\left[E(X_t|N_t)\right]+E_{N_t}\left[Var(X_t|N_t)\right]$$

$$=Var(\lambda N_t)+E[\lambda(1-\lambda)N_t]$$

$$=\lambda^2\left\{\mu+(1-\lambda)\left[1-(1-\lambda)^{t-1}\right]\left[\frac{\mu}{\lambda}+(1-\lambda)^{t-2}(N_0-X_0)\right]\right\}$$

$$+\lambda(1-\lambda)\left\{\frac{\mu}{\lambda}-\frac{\mu}{\lambda}(1-\lambda)^t+(1-\lambda)^{t-1}(N_0-X_0)\right\}$$

$$=\lambda^2\mu+(1-\lambda)\mu\left\{\left(\lambda-\lambda(1-\lambda)^{t-1}\right)+\left(1-(1-\lambda)^t\right)\right\}$$

$$+\left\{\lambda^2(1-\lambda)^{t-1}\left(1-(1-\lambda)^{t-1}\right)+\lambda(1-\lambda)^t\right\}(N_0-X_0)$$

$$=\mu-\mu(1-\lambda)^t+\lambda(1-\lambda)^{t-1}\left(1-\lambda(1-\lambda)^{t-1}\right)(N_0-X_0).$$

Note that

$$\lim_{t\to\infty} Var(X_t)=\mu=\lim_{t\to\infty} E(X_t).$$

Finally, as in Subsection 2.1.2, the formula given for the variance of X_t, N_t and Y_t may be stated more generally in the following way. For $0<r\le t$,

$$Var\left(X_t \mid \mathcal{F}^*_{t-r}\right) = \mu\left[1-(1-\lambda)^r\right] + \lambda(1-\lambda)^{r-1}\left[1-\lambda(1-\lambda)^{r-1}\right](N_{t-r} - X_{t-r}), \quad (2.1.4')$$

$$Var\left(N_t \mid \mathcal{F}^*_{t-r}\right) = \mu + (1-\lambda)\left[1-(1-\lambda)^{r-1}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{r-2}(N_{t-r} - X_{t-r})\right], \quad (2.1.5')$$

$$Var\left(Y_t \mid \mathcal{F}^*_{t-r}\right) = (1-\lambda)\left[1-(1-\lambda)^r\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{r-1}(N_{t-r} - X_{t-r})\right], \quad (2.1.6')$$

where $\mathcal{F}^*_{t-r}$ is the σ-field generated by $\{N_0, X_0, N_1, X_1, \ldots, N_{t-r}, X_{t-r}\}$.

2.1.4 Covariances

We begin by calculating the covariance of X_t and Y_t, that is, we establish equation (2.1.7). Since

$$X_t + Y_t = N_t$$

then this covariance is necessarily negative.

$$Cov(X_t, Y_t) = E[(N_t - X_t)X_t] - E[(N_t - X_t)]E(X_t)$$

$$= \lambda E(N_t^2) - E(X_t^2) - \lambda(1-\lambda)[E(N_t)]^2$$

$$= \lambda E(N_t^2) - Var(X_t) - \lambda^2[E(N_t)]^2 - \lambda(1-\lambda)[E(N_t)]^2.$$

Since

$$Var(X_t) = E_{N_t}\left[Var(X_t \mid N_t)\right] + Var_{N_t}\left[E(X_t \mid N_t)\right]$$

$$= \lambda(1-\lambda)E(N_t) + \lambda^2 E(N_t^2) - \lambda^2[E(N_t)]^2$$

then

$$Cov(X_t, Y_t) = \lambda(1-\lambda)\left(E[(N_t^2)] - [E(N_t)]^2\right) - \lambda(1-\lambda)E(N_t)$$

$$= \lambda(1-\lambda)\left(Var(N_t) - E(N_t)\right) \qquad (2.1.12)$$

$$= -\lambda(1-\lambda)^{2t-1}(N_0 - X_0).$$

By the same argument we have, for $0 < r \le t$,

$$Cov\left(X_t, Y_t | \mathcal{I}_{t-r}^*\right) = -\lambda(1-\lambda)^{2r-1}(N_{t-r} - X_{t-r}).$$

Also note that

$$\lim_{t\to\infty} Cov\left(X_t, Y_t | \mathcal{I}_0\right) = 0.$$

The formula for the covariance of X_t and Y_t can be used to derive the covariance of X_t and X_{t-r} (for $0 < r < t$). Before calculating this covariance we point out that we should anticipate a negative value for this quantity. To see this, consider the case when r is equal to one. If X_{t-1} is large then (assuming the number of migrants entering during the interval $(t-1, t]$ is not extraordinarily large) the number of susceptibles at time t will be much smaller than that at time $t-1$. Since we are assuming a constant rate of infection, λ, then the expected number of new infecteds at time t will also be comparatively small. Similarly, when X_{t-1} is small we expect the value of X_t to be large.

$$Cov\left(X_t, X_{t-r} | \mathcal{I}_0\right) = E_{\mathcal{I}_{t-r}^*}\left(Cov\left(X_t, X_{t-r} | \mathcal{I}_{t-r}^*\right)\right) + Cov\left(E\left(X_t | \mathcal{I}_{t-r}^*\right), E\left(X_{t-r} | \mathcal{I}_{t-r}^*\right)\right)$$

$$= 0 + Cov\left(\mu - \mu(1-\lambda)^r + \lambda(1-\lambda)^{r-1}(N_{t-r} - X_{t-r}), X_{t-r}\right)$$

$$= \lambda(1-\lambda)^{r-1} Cov(N_{t-r} - X_{t-r}, X_{t-r})$$

$$= -\lambda^2(1-\lambda)^{2t-r-2}(N_0 - X_0).$$

We comment that this formula is consistent with our intuitive belief that the correlation between X_t and X_{t-r} is negative for all $0 < r < t$.

In Section 2.2 this formula will be used in the alternative form, for $j > i \geq 1$,

$$Cov\left(X_i, X_j | \mathcal{I}_0\right) = -\lambda^2(1-\lambda)^{i+j-2}(N_0 - X_0).$$

In general, for $r < k \leq t$,

$$Cov\left(X_t, X_{t-r} | \mathcal{I}^*_{t-k}\right) = -\lambda^2(1-\lambda)^{2(k-1)-r}(N_{t-k} - X_{t-k}).$$

To establish this result we apply the argument which gave equation (2.1.12) to $Cov\left(X_t, X_{t-r} | \mathcal{I}^*_{t-k}\right)$.thus we obtain

$$Cov\left(X_t, X_{t-r} | \mathcal{I}^*_{t-k}\right) = \lambda^2(1-\lambda)^r\left(Var\left(N_{t-r} | \mathcal{I}^*_{t-k}\right) - E\left(N_{t-r} | \mathcal{I}^*_{t-k}\right)\right).$$

The result now follows from equations (2.1.2′) and (2.1.5′).

We now calculate the covariance of the susceptible counts using a similar technique to that used in calculating the covariance of the infected counts. Because of the constant infection rate we anticipate that the covariance of N_t and N_{t-r} will be positive for $0 < r < t$.

$$Cov\left(N_t, N_{t-r} | \mathcal{I}_0\right) = E_{\mathcal{I}^*_{t-r}}\left(Cov\left(N_t, N_{t-r} | \mathcal{I}^*_{t-r}\right)\right) + Cov\left(E\left(N_t | \mathcal{I}^*_{t-r}\right), E\left(N_{t-r} | \mathcal{I}^*_{t-r}\right)\right)$$

$$= 0 + Cov\left(\frac{\mu}{\lambda} - \frac{\mu}{\lambda}(1-\lambda)^r + (1-\lambda)^{r-1}(N_{t-r} - X_{t-r}), N_{t-r}\right)$$

$$= (1-\lambda)^{r-1} Cov(N_{t-r} - X_{t-r}, N_{t-r})$$

$$= (1-\lambda)^{r-1} E[(N_{t-r} - X_{t-r})N_{t-r}] - (1-\lambda)^r [E(N_t)]^2 .$$

Since

$$E[(N_{t-r} - X_{t-r})N_{t-r}] = (1-\lambda)E(N_{t-r}^2)$$

$$= (1-\lambda)\left(Var(N_{t-r}) + [E(N_{t-r})]^2\right)$$

then

$$Cov\left(N_t, N_{t-r} | \mathcal{J}_0\right) = (1-\lambda)^r Var(N_{t-r}),$$

which is equal to

$$(1-\lambda)^r \left\{\mu + (1-\lambda)\left[1-(1-\lambda)^{t-r-1}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{t-r-2}(N_0 - X_0)\right]\right\}.$$

Alternatively, for $j > i \geq 1$,

$$Cov\left(N_i, N_j | \mathcal{J}_0\right) = (1-\lambda)^{j-i}\left\{\mu + (1-\lambda)\left[1-(1-\lambda)^{i-1}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{i-2}(N_0 - X_0)\right]\right\}$$

In general, for $r < k \leq t$,

$$Cov\left(N_t, N_{t-r} | \mathcal{J}^*_{t-k}\right) = (1-\lambda)^r Var(N_{t-r} | \mathcal{J}^*_{t-k})$$

and this is equal to

$$(1-\lambda)^r \left\{\mu + (1-\lambda)\left[1-(1-\lambda)^{k-r-1}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{k-r-2}(N_0 - X_0)\right]\right\}$$

The assumption of a constant infection rate implies that the covariance of Y_t and Y_{t-r}

is positive for all positive values of r. To see why this is true we again consider the case where r is equal to one.

$$N_t = Y_{t-1} + M_t,$$

where M_t is the number of new susceptibles entering during the time interval $(t-1, t]$. Suppose Y_{t-1} is large then N_t will also be large. Then, as individuals escape infection at a constant rate (equal to $1-\lambda$), we expect Y_t will be relatively large. Similarly, suppose Y_{t-1} is small then N_t is approximately equal to M_t. Thus, N_t is small so we expect Y_t to be small.

We now derive a formula for the covariance of Y_t and Y_{t-r}.

$$Cov(Y_t, Y_{t-r} | \mathcal{I}_0) = E_{\mathcal{I}^*_{t-r}}\left(Cov\left(Y_t, Y_{t-r} | \mathcal{I}^*_{t-r}\right)\right) + Cov\left(E\left(Y_t | \mathcal{I}^*_{t-r}\right), E\left(Y_{t-r} | \mathcal{I}^*_{t-r}\right)\right)$$

$$= Cov\left(\frac{\mu}{\lambda}(1-\lambda) - \frac{\mu}{\lambda}(1-\lambda)^{r+1} + (1-\lambda)^r (N_{t-r} - X_{t-r}), N_{t-r} - X_{t-r}\right)$$

$$= (1-\lambda)^r Var(N_{t-r} - X_{t-r})$$

$$= (1-\lambda)^{r+1}\left[1 - (1-\lambda)^{t-r}\right]\left[\frac{\mu}{\lambda} + (1-\lambda)^{t-r-1}(N_0 - X_0)\right].$$

In Section 2.2 we will use this result in the alternative form, for $j > i \geq 1$,

$$Cov\left(Y_i, Y_j | \mathcal{I}_0\right) = (1-\lambda)^{j-i+1}\left(1 - (1-\lambda)^i\right)\left(\frac{\mu}{\lambda} + (1-\lambda)^{i-1}(N_{t-k} - X_{t-k})\right).$$

In general, for $r < k \leq t$,

$$Cov\left(Y_t, Y_{t-r} | \mathcal{I}^*_{t-k}\right) = (1-\lambda)^{r+1}\left(1 - (1-\lambda)^{k-r}\right)\left(\frac{\mu}{\lambda} + (1-\lambda)^{k-r-1}(N_{t-k} - X_{t-k})\right).$$

2.2 The Averages of X_t, N_t and Y_t

This section is concerned with the moments of $\overline{X}_n$, $\overline{N}_n$ and $\overline{Y}_n$; the averages of X_t, N_t and Y_t over the time points $t = 1, 2, 3, \ldots$. Implications for the asymptotic properties of these random variables will be investigated.

In Subsection 2.2.1 we give formulae for the expectations of these averages and the variance of $\overline{X}_n$. The limiting values of $Var(\overline{N}_n)$ and $Var(\overline{Y}_n)$ are also given. In Subsection 2.2.2 we show that Chebychev's inequality is sufficient to establish a weak law of large numbers for $\{X_t : t = 1, 2, 3, \ldots\}$, $\{N_t : t = 1, 2, 3, \ldots\}$ and $\{Y_t : t = 1, 2, 3, \ldots\}$. That is, $\overline{X}_n$ converges in probability to μ, $\overline{N}_n$ converges in probability to $\frac{\mu}{\lambda}$ and $\overline{Y}_n$ converges in probability to $\frac{\mu(1-\lambda)}{\lambda}$. In Subsection 2.2.3 we show each of these results can be extended to a strong law of large numbers. That is, $\overline{X}_n$ converges almost surely to μ, $\overline{N}_n$ converges almost surely to $\frac{\mu}{\lambda}$ and $\overline{Y}_n$ converges almost surely to $\frac{\mu(1-\lambda)}{\lambda}$.

2.2.1 Expectations and Variances of $\overline{X}_n$, $\overline{N}_n$ and $\overline{Y}_n$

Using equation (2.1.1) we have

$$E(\overline{X}_n) = \frac{1}{n}\sum_{t=1}^{n} E(X_t)$$

$$= \mu - \frac{\mu(1-\lambda)}{n\lambda}\left(1-(1-\lambda)^n\right) + \frac{1-(1-\lambda)^n}{n}(N_0 - X_0).$$

Note that

$$\lim_{n\to\infty} E(\overline{X}_n) = \mu.$$

That is, the mean of the count of newly infected individuals is a consistent estimator of the rate at which new susceptibles enter the population at risk of infection.

$$Var\left(\overline{X}_n\right)=\frac{1}{n^2}\left(\sum_{t=1}^{n} Var(X_t)+2\sum_{1\le s<t\le n} Cov(X_s,X_t)\right).$$

From equation (2.1.8),

$$\sum_{1\le s<t\le n} Cov(X_s,X_t)=-\frac{\lambda^2(N_0-X_0)}{(1-\lambda)^2}\sum_{1\le s<t\le n}(1-\lambda)^{s+t}.$$

Since

$$\sum_{1\le s<t\le n}(1-\lambda)^{s+t}=\left((1-\lambda)^3+(1-\lambda)^4+(1-\lambda)^5+\ldots+(1-\lambda)^{n+1}\right)$$

$$+\left((1-\lambda)^5+(1-\lambda)^6+(1-\lambda)^7+\ldots+(1-\lambda)^{n+2}\right)$$

$$+\ldots+\left((1-\lambda)^{2n-3}+(1-\lambda)^{2n-2}\right)+(1-\lambda)^{2n-1}$$

$$=\frac{1}{\lambda}\left((1-\lambda)^3+(1-\lambda)^5+(1-\lambda)^7+\ldots+(1-\lambda)^{2n-1}\right)$$

$$-\frac{1}{\lambda}\left((1-\lambda)^{n+2}+(1-\lambda)^{n+3}+(1-\lambda)^{n+4}+\ldots+(1-\lambda)^{2n}\right)$$

$$=\frac{(1-\lambda)^3}{\lambda^2(2-\lambda)}\left(1-(1-\lambda)^{n-1}-(1-\lambda)^n+(1-\lambda)^{2n-1}\right)$$

then

$$\sum_{1\le s<t\le n} Cov(X_s,X_t)=\frac{\lambda-1}{2-\lambda}(N_0-X_0)\left(1-(1-\lambda)^{n-1}-(1-\lambda)^n+(1-\lambda)^{2n-1}\right)$$

$$= \frac{\lambda-1}{2-\lambda}(N_0 - X_0)\left(1-(1-\lambda)^{n-1}\right)\left(1-(1-\lambda)^{n}\right).$$

From equation (2.1.4),

$$Var(X_t) = \mu - \mu(1-\lambda)^t + \lambda(N_0 - X_0)(1-\lambda)^{t-1} - \lambda^2(N_0 - X_0)(1-\lambda)^{2t-2}.$$

So

$$\sum_{t=1}^{n} Var(X_t) = n\mu - \frac{\mu(1-\lambda)}{\lambda}\left(1-(1-\lambda)^{n}\right) + (N_0 - X_0)\left(1-(1-\lambda)^{n}\right) - \frac{\lambda}{2-\lambda}(N_0 - X_0)\left(1-(1-\lambda)^{2n}\right)$$

$$= n\mu - \frac{\mu(1-\lambda)}{\lambda}\left(1-(1-\lambda)^{n}\right) + \frac{N_0 - X_0}{2-\lambda}\left(1-(1-\lambda)^{n}\right)\left(2-2\lambda-\lambda(1-\lambda)^{n}\right).$$

Now

$$Var(\overline{X}_n) = \frac{1}{n^2}\left[n\mu - \frac{\mu(1-\lambda)}{\lambda}\left(1-(1-\lambda)^{n}\right) + (N_0 - X_0)(1-\lambda)^{n}\left(1-(1-\lambda)^{n}\right)\right]$$

$$= \frac{1}{n^2}\left[n\mu - \left(1-(1-\lambda)^{n}\right)\left((N_0 - X_0)(1-\lambda)^{n} + \mu - \frac{\mu}{\lambda}\right)\right].$$

Note that

$$\lim_{n\to\infty} Var(\overline{X}_n) = 0.$$

For the average susceptible count we have

$$E(\overline{N}_n) = \frac{1}{n}\sum_{t=1}^{n} E(N_t)$$

$$= \frac{\mu}{\lambda} - \frac{1}{n\lambda}\left(1-(1-\lambda)^{n}\right)\left(\frac{\mu}{\lambda} - \mu - (N_0 - X_0)\right).$$

We note that as n tends to infinity, $E(\overline{N}_n)$ converges to $\frac{\mu}{\lambda}$. We now show that $Var(\overline{N}_n)$ converges to zero as n tends to infinity.

