MIGRATION MODELS FOR ANIMAL POPULATIONS

by

Arni Neil Arnason

Thesis Submitted for the Degree of Doctor of Philosophy

University of Edinburgh

January 1971

Declaration

The following record of research work is submitted as a thesis for the degree of Doctor of Philosophy of the University of Edinburgh, having been submitted for no other degree.

The research was carried out under the supervision of Dr. R. M. Cormack. Except where otherwise indicated, the work is original.

Abstract of Thesis

In this thesis, the theory of markov processes is used to develop models for animal populations which are subject to migration: that is, animals are capable of moving among a number of discrete areas. Models are developed for populations that are closed to birth and death, but may migrate back or forth between two areas. Models are developed for situations where the probability that an animal migrates is density dependent, or time dependent, or when migration occurs independently of either of these factors. The models are then generalized to consider stochastic birth, death and immigration, and to include migration among many areas. For some of these models the distribution of animals is more conveniently represented as the convolution of multinomial distributions. This representation of the models is developed and its relation to the equivalent markov process model is completely specified. The properties of all the models are developed and some comparisons of the different models are given by means of tables and graphs.

The second half of the thesis is devoted to the development of estimates for the parameters of the models (including, in some instances, population size). For populations subject to migration between two areas, estimates are derived for use with step function observations (the times at which each migration occurs). This method of estimation is investigated for the situation where initial population sizes are known, and where they are not known. Estimation procedures based on complete counts, (with and without identification of each individual in the two populations) are also developed using least squares methods.

A likelihood expression (after the manner of Jolly (1965)) is developed for estimates of population size, survival and migration rates, using mark-recapture sampling on n occasions in 2 populations subject to interchange. It was not possible to derive estimates for the general case, but estimates were derived for the 3-sample case.

A critical review of the literature both of markov process models, and of estimation procedures applicable to markov processes and to animal populations is provided. In particular it is shown that the only extant method for forming estimates from the 3-sample experiment (Iwao, (1963)) incorporates assumptions which make its use inappropriate in certain circumstances. These assumptions are not inherent in the 3-sample estimates derived in this thesis.

Acknowledgements

I wish to record my deepest thanks to my supervisor, Dr. Richard M. Cormack, for his guidance, encouragement and helpful criticisms throughout the period of work on this project. I am also grateful to George M. Jolly for his suggestions and useful discussions about the work described in this thesis.

The work was supported by the National Research Council of Canada whose aid I gratefully acknowledge.

I am also indebted to the University of Edinburgh for providing generous computing time at the Edinburgh Regional Computing Centre.

Table of Contents

CHAPTER ONE

INTRODUCTION

There is a great deal of literature, developed mainly over the last thirty years, devoted to mathematical models for animal populations, and to methods of estimating the parameters of these models using various types of direct observations or data from sampling experiments. One of the characteristics of animal populations is their mobility: their ability to move about within the area under observation, to move into and out of the area during the course of observation, or to move among discrete observation areas. The emphasis, in much of this work in statistical ecology has been on obtaining estimates for population size, birth rates and death rates. The presence of animal movement such as that described above complicates the estimation procedures considerably, so most workers have ignored the problem by assuming it away, or in the very few cases where the problem has been confronted, the attempt is to keep assumptions about the nature of movement to a minimum and derive estimates of population size which take account of these assumptions.

Increasingly, however, animal ecologists are becoming interested in animal behaviour with a view to applying such knowledge to questions of management and control. In such situations, the nature of animal movement is itself of interest. Biologists may be concerned to discover the rate and extent of movement or population interchange, and will not be content simply with estimates of other features of the population even if they are "robust" to certain types of animal movement. Except for a few notable exceptions, (outlined in section 2.3) there has been little attention paid to the development of a fairly general representation of populations subject to movement. This thesis is an attempt to carry out just such an investigation: to develop a class of models representing animal movement, and to see how well such models could be extended to cover a wide range of types of migration that might occur in animal populations: where animals move independently of one another, where movement is density dependent, or time dependent, and extensions to many dimensions (movement among several areas). As well, the models should be able to take account of birth and death.