$$Var(\overline{N}_n) = \frac{1}{n^2}\left(\sum_{t=1}^{n} Var(N_t) + 2\sum_{1\le s<t\le n} Cov(N_s, N_t)\right).$$

Now, from equation (2.1.9),

$$\sum_{1\le s<t\le n} Cov(N_s, N_t) \le \left(\mu + \frac{\mu}{\lambda} + N_0 - X_0\right)\sum_{1\le s<t\le n}(1-\lambda)^{t-s}$$

which is equal to

$$\left(\mu + \frac{\mu}{\lambda} + N_0 - X_0\right)\left((n-1)(1-\lambda) + (n-2)(1-\lambda)^2 + \ldots + (1-\lambda)^{n-1}\right).$$

Therefore,

$$\sum_{1\le s<t\le n} Cov(N_s, N_t) \le n\left(\mu + \frac{\mu}{\lambda} + N_0 - X_0\right)\sum_{j=1}^{n-1}(1-\lambda)^j$$

$$= \frac{n(1-\lambda)}{\lambda}\left(\mu + \frac{\mu}{\lambda} + N_0 - X_0\right)\left(1 - (1-\lambda)^{n-1}\right). \qquad (2.2.1)$$

Thus, as n tends to infinity,

$$\frac{1}{n^2}\sum_{1\le s<t\le n} Cov(N_s, N_t) \to 0.$$

Also, from equation (2.1.5),

$$\frac{1}{n^2}\sum_{t=1}^{n} Var(N_t) \le \frac{1}{n}\left(\mu + \frac{\mu}{\lambda} + N_0 - X_0\right) \to 0 \quad \text{as } n \to \infty.$$

Hence,

$$Var(\overline{N}_n) \to 0 .$$

For the average number of individuals escaping infection over the time points $t = 1, 2, 3, \ldots, n,$ we have

$$E(\overline{Y}_n) = \frac{1}{n}\sum_{t=1}^{n} E(Y_t)$$

$$= \frac{\mu(1-\lambda)}{\lambda} + \frac{1-\lambda}{n\lambda}\left(\mu - \frac{\mu}{\lambda} + (N_0 - X_0)\right)\left(1 - (1-\lambda)^n\right)$$

$$\to \frac{\mu(1-\lambda)}{\lambda} .$$

The variance of this average also converges to zero as n tends to infinity. To see this we again use

$$Var(\overline{Y}_n) = \frac{1}{n^2}\left(\sum_{t=1}^{n} Var(Y_t) + 2\sum_{1\le s<t\le n} Cov(Y_s, Y_t)\right).$$

By equation (2.1.10),

$$Cov(Y_s, Y_t) \le (1-\lambda)\left(\frac{\mu}{\lambda} + (N_0 - X_0)\right)(1-\lambda)^{t-s}$$

so, by the argument used to show

$$\frac{1}{n^2}\sum_{1\le s<t\le n} Cov(N_s, N_t) \to 0 ,$$

we also have

$$\frac{1}{n^2}\sum_{1\le s<t\le n} Cov(Y_s, Y_t) \to 0.$$

From equation (2.1.6),

$$\sum_{t=1}^{n} Var(Y_t) \le n\left(\frac{\mu}{\lambda} + (N_0 - X_0)\right).$$

Thus,

$$Var(\overline{Y}_n) \to 0.$$

2.2.2 Weak Laws of Large Numbers

We now show that $\overline{X}_n$ converges in probability to the migration rate μ. By Chebychev's inequality, for $\varepsilon > 0$,

$$P\left(|\overline{X}_n - \mu| > \varepsilon\right) \le \frac{1}{\varepsilon^2} E\left((\overline{X}_n - \mu)^2\right)$$

$$= \frac{1}{\varepsilon^2}\left(Var(\overline{X}_n) + \left(E(\overline{X}_n) - \mu\right)^2\right).$$

Since $Var(\overline{X}_n)$ and $E(\overline{X}_n) - \mu$ converge to zero then the result holds.

A weak law of large numbers also holds for $\{N_t : t = 1, 2, 3, \ldots\}$. That is, $\overline{N}_n$ converges in probability to $\frac{\mu}{\lambda}$. This is clearly true as $E(\overline{N}_n)$ converges to $\frac{\mu}{\lambda}$ and $Var(\overline{N}_n)$ converges to zero.

Since

$$\overline{Y}_n = \overline{N}_n - \overline{X}_n$$

then $\overline{Y}_n$ converges in probability to

$$\frac{\mu}{\lambda} - \mu = \frac{\mu(1-\lambda)}{\lambda}.$$

2.2.3 Strong Laws of Large Numbers

In this subsection we use the Martingale Convergence Theorem (see page 242 of Feller (1971)) to show that the weak laws of large numbers established in the previous subsection can be extended to strong laws of large numbers.

To prove this property for $\{X_t : t = 1, 2, 3, \ldots\}$ we consider $\{U_n : n = 1, 2, 3, \ldots\}$, where

$$U_n = \sum_{t=1}^{n} \tfrac{1}{t}\left(X_t - E(X_t)\right).$$

In U_n the summand has zero expectation. Furthermore,

$$E\left(U_n^2\right) = \sum_{t=1}^{n} \tfrac{1}{t^2} Var(X_t) + 2 \sum_{1 \le s < t \le n} \tfrac{1}{st} Cov(X_s, X_t).$$

From equation (2.1.8),

$$Cov(X_s, X_t) < 0$$

so

$$E\left(U_n^2\right) \le \sum_{t=1}^{n} \tfrac{1}{t^2} Var(X_t).$$

Equation (2.1.4) states that

$$Var(X_t) = \mu\left[1-(1-\lambda)^t\right] + \lambda(1-\lambda)^{t-1}\left[1-\lambda(1-\lambda)^{t-1}\right](N_0 - X_0),$$

which implies

$$Var(X_t) \le \mu + (N_0 - X_0).$$

Hence,

$$E\left(U_n^2\right) \le (\mu + (N_0 - X_0))\sum_{t=1}^{n}\frac{1}{t^2} \le \frac{\pi^2}{6}(\mu + (N_0 - X_0)).$$

This implies,

$$E\left((U_n - U_m)^2\right) \to 0 \qquad \text{as } m, n \to \infty.$$

By the Martingale Convergence Theorem (see pages 242 and 243 of Feller (1971)), $\{U_n : n = 1, 2, 3, \ldots\}$ converges almost surely. Therefore, by Kronecker's lemma (see page 239 of Feller (1971)),

$$\frac{1}{n}\sum_{t=1}^{n}\left(X_t - E(X_t)\right) \to 0 \qquad \text{almost surely.}$$

The result now follows since

$$\frac{1}{n}\sum_{t=1}^{n}E(X_t) \to \mu.$$

In order to establish the strong law of large numbers for $\{N_t : t = 1, 2, 3, \ldots\}$ we use $\{U_n' : n = 1, 2, 3, \ldots\}$, where

$$U_n' = \sum_{t=1}^{n} \frac{1}{t}\left(N_t - E(N_t)\right).$$

$$E\left((U_n')^2\right) = \sum_{t=1}^{n} \frac{1}{t^2} Var(N_t) + 2 \sum_{1 \le s < t \le n} \frac{1}{st} Cov(N_s, N_t).$$

By equation (2.1.5),

$$Var(N_t) \le \mu + \frac{\mu}{\lambda} + (N_0 - X_0)$$

so

$$\sum_{t=1}^{n} \frac{1}{t^2} Var(N_t)$$

is a convergent series.

Also, by equation (2.1.9),

$$\sum_{1 \le s < t \le n} \frac{1}{st} Cov(N_s, N_t) \le \left(\mu + \frac{\mu}{\lambda} + N_0 - X_0\right) \sum_{1 \le s < t \le n} \frac{(1-\lambda)^{t-s}}{st}$$

$$\le \left(\mu + \frac{\mu}{\lambda} + N_0 - X_0\right) \sum_{k=1}^{n-1} (1-\lambda)^k < \frac{1-\lambda}{\lambda}.$$

Hence,

$$E\left((U_n' - U_m')^2\right) \to 0 \qquad \text{as } m, n \to \infty.$$

Thus,

$$\frac{1}{n}\sum_{t=1}^{n}\left(N_t - E(N_t)\right) \to 0 \qquad \text{almost surely.}$$

Therefore, the result holds.

Since $\overline{X}_n$ converges to μ almost surely and $\overline{N}_n$ converges to $\frac{\mu}{\lambda}$ almost surely then

$$\overline{Y}_n = \overline{N}_n - \overline{X}_n$$

converges almost surely to $\frac{\mu(1-\lambda)}{\lambda}$.

2.3 Central Limit Theorems

We now combine the formulae derived in Section 2.1 with a result from the limit theory of martingales to prove Theorem 2.3. That is, we establish central limit theorems for

$$\sum_{t=1}^{n} X_t \,,\ \sum_{t=1}^{n} N_t \ \text{ and } \sum_{t=1}^{n} Y_t \,.$$

In particular,

$$\sum_{t=1}^{n} E(N_t - X_t) = \frac{n\mu(1-\lambda)}{\lambda} + o(n)$$

and

$$\sum_{j=1}^{n} Var(N_t - X_t) = o(n^2)$$

are key results used in establishing the asymptotic normality of these random variables.

We begin by re-stating the theorem.

Theorem 2.3 If $\mathcal{F}_t^*$ is the σ-field generated by $\{N_0, X_0, N_1, X_1, \ldots, N_t, X_t\}$ then, for the model described in equation (1.4.2), we have the following central limit theorems.

(i) $$\frac{1}{\sqrt{n}}\left[\sum_{t=1}^{n} X_t - \sum_{t=1}^{n} E\left(X_t | \mathcal{F}_{t-1}^*\right)\right]$$

converges in distribution to a normal random variable with mean zero and variance $\mu(1-\lambda+\lambda^2)$.

(ii) $$\frac{1}{\sqrt{n}}\left[\sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} E\left(Y_t | \mathcal{F}_{t-1}^*\right)\right]$$

converges in distribution to a normal random variable with mean zero and variance $\mu(1-\lambda)(2-\lambda)$.

(iii) $$\frac{1}{\sqrt{n}}\left[\sum_{t=1}^{n} N_t - \sum_{t=1}^{n} E\left(N_t | \mathcal{F}_{t-1}^*\right)\right]$$

converges in distribution to a normal random variable with mean zero and variance μ.

The proofs of parts (i), (ii) and (iii) are given in Subsections 2.3.1, 2.3.2 and 2.3.3, respectively.

2.3.1 Proof of the Asymptotic Normality of $\overline{X}_n$

Consider the array $\{S_{ni}, \mathcal{F}_{ni}^* : n = 1, 2, 3, \ldots, 1 \le i \le n\}$, where $\mathcal{F}_{ni}^* = \mathcal{F}_i^*$ (the σ-field generated by $\{N_0, X_0, N_1, X_1, \ldots, N_i, X_i\}$) and

$$S_{ni} = \frac{1}{\sqrt{n}} \sum_{j=1}^{i} \left(X_j - E(X_j | \mathcal{F}^*_{nj-1}) \right).$$

This is a martingale array as

$$E\left(S_{ni} | \mathcal{F}^*_{ni-1}\right) = S_{ni-1} + \frac{1}{\sqrt{n}} E\left(X_i - E(X_i | \mathcal{F}^*_{i-1}) | \mathcal{F}^*_{i-1}\right)$$

$$= S_{ni-1}.$$

Furthermore, since $E\left(S_{n1} | \mathcal{F}^*_{n0}\right)$ is zero then it is a zero mean martingale array.

Let $\{W_{ni}: n = 1, 2, 3, \ldots, 1 \le i \le n\}$ be the martingale differences. That is, for $n = 1, 2, 3, \ldots,$

$$W_{n1} = S_{n1}$$

$$W_{ni} = S_{ni} - S_{ni-1} \qquad \text{for } 2 \le i \le n.$$

By Theorem 3.5 of Hall and Heyde (1980) (see page 71), S_{nn} converges in distribution to a normally distributed random variable with mean zero and variance $\mu(1 - \lambda + \lambda^2)$ when the following three conditions are satisfied.

(A) $$E\left(|V_n^2 - \mu(1 - \lambda + \lambda^2)|\right) \to 0,$$

where

$$V_n^2 = \sum_{j=1}^{n} E\left(W_{nj}^2 | \mathcal{F}^*_{nj-1}\right).$$

(B) $$\max_{1 \le j \le n} E\left(W_{nj}^2 | \mathcal{F}^*_{nj-1}\right) \to 0 \qquad \text{in probability.}$$

(C) For any $\varepsilon > 0$,

$$\sum_{j=1}^{n} E\left(W_{nj}^2 I\left(\{|W_{nj}| > \varepsilon\}\right)\right) \to 0,$$

where $I(.)$ is the indicator function.

We now show that each of these conditions is satisfied. Firstly, we show that

$$E\left(|V_n^2 - \mu(1-\lambda+\lambda^2)|\right) \to 0.$$

By definition,

$$V_n^2 = \sum_{j=1}^{n} E\left(\frac{1}{n}(X_j - E(X_j|\mathcal{F}_{j-1}^*))^2 | \mathcal{F}_{j-1}^*\right)$$

$$= \frac{1}{n}\sum_{j=1}^{n} Var\left(X_j|\mathcal{F}_{j-1}^*\right)$$

$$= \lambda\mu + \frac{\lambda(1-\lambda)}{n}\sum_{j=1}^{n}(N_{j-1} - X_{j-1}).$$

Hence,

$$s_n^2 \equiv E\left(V_n^2\right) = \lambda\mu + \frac{\lambda(1-\lambda)}{n}\sum_{j=1}^{n} E(N_{j-1} - X_{j-1})$$

$$= \lambda\mu + \frac{\lambda(1-\lambda)}{n}\sum_{j=1}^{n}\left[\frac{\mu(1-\lambda)}{\lambda}\left(1-(1-\lambda)^{j-1}\right) + (1-\lambda)^{j-1}(N_0 - X_0)\right]$$

$$= \lambda\mu + \frac{\lambda(1-\lambda)}{n}\left[\frac{\mu(1-\lambda)}{\lambda}\left(n - \frac{1-(1-\lambda)^n}{\lambda}\right) + \frac{1-(1-\lambda)^n}{\lambda}(N_0 - X_0)\right]$$

$$\rightarrow \lambda\mu + \mu(1-\lambda)^2 = \mu(1-\lambda+\lambda^2).$$

So to show condition (A) is satisfied we need to show $E\left(\left|V_n^2 - s_n^2\right|\right)$ converges to zero. We do this by showing

$$E\left(\left(V_n^2 - s_n^2\right)^2\right) \rightarrow 0.$$

Note that

$$E\left(\left(V_n^2 - s_n^2\right)^2\right) = Var\left(V_n^2\right)$$

$$= Var\left(\lambda\mu + \frac{\lambda(1-\lambda)}{n}\sum_{j=1}^{n-1}(N_j - X_j)\right)$$

$$= \frac{\lambda^2(1-\lambda)^2}{n^2} Var\left(\sum_{j=1}^{n-1}(N_j - X_j)\right).$$

From equation (2.1.6),

$$Var(N_j - X_j) \le (1-\lambda)\left(\frac{\mu}{\lambda} + N_0 - X_0\right) \qquad \text{for all } j.$$

From equation (2.1.10), for $j > i \ge 1$,

$$Cov(N_i - X_i, N_j - X_j) \le (1-\lambda)^{j-i+1}\left(\frac{\mu}{\lambda} + N_0 - X_0\right).$$

Hence,

$$E\left(\left(V_n^2 - s_n^2\right)^2\right) \le \frac{\lambda^2(1-\lambda)^2}{n^2}\left(\frac{\mu}{\lambda} + N_0 - X_0\right)\left((n-1)(1-\lambda) + 2(1-\lambda)\sum_{1\le i<j\le n-1}(1-\lambda)^{j-i}\right).$$

Now

$$\sum_{1\le i<j\le n-1}(1-\lambda)^{j-i} \le (n-2)\left((1-\lambda) + (1-\lambda)^2 + \ldots + (1-\lambda)^{n-2}\right)$$

so

$$E\left(\left(V_n^2 - s_n^2\right)^2\right) \le \frac{\lambda^2(1-\lambda)^3(n-1)}{n^2}\left(\frac{\mu}{\lambda} + N_0 - X_0\right)\left(1 + \frac{2(1-\lambda)}{\lambda}\left(1 - (1-\lambda)^{n-2}\right)\right)$$

$$\to 0\,.$$

For condition (B) which requires

$$\max_{1\le j\le n} E\left(W_{nj}^2 | \mathcal{F}_{nj\text{-}1}^*\right) \to 0 \qquad \text{in probability,}$$

we observe that

$$E\left(W_{nj}^2 | \mathcal{F}_{nj\text{-}1}^*\right) = \frac{1}{n} Var\left(X_j | \mathcal{F}_{j-1}^*\right) = \frac{1}{n}\left(\lambda\mu + \lambda(1-\lambda)(N_{j-1} - X_{j-1})\right).$$

Therefore, we need to show

$$\max_{1\le j\le n-1} \frac{N_j - X_j}{n} \to 0 \qquad \text{in probability.}$$

For $\varepsilon > 0$,

$$\mathrm{P}\left(N_j - X_j > n\varepsilon \text{ for some } 1 \le j \le n-1\right) = \mathrm{P}\left(\sum_{j=1}^{n-1}(N_j - X_j)I\left(\left\{N_j - X_j > n\varepsilon\right\}\right) > n\varepsilon\right)$$

$$\le \frac{E\left(\sum_{j=1}^{n-1}(N_j - X_j)I\left(\left\{N_j - X_j > n\varepsilon\right\}\right)\right)}{n\varepsilon}$$

$$\le \frac{\sum_{j=1}^{n-1}E\left((N_j - X_j)^2\right)}{n^2\varepsilon^2}$$

$$= \frac{\sum_{j=1}^{n-1}Var(N_j - X_j)}{n^2\varepsilon^2} + \frac{\sum_{j=1}^{n-1}\left(E(N_j - X_j)\right)^2}{n^2\varepsilon^2}$$

$$\le \frac{1}{n\varepsilon^2}\left[\left(\frac{\mu}{\lambda} + N_0 - X_0\right) + \left(\frac{\mu}{\lambda} + N_0 - X_0\right)^2\right]$$

$$\to 0.$$

Finally, we show that, for $\varepsilon > 0$,

$$\sum_{j=1}^{n}E\left(W_{nj}^2 I\left(\left\{|W_{nj}| > \varepsilon\right\}\right)\right) \equiv E\left(\sum_{j=1}^{n}\frac{(X_j - E(X_j|\mathcal{F}_{j-1}^*))^2}{n}I\left(\left\{|X_j - E(X_j|\mathcal{F}_{j-1}^*)| > \varepsilon\sqrt{n}\right\}\right)\right)$$

$$\to 0.$$

Now

$$E\left(\sum_{j=1}^{n}\frac{(X_j - E(X_j|\mathcal{F}_{j-1}^*))^2}{n}I\left(\left\{|X_j - E(X_j|\mathcal{F}_{j-1}^*)| > \varepsilon\sqrt{n}\right\}\right)\right)$$

$$= \frac{1}{n} \sum_{j=1}^{n} E\left[z_j^2 I\left(\left\{ |z_j| > \varepsilon \sqrt{n} \right\} \right) \right],$$

where

$$z_j = X_j - E(X_j | \mathcal{F}_{j-1}^{*}) .$$

We anticipate the result in Section 3.3 that $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ is an irreducible, aperiodic, positive recurrent Markov chain and so has a proper limiting distribution. The sequences $\{X_t: t = 0, 1, 2, \ldots\}$ and $\{Y_t: t = 0, 1, 2, \ldots\}$ both converge in distribution to Poisson random variables. This implies that, for $t = 0, 1, 2, \ldots$, all moments of X_t and Y_t are finite. Therefore, for $j = 1, 2, 3, \ldots$,

$$E\left(|z_j|^3 \right)$$

is also finite.

Hence, it follows that

$$E\left[\sum_{j=1}^{n} \frac{z_j^2}{n} I\left(\left\{ |z_j| > \varepsilon \sqrt{n} \right\} \right) \right] \leq \frac{1}{\varepsilon n \sqrt{n}} E\left[\sum_{j=1}^{n} |z_j|^3 \right]$$

$$\rightarrow 0 .$$

This completes the proof of part (i) of Theorem 2.3.

Note that since

$$\frac{V_n^2}{\mu(1-\lambda+\lambda^2)} \rightarrow 1 \qquad \text{in } L_1$$

then the convergence of

$$\frac{S_{nn}}{\sqrt{\mu(1-\lambda+\lambda^2}}$$

to a standard normal random variable is equivalent to the convergence (in distribution) of

$$\frac{S_{nn}}{\sqrt{V_n^2}}$$

to a standard normal random variable. That is, we have an analogous result where the norming factor, V_n^2, is a random variable. Furthermore, the randomness in V_n^2 is a function of the randomness in the history of the process. Recall that

$$V_n^2 = \lambda\mu + \frac{\lambda(1-\lambda)}{n}\sum_{j=1}^{n}(N_{j-1} - X_{j-1}).$$

Hence, the norming factor at time n is representative of the history of the process up to time $n-1$.

2.3.2 Proof of the Asymptotic Normality of $\overline{Y}_n$

We prove part (ii) of Theorem 2.3 in exactly the same way we proved part (i). In this case we use the zero mean martingale array $\{S'_{ni}, \mathcal{F}^*_{ni} : n = 1, 2, 3, \ldots, 1 \le i \le n\}$, where

$$S'_{ni} = \frac{1}{\sqrt{n}}\sum_{j=1}^{i}\left(Y_j - E(Y_j|\mathcal{F}^*_{nj-1})\right).$$

We denote the martingale differences by $\{W'_{ni} : n = 1, 2, 3, \ldots, 1 \le i \le n\}$. To establish this result we need to show the following conditions are satisfied.

(A′) $$E\left(\left|(V_n')^2 - \mu(1-\lambda)(2-\lambda)\right|\right) \to 0,$$

where

$$(V_n')^2 = \sum_{j=1}^{n} E\left((W_{nj}')^2 \middle| \mathcal{F}_{nj-1}^*\right).$$

(B′) $$\max_{1 \le j \le n} E\left((W_{nj}')^2 \middle| \mathcal{F}_{nj-1}^*\right) \to 0 \quad \text{in probability.}$$

(C′) For any $\varepsilon > 0$,

$$\sum_{j=1}^{n} E\left((W_{nj}')^2 I\left(\left\{|W_{nj}'| > \varepsilon\right\}\right)\right) \to 0.$$

Condition (A′)

$$(V_n')^2 = \frac{1}{n} \sum_{j=1}^{n} Var\left(Y_j \middle| \mathcal{F}_{j-1}^*\right)$$

$$= \mu(1-\lambda) + \frac{\lambda(1-\lambda)}{n} \sum_{j=1}^{n} (N_{j-1} - X_{j-1}).$$

Therefore,

$$(s_n')^2 \equiv E\left((V_n')^2\right) = \mu(1-\lambda) + \frac{\lambda(1-\lambda)}{n} \sum_{j=1}^{n} E(N_{j-1} - X_{j-1})$$

$$= \mu(1-\lambda) + \frac{\lambda(1-\lambda)}{n}\left[\frac{\mu(1-\lambda)}{\lambda}\left(n - \frac{1-(1-\lambda)^n}{\lambda}\right) + \frac{1-(1-\lambda)^n}{\lambda}(N_0 - X_0)\right]$$

$$\rightarrow \mu(1-\lambda)(2-\lambda).$$

The argument used to show $E\left(|V_n^2 - s_n^2|\right)$ converges to zero can also be used to show $E\left(|(V_n')^2 - (s_n')^2|\right)$ converges to zero. Therefore, condition (A′) is satisfied.

Condition (B′)

$$E\left((W_{nj}')^2 | \mathcal{F}_{nj-1}^*\right) = \frac{1}{n} Var\left(Y_j | \mathcal{F}_{j-1}^*\right) = \frac{1}{n}\left((1-\lambda)\mu + \lambda(1-\lambda)(N_{j-1} - X_{j-1})\right)$$

and we have already shown (when proving condition (B) holds) that

$$\max_{1 \le j \le n-1} \frac{N_j - X_j}{n} \rightarrow 0 \qquad \text{in probability.}$$

Condition (C′)

$$\sum_{j=1}^{n} E\left((W_{nj}')^2 I\left(\left\{|W_{nj}'| > \varepsilon\right\}\right)\right) = \frac{1}{n}\sum_{j=1}^{n} E\left[(z_j')^2 I\left(\left\{|z_j'| > \varepsilon\sqrt{n}\right\}\right)\right],$$

where

$$z_j' = Y_j - E(Y_j | \mathcal{F}_{j-1}^*).$$

As in Subsection 2.3.1, we anticipate from Section 3.3 that $\{Y_t : t = 0, 1, 2, \ldots\}$ converges in distribution to a Poisson random variable. This implies that, for $t = 0, 1, 2, \ldots$, all moments of Y_t are finite. Therefore, for $j = 1, 2, 3, \ldots$,

$$E\left(|z_j'|^3\right)$$

is also finite.

Thus, the following inequality completes the proof.

$$E\left[\sum_{j=1}^{n}\frac{(z_j')^2}{n}I\left(\left\{|z_j'|>\varepsilon\sqrt{n}\right\}\right)\right]\leq\frac{1}{\varepsilon n\sqrt{n}}E\left[\sum_{j=1}^{n}|z_j'|^3\right]$$

$$\rightarrow 0\,.$$

2.3.3 Proof of the Asymptotic Normality of $\overline{N}_n$

In this case we use the zero mean martingale array $\{S_{ni}'', \mathcal{F}_{ni}^*: n=1,2,3,\ldots, 1\leq i\leq n\}$, where

$$S_{ni}''=\frac{1}{\sqrt{n}}\sum_{j=1}^{i}\left(N_j-E(N_j|\mathcal{F}_{nj-1}^*)\right).$$

We denote the martingale differences by $\{W_{ni}'': n=1,2,3,\ldots,1\leq i\leq n\}$. To prove part (iii) of Theorem 2.3 we need to show the following conditions are satisfied.

(A″) $\qquad E\left(|(V_n'')^2-\mu|\right)\rightarrow 0\,,$

where

$$(V_n'')^2=\sum_{j=1}^{n}E\left((W_{nj}'')^2|\mathcal{F}_{nj-1}^*\right).$$

(B″) $\qquad \max_{1\leq j\leq n}E\left((W_{nj}'')^2|\mathcal{F}_{nj-1}^*\right)\rightarrow 0$ in probability.

(C″) For any $\varepsilon > 0$,

$$\sum_{j=1}^{n} E\left((W''_{nj})^2 I\left(\left\{|W''_{nj}| > \varepsilon\right\}\right)\right) \to 0.$$

Condition (A″)

$$(V''_n)^2 = \frac{1}{n}\sum_{j=1}^{n} Var\left(N_j | \mathcal{F}^*_{j-1}\right)$$

$$= \frac{1}{n}\sum_{j=1}^{n} \mu = \mu.$$

Therefore,

$$(s''_n)^2 \equiv E\left((V''_n)^2\right) = \mu.$$

Condition (B″)

$$E\left((W''_{nj})^2 | \mathcal{F}^*_{nj-1}\right) = \frac{1}{n} Var\left(N_j | \mathcal{F}^*_{j-1}\right) = \frac{\mu}{n} \to 0.$$

Condition (C″)

$$\sum_{j=1}^{n} E\left((W''_{nj})^2 I\left(\left\{|W''_{nj}| > \varepsilon\right\}\right)\right) = \frac{1}{n}\sum_{j=1}^{n} E\left[(z''_j)^2 I\left(\left\{|z''_j| > \varepsilon\sqrt{n}\right\}\right)\right],$$

where

$$z''_j = N_j - E(N_j | \mathcal{F}^*_{j-1}).$$

In order to show this condition holds we again use the result from Section 3.3 that

$\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ has a proper limiting distribution; with $\{Y_t: t = 0, 1, 2, \ldots\}$ and $\{N_t: t = 0, 1, 2, \ldots\}$ both converging (in distribution) to Poisson random variables. This implies that, for $j = 1, 2, 3, \ldots,$

$$E\left(|z_j''|^3\right)$$

is finite. It follows that

$$E\left[\sum_{j=1}^{n} \frac{(z_j'')^2}{n} I\left(\left\{|z_j''| \geq \varepsilon\sqrt{n}\right\}\right)\right] \leq \frac{1}{\varepsilon n\sqrt{n}} E\left[\sum_{j=1}^{n} |z_j''|^3\right]$$

$$\rightarrow 0 .$$

CHAPTER THREE

THE LIMITING BEHAVIOUR OF THE INFECTED AND SUSCEPTIBLE COUNTS

3.0 Introduction

This chapter examines the long run behaviour of X_t, N_t and Y_t (the number of individuals escaping infection at time t). We show that of these three random variables only Y_t forms a univariate Markov chain. A bivariate Markov chain is formed by any of the pairs (X_t, Y_t), (N_t, X_t) or (N_t, Y_t). Our attention will be focussed on determining the limiting distributions of $\{Y_t: t = 0, 1, 2, \ldots\}$ and $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$.

In Section 3.1 we derive the joint probability generating function for the pairs (X_t, Y_t), (N_t, X_t) and (N_t, Y_t) conditional on $\mathcal{I}_{t-1}^*$. These functions are then used in Section 3.2 and 3.3 to determine the limiting behaviour of $\{Y_t: t = 0, 1, 2, \ldots\}$ and $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$. In Section 3.2 we consider the Markov chain $\{Y_t: t = 0, 1, 2, \ldots\}$ and show that it can be regarded as the convolution of a branching process and a Poisson process. We use this description to determine its limiting behaviour. In fact, using the theory of Heathcote (1966) we show the existence of a proper limiting distribution for $\{Y_t: t = 0, 1, 2, \ldots\}$ is guaranteed. In Section 3.3 we use the limiting distribution of $\{Y_t: t = 0, 1, 2, \ldots\}$ to show $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ also has a proper limiting distribution. In Section 3.4 we give a discussion of results from the estimation theory of branching processes that are pertinent to $\{Y_t: t = 0, 1, 2, \ldots\}$. These results are then used to derive estimators for the infection rate λ and the migration rate μ.

We conclude this introduction with a summary of the main results presented in this chapter.

Theorem 3.1

The probability generating function $Y_t(z) = E\left[z^{Y_t}\right]$ for the process $\{Y_t: t = 0, 1, 2, \ldots\}$ is equal to

$$e^{-\frac{\mu}{\lambda}(1-\lambda)[1-(1-\lambda)^t](1-z)}[1-(1-\lambda)^t+(1-\lambda)^t z]^{Y_0}.$$

Consequently, $\{Y_t: t = 0, 1, 2, \ldots\}$ converges in distribution to Y_∞, where Y_∞ is a Poisson random variable with parameter

$$\frac{\mu(1-\lambda)}{\lambda}.$$

From this theorem we deduce that

$$P(Y_t = 0 | \mathcal{F}_0) = Y_t(0) = e^{-\frac{\mu}{\lambda}(1-\lambda)[1-(1-\lambda)^t]}[1-(1-\lambda)^t]^{Y_0}.$$

The event $Y_t = 0$ occurs when there are no susceptibles left to be infected. In this case, the epidemic will only be perpetuated if there is subsequent migration of individuals into the susceptible population.

Theorem 3.2

The bivariate Markov chain $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ has a proper limiting distribution (X_∞, Y_∞), whose probability generating function is given by

$$E\left[z_1^{X_\infty} z_2^{Y_\infty}\right] = e^{-\frac{\mu}{\lambda}(1-\lambda z_1 - z_2 + \lambda z_2)}.$$

Theorem 3.3

We consider the estimators $\overline{\lambda}$ and $\overline{\mu}$ of λ and μ defined by

$$\overline{\lambda} = \frac{(Y_0 - Y_T)\sum_{t=1}^{T} Y_{t-1} + T\left(\sum_{t=1}^{T} Y_t Y_{t-1} - \sum_{t=1}^{T} Y_{t-1}^2\right)}{\left(\sum_{t=1}^{T} Y_{t-1}\right)^2 - T\sum_{t=1}^{T} Y_{t-1}^2},$$

$$\overline{\mu} = \frac{\sum_{t=1}^{T} Y_{t-1} Y_t \sum_{t=1}^{T} Y_{t-1} - \sum_{t=1}^{T} Y_{t-1}^2 \sum_{t=1}^{T} Y_t}{\sum_{t=1}^{T} Y_t \sum_{t=1}^{T} Y_{t-1} - T\sum_{t=1}^{T} Y_{t-1} Y_t}.$$

The fact that the value of λ lies strictly between zero and one implies that both estimators are strongly consistent and their joint distribution is asymptotically normal. Another pair of strongly consistent estimators whose joint distribution is asymptotically normal is

$$\overline{\overline{\lambda}} = \frac{(Y_0 - Y_T + T)\sum_{t=1}^{T} \frac{1}{Y_{t-1}+1} + T\sum_{t=1}^{T} \frac{Y_t}{Y_{t-1}+1} - T^2}{\sum_{t=1}^{T} (Y_{t-1}+1)\sum_{t=1}^{T} \frac{1}{Y_{t-1}+1} - T^2},$$

$$\overline{\overline{\mu}} = \frac{\sum_{t=1}^{T} Y_{t-1} \sum_{t=1}^{T} \frac{Y_t}{Y_{t-1}+1} - \sum_{t=1}^{T} Y_t \sum_{t=1}^{T} \frac{Y_{t-1}}{Y_{t-1}+1}}{\sum_{t=1}^{T} Y_t \sum_{t=1}^{T} \frac{1}{Y_{t-1}+1} - T\sum_{t=1}^{T} \frac{Y_t}{Y_{t-1}+1}}.$$

These estimators have the advantage that their asymptotic properties do not depend on prior knowledge of the true value of λ. Therefore, they are applicable to a much broader class of practical problems.

Theorems 3.1, 3.2 and 3.3 are proved in Sections 3.2, 3.3 and 3.4, respectively.

3.1 Probability Generating Functions Associated with the Greenwood Model

We begin by noting that the joint probability density function of N_t and Y_t, given $\mathcal{I}_{t-1}^*$, is of a form which enables a very straightforward calculation of the corresponding probability generating function. The conditional density

$$P(N_t = n_t, Y_t = y_t \mid N_{t-1} = n_{t-1}, Y_{t-1} = y_{t-1})$$

is zero unless $0 \le y_t \le n_t$ and $y_{t-1} \le n_t$. Within this range, its value is given by

$$\binom{n_t}{y_t} p(t)^{n_t - y_t} (1 - p(t))^{y_t} \frac{e^{-\mu} \mu^{n_t - y_{t-1}}}{(n_t - y_{t-1})!}.$$

This expression is obtained from equation (1.4.1) by setting

$$\mu(t) = \mu$$

for all $t = 1, 2, 3, \ldots$.