Naturally there are many ways of representing animal movement. The models may be deterministic or stochastic; in the latter case, the models may be based on diffusion processes, or markov processes or a matrix representation might be used. It is not possible, in one thesis, to investigate all these approaches. Therefore, we chose to concentrate only on the markov process models. A discussion of this and other approaches; to what extent they have been developed by others and how they might be extended further, is given in section 2.3. Even within the class of markov process models, it was not possible to investigate many types of models that might be relevant to animal populations. These models are not really appropriate, for instance, for describing movement within an area, but are most readily adapted to describing movement among distinct areas. This type of movement is what is meant, in this thesis, by migration. Even within this more restricted class of models, there are situations which were not investigated in this thesis (e.g. "flow" models, where the probability that an individual moves from one area to another is dependent upon the number of individuals in the area to which the animal might migrate).

That statistical ecologists have not developed models for migration to any great extent does not mean that such models have been entirely neglected. There is a large and growing literature on migration models, both as a development of markov processes in their own right, and as an application as models of actual phenomena (mainly in statistical mechanics, epidemiology, immunology and compartmental analysis). A brief survey of these developments is given in section 2.2. For the most part, however, these models are not suitable for application to animal populations, or they require some adaptation or elucidation of the properties of the model before they can be applied to animal populations. The only useful models that have been extensively developed are the simple linear model (Ehrenfest model) and the n-dimensional linear model with death (evolutionary model). For completeness, these models are described in sections 3.3 and 4.3 (respectively), though their development is not original to this thesis. However, the reasonably complete investigation of their properties, the delineation of their relevance to animal populations, and the later development of estimation procedures for these models, justifies their inclusion.

Thus, the first half of the thesis is devoted to an investigation of the suitability of markov processes for describing migration of animal populations. It was found that these models can conveniently be used to describe several types of migration: independent, or density dependent migration between two areas; they are easily adapted to account for death, with more difficulty, to account for birth, and with extreme difficulty to account for stochastic immigration. It would seem that birth and immigration can most easily be introduced into migration models in some deterministic fashion (as in Jolly (1965), for example). It is a simple matter to extend the markov process models to several dimensions (in the absence of birth or immigration), but difficulties are encountered when death rates differ from area to area. It was also found that it is difficult to investigate, analytically, the properties of most models where migration is time dependent. The difficulties of both these situations are overcome by an alternate formulation of the models as convolutions of multinomial distributions. This formulation is developed in section 4.3 and the relation of such models to the equivalent markov process model is completely specified. The importance of this formulation for estimation from sampling experiments is outlined.

It appears, then, that markov process models, or equivalent "convolution" models can be used to describe a wide range of migration mechanisms. However, it is not sufficient for the animal ecologist to know that such models can be constructed and manipulated analytically. He must be able to test if the models are appropriate for describing particular aspects of an actual population, and he must be able to estimate all unknown parameters of the model. The problem of testing the appropriateness of the model was too large a question to be undertaken in this thesis, but we have devoted the second part of the thesis to the investigation of how the parameters of the models can be estimated (a prerequisite, in any case, to undertaking the problem of model testing).

It was felt that estimation procedures should be developed keeping in mind a number of considerations. First, it is important to investigate the traditional problem of deriving estimates for population size, survival rates etc.; the situation where migration essentially constitutes a nuisance which must be accounted for, but is not itself of great interest. Second, there may well be situations where the nature of migration is the question of primary interest, and it is desirable to get very precise estimates of migration rates, or to investigate the way migration rates vary over time. In such cases, it might be reasonable to assume that population sizes are known, or that complete counts or estimates of population size can be obtained. Third, one must keep in mind what sorts of observations can be made on animal populations. Traditionally, the methods used have been complete counts and various sampling schemes, with or without some means of identifying individual animals (usually by means of marking) as they are observed. Fourth, estimation should be looked at from the point of view of the models used, rather than from the point of view of the data which is usually available. There may well be means of deriving very precise estimates, or estimates which are eminently suitable for revealing the nature of the migration process but which require observations which are more extensive than or different from the observations usually obtained. The models should be investigated to see what sort of information they might be able to provide, and how much data must be collected to give useful estimates. It is then up to the biologist to determine whether the type of observations demanded can feasibly be obtained, and whether he can afford to obtain enough data to get the precision he requires.

This last consideration prompted the work of Chapter 6, where estimates from step function observations are developed. This type of data is not usually obtained by biologists, and it is questionable whether such data could be obtained for many types of animal populations. However, step function observations, given population initial sizes, provide complete information about the markov process, and so this type of estimation can hardly be neglected. It is also important to know if step function observations are useable in the absence of knowledge of population initial sizes.

Estimation procedures using the more common types of observation (complete counts, marking, sampling) are developed in Chapters 7 and 8.

The estimates developed in these chapters (6, 7 and 8) are entirely new, though in some cases, based on well known techniques. These techniques are described in section 5.2. Few techniques of any real use for animal ecology have been developed by workers in other fields, but a brief review is nevertheless given in section 5.3. There are only two main methods of estimation developed to deal specifically with animal populations subject to migration. Of these, one incorporates implicitly a number of assumptions that were not pointed out by the original author (Iwao, (1963)), and so section 5.4 includes not only a review of extant techniques, but also a detailed reworking of the method of Iwao to reveal its shortcomings.

In the absence of real data (and in the absence of tests to determine whether real data can be considered to come from a particular model), investigation of the properties of the estimates has been carried out using simulated data, whenever analytic investigations proved impossible.

This thesis is by no means an exhaustive study even of the limited problem that I have outlined above. In particular, the work of Chapter 8 could be extended considerably. The problem of forming estimates when migration occurs among m areas, from sampling and marking data at n sampling occasions, remains to be solved. Only a general likelihood expression is given in this thesis. Much of the work that remains to be done, however, demands considerable "algebraic heroics". Nevertheless the thesis does tend to show that the markov process model is a tractable and adaptable representation for migrating animal populations, and that the parameters of these models can be estimated using a wide variety of observational data.

CHAPTER TWO

MIGRATION MODELS:

SURVEY OF PREVIOUS WORK

2.1 Introduction

The phenomenon of a population of individuals that is subject to migration among a number of distinct areas is a commonly occurring situation in nature. As a result, a great deal of work has been done in constructing probabilistic and deterministic models to describe this situation, usually adapted to incorporate the essential features of some particular population. In addition, other influences may be operating on the population simultaneously: birth and immigration may be adding to the population, death and emigration depleting it; the population may be composed of several different classes of individual which interact in some way. Depending on the application, the investigator will want to manipulate these models to answer different questions. Usually interest centres on making statements about the numbers of individuals to be found in each area or class after a given time lapse, and given specific initial conditions. However, other questions such as the probability of extinction of a given class or area, or the limiting behaviour of the system may be of interest.

In order to construct adequate models and manipulate them to answer such questions, investigators have developed and used a wide variety of techniques. Section 2.2 is a brief review of work that has been done in applying migration models to fields other than animal ecology. This will serve to illustrate the widespread use of migration models, the particular questions that arise in the various fields of application, and the techniques that have been developed to cope with these problems. In section 2.3, there is a more detailed review of the development of models which are suitable for application to animal populations. Even within this more restricted field, it will become evident that a wide variety of models has been used. In this thesis, we are concerned mainly with stochastic models and their properties, particularly when small numbers of animals are involved. Therefore the most detailed review will be reserved for similar work that has been carried out in the past.

2.2 Migration Models Applied to Fields other than Animal Ecology.

Perhaps the first use of a stochastic model to describe a system involving migratory behaviour was the model, derived by Ehrenfest (1907) in an application of statistical mechanics to molecular diffusion. Ehrenfest describes a conceptual experiment in which N molecules are distributed between two containers, A and B. At discrete points of time, one of the N molecules is chosen at random and moved to the other container. Thus, after n such steps, the state of the system can be described by the number of molecules in A, and probability statements about this number, given any specified initial division of the N molecules between A and B, and after any specified number of steps n, can be derived using the theory of markov chains. The theory for the Ehrenfest model has been worked out in detail by Kac (1947) and is described in many standard text books (Feller, chapter XV, Cox & Miller, chapter III). The analagous situation in continuous time is a markov process. In fact, it is a finite birth and death process with $\lambda_x = (N-x)\alpha$ $\mu_x = x\beta$ for $0 \leq x \leq N$. As this model is particularly suitable for describing animal populations, it is developed again in chapter 3, along with some properties of the distribution.