Now specialising to the case of the Greenwood model, with $p(t) = \lambda$ for all t, and assuming a constant migration rate $\mu(t) \equiv \mu$, the joint conditional density becomes

$$P(N_t = n_t, Y_t = y_t \mid \mathcal{I}_{t-1}^*) = \binom{n_t}{y_t} \lambda^{n_t - y_t} (1-\lambda)^{y_t} \frac{e^{-\mu} \mu^{n_t - y_{t-1}}}{(n_t - y_{t-1})!}$$

$$\text{for} \quad 0 \le y_t \le n_t, \; y_{t-1} \le n_t$$

$$= 0 \qquad \text{otherwise.}$$

Hence,

$$E[z_1^{N_t} z_2^{Y_t} \mid \mathcal{I}_{t-1}^*] = E[z_1^{N_t} (\lambda + (1-\lambda) z_2)^{N_t} \mid \mathcal{I}_{t-1}^*]$$

$$= e^{-\mu} \sum_{n_t = y_{t-1}}^{\infty} \frac{(z_1[\lambda + (1-\lambda)z_2])^{n_t} \mu^{n_t - y_{t-1}}}{(n_t - y_{t-1})!}$$

$$= e^{-\mu} (z_1[\lambda + (1-\lambda)z_2])^{y_{t-1}} \sum_{n_t = y_{t-1}}^{\infty} \frac{(\mu z_1[\lambda + (1-\lambda)z_2])^{n_t - y_{t-1}}}{(n_t - y_{t-1})!}.$$

That is,

$$E[z_1^{N_t} z_2^{Y_t} \mid \mathcal{J}_{t-1}^*] = (z_1[\lambda + (1-\lambda)z_2])^{Y_{t-1}} e^{-\mu[1-\lambda z_1 - z_1 z_2 + \lambda z_1 z_2]}. \qquad (3.1.1)$$

We now derive the other two probability generating functions (for (X_t, Y_t) and (N_t, X_t)). The former is used in Section 3.3 to show that $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ has a proper limiting distribution. We include the probability generating function of (N_t, X_t) for completeness.

$$E[z_1^{N_t} z_2^{Y_t} \mid \mathcal{J}_{t-1}^*] = E[z_1^{X_t + Y_t} z_2^{Y_t} \mid \mathcal{J}_{t-1}^*] = E[z_1^{X_t} (z_1 z_2)^{Y_t} \mid \mathcal{J}_{t-1}^*].$$

Putting $z = z_1$ and $w = z_1 z_2$ in equation (3.1.1) we obtain

$$E[z^{X_t} w^{Y_t} \mid \mathcal{J}_{t-1}^*] = [\lambda z + (1-\lambda)w]^{Y_{t-1}} e^{-\mu[1-\lambda z - w + \lambda w]}.$$

Since

$$E[z_1^{N_t} z_2^{Y_t} \mid \mathcal{J}_{t-1}^*] = E[z_1^{N_t} z_2^{N_t - X_t} \mid \mathcal{J}_{t-1}^*] = E[(z_1 z_2)^{N_t} z_2^{-X_t} \mid \mathcal{J}_{t-1}^*]$$

then putting $u = z_1 z_2$ and $v = z_2^{-1}$ in equation (3.1.1) we have

$$E[u^{N_t} v^{X_t} \mid \mathcal{J}_{t-1}^*] = [u(\lambda v + 1 - \lambda)]^{Y_{t-1}} e^{-\mu(1-u+\lambda u - \lambda uv)}. \qquad (3.1.2)$$

The probability generating functions of the marginal distributions are also readily obtained from equation (3.1.1).

$$E[z^{N_t} \mid \mathcal{I}^*_{t-1}] = z^{Y_{t-1}} e^{-\mu(1-z)}. \tag{3.1.3}$$

$$E[z^{Y_t} \mid \mathcal{I}^*_{t-1}] = [\lambda + (1-\lambda)z]^{Y_{t-1}} e^{-\mu(1-\lambda)(1-z)}. \tag{3.1.4}$$

$$E[z^{X_t} \mid \mathcal{I}^*_{t-1}] = [1-\lambda+\lambda z]^{Y_{t-1}} e^{-\mu\lambda(1-z)}. \tag{3.1.5}$$

3.2 The Markov Chain $\{Y_t : t = 0, 1, 2, \ldots\}$

From equations (3.1.3), (3.1.4) and (3.1.5) we see that it is possible to calculate the probability that Y_t takes a particular value given the value of Y_{t-1}. However, this is not the case for X_t given X_{t-1} nor for N_t given N_{t-1}. That is, of $\{X_t : t = 0, 1, 2, \ldots\}$ $\{Y_t : t = 0, 1, 2, \ldots\}$ and $\{N_t : t = 0, 1, 2, \ldots\}$ only $\{Y_t : t = 0, 1, 2, \ldots\}$ forms a univariate Markov chain. Also note that equation (3.1.4) is a re-statement of the recurrence relation for the probability generating functions associated with $\{Y_t : t = 0, 1, 2, \ldots\}$ originally derived in Section 2.1

We assume Y_0 is a known constant and show that $\{Y_t : t = 0, 1, 2, \ldots\}$ can be regarded as the convolution of a Galton-Watson branching process and a Poisson process. Becker (1977) first applied branching processes to epidemiology when he used such a model for the number of infected individuals in the early stages of an epidemic. This analysis did not allow for the possibility of migration into the population at risk of infection.

This interpretation of $\{Y_t : t = 0, 1, 2, \ldots\}$ enables us to make use of the Theorem of Heathcote (1966) which guarantees the existence of a proper limiting distribution for this process. The form of this limiting distribution is also given. We point out that Heathcote's result was generalised slightly by Seneta (1969), however, we will not need this improved result for our purposes.

3.2.1 Interpreting $\{Y_t : t = 0, 1, 2. . .\}$ as a Branching Process with Immigration

We now show that $\{Y_t : t = 1, 2, 3, \ldots\}$ can be formulated as the convolution of a branching process and a Poisson process. Let U_t be the number of susceptible individuals migrating into the population at risk of infection during the time interval $(t-1, t]$ who do not become infected at time t. Let S_t be the number of individuals escaping infection at time t who also escaped infection at time $t-1$. That is, S_t is the number of individuals in the susceptible population at time $t-1$ who escape infection at both time $t-1$ and time t. For $t = 1, 2, 3, \ldots,$

$$Y_t = S_t + U_t .$$

We repeat that Y_0 is assumed to be a known constant.

Each individual escaping infection at time $t-1$ either escapes infection at time t, and so contributes to S_t, or becomes infected and does not contribute to S_t. Thus, we regard individuals comprising the count S_t as offspring of the Y_{t-1} individuals escaping infection at time $t-1$. That is,

$$S_t = \sum_{j=1}^{Y_{t-1}} A_j .$$

where A_j is the number of offspring produced by individual j. For this interpretation the offspring counts $A_1, A_2, \ldots, A_{Y_{t-1}}$ are independent Bernoulli random variables with parameter $1-\lambda$. That is, the offspring counts are independent and identically distributed random variables and if A denotes a random variable with their common distribution then

$$\begin{aligned} P(A = a) &= (1-\lambda)^a \lambda^{1-a} && \text{for } a = 0, 1 \\ &0 && \text{otherwise.} \end{aligned}$$

The probability generating function for the offspring distribution is

$$A(z) = \lambda + (1-\lambda)z .$$

Since migration counts during disjoint time intervals are independent and infected counts at distinct time points are also independent then $\{U_t : t = 1, 2, 3, \ldots\}$ is a sequence of independent random variables. Furthermore, the infection process is independent of time so $\{U_t : t = 1, 2, 3, \ldots\}$ is a sequence of independent and identically distributed random variables. Let U be a random variable with their common distribution. We subsequently show that U is Poisson with parameter $\mu(1-\lambda)$.

Thus, $\{Y_t : t = 1, 2, 3, \ldots\}$ is the sum of two independent processes; the branching process $\{S_t : t = 1, 2, 3, \ldots\}$ and the process $\{U_t : t = 1, 2, 3, \ldots\}$. For the branching process the offspring counts at time points $t = 1, 2, 3, \ldots$ are independent and identically distributed. This is also true of $U_1, U_2, U_3, \ldots$. That is, $\{Y_t : t = 1, 2, 3, \ldots\}$ together with the initial condition $Y_0 = y_0$ forms an example of what is known in the literature as a *branching process with immigration.* In this context, $\{U_t : t = 1, 2, 3, \ldots\}$ is the immigration process. The problem of estimating the mean of the offspring distribution and the mean of the immigration distribution (whose value in this case is $\mu(1-\lambda)$) has received considerable attention. Research in this area will be discussed in Section 3.4. The formulation of $\{Y_t : t = 0, 1, 2, \ldots\}$ as a branching process with immigration allows us to use the work of Heathcote (1966) and give a very simple derivation of the recurrence relation for the probability generating function of Y_t (originally derived in equation (2.1.11)).

From page 1 of Heathcote (1966) and the fact that $\{S_t : t = 1, 2, 3, \ldots\}$ and $\{U_t : t = 1, 2, 3, \ldots\}$ are independent we have

$$Y_t(z) = U(z)Y_{t-1}(\lambda + (1-\lambda)z), \qquad (3.2.1)$$

where $U(z)$ is the probability generating function for the random variable U.

For $t = 1, 2, 3, \ldots$, let M_t be the number of susceptible migrants entering the population at risk of infection during the time interval $(t-1, t]$ then $M_1, M_2, M_3, \ldots$, are independent and identically distributed Poisson random variables with parameter μ. For $k = 0, 1, 2, \ldots$,

$$P(U = k) = \sum_{m=0}^{\infty} P(M = m)P(U = k \mid M = m), \qquad (3.2.2)$$

where M is Poisson with parameter μ. Since the second term in this summation is zero for m strictly less than k then

$$P(U = k) = \sum_{m=k}^{\infty} \frac{e^{-\mu}\mu^m}{m!}\binom{m}{k}(1-\lambda)^k\lambda^{m-k}$$

$$= \sum_{m=k}^{\infty} \frac{e^{-\mu}\mu^m}{k!(m-k)!}(1-\lambda)^k\lambda^{m-k}$$

$$= \frac{e^{-\mu}[\mu(1-\lambda)]^k}{k!}\sum_{m=k}^{\infty}\frac{(\mu\lambda)^{m-k}}{(m-k)!}$$

$$= \frac{e^{-\mu(1-\lambda)}[\mu(1-\lambda)]^k}{k!}.$$

That is, U is Poisson with parameter $\mu(1-\lambda)$.

Therefore,

$$U(z) = e^{-\mu(1-\lambda)(1-z)}.$$

Substituting this value of $U(z)$ into equation (3.2.1) we obtain

$$Y_t(z) = e^{-\mu(1-\lambda)(1-z)}Y_{t-1}(\lambda + (1-\lambda)z),$$

which is equation (2.1.11).

The Theorem of Heathcote (1966) gives the following necessary and sufficient conditions for the existence of a proper limiting distribution for a branching process with immigration which forms an irreducible aperiodic Markov chain.

(i) The mean of the offspring distribution is strictly less than one.

(ii) $$\sum_{j=0}^{\infty} r_j (j+1)^{-1} < \infty,$$

where

$$r_j = \sum_{j+1}^{\infty} \mathrm{P}(U = k).$$

$\{Y_t : t = 0, 1, 2, \ldots\}$ is clearly an irreducible Markov chain on the non-negative integers so to show it is aperiodic it is sufficient to show that state zero is aperiodic. This follows from equation (3.1.4) which states,

$$\sum_{j=0}^{\infty} z^j P(Y_t = j \mid Y_{t-1} = 0) = e^{-\mu(1-\lambda)(1-z)}.$$

Whence,

$$P(Y_t = 0 \mid Y_{t-1} = 0) = e^{-\mu(1-\lambda)} \neq 0.$$

Condition (i) is clearly satisfied as the mean of the offspring distribution is $1-\lambda$ which, in any practical situation, is strictly less than one. For condition (ii) we note that

$$\sum_{j=0}^{\infty} r_j (j+1)^{-1} \leq \sum_{j=0}^{\infty} r_j = E(U) = \mu(1-\lambda).$$

Hence, $\{Y_t : t = 0, 1, 2, \ldots\}$ has a proper limiting distribution Y_∞.

The probability density of Y_∞ can be found by taking limits of both sides of equation (2.1.11). This gives

$$Y_\infty(z) = e^{-\mu(1-\lambda)(1-z)} Y_\infty(\lambda + (1-\lambda)z).$$

It is readily checked that the probability generating function

$$z \mapsto e^{-\frac{\mu}{\lambda}(1-\lambda)(1-z)}$$

is a solution of this equation. Thus, by the convergence property of probability generating functions,

$$Y_\infty(z) = e^{-\frac{\mu}{\lambda}(1-\lambda)(1-z)}. \qquad (3.2.3)$$

Since a probability generating function uniquely defines the associated probability density function then we conclude Y_∞ is Poisson with parameter

$$\frac{\mu(1-\lambda)}{\lambda}.$$

Alternatively, we can use equation (2.1.11) to obtain an expression for $Y_t(z)$ and take the pointwise limit as t goes to infinity. That is,

$$Y_\infty(z) = \lim_{t \to \infty} Y_t(z).$$

$Y_t(z)$ is given by iterating the recurrence relation

$$\begin{aligned} Y_t(z) &= U(z) Y_{t-1}(A(z)). \\ &= U(z) U(A(z) Y_{t-2}(A_2(z)), \end{aligned}$$

where

$$A_2(z) = A(A(z)).$$

In general, let $A_k(z)$ be the k-fold composition of the function $A(z)$ with itself. That is,

$$A_k(z) = A(A_{k-1}(z)) = A(A(A_{k-2}(z)))$$

Then

$$Y_t(z) = U(z)U(A(z)) \times \ldots \times U(A_{t-1}(z))Y_0(A_t(z)).$$

Now,

$$\begin{aligned} A_2(z) &= A(\lambda + (1-\lambda)z) \\ &= \lambda + (1-\lambda)[\lambda + (1-\lambda)z] \\ &= \lambda + (1-\lambda)\lambda + (1-\lambda)^2 z. \end{aligned}$$

Similarly,

$$\begin{aligned} A_3(z) &= \lambda[1 + (1-\lambda) + (1-\lambda)^2] + (1-\lambda)^3 z \\ &= [1 - (1-\lambda)^3] + (1-\lambda)^3 z. \end{aligned}$$

In general,

$$A_t(z) = [1 - (1-\lambda)^t] + (1-\lambda)^t z.$$

Thus,

$$\begin{aligned} Y_t(z) &= e^{-\mu(1-\lambda)(1-z)[1+(1-\lambda)+\ldots+(1-\lambda)^{t-1}]} Y_0(A_t(z)) \\ &= e^{-\frac{\mu}{\lambda}(1-\lambda)[1-(1-\lambda)^t](1-z)} (A_t(z))^{Y_0} \\ &= e^{-\frac{\mu}{\lambda}(1-\lambda)[1-(1-\lambda)^t](1-z)} [1 - (1-\lambda)^t + (1-\lambda)^t z]^{Y_0}. \end{aligned}$$

In particular,

$$P(Y_t = 0 | \mathcal{F}_0) = Y_t(0) = e^{-\frac{\mu}{\lambda}(1-\lambda)[1-(1-\lambda)^t]}[1-(1-\lambda)^t]^{Y_0}. \qquad (3.2.4)$$

This is the probability that all members of the susceptible population at time t become infected at time t. We remark that, in the absence of immigration, this would mean that the epidemic had come to an end.

Finally, we note that,

$$Y_\infty(z) = \lim_{t\to\infty} Y_t(z) = e^{-\frac{\mu}{\lambda}(1-\lambda)(1-z)},$$

which agrees with equation (3.2.3).

3.3 The Limiting Behaviour of the Markov Chain $\{(X_t, Y_t) : t = 0, 1, 2, \ldots\}$

It is clear that $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ is a Markov chain on the Cartesian product of the set of non-negative integers with itself. In Subsection 3.3.1 we show that this Markov chain is irreducible, aperiodic and positive recurrent. Therefore, it has a proper limiting distribution (X_∞, Y_∞) which we calculate in Subsection 3.3.2. Of course, the moments of X_∞ and $N_\infty \equiv X_\infty + Y_\infty$ are readily calculated from the joint probability generating function of X_∞ and Y_∞. In Subsection 3.3.3 we show that deriving the moments in this way gives results which are consistent with those given in Chapter Two.

3.3.1 Properties of $\{(X_t, Y_t) : t = 0, 1, 2, \ldots\}$

Since this chain is irreducible then all states are of the same type. Hence, to show all states are aperiodic and positive recurrent it is sufficient to show that any particular state has these properties. The state $(0, 0)$ is convenient to work with.

For $t = 1, 2, 3, \ldots,$

$$P(X_t = 0, Y_t = 0 | X_0 = 0, Y_0 = 0) = P(X_t = 0, Y_t = 0 | Y_0 = 0).$$

Since X_t and Y_t are non-negative random variables then this is equal to

$$P(N_t = 0 | Y_0 = 0).$$

Now

$$P(N_t = 0 | Y_0 = 0) = P(Y_{t-1} = 0 | Y_0 = 0) P(M_t = 0),$$

where M_t is the number of individuals entering the susceptible population during the time interval $(t-1, t]$.

From equation (3.2.4),

$$P(N_t = 0 | Y_0 = 0) = e^{-\frac{\mu}{\lambda}[1-(1-\lambda)^t]}.$$

Thus, as t tends to infinity

$$P(N_t = 0 | Y_0 = 0) \to e^{-\frac{\mu}{\lambda}},$$

which implies that state $(0, 0)$ is positive recurrent. Hence, we have shown $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$ is an irreducible, aperiodic, positive recurrent Markov chain.