The continuous time analogue of the Ehrenfest model for use in chemical kinetics has been included in a comprehensive review of stochastic models in chemical kinetics by McQuarrie (1967). It arises in describing the reversible, unimolecular reaction

$$A \rightleftharpoons B \quad .$$

He shows that the distribution of A molecules at time t , $P_x(t)$, has the binomial distribution, given the initial condition:

$$P_{x_0}(0) = 1$$

$$P_x(0) = 0 \qquad x \neq x_0$$

McQuarrie's review also considers unidirectional migration systems: $A \rightarrow B$, but these have been treated in more detail with reference to animal populations by Pyke (1956), and will be discussed in the next section. McQuarrie considers an interesting initial condition for the unidirectional system; that is, when x_0 is binomially distributed with given mean and variance. This allows for indeterminacy in ascertaining the initial state, as would happen if x_0 is determined by some process (such as weighing) which is subject to error. This may also be the case in animal population work where the initial size of a population may have been determined by some sampling procedure. For this model with x_0 known, $X(t)$ is binomially distributed. If x_0 is also considered to be binomially distributed, the binomial form of $X(t)$ is preserved, the mean is the same (if x_o is replaced by $E(x_o)$), but higher moments are altered.

McQuarrie's review then goes on to consider more complex reactions involving two or more molecules:

$$\text{e.g.} \qquad A + B \rightarrow C \tag{2.2.1}$$

More often than not, these lead to markov processes whose p.g.f. is defined by a differential equation of second or higher order. These models have little relevance to animal populations, but the difficulties of solving the equation for the p.g.f. of such systems, or even of finding analytic expressions for the mean and variance, lead him to a review of standard approximating techniques. These include

(i) The Deterministic Analogue.

The deterministic analogue of a markov process may be obtained by assuming that the variance (and higher cumulants) of $X(t)$, the state of the system at time t, is zero for all t. Thus all terms in the c.g.f. except for those involving first moments disappear. The differential equation for the c.g.f., thus simplified, can usually be solved easily to give the deterministic solution for the state of the system at time t. The deterministic analogue of a process can also be derived by defining a differential equation for rate of change of the system; e.g. for the system in (2.2.1) above, the rate of formation of C molecules is proportional to the number of A and B molecules (given appropriate independence assumptions) which leads to the deterministic expression for the expected number of C molecules at time t :

$$\frac{\partial C(t)}{dt} = kAB$$

where k is a rate constant, and A and B the number of molecules of each type. Deterministic descriptions of a system may be of interest in their own right when it is known that stochastic variation about the mean is very small in relation to the size of the mean itself, or when this variation is very small when compared to variation about the mean introduced in estimating the rate constants. The first situation may frequently occur when very large populations are involved. This, of course, is frequently the case in chemical kinetics and other applications of statistical mechanics, and may be true of some animal populations (insects, fish). In this case, another approach to describing the population is to make use of the theory of Diffusion Processes (markov processes in continuous time with continuous state spaces). A great deal of work has been done on the application of these techniques to migrating systems in physics and biology (e.g. genetics), and one is referred to Bharucha-Reid (1960) for a review of the theory and applications. However, as this thesis is concerned with small or moderate sized populations, diffusion processes will not be resorted to.

McQuarrie found, in applying deterministic methods to processes defined by first order partial differential-difference equations for the p.g.f., that the deterministic analogue gives the exact equation for defining the mean of the equivalent markov process. This is probably true in general, but no general proof seems to be known. McQuarrie describes such deterministic analogues as being "consistent in the mean" with the related markov process. A process whose p.g.f. is defined by a partial differential-difference equation of second order, known as a second order process or system, does not appear to have a deterministic analogue that is "consistent in the mean".

(ii) Approximations for the variance.

Using standard techniques, one can convert the partial differential-difference equation for the p.g.f. of a markov process to a differential equation for the mean or for the second moment of $X(t)$. These equations, especially those derived from second or higher order systems, may involve other moments and be very difficult to solve unless approximation is resorted to. McQuarrie suggests, on empirical grounds, that the equations be simplified using relations between the moments: e.g.

$$\frac{\text{var}(X(t))}{[E(X(t))]^2} = \exp(pt) - 1$$

where p is a constant determined from the initial condition $E[(X(0)]^n = x_o{}^n$.

Ishida (1960) suggests that a second order system with constant rate parameters may be approximated, for purposes of deriving the mean and variance, by a first order system with time dependent rates.

Since the distribution of $X(t)$ may quickly approximate to a normal distribution (convergence is related to the expected number of changed of state up to time t and, so, the approximation improves with increasing elapsed time, population size, and transition rates), it may be reasonable to assume that third and higher cumulants are zero.