3.3.2 The Limiting Behaviour of $\{(X_t, Y_t) : t = 0, 1, 2, \ldots\}$

Since $\{(X_t, Y_t) : t = 0, 1, 2, \ldots\}$ is irreducible, aperiodic and positive recurrent then (for example, by Theorem III.2.2 of Isaacson and Madsen (1976)) it has a proper limiting distribution, (X_∞, Y_∞), which is characterised as the unique solution to

$$P(X_\infty = j_1, Y_\infty = j_2) = \sum_{i_1, i_2 = 0}^{\infty} P(X_\infty = i_1, Y_\infty = i_2) P(X_t = j_1, Y_t = j_2 \mid X_{t-1} = i_1, Y_{t-1} = i_2)$$

which also satisfies

$$\sum_{j_1, j_2 = 0}^{\infty} P(X_\infty = j_1, Y_\infty = j_2) = 1 \ .$$

The first of these conditions implies $E[{z_1}^{X_\infty} {z_2}^{Y_\infty}]$ is equal to

$$\sum_{j_1, j_2 = 0}^{\infty} {z_1}^{j_1} {z_2}^{j_2} \sum_{i_1, i_2 = 0}^{\infty} P(X_\infty = i_1, Y_\infty = i_2) P(X_t = j_1, Y_t = j_2 \mid X_{t-1} = i_1, Y_{t-1} = i_2) .$$

By Fubini's Theorem, this is equal to

$$\sum_{i_1, i_2 = 0}^{\infty} P(X_\infty = i_1, Y_\infty = i_2) \sum_{j_1, j_2 = 0}^{\infty} {z_1}^{j_1} {z_2}^{j_2} P(X_t = j_1, Y_t = j_2 \mid X_{t-1} = i_1, Y_{t-1} = i_2) .$$

Thus we have

$$E[{z_1}^{X_\infty} {z_2}^{Y_\infty}] = \sum_{i_1, i_2 = 0}^{\infty} P(X_\infty = i_1, Y_\infty = i_2) \{\lambda z_1 + (1 - \lambda) z_2\}^{i_2} e^{-\mu(1 - \lambda z_1 - z_2 + \lambda z_2)}$$

$$= e^{-\mu(1 - \lambda z_1 - z_2 + \lambda z_2)} \sum_{i_2 = 0}^{\infty} \sum_{i_1 = 0}^{\infty} P(X_\infty = i_1, Y_\infty = i_2) \{\lambda z_1 + (1 - \lambda) z_2\}^{i_2}$$

$$= e^{-\mu(1-\lambda z_1 - z_2 + \lambda z_2)} \sum_{i_2=0}^{\infty} \mathrm{P}(Y_\infty = i_2)\{\lambda z_1 + (1-\lambda) z_2\}^{i_2}$$

$$= e^{-\mu(1-\lambda z_1 - z_2 + \lambda z_2)} \, E[\lambda z_1 + (1-\lambda) z_2]^{Y_\infty} .$$

From equation (3.2.3) we obtain

$$E[{z_1}^{X_\infty} {z_2}^{Y_\infty}] = e^{-\mu(1-\lambda z_1 - z_2 + \lambda z_2)} \, e^{-\mu(\lambda^{-1}-1)(1-\lambda z_1 - z_2 + \lambda z_2)}$$

$$= e^{-\frac{\mu}{\lambda}(1-\lambda z_1 - z_2 + \lambda z_2)} .$$

Putting $z_1 = z_2 = 1$ gives

$$\sum_{i_1, i_2=0}^{\infty} \mathrm{P}(X_\infty = i_1, Y_\infty = i_2) = 1 .$$

Thus the probability distribution with probability generating function

$$(z_1, z_2) \mapsto e^{-\frac{\mu}{\lambda}(1-\lambda z_1 - z_2 + \lambda z_2)} . \qquad (3.3.1)$$

is the proper limiting distribution of $\{(X_t, Y_t): t = 0, 1, 2, \ldots\}$.

3.3.3 The Limiting Behaviour of the Associated Marginal Distributions

We now derive the distributions of X_∞ and $N_\infty = X_\infty + Y_\infty$ from the probability generating function of the joint distribution of X_∞ and Y_∞. We compare the means and variances of these quantities with the limiting values of the means and variances of X_t, Y_t and N_t given in Chapter Two. Recall that in Section 2.1 we showed that

$$\lim_{t\to\infty} E[X_t] = \lim_{t\to\infty} Var[X_t] = \mu . \qquad (3.3.2)$$

$$\lim_{t\to\infty} E[Y_t] = \lim_{t\to\infty} Var[X_t] = \frac{\mu}{\lambda}(1-\lambda). \qquad (3.3.3)$$

and

$$\lim_{t\to\infty} E[N_t] = \lim_{t\to\infty} Var[N_t] = \frac{\mu}{\lambda}. \qquad (3.3.4)$$

As a check on the consistency of our calculations we show that these equations are also satisfied by the moments of the corresponding limiting distributions.

Putting $z_2 = 1$ in the probability generating function of (X_∞, Y_∞), given in equation (3.3.1), we obtain

$$E[z_1^{X_\infty}] = e^{-\mu(1-z_1)}$$

so X_∞ is Poisson with parameter μ. That is, the distribution of X_∞ is exactly the same as that of the count of migrants entering the susceptible population during a unit time interval. In other words, at equilibrium, the count of new infecteds is 'balanced' by 'new' individuals migrating into the susceptible population.

Similarly, for Y_∞ we have

$$E[z_2^{Y_\infty}] = e^{-\frac{\mu}{\lambda}(1-\lambda)(1-z_2)}$$

so, as shown in Subsection 3.2.1, Y_∞ is Poisson with parameter $\frac{\mu}{\lambda}(1-\lambda)$. Putting $z_1 = z_2 = z$ in the probability generating function of (X_∞, Y_∞) we have

$$E[z^{N_\infty}] = e^{-\frac{\mu}{\lambda}(1-z)}$$

so N_∞ is Poisson with parameter $\frac{\mu}{\lambda}$.

Clearly equations (3.3.2), (3.3.3) and (3.3.4) are satisfied when the limits of the moments are replaced with the moments of the corresponding limiting distributions.

Note also that

$$Var\,[N_\infty] = Var\,[X_\infty] + Var[Y_\infty].$$

Hence, X_t and Y_t are asymptotically uncorrelated. This is consistent with equation (2.1.12) which states that

$$Cov[X_t, Y_t] = \lambda(1-\lambda)(Var[N_t] - E[N_t]).$$

Alternatively, the covariance of X_∞ and Y_∞ can be calculated from their joint probability generating function.

$$\frac{\partial^2}{\partial z_1 \partial z_2} E[z_1^{X_\infty} z_2^{Y_\infty}] = \frac{\partial^2}{\partial z_1 \partial z z_2}\left[e^{-\frac{\mu}{\lambda}(1-\lambda z_1 - z_2 + \lambda z_2)}\right]$$

$$= \frac{\mu^2}{\lambda}(1-\lambda) e^{-\frac{\mu}{\lambda}(1-\lambda z_1 - z_2 + \lambda z_2)}.$$

So $E[X_\infty Y_\infty] = \frac{\mu^2}{\lambda}(1-\lambda)$ and hence the covariance of X_∞ and Y_∞ is zero, as expected.

We remark that since X_∞ and Y_∞ are uncorrelated then the value of X_∞ provides no information about the value of Y_∞. That is, we can not predict Y_∞ from X_∞.

3.4 Branching Process Estimators for the Infection and Migration Rates

In Section 3.2 we showed that $\{Y_t : t = 0, 1, 2, \ldots\}$ can be formulated as a branching process with immigration. Y_0 is assumed to be a known constant and, for $t = 1, 2, 3, \ldots,$

Y_t is the sum of the offspring count S_t and immigrant count U_t For $\{S_t: t = 1, 2, 3, \ldots\}$ the offspring distribution is Bernoulli with parameter $1-\lambda$ and $\{U_t: t = 1, 2, 3, \ldots\}$ is a Poisson process with parameter $\mu(1-\lambda)$. Subsection 3.4.1 deals with the estimation theory for these processes and we obtain estimators for the mean of the offspring distribution $\theta_1 = 1-\lambda$ and the mean of the immigration process $\theta_2 = \mu(1-\lambda)$.

λ and μ are functions of θ_1 and θ_2 which have continuous partial derivatives of all orders on $R^2 \setminus (0, 0)$. In Subsection 3.4.2 we use this observation to derive estimators for these parameters from the estimators of θ_1 and θ_2. Results concerning the asymptotic behaviour of these estimators are also given.

Of course, these estimators assume we can observe $\{Y_t: t = 0, 1, 2, \ldots\}$ from time $t = 0$ to time $t = T$. That is, in addition to knowing N_0 and X_0 we must be able to observe the newly infected counts $X_1, X_2, \ldots, X_T$ and the susceptible population sizes $N_1, N_2, \ldots, N_T$. In contrast, in Chapter Four we give estimators of λ and μ based on just a realisation of $X_1, X_2, \ldots, X_T$ and the initial conditions $N_0 = n_0$ and $X_0 = x_0$.

3.4.1 A Least Squares Approach to Branching Processes with Immigration

Work on estimating the means of the offspring and immigration distributions dates back to Patankar and Bartlet. Their estimates, derived using a maximum likelihood approach, are discussed in Bartlet (1955). For our process $\{Y_t: t = 0, 1, 2, \ldots\}$, the mean of the offspring distribution is strictly less than one. In this case estimators of both means have usually been derived using the so-called *conditional least squares* approach. Hence, we will concentrate on estimators derived using this method.

From Subsection 3.2.1 we have, for $t = 1, 2, 3, \ldots$,

$$Y_t = S_t + U_t = \sum_{j=1}^{Y_{t-1}} A_j + U_t,$$

where $A_1, A_2, \ldots, A_{Y_{t-1}}$ are independent and identically distributed Bernoulli random variables with parameter $\theta_1 = 1-\lambda$ and U_t is a Poisson random variable with parameter $\theta_2 = \mu(1-\lambda)$. Hence, if $\mathcal{F}^*_{t-1}$ is the σ-field generated by $\{N_0, X_0, N_1, X_1, \ldots, N_{t-1}, X_{t-1}\}$ then

$$E[Y_t|\mathcal{F}^*_{t-1}] = \theta_1 Y_{t-1} + \theta_2 . \qquad (3.4.1)$$

Klimko and Nelson (1978) defined the conditional least squares estimators $\overline{\theta_1}$ and $\overline{\theta_2}$ as the values of θ_1 and θ_2 which minimise

$$\sum_{t=1}^{T} (Y_t - E[Y_t|\mathcal{F}^*_{t-1}])^2 .$$

That is,

$$\overline{\theta_1} = \frac{\sum_{t=1}^{T} Y_t \sum_{t=1}^{T} Y_{t-1} - T\sum_{t=1}^{T} Y_t Y_{t-1}}{\left(\sum_{t=1}^{T} Y_{t-1}\right)^2 - T\sum_{t=1}^{T} Y_{t-1}^2}$$

and

$$\overline{\theta_2} = \frac{\sum_{t=1}^{T} Y_t Y_{t-1} \sum_{t=1}^{T} Y_{t-1} - \sum_{t=1}^{T} Y_{t-1}^2 \sum_{t=1}^{T} Y_t}{\left(\sum_{t=1}^{T} Y_{t-1}\right)^2 - T\sum_{t=1}^{T} Y_{t-1}^2} .$$

On page 638 these authors show that $\overline{\theta_1}$ and $\overline{\theta_2}$ are strongly consistent and

$$\sqrt{T}\left(\overline{\theta_1} - \theta_1, \overline{\theta_2} - \theta_2\right).$$

is asymptotically normal with variance-covariance matrix expressed in terms of the infection and migration rates, λ and μ, as follows

$$\begin{bmatrix} \lambda^2(\mu^{-1}-1)+2\lambda & -\mu(1-\lambda)(2-\lambda) \\ -\mu(1-\lambda)(2-\lambda) & \frac{\mu(1-\lambda)}{\lambda}[\mu(1-\lambda)(2-\lambda)+\lambda] \end{bmatrix}.$$

Venkataraman (1982) uses the analogy between equation (3.4.1) and the first order autoregression model for time series to establish the asymptotic normality of

$$\sqrt{T}\left(\overline{Y}-E(Y_\infty), \overline{\theta_1}-\theta_1, \overline{\theta_2}-\theta_2\right),$$

where

$$\overline{Y}=\frac{1}{T}\sum_{t=1}^{T} Y_t .$$

This analogy was first discussed in Heyde and Seneta (1972).

The results of Klimko and Nelson (1978) and Venkataraman (1982) are only applicable when the mean of the offspring distribution is strictly less than one. Of course, this is sufficient for our purposes because $\theta_1 = 1-\lambda$ is always strictly less than one. However, Wei and Winnicki (1987), (1989) and (1990) were concerned with using conditional least squares estimators for the other two cases $\theta_1 \geq 1$.

In Wei and Winnicki (1987) and (1989) it is shown that $\overline{\theta_1}$ is strongly consistent when $\theta_1 > 1$ and weakly consistent when $\theta_1 = 1$. With the goal of obtaining unified results for the asymptotic properties of the estimators of both θ_1 and θ_2, Wei and Winnicki (1990) define alternative estimators $\overline{\overline{\theta_1}}$ and $\overline{\overline{\theta_2}}$. These estimators are derived using *weighted conditional least squares*; a technique introduced by Nelson (1980).

Multiplying both sides of equation (3.4.1) by the weight

$$\frac{1}{\sqrt{Y_{t-1}+1}}$$

we have

$$E\left[\frac{Y_t}{Y_{t-1}+1}\middle|\mathcal{F}_{t-1}^*\right]=\frac{\theta_1 Y_{t-1}}{\sqrt{Y_{t-1}+1}}+\frac{\theta_2}{\sqrt{Y_{t-1}+1}}.$$

$\overline{\overline{\theta_1}}$ and $\overline{\overline{\theta_2}}$ .are defined as the values of θ_1 and θ_2 which minimise

$$\sum_{t=1}^{T}\left[\frac{Y_t}{\sqrt{Y_{t-1}+1}}-\frac{\theta_1 Y_{t-1}}{\sqrt{Y_{t-1}+1}}-\frac{\theta_2}{\sqrt{Y_{t-1}+1}}\right]^2.$$

That is,

$$\overline{\overline{\theta_1}}=\frac{\left[\sum_{t=1}^{T}Y_t\sum_{t=1}^{T}\frac{1}{Y_{t-1}+1}-T\sum_{t=1}^{T}\frac{Y_t}{Y_{t-1}+1}\right]}{\sum_{t=1}^{T}(Y_{t-1}+1)\sum_{t=1}^{T}\frac{1}{Y_{t-1}+1}-T^2}$$

and

$$\overline{\overline{\theta_2}}=\frac{\left[\sum_{t=1}^{T}Y_{t-1}\sum_{t=1}^{T}\frac{Y_t}{Y_{t-1}+1}-\sum_{t=1}^{T}Y_t\sum_{t=1}^{T}\frac{Y_{t-1}}{Y_{t-1}+1}\right]}{\sum_{t=1}^{T}(Y_{t-1}+1)\sum_{t=1}^{T}\frac{1}{Y_{t-1}+1}-T^2}.$$

They also establish the following asymptotic properties; when $\theta_1<1$ both estimators are strongly consistent, when $\theta_1=1$ both estimators are weakly consistent and, for the case $\theta_1>1$, $\overline{\overline{\theta_1}}$ is strongly consistent and it is impossible to have a consistent estimator of θ_2. Furthermore, the asymptotic distributions of $\overline{\overline{\theta_1}}$ and $\overline{\overline{\theta_2}}$ in these three cases are given. In particular, they show that, for $\theta_1<1$, the joint distribution of

$$\left[\sum_{t=1}^{T}(Y_{t-1}+1)\right]^{1/2}\left(\overline{\theta_1}-\theta_1\right)$$

and

$$\left[\sum_{t=1}^{T}\frac{1}{Y_{t-1}+1}\right]^{1/2}\left(\overline{\theta_2}-\theta_2\right)$$

is asymptotically normal with variance-covariance matrix given by

$$V^{-1}W(V^{-1})^T,$$

where V and W can be expressed in terms of the infection rate λ, the migration rate μ and the moments of the limiting distribution Y_∞ of $\{Y_t: t = 0, 1, 2, \ldots\}$ as follows.

$$V=\begin{bmatrix} \dfrac{E(Y_\infty)}{\sqrt{E(Y_\infty)+1}} & \left(E\left[\dfrac{1}{Y_\infty+1}\right]\right)^{-1/2} \\ \dfrac{E\left[Y_\infty(Y_\infty+1)^{-1}\right]}{\sqrt{E(Y_\infty)+1}} & \left(E\left[\dfrac{1}{Y_\infty+1}\right]\right)^{1/2} \end{bmatrix}.$$

$$W=\begin{bmatrix} E[\lambda(1-\lambda)Y_\infty+\mu(1-\lambda)] & E\left[\dfrac{\lambda(1-\lambda)Y_\infty+\mu(1-\lambda)}{Y_\infty+1}\right] \\ E\left[\dfrac{\lambda(1-\lambda)Y_\infty+\mu(1-\lambda)}{Y_\infty+1}\right] & E\left[\dfrac{\lambda(1-\lambda)Y_\infty+\mu(1-\lambda)}{(Y_\infty+1)^2}\right] \end{bmatrix}.$$

In Section 3.2 we showed that Y_∞ is Poisson with parameter

$$\alpha=\frac{\mu(1-\lambda)}{\lambda}.$$

Therefore,

$$E\left[\frac{1}{Y_\infty+1}\right]=\sum_{x=0}^{\infty}\frac{e^{-\alpha}\alpha^x}{x!(x+1)}=\frac{1-e^{-\alpha}}{\alpha}$$

and

$$E\left[\frac{1}{(Y_\infty+1)^2}\right]=\sum_{x=0}^{\infty}\frac{e^{-\alpha}\alpha^x}{x!(x+1)^2}=\frac{e^{-\alpha}}{\alpha}\sum_{k=1}^{\infty}\frac{\alpha^k}{k!k}.$$

From page 61 of Hansen (1975)

$$K(\alpha)\equiv\sum_{k=1}^{\infty}\frac{\alpha^k}{k!k}=-\chi-\log\alpha+\int_{-\infty}^{\alpha}\frac{e^t}{t}dt=\int_{0}^{\alpha}\frac{e^t-1}{t}dt,$$

where χ is Euler's constant, that is,

$$\chi=\lim_{N\to\infty}\left[\sum_{j=1}^{N}j^{-1}-\log N\right].$$

Hence,

$$V=\begin{bmatrix}\frac{\alpha}{\sqrt{\alpha+1}} & \sqrt{\frac{\alpha}{1-e^{-\alpha}}}\\ \frac{\alpha-1+e^{-\alpha}}{\alpha\sqrt{\alpha+1}} & \sqrt{\frac{1-e^{-\alpha}}{\alpha}}\end{bmatrix}$$

and

$$W=\begin{bmatrix}\mu(1-\lambda)(2-\lambda) & \lambda\left[2-\lambda-\frac{\lambda}{\mu}+\frac{(\lambda-\mu)e^{-\alpha}}{\mu}\right]\\ \lambda\left[2-\lambda-\frac{\lambda}{\mu}+\frac{(\lambda-\mu)e^{-\alpha}}{\mu}\right] & \lambda\left[\frac{\lambda}{\mu}\left(1-e^{-\alpha}\right)+\left(1-\frac{\lambda}{\mu}\right)e^{-\alpha}K(\alpha)\right]\end{bmatrix}.$$

Finally, we show this particular V has a non-zero determinant, so its inverse exists.

$$\det V = \sqrt{\frac{\alpha\left(1-e^{-\alpha}\right)}{\alpha+1}} - \frac{\sqrt{\alpha}\left[1-\frac{1}{\alpha}\left(1-e^{-\alpha}\right)\right]}{\sqrt{(\alpha+1)\left(1-e^{-\alpha}\right)}}$$

$$= \frac{\left(1-e^{-\alpha}-\alpha e^{-\alpha}\right)}{\sqrt{\alpha(\alpha+1)\left(1-e^{-\alpha}\right)}} \neq 0.$$

3.4.2 Formulae and Properties for these Estimators

We now derive estimators $\left(\overline{\lambda},\overline{\mu}\right)$ from $\left(\overline{\theta_1},\overline{\theta_2}\right)$ and $\left(\overline{\overline{\lambda}},\overline{\overline{\mu}}\right)$ from $\left(\overline{\overline{\theta_1}},\overline{\overline{\theta_2}}\right)$. Since

$$\lambda = 1-\theta_1$$

and

$$\mu = \frac{\theta_2}{\theta_1}$$

then, using the formulae for $\overline{\theta_1}$ and $\overline{\theta_2}$, we have

$$\overline{\lambda} = \frac{(Y_0 - Y_T)\sum_{t=1}^{T} Y_{t-1} + T\left(\sum_{t=1}^{T} Y_t Y_{t-1} - \sum_{t=1}^{T} Y_{t-1}^2\right)}{\left(\sum_{t=1}^{T} Y_{t-1}\right)^2 - T\sum_{t=1}^{T} Y_{t-1}^2}$$

and

$$\overline{\mu} = \frac{\sum_{t=1}^{T} Y_{t-1} Y_t \sum_{t=1}^{T} Y_{t-1} - \sum_{t=1}^{T} Y_{t-1}^2 \sum_{t=1}^{T} Y_t}{\sum_{t=1}^{T} Y_t \sum_{t=1}^{T} Y_{t-1} - T\sum_{t=1}^{T} Y_{t-1} Y_t}.$$

Similarly, using the formulae for $\overline{\overline{\theta_1}}$ and $\overline{\overline{\theta_2}}$, we derive

$$\overline{\overline{\lambda}} = \frac{(Y_0 - Y_T + T)\sum_{t=1}^{T} \frac{1}{Y_{t-1}+1} + T\sum_{t=1}^{T} \frac{Y_t}{Y_{t-1}+1} - T^2}{\sum_{t=1}^{T}(Y_{t-1}+1)\sum_{t=1}^{T} \frac{1}{Y_{t-1}+1} - T^2}$$

and

$$\overline{\overline{\mu}} = \frac{\sum_{t=1}^{T} Y_{t-1}\sum_{t=1}^{T} \frac{Y_t}{Y_{t-1}+1} - \sum_{t=1}^{T} Y_t \sum_{t=1}^{T} \frac{Y_{t-1}}{Y_{t-1}+1}}{\sum_{t=1}^{T} Y_t \sum_{t=1}^{T} \frac{1}{Y_{t-1}+1} - T\sum_{t=1}^{T} \frac{Y_t}{Y_{t-1}+1}}.$$

We remark that since $\overline{\theta_1}$ and $\overline{\theta_2}$ are strongly consistent then $\overline{\lambda}$ and $\overline{\mu}$ are also strongly consistent. Similarly, $\overline{\overline{\lambda}}$ and $\overline{\overline{\mu}}$ are strongly consistent.

The asymptotic distributions of these estimators can be derived from the Taylor series expansions of the functions $g_1: R^2 \to R$ and $g_2: R \setminus \{0\} \times R \to R$ defined by

$$g_1(\theta_1, \theta_2) = 1 - \theta_1$$

and

$$g_2(\theta_1, \theta_2) = \frac{\theta_2}{\theta_1}.$$

Now

$$\sqrt{T}\begin{bmatrix} \overline{\lambda} - \lambda \\ \overline{\mu} - \mu \end{bmatrix} = \sqrt{T}\begin{bmatrix} g_1(\overline{\theta_1}, \overline{\theta_2}) - g_1(\theta_1, \theta_2) \\ g_2(\overline{\theta_1}, \overline{\theta_2}) - g_2(\theta_1, \theta_2) \end{bmatrix}$$

$$= \sqrt{T}\begin{bmatrix} \partial g_1/\partial\theta_1 & \partial g_1/\partial\theta_2 \\ \partial g_2/\partial\theta_1 & \partial g_2/\partial\theta_2 \end{bmatrix}\begin{bmatrix} \overline{\theta_1} - \theta_1 \\ \overline{\theta_2} - \theta_2 \end{bmatrix}$$

$$+\frac{\sqrt{T}}{2}\begin{bmatrix} \frac{\partial^2 g_1}{\partial\theta_1^2} & \frac{\partial^2 g_1}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 g_1}{\partial\theta_2^2} \\ \frac{\partial^2 g_2}{\partial\theta_1^2} & \frac{\partial^2 g_2}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 g_2}{\partial\theta_2^2} \end{bmatrix}\begin{bmatrix} (\overline{\theta_1} - \theta_1)^2 \\ (\overline{\theta_1} - \theta_1)(\overline{\theta_2} - \theta_2) \\ (\overline{\theta_2} - \theta_2)^2 \end{bmatrix} + \ldots$$

If we assume $\sqrt{T}(\overline{\theta_1} - \theta_1)^2$, $\sqrt{T}(\overline{\theta_1} - \theta_1)(\overline{\theta_2} - \theta_2)$ and $\sqrt{T}(\overline{\theta_2} - \theta_2)^2$ converge to zero in probability then, by Slutzky's Theorem, the asymptotic distribution of

$$\sqrt{T}(\overline{\lambda} - \lambda, \overline{\mu} - \mu)$$

is the same as that of

$$\sqrt{T}\begin{bmatrix} -1 & 0 \\ -\frac{\theta_2}{\theta_1^2} & \frac{1}{\theta_1} \end{bmatrix}\begin{bmatrix} \overline{\theta_1} - \theta_1 \\ \overline{\theta_2} - \theta_2 \end{bmatrix} = \sqrt{T}\begin{bmatrix} -1 & 0 \\ -\frac{\mu}{1-\lambda} & \frac{1}{1-\lambda} \end{bmatrix}\begin{bmatrix} \overline{\theta_1} - \theta_1 \\ \overline{\theta_2} - \theta_2 \end{bmatrix}.$$

That is, the asymptotic distribution is the bivaiate normal distribution with mean zero and variance-covariance matrix given by

$$\begin{bmatrix} -1 & 0 \\ -\frac{\mu}{1-\lambda} & \frac{1}{1-\lambda} \end{bmatrix}\begin{bmatrix} \lambda^2(\mu^{-1} - 1) + 2\lambda & -\mu(1-\lambda)(2-\lambda) \\ -\mu(1-\lambda)(2-\lambda) & \frac{\mu(1-\lambda)}{\lambda}[\mu(1-\lambda)(2-\lambda) + \lambda] \end{bmatrix}\begin{bmatrix} -1 & -\frac{\mu}{1-\lambda} \\ 0 & \frac{1}{1-\lambda} \end{bmatrix}.$$

Similarly, if

$$\left[\sum_{t=1}^{T}(Y_{t-1}+1)\right]^{1/2}\left(\overline{\overline{\theta_1}}-\theta_1\right)^2,$$

$$\left[\sum_{t=1}^{T}(Y_{t-1}+1)\sum_{t=1}^{T}\frac{1}{Y_{t-1}+1}\right]^{1/2}\left(\overline{\overline{\theta_1}}-\theta_1\right)\left(\overline{\overline{\theta_2}}-\theta_2\right)$$

and

$$\left[\sum_{t=1}^{T}\frac{1}{Y_{t-1}+1}\right]^{1/2}\left(\overline{\overline{\theta_2}}-\theta_2\right)^2$$

all converge to zero in probability then the joint distribution of

$$\left[\sum_{t=1}^{T}(Y_{t-1}+1)\right]^{1/2}\left(\overline{\overline{\lambda}}-\lambda\right)$$

and

$$\left[\sum_{t=1}^{T}\frac{1}{Y_{t-1}+1}\right]^{1/2}\left(\overline{\overline{\mu}}-\mu\right)$$

is asymptotically normal with mean zero and variance- covariance matrix

$$\begin{bmatrix} -1 & 0 \\ -\frac{\mu}{1-\lambda} & \frac{1}{1-\lambda} \end{bmatrix} V^{-1}W(V^{-1})^T \begin{bmatrix} -1 & -\frac{\mu}{1-\lambda} \\ 0 & \frac{1}{1-\lambda} \end{bmatrix}.$$

CHAPTER FOUR

USE OF THE DATA AUGMENTATION METHODS TO ESTIMATE INFECTION AND MIGRATION RATES

4.0 Introduction

In this chapter we consider the use of data augmentation methods to locate the mode of the likelihood

$$L(\lambda, \mu | N_0, X_0, X_1, \ldots, X_T).$$

We are concerned with the situation where the susceptible counts $N_1, N_2, \ldots, N_T$ are unobservable. The calculation of the likelihood is difficult since it involves summing the expression for the augmented likelihood

$$L(\lambda, \mu | N_0, X_0, X_1, \ldots, X_T, N_1, N_2, \ldots, N_T)$$

(given in equation (1.4.2)) over all possible values of the latent variables.

The *data augmentation methods* alleviate this difficulty by imputing for the latent variables so that the augmented likelihood can be used for making inference about the infection rate λ and the migration rate μ.

We give a general discussion of the *EM algorithm*, the *data augmentation algorithm*, *chained data augmentation* and the *Gibbs sampler*. These algorithms are discussed in Sections 4.1, 4.2.2, 4.2.3 and 4.2.4, respectively. When discussing each of these methods we will consider whether or not it is applicable to our particular problem. That

is, we decide whether each of these algorithms can be applied to the likelihood function specified in equation (1.4.2).

In Section 4.4 we give estimates of λ and μ. The data on which these estimates are based are monthly counts of diagnosed AIDS cases in Australia from January 1985 to December 1994 (see Appendix A). We emphasise that the use of AIDS data is for illustrative purposes only. In Section 1.1 we stated that chain binomial models are suitable for diseases where the incubation and latent periods are of low variability and the infectious period is short. Clearly, these assumptions do not apply to the AIDS epidemic.

4.1 The EM Algorithm

In essence, the EM algorithm imputes for the latent variables by taking the expectation of the augmented likelihood with respect to the joint conditional distribution of the latent variables given the data.

The EM algorithm requires two steps at each iteration; the E-step (E standing for expectation) and the M-step (M standing for maximisation). In discussing this algorithm we will use $\lambda^{(i)}$ and $\mu^{(i)}$ to denote the current estimates of λ and μ at the end of the i-th iteration.

In the E-step of the $(i+1)$-th iteration we calculate the expectation of

$$\log L(\lambda, \mu, N_1, N_2, \ldots, N_T \mid N_0, X_0, X_1, \ldots, X_T).$$

That is, we take the expectation of the log-likelihood of the unknowns, $\lambda, \mu, N_1, N_2, \ldots, N_T$, with respect to the joint conditional distribution of $N_1, N_2, \ldots, N_T$ given the data $N_0, X_0, X_1, \ldots, X_T$.

In the M-step of the $(i+1)$-th iteration we maximise the expectation calculated in the E-step with respect to λ and μ. The maximising values become the updated estimates and so are used in the E-step of the next iteration.

The EM algorithm increases the value of the likelihood

$$L(\lambda, \mu | N_0, X_0, X_1, \ldots, X_T)$$

at each iteration. That is,

$$L(\lambda^{(i+1)}, \mu^{(i+1)} | N_0, X_0, X_1, \ldots, X_T) \geq L(\lambda^{(i)}, \mu^{(i)} | N_0, X_0, X_1, \ldots, X_T).$$

Thus the EM algorithm converges to a local maximum of the likelihood function. For the algorithm to converge to the mode of the likelihood the initial values, $\lambda^{(0)}$ and $\mu^{(0)}$, must be sufficiently close to the mode. Otherwise, there is the possibility of convergence to some other point of local maxima.

Note that since

$$N_t = N_{t-1} - X_{t-1} + M_t$$

then calculating the expectation of the log-likelihood with respect to $(N_1, N_2, \ldots, N_T)$ will be problematic. The data augmentation algorithms, discussed in the next section, do not require the calculation of this expectation. Thus, we look at these algorithms as possible methods of estimating λ and μ.

4.2 The Data Augmentation Algorithms

In contrast to the EM algorithm, the methods discussed in this section are not aimed at approximating the mode of the likelihood but at approximating the shape of the entire function. These techniques were developed in the Bayesian context. If we have flat priors for our parameters λ and μ then the likelihood is proportional to the conditional probability density

$$P(\theta | N_0, X_0, X_1, \ldots, X_T),$$

where $\theta = (\lambda, \mu)$. This density is approximated by these data augmentation algorithms.

In this section we discuss the *data augmentation algorithm*, *chained data augmentation*, and the *Gibbs sampler*. Chained data augmentation is just a slight modification of the data augmentation algorithm and the Gibbs sampler extends chained data augmentation from the univariate to the multivariate case.

Our data consists of the newly infected counts, $X_1, X_2, \ldots, X_T$, together with the initial conditions N_0 and X_0. So we will use $\mathbf{X}$ (slightly ambiguously) to denote the vector $(N_0, X_0, X_1, \ldots, X_T)$ and $\mathbf{N}$ to denote the vector of latent variables, $(N_1, N_2, \ldots, N_T)$. Throughout this section we closely follow Tanner (1993). Subsections 4.2.1, 4.2.2, 4.2.3 and 4.2.4 are based on Sections 3.3, 5.1, 6.1.1 and 6.1.2 of this text, respectively. We begin our discussion with a statement of the method of Monte Carlo which is central to the theory of the techniques discussed in this section.

4.2.1 Monte Carlo Methods

Suppose $\mathbf{x}$ and $\mathbf{y_0}$ are vectors, g is a probability density function and f is an arbitrary function. Consider the integral

$$J(\mathbf{y_0}) \equiv \int f(\mathbf{y_0}|\mathbf{x}) g(\mathbf{x}) d\mathbf{x} = \mathrm{E}_g[f(\mathbf{y_0}|\mathbf{x})].$$

Assuming this expectation is finite, this integral can be approximated using the following steps.

Step 1 - generate an independent and identically distributed sample, $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_m}$, from g.

Step 2 - approximate $J(\mathbf{y_0})$ by $\frac{1}{m}\sum_{j=1}^{m} f(\mathbf{y_0}|\mathbf{x_j})$.

If, for $j = 1, 2, 3, \ldots,$ we define the random variable Y_j by

$$Y_j = f(\mathbf{y_0}|\mathbf{x_j})$$

then the Y_j form a sequence of independent and identically distributed random variables with finite mean $J(\mathbf{y_0})$. Therefore, by the Law of Large Numbers, as m tends to infinity,

$$\frac{1}{m}\sum_{j=1}^{m} f(\mathbf{y_0}|\mathbf{x_j}) \to J(\mathbf{y_0}) \qquad \text{almost surely.}$$

This technique of estimation is known as the *method of Monte Carlo.*

In the special case that $\mathbf{y} \mapsto f(\mathbf{y}|\mathbf{x})$ is a probability density function for all $\mathbf{x}$,

$$J(\mathbf{y}) = \int f(\mathbf{y}|\mathbf{x})g(\mathbf{x})d\mathbf{x}$$

is a probability density function. We may obtain an independent and identically distributed sample of size m from this density as follows.

Step 1 - generate an independent and identically distributed sample, $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_m}$, from g.

Step 2 - for each $j = 1, 2, 3, \ldots, m$, generate $\mathbf{y_j}$ from the probability density $f(\mathbf{y}|\mathbf{x_j})$.

The pairs $(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), (\mathbf{x_3}, \mathbf{y_3}), \ldots, (\mathbf{x_m}, \mathbf{y_m})$ form an independent and identically distributed sample from the joint density. Hence, $\{\mathbf{y_1}, \mathbf{y_2}, \mathbf{y_3}, \ldots, \mathbf{y_m}\}$ is an independent and identically distributed sample from the marginal distribution J. This technique is known as the *method of composition.*

4.2.2 The Data Augmentation Algorithm

This method is based on the following two identities

$$P(\theta|\mathbf{X}) = \int P(\theta|\mathbf{X},\mathbf{N})P(\mathbf{N}|\mathbf{X})d\mathbf{N} \qquad (4.2.1)$$

and

$$P(\mathbf{N}|\mathbf{X}) = \int P(\mathbf{N}|\phi,\mathbf{X})P(\phi|\mathbf{X})d\phi\,. \qquad (4.2.2)$$

Equation (4.2.1) is known as the *posterior identity* and equation (4.2.2) is known as the *predictive identity*. If we substitute equation (4.2.2) into equation (4.2.1) we have

$$P(\theta|\mathbf{X}) = \int_{\mathbf{N}}\int_{\phi} P(\theta|\mathbf{X},\mathbf{N})P(\mathbf{N}|\phi,\mathbf{X})P(\phi|\mathbf{X})d\phi d\mathbf{N}\,.$$

Interchanging the order of integration gives

$$P(\theta|\mathbf{X}) = \int_{\phi}\int_{\mathbf{N}} P(\theta|\mathbf{X},\mathbf{N})P(\mathbf{N}|\phi,\mathbf{X})d\mathbf{N}.P(\phi|\mathbf{X})d\phi\,.$$

That is, the density $P(\theta|\mathbf{X})$ is a solution of the integral equation

$$g(\theta) = \int K(\theta,\phi)g(\phi)d\phi\,, \qquad (4.2.3)$$

where

$$K(\theta,\phi) = \int P(\theta|\mathbf{X},\mathbf{N})P(\mathbf{N}|\phi,\mathbf{X})d\mathbf{N}\,.$$

Thus, the solution of the integral equation (4.2.3) is the (unique) limiting distribution of the Markov chain with transition function $K(\theta,\phi)$. The existence of this limiting distribution is guaranteed under the following regularity conditions.

(i) $K(\theta,\phi)$ is uniformly bounded.

(ii) The mapping $\phi \mapsto K(\theta,\phi)$ is equicontinuous in θ. That is, given $\varepsilon > 0$ there exists $\delta_1(\varepsilon,\phi_0) > 0$ such that

$$\|\phi - \phi_0\| < \delta_1 \text{ implies } |K(\theta,\phi) - K(\theta,\phi_0)| < \varepsilon .$$

(iii) For any θ_0, there exists $\delta_2(\theta_0) > 0$ such that $\|\theta - \theta_0\| < \delta_2$ and $\|\phi - \theta_0\| < \delta_2$ together imply

$$K(\theta,\phi) > 0 .$$

In order to generate a realisation, $\{g_0(\theta), g_1(\theta), g_2(\theta), \ldots\}$, of this Markov chain we proceed iteratively. Given θ^* from $g_i(\theta)$, construct $g_{i+1}(\theta)$ as follows. First, sample from $P(\mathbf{N}|\mathbf{X})$ by applying the method of composition to the predictive identity. That is, generate $\mathbf{N}_1^*$ from $P(\mathbf{N}|\theta^*,\mathbf{X})$. Repeat this a further $m-1$ times to obtain

$$\mathbf{N}_1^*, \mathbf{N}_2^*, \mathbf{N}_3^*, \ldots, \mathbf{N}_m^* \quad \text{from} \quad P(\mathbf{N}|\theta^*,\mathbf{X}).$$

Now, following Tanner and Wong (1987), we apply the method of Monte Carlo to the posterior identity to obtain

$$g_{i+1}(\theta) = \frac{1}{m}\sum_{j=1}^{m} P(\theta|\mathbf{X},\mathbf{N}_j^*) .$$

At each iterate of this procedure we generate a value θ^* from the current approximation to the density $P(\theta|\mathbf{X})$. We obtain a geometric approximation to this density by smoothing a frequency histogram of these θ^* s.

Note that for this algorithm we need to calculate the joint conditional density of

$$\mathbf{N} = (N_1, N_2, N_3, \ldots, N_T)$$

given $\mathbf{X}$, λ and μ.

Since

$$N_t = N_{t-1} - X_{t-1} + M_t$$

then calculating this joint conditional density is difficult. Hence, this method is not applicable to our problem.

4.2.3 Chained Data Augmentation

The Gibbs sampler is a multivariate version of a special case of the data augmentation algorithm known as *chained data augmentation*. We discuss this algorithm as a way of introducing the Gibbs sampler.

In each iteration of the data augmentation algorithm we apply the method of composition to the predictive identity (m times) to generate m latent data patterns and then apply the method of Monte Carlo to the posterior identity. The chained data augmentation algorithm is the special case where $m = 1$.

In chained data augmentation the θ^* s which provide the approximation to the density $P(\theta|\mathbf{X})$ are generated as follows. Let θ_j^* be the value generated from the current estimate of $P(\theta|\mathbf{X})$. We generate one latent data pattern N_1^* from

$$P(\mathbf{N}|\theta_j^*, \mathbf{X}).$$

Now rather than apply the method of Monte Carlo to the posterior identity, as we did in Subsection 4.2.2, we sample directly from the updated estimate of $P(\theta|\mathbf{X})$ by applying the method of composition to the posterior identity. That is, we generate θ_{j+1}^* from the distribution

$$P(\theta|\mathbf{X}, \mathbf{N}_1^*).$$

We note that chained data augmentation also requires knowledge of the joint conditional density $P(\mathbf{N}|\theta, \mathbf{X})$. However, the Gibbs sampler, discussed in the next subsection, imputes the latent variables sequentially. That is, imputation is carried out using the densities

$$P(\mathbf{N}_i|\theta, \mathbf{X}, \mathbf{N} \setminus \{\mathbf{N}_i\}),$$

where $i = 1, 2, 3, \ldots, T$. In Section 4.3 it is shown that we can readily sample from each of these distributions. Thus, we have chosen to use the Gibbs sampler to estimate λ and μ.

4.2.4 The Gibbs Sampler

As mentioned in Subsection 4.2.3, the Gibbs sampler is a multivariate extension of chained data augmentation. In chained data augmentation the latent data is considered as a single vector **N**. That is, at each iterate, the latent variables are generated simultaneously. However, the Gibbs sampler generates each of the latent variables separately in a sequential manner.

Let $\psi = (\psi_1, \psi_2, \psi_3, \ldots, \psi_d)$ be a vector consisting of the latent variables **N** and the parameters λ and μ. At the end of the i-th iterate of the Gibbs sampler we have estimates,

$$\psi_1^{(i)}, \psi_2^{(i)}, \psi_3^{(i)}, \ldots, \psi_d^{(i)}.$$

In the $(i+1)$-th iterate of the Gibbs sampler these estimates are updated using the procedure outlined below.

(1) Generate $\psi_1^{(i+1)}$ from $P(\psi_1 | \psi_2^{(i)}, \psi_3^{(i)}, \psi_4^{(i)}, \ldots, \psi_d^{(i)}, \mathbf{X})$.

(2) Generate $\psi_2^{(i+1)}$ from $P(\psi_2 | \psi_1^{(i+1)}, \psi_3^{(i)}, \psi_4^{(i)}, \ldots, \psi_d^{(i)}, \mathbf{X})$.

Continue till the following step is completed.

(d) Generate $\psi_d^{(i+1)}$ from $P(\psi_d | \psi_1^{(i+1)}, \psi_2^{(i+1)}, \psi_3^{(i+1)}, \ldots, \psi_{d-1}^{(i+1)}, \mathbf{X})$.

The vectors $(\psi_1^{(i)}, \psi_2^{(i)}, \psi_3^{(i)}, \ldots, \psi_d^{(i)})$ are a realisation of the Markov chain with transition function

$$K(\psi, \psi') = P(\psi_1' | \psi_2, \psi_3, \psi_4, \ldots, \psi_d, \mathbf{X}) P(\psi_2' | \psi_1', \psi_3, \psi_4, \ldots, \psi_d, \mathbf{X})$$

$$\times \ldots \times P(\psi_d' | \psi_1', \psi_2', \psi_3', \ldots, \psi_{d-1}', \mathbf{X}).$$

Geman and Geman (1984) give conditions such that $\psi_j^{(i)}$ converges in distribution to ψ_j for all j.

4.3 Probability Densities of the Latent Variables

In Subsection 4.3.1 we prove the following theorem which provides us with the probability density functions needed to implement the Gibbs sampler.

Theorem 4.3 For the model described in equation (1.4.2):

(i) For $t = 1, 2, 3, \ldots, T-1$, the conditional distribution of $M_t = N_t - Y_{t-1}$ given $\mathbf{X} = (N_0, X_0, X_1, \ldots, X_T), N_{t-1}$ and N_{t+1} is binomial with parameters K_t and $\frac{1}{2}$, where

$$K_t = N_{t+1} - N_{t-1} + X_{t-1} + X_t.$$

(ii) For $M_T = N_T - Y_{T-1}$, $\mathrm{P}(M_T = m | \mathbf{X}, \mathbf{N} \setminus \{N_T\})$ is proportional to

$$\binom{Y_{T-1} + m}{X_T} \lambda^{X_T} (1-\lambda)^{Y_{T-1}+m-X_T} \frac{e^{-\mu}\mu^m}{m!},$$

where $m = 0, 1, 2, \ldots$.

From part (ii) of this theorem we see that the conditional distribution of M_T given $\mathbf{X}$ and N_{T-1} is not a standard distribution and in order to calculate this distribution explicitly we need the value of the sum

$$\sum_{m=0}^{\infty} \binom{Y_{T-1} + m}{X_T} \lambda^{X_T} (1-\lambda)^{Y_{T-1}+m-X_T} \frac{e^{-\mu}\mu^m}{m!}.$$

It can be shown that this sum is equal to

$$e^{-\mu} \binom{Y_{T-1}}{X_T} \lambda^{X_T} (1-\lambda)^{Y_{T-1}-X_T} H\big(Y_{T-1}+1, Y_{T-1}-X_T+1, \mu(1-\lambda)\big),$$

where H is the *confluent hypergeometric function* (see page 8 of Johnson and Kotz (1969)).

However, as mentioned above, we have chosen to use the Gibbs sampler. Therefore, we will need to be able to generate a value from this non-standard distribution. In order to

do this we use the *acceptance/rejection algorithm* (see, for example, page 34 of Tanner (1993)). A detailed discussion of this algorithm is given in Subsection 4.3.2.

4.3.1 Proof of Theorem 4.3

For $t = 1, 2, 3, \ldots, T-1$,

$$\begin{aligned} N_{t+1} &= N_t - X_t + M_{t+1} \\ &= N_{t-1} - X_{t-1} - X_t + M_t + M_{t+1}. \end{aligned}$$

That is,

$$M_t + M_{t+1} = N_{t+1} - N_{t-1} + X_{t-1} + X_t = K_t.$$

Now, for $t = 1, 2, 3, \ldots, T-1$,

$$\begin{aligned} \mathrm{P}(M_t = m | \mathbf{X}, \mathbf{N} \setminus \{N_t\}) &= \mathrm{P}(M_t = m | M_t + M_{t+1} = K_t) \\ &= \frac{\mathrm{P}(M_t = m, M_t + M_{t+1} = K_t)}{\mathrm{P}(M_t + M_{t+1} = K_t)}. \end{aligned}$$

Since M_t and M_{t+1} are independent Poisson random variables with parameter μ then $M_t + M_{t+1}$ is a Poisson random variable with parameter 2μ. Hence,

$$\mathrm{P}(M_t = m, M_t + M_{t+1} = K_t) = \frac{e^{-\mu}\mu^m}{m!} \times \frac{e^{-\mu}\mu^{K_t - m}}{(K_t - m)!}$$

and

$$\mathrm{P}(M_t + M_{t+1} = K_t) = \frac{e^{-2\mu}(2\mu)^{K_t}}{K_t!}.$$

Hence,

$$\mathrm{P}(M_t = m|\mathbf{X}, \mathbf{N} \setminus \{N_t\}) = \binom{K_t}{m}\left(\tfrac{1}{2}\right)^{K_t}.$$

(ii) $$\mathrm{P}(M_T = m|\mathbf{X}, \mathbf{N} \setminus \{N_T\}) = \mathrm{P}(M_T = m|\mathbf{X}, N_{T-1})$$

$$= \frac{\mathrm{P}(M_T = m, X_T = x_T|\mathbf{X} \setminus \{X_T\}, N_{T-1})}{\mathrm{P}(X_T = x_T|\mathbf{X} \setminus \{X_T\}, N_{T-1})},$$

which is proportional to

$$\binom{Y_{T-1}+m}{X_T}\lambda^{X_T}(1-\lambda)^{Y_{T-1}+m-X_T}\frac{e^{-\mu}\mu^m}{m!}.$$

4.3.2 Sampling $\mathbf{M_T}$ - The Acceptance/Rejection Algorithm

In this subsection we give a method of generating an observation from the conditional probability density of M_T given $\mathbf{X}$ and $\mathbf{N} \setminus \{N_T\}$. That is, from the non-standard probability density function

$$\frac{h(m)}{\sum_{m=0}^{\infty} h(m)},$$

where

$$h(m) = \begin{cases} \binom{Y_{T-1}+m}{X_T}\lambda^{X_T}(1-\lambda)^{Y_{T-1}+m-X_T}\dfrac{e^{-\mu}\mu^m}{m!} & m = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

The device used is the *acceptance/rejection algorithm.* From page 34 of Tanner (1993), this algorithm can be implemented if we can find a probability density $I(m)$ and a constant $M > 0$ such that, for $m = 0, 1, 2, \ldots,$

$$h(m) \leq MI(m).$$

Once M and the density $I(m)$ have been determined, the algorithm proceeds as follows.

(i) Generate m from $I(m)$.

(ii) Generate u from the uniform distribution on $[0, 1]$.

(iii) If

$$u \leq \frac{h(m)}{MI(m)}$$

then accept m as a value from the conditional distribution of M_T. Otherwise, return to Step (i) and repeat this procedure.

Of course, some work is required to determine the constant M and density $I(m)$. We do this by finding a probability density I whose mode is equal to m_h, the value of m which maximises $h(m)$, and whose slope is not as steep as the slope of h. That is, we require

$$\frac{h(m+1)}{h(m)} \leq \frac{I(m+1)}{I(m)} \qquad \text{for } m \geq m_h$$

and

$$\frac{h(m+1)}{h(m)} \geq \frac{I(m+1)}{I(m)} \qquad \text{for } m \leq m_h - 1.$$

For such a distribution I,

$$h(m) \leq MI(m),$$

where

$$M = \frac{h(m_h)}{I(m_h)}.$$

We begin by calculating the value of m_h.

$$\frac{h(m+1)}{h(m)} = \frac{(Y_{T-1} + m + 1)\beta}{(Y_{T-1} - X_T + m + 1)(m+1)},$$

where

$$\beta = \mu(1-\lambda).$$

Hence,

$$\frac{h(m+1)}{h(m)} \geq 1$$

is equivalent to

$$(Y_{T-1} + m + 1)\beta \geq (Y_{T-1} - X_T + m + 1)(m+1).$$

That is,

$$(m+1)^2 + (Y_{T-1} - X_T - \beta)(m+1) - \beta Y_{T-1} \leq 0.$$

This inequality is satisfied when

$$0 \leq m \leq -1 + \tfrac{1}{2}\left(-(Y_{T-1} - X_T - \beta) + \sqrt{(Y_{T-1} - X_T - \beta)^2 + 4\beta Y_{T-1}}\right).$$

Therefore,

$$m_h = \text{integer part of}\left[\tfrac{1}{2}\left(-(Y_{T-1} - X_T - \beta) + \sqrt{(Y_{T-1} - X_T - \beta)^2 + 4\beta Y_{T-1}}\right)\right].$$

The majorising distribution we have chosen for I is the negative binomial with parameters S (a real number) and p (from the interval $(0,1)$). Our aim is to choose S and p such that the above conditions (on mode and slope) are satisfied. For $m = 0, 1, 2, \ldots,$

$$I(m) = \binom{S+m-1}{S-1} p^S (1-p)^m.$$

It is readily checked that the mode of I occurs at the integer part of

$$(S-1)(p^{-1} - 1).$$

Thus, we can make the mode equal to m_h by choosing

$$S = 1 + \frac{pm_h}{1-p}. \tag{4.3.1}$$

We now compare the slope of h with that of I.

$$\frac{h(m+1)}{h(m)} = \frac{(Y_{T-1} + m + 1)\beta}{(Y_{T-1} - X_T + m + 1)(m+1)}$$

$$= \left(1 + \frac{X_T}{(Y_{T-1} - X_T + m + 1)}\right)\frac{\beta}{m+1}.$$

For I we have

$$\frac{I(m+1)}{I(m)} = \left(1 + \frac{S-1}{m+1}\right)(1-p).$$

For $m \geq m_h$, we require

$$\left(1+\frac{S-1}{m+1}\right)(1-p) \geq \left(1+\frac{X_T}{Y_{T-1}-X_T+m+1}\right)\frac{\beta}{m+1}.$$

Assuming equation (4.3.1) is satisfied, this inequality becomes

$$(m+1)(1-p)+pm_h \geq \beta+\frac{X_T\beta}{(Y_{T-1}-X_T+m+1)}.$$

For $m \geq m_h$, the left hand side of this inequality is bounded below by

$$(m_h+1)(1-p)+pm_h = m_h+1-p.$$

The right hand side is bounded above by

$$\beta+\frac{X_T\beta}{Y_{T-1}-X_T+m_h+1}.$$

Hence, p should be chosen such that

$$m_h+1-p \geq \beta+\frac{X_T\beta}{Y_{T-1}-X_T+m_h+1}.$$

That is,

$$p \leq m_h+1-\beta-\frac{X_T\beta}{Y_{T-1}-X_T+m_h+1}.$$

For $m \leq m_h-1$, we require

$$\frac{h(m+1)}{h(m)} \geq \frac{I(m+1)}{I(m)}.$$

Again assuming equation (4.3.1) is satisfied, this inequality becomes

$$\beta + \frac{X_T \beta}{(Y_{T-1} - X_T + m + 1)} \geq (m+1)(1-p) + pm_h .$$

The right hand side is bounded above by

$$m_h(1-p) + pm_h = m_h .$$

The left hand side is bounded below by

$$\beta + \frac{X_T \beta}{Y_{T-1} - X_T + m_h} .$$

Thus the slope condition is satisfied if

$$\beta \frac{Y_{T-1} + m_h}{Y_{T-1} - X_T + m_h} \geq m_h .$$

That is,

$$0 \geq m_h^2 + (Y_{T-1} - X_T - \beta) m_h - \beta Y_{T-1} .$$

This is true because, when calculating m_h, we showed it was the integer part of the solution of the corresponding equality.

Now we have

$$h(m) \leq MI(m) ,$$

where

$$M = \frac{h(m_h)}{I(m_h)}$$

and $I(m)$ is the negative binomial distribution with

$$p = m_h + 1 - \beta - \frac{X_T \beta}{Y_{T-1} - X_T + m_h + 1},$$

$$S = 1 + \frac{p}{1-p} m_h .$$

However, at present the value of S may not be a strictly positive integer. We can overcome this problem by setting

$$S' = \text{integer part of } (S+1).$$

Then

$$S' = S + \alpha$$

for some $0 < \alpha \leq 1$.

Let I' be the negative binomial distribution with parameters p and S'. Then, for $m = 0, 1, 2, \ldots,$

$$\frac{I(m)}{I'(m)} = p^{-\alpha} \frac{S+m+1}{S+\alpha+m+1} \times \frac{S+m}{S+\alpha+m} \times \ldots \times \frac{S}{S+\alpha}$$

$$\leq p^{-\alpha}.$$

Thus we have

$$h(m) \leq p^{-\alpha} \frac{h(m_h)}{I(m_h)} I'(m).$$

4.4 Estimation Using the Gibbs Sampler

In Subsection 4.4.1 we show approximations to the probability densities

$$P(\lambda | N_0, X_0, X_1, \ldots, X_T)$$

and

$$P(\mu | N_0, X_0, X_1, \ldots, X_T)$$

obtained by plotting values from 1,000 iterates of the Gibbs sampler. We estimate λ and μ as the modes of these distributions.

The data used was monthly counts of diagnosed AIDS cases in Australia from January 1985 to December 1994 (see Appendix A). As mentioned in Section 4.0, the use of AIDS data is strictly for illustrative purposes. The high variability of the incubation time of this disease and the fact that infected individuals usually remain infectious for long periods imply that a continuous time model should be used for this epidemic.

4.4.1 Discussion of the Estimates

Smoothed histograms of the values generated from 1,000 iterates of the Gibbs sampler (with starting values $\lambda^{(0)} = 0.1$, $\mu^{(0)} = 50$ and $N_t^{(0)} = N_0 = 540$ for $t = 1, 2, \ldots, 120$) for λ and μ are given in Figures 4.1 (page 117) and 4.2 (page 118), respectively.

We observe that the mode for λ is approximately 0.088 and the mode for μ is approximately 47.6.

The generated values for λ ranged from 0.083 to 0.092. For μ, the range was [45.7, 49.2]. In both cases the generated values were 'bunched' around the mode. For λ there were over 500 values within 0.001 of the mode, while for μ, there were 470

values within 0.5 of the mode. Hence, the HPD (highest posterior density) regions for both posterior densities will be very narrow.

Sensitivity of the mode with respect to the initial values $\lambda^{(0)}$ and $\mu^{(0)}$ was investigated using $N_t^{(0)} = 540$, for $t = 1, 2, \ldots, 120$, and the following values for $(\lambda^{(0)}, \mu^{(0)})$.

$(0.01, 50)$, $(0.02, 50)$, $(0.1, 45)$, $(0.1, 60)$, $(0.01, 45)$, $(0.01, 60)$, $(0.02, 45)$, $(0.2, 60)$. We found, with 1,000 iterates of the Gibbs sampler, that the mode of μ ranged between 47.5 and 47.8 while the mode of λ was always between 0.087 and 0.089. We then looked at the sensitivity with respect to the initial values $N_1^{(0)}, N_2^{(0)}, \ldots, N_{120}^{(0)}$. Several samples, of size 120, were generated from different Poisson distributions. The parameters of these distributions ranged from 500 to 600. With $\lambda^{(0)} = 0.1$, $\mu^{(0)} = 50$ and the initial values for the N_t generated in this fashion we again obtained estimates of λ between 0.087 and 0.089. The range for μ was [47.2, 47.8].

From equation (3.2.4), the probability of the epidemic coming to an end at time t is

$$e^{-\frac{\mu}{\lambda}(1-\lambda)[1-(1-\lambda)^t]}[1-(1-\lambda)^t]^{Y_0},$$

where Y_0 is the initial number of susceptible individuals. For fixed values of λ, μ and Y_0, this expression is bounded above by

$$e^{-\min\left\{\frac{\mu}{\lambda}(1-\lambda), Y_0\right\}}$$

for all non-negative t. Hence, our estimated values of λ and μ indicate a very low probability of the disease dying out. However, when drawing this conclusion we must keep in mind that we are using AIDS data only to illustrate our methods and that this epidemic does not satisfy the assumptions of a discrete time model.

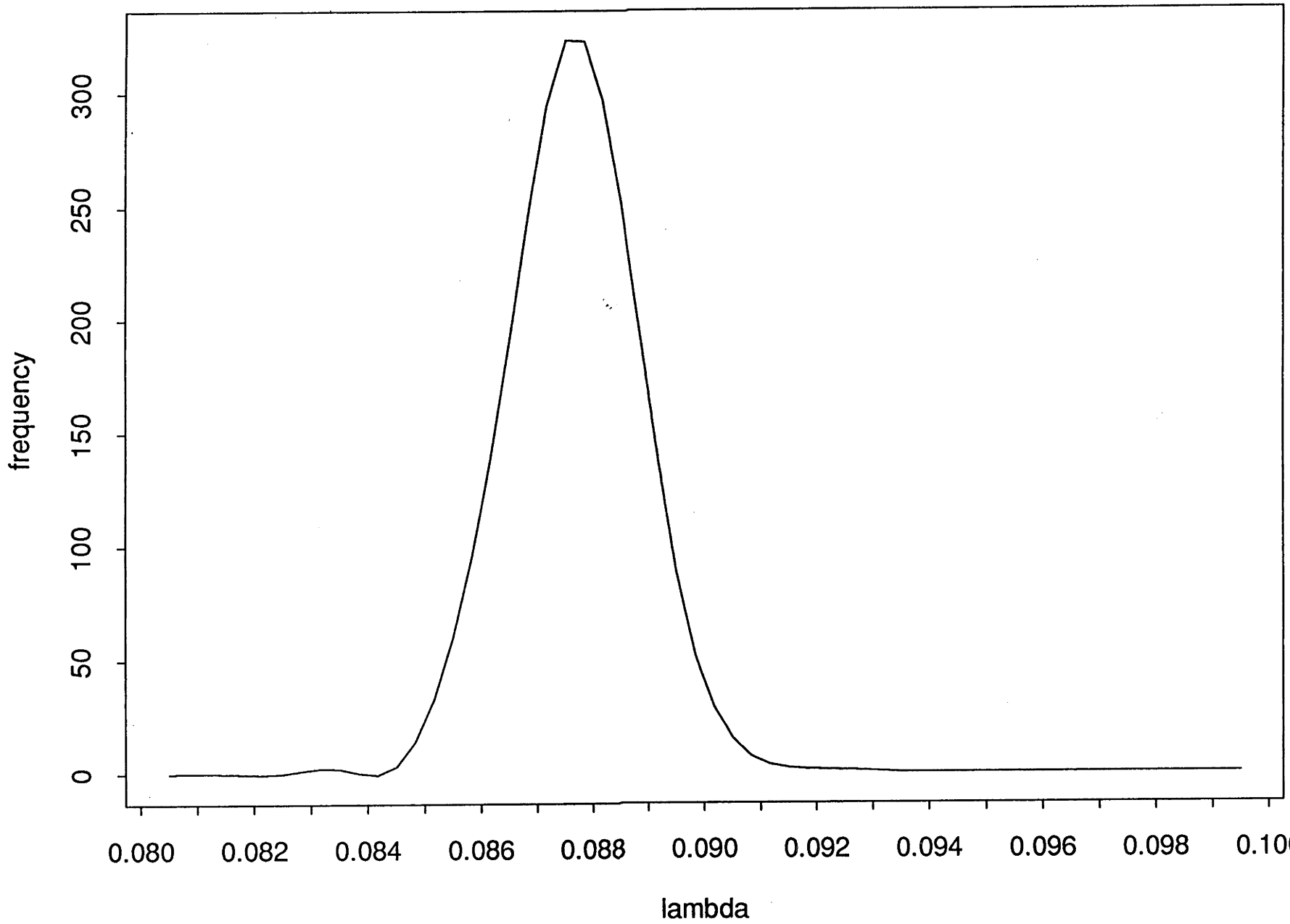


Figure 4.1 Smoothed Histogram of Values from Gibbs Sampler

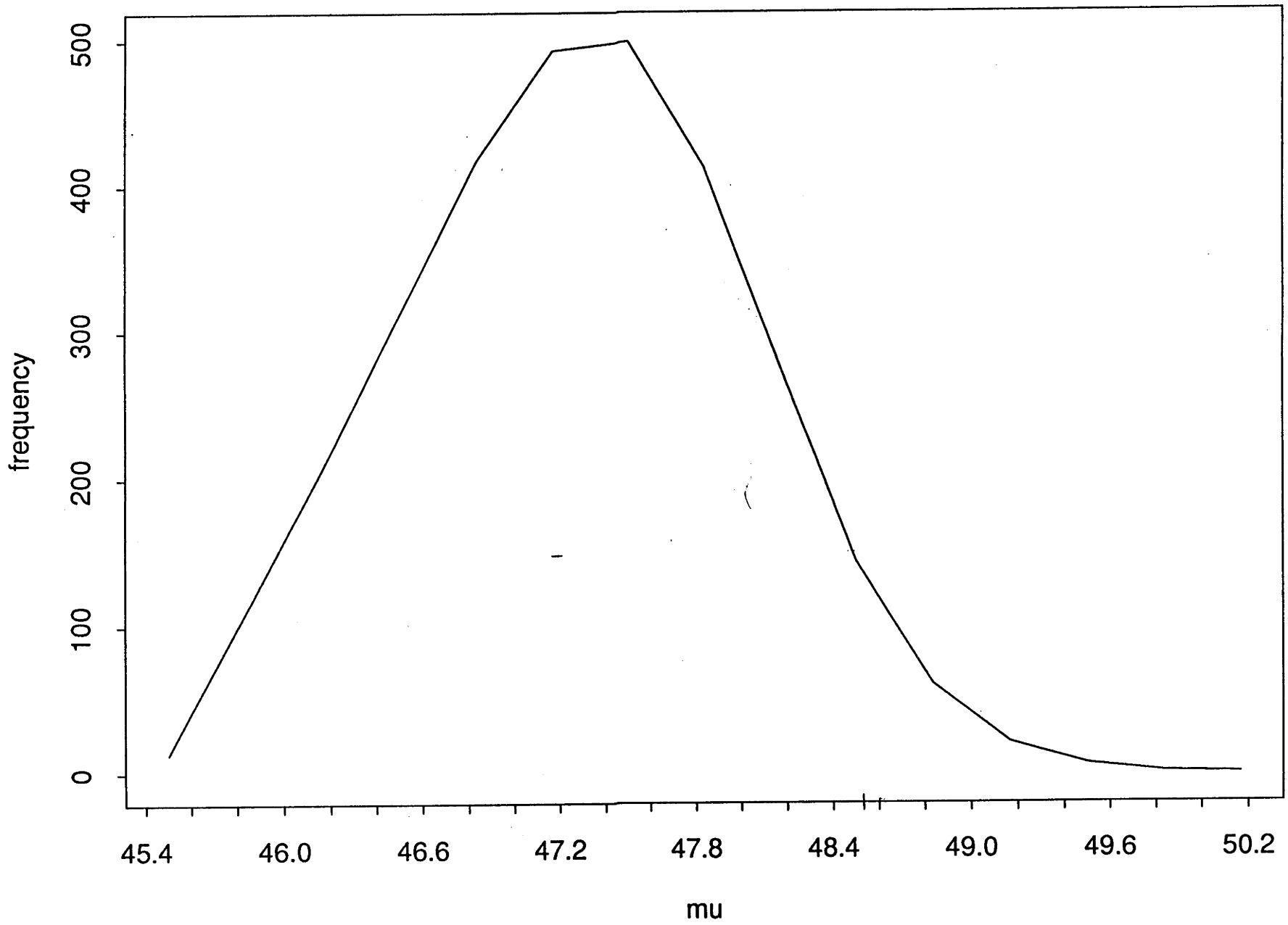


Figure 4.2 Smoothed Histogram of Values from Gibbs Sampler

APPENDIX A

Cases of AIDS in Australia by Month of Diagnosis : 1985 - 1994

The following data are taken from Volume 11 of the *Australian HIV surveillance report* (see page 24) which is produced by the National Centre in HIV Epidemiology and Clinical Research.

	Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	
1985	10	10	7	8	21	10	12	4	15	10	10	10	127
1986	14	15	14	14	19	19	17	24	24	32	26	13	231
1987	29	27	33	20	43	34	28	26	38	30	45	29	382
1988	42	43	24	35	34	45	56	50	44	52	59	49	533
1989	62	47	41	31	47	55	48	57	56	63	50	53	610
1990	62	46	57	50	45	52	59	59	65	69	49	50	663
1991	65	66	65	70	60	61	54	66	85	77	67	60	796
1992	55	67	65	61	75	64	71	73	59	63	62	57	772
1993	67	66	64	65	48	63	71	78	67	71	61	61	782
1994	69	63	74	73	55	67	50	75	85	85	46	45	787

BIBLIOGRAPHY

Australian HIV surveillance report. Technical Report Vol. 11 No. 2, National Centre in HIV Epidemiology and Clinical Research, Sydney, April 1995.

Bailey, N. T. J. (1957). *The Mathematical Theory of Epidemics.* Griffin, London.

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications.* Griffin, London.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes.* Cambridge University Press, Cambridge.

Becker, N. G. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics*, **33**, 515-522.

Brown, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.*, **42**, 59-66.

Chung, K. L. (1974). *A Course in Probability Theory.* 2nd edition. Academic Press, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Vol. **2**, 2nd edition. Wiley, New York.

Gani, J. and Jerwood, D. (1971). Markov chain methods in chain binomial epidemic models. *Biometrics*, **27**, 591-603.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* Academic Press, London.

Hansen, E. R. (1975). *A Table of Series and Products.* Prentice-Hall Inc., USA.

Heathcote, C. R. (1965). A branching process allowing immigration. *Journal of the Royal Statistical Society, Series B*, **27**, 138-143.

Heathcote, C. R. (1966). Corrections and comments on the paper "A branching process allowing immigration". *Journal of the Royal Statistical Society, Series B*, **28**, 213-217.

Heyde, C. C. (1970). Extensions of a result of Seneta for the supercritical Galton-Watson process. *Ann. Math. Statist.*, **41**, 739-742.

Heyde, C. C. and Seneta, E. (1971). Analogues of classical limit theorems for the supercritical Galton-Watson process with immigration. *Math. Biosci.*, **11**, 249-259.

Heyde, C. C. and Seneta, E. (1972). Estimation theory for growth and immigration rates in a multiplicative process. *Journal of Applied Probability*, **9**, 235-258.

Heyde, C. C. and Seneta, E. (1974). Notes on "Estimation theory for growth and immigration rates in a multiplicative process". *Journal of Applied Probability*, **11**, 572-577.

Huggins, R. M. (1993). Asymptotic inference for a class of chain binomial models. *Austral. J. Statist.*, **35** (**1**), 81-87.

Isaacson, D. L. and Madsen, R. W. (1976). *Markov chains - Theory and Applications.* Wiley, New York.

Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions.* Houghton Mifflin, Boston.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.*, **6**, 629-642.

Nelson, P. I. (1980). A note on strong consistency of least squares estimators in regression models with martingale difference errors. *Ann. Statist.*, **8**, 1057-1064.

Port, S. C. (1994). *Theoretical Probability for Applications.* Wiley, New York.

Saunders, I. W. (1980a). A model for myxomatosis. *Math. Biosci.*, **48**, 1-15.

Saunders, I. W. (1980b). An approximate maximum likelihood estimator for chain binomial models. *Austral. J. Statist.*, **22** (**3**), 307-316.

Seber, G. A. F. (1982). *The Estimation of Animal Abundance : and Related Parameters.* 2nd edition. Griffin, London.

Seneta, E. (1969). Functional equations and the Galton-Watson process. *Adv. in Appl. Prob.*, **1**, 1-42.

Sriram, T. N. (1991). On the uniform strong consistency of an estimator of the offspring mean in a branching process with immigration. *Statistics and Probability Letters*, **12**, 151-155.

Sriram, T. N., Basawa, I. V. and Huggins R. M. (1991). Sequential estimation for branching processes with immigration. *Ann. Statist.*, **19**, 2232-2243.

Tanner, M. A. (1993). *Tools for Statistical Inference - Methods for the Exploration of Posterior Distributions and Likelihood Functions.* 2nd edition. Springer-Verlag, New York.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, **82**, 528-550.

Venkataraman, K. N. (1982). A time series approach to the simple subcritical Galton-Watson process with immigration. *Adv. in Appl. Prob.*, **14**, 1-20.

Wei, C. Z. and Winnicki, J. (1987). *A Unified Estimation Theory for the Branching Process with Immigration*. Technical Report, Univ. Maryland.

Wei, C. Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with migration.. *Stochastic Processes and Their Applications*, **31**, 261-282.

Wei, C. Z. and Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *Ann. Statist.*, **18**, 1757-1773.

Yip, P. (1991). A martingale estimating equation for a capture-recapture experiment in discrete time. *Biometrics*, **47**, 1081-1088