In working with any such approximations, of course, one must ascertain, by simulation if necessary, that the approximations are satisfactory for the range of population sizes, and other parameter values likely to be encountered.

Another very early application of migration models was the work of Bateman (1910) to represent the Rutherford theory of radioactive transformations. An unstable atom will disintegrate, in time δt, with probability $\lambda\delta t$ to form a new atom. This may in turn be unstable to form a one way, multidimensional migration system, which terminates in a stable state at the n^{th} stage. If there are $N_i(t)$ atoms of the i^{th} type at time t, this may be represented as

$$\begin{tikzcd}
\boxed{N_1(t)} \arrow[r, "\lambda_1"] & \boxed{N_2(t)} \arrow[r, "\lambda_2"] & \boxed{N_3(t)} & \cdots \arrow[r, "\lambda_{n-1}"] & \boxed{N_n(t)}
\end{tikzcd}$$

Bateman solved the deterministic equations for this process, using various appropriate initial conditions. [E.g. $N_1(0) = x_o$ $N_j(0) = 0 \quad j \neq 1$] Bharucha-Reid (1960, chapter 6) has pointed out that the same equations define the probability that a single atom that started in state 1 at time 0 is in state i at time t, if $x_o = 1$ and $N_i(t)$ is interpreted as this probability instead of as the deterministic number of atoms of the i^{th} type.

Neither Bateman nor Bharucha-Reid attempted to deal with the stochastic model implied by Rutherford's theory. This has been treated, in another context, by Ishida (1960). He uses the same model to describe a volume of gas whose molecules have been excited to a higher energy state than the basal internal energy (the n^{th} state), but the molecular energy degenerates, in order through the lower states, to the basal state. Ishida has to consider a more general initial condition [$N_j(0) = x_{oj} \quad j = 1,\ldots,n$] and his interest is to derive an expression for $X(t) = \sum_{i=1}^{n-1} N_i(t)$, the number of molecules in excited state at time t. He is able to derive this expression, even for the case that λ_i is a time dependent function and finds that the result accords well with accepted empirical laws.

Stochastic migration models have been widely used by workers in the field of epidemiology. An early paper by Bartlett (1949) illustrated a wide variety of techniques for developing markov chains and markov processes with particular reference to epidemics. Among the markov processes which he develops, is a multidimensional generalisation of the Ehrenfest process: that is, individuals move independently of each other from one state to another, the probability of moving from state r to state s in a time interval δt being $\lambda_{rs}\delta t$ $(r,s = 1,\ldots,n$ and $\lambda_{rr} = 0)$. Death may be incorporated as a further state from which no return is possible. Immigration can also be provided for but not births. He is able to write down the partial differential equation for the P.G.F. of this system and to give the form of the solution. The solution involves the latent roots of the matrix whose i,j^{th} element is λ_{ij} $(i \neq j)$ and whose i^{th} diagonal element is $-\sum_j \lambda_{ij}$.

The use of migration models in describing epidemics arises because two or more states may be recognized in an infected population. In the simplest models there are two states: "infected" and "resistant". Migration is one way, from the susceptible to the infected state, but individuals may pass through a number of intermediate "incubation" states, and in some diseases may, after infection, revert to the susceptible state. A lot of early work was devoted to the properties of deterministic models describing this situation (for example, Kermack and McKendrick (1933)) and to using these properties to fit values for the parameters of the models (e.g. Soper (1929)). However, as Bartlett (1949, 1956) has pointed out, these are not entirely satisfactory. Deterministic models may adequately describe the growth and spread of epidemics when numbers are large, but may not be satisfactory when numbers are small, as in the early states, or in endemic situations. Thus, more recent work (Bartlett (1956), Dietz and Downton (1968)) has concentrated on markov process representations of epidemics, and the intractibility of such models has been eased by the use of computers for investigating the properties of the models.

Models such as the one described above are of little use in describing animal populations because the migration probability, being a result of contact between infecteds and susceptibles, is taken as proportional to the product of the numbers in these two classes. Models with more possibility of application to our study have arisen out of work in immunology and bacteriology. Armitage (1952) gives a review of early work and develops a number of models to describe changes in a population of bacteria which are subject to birth and to mutation. Mutation may be considered to be migration to a different class, this class also being capable of birth (growth) and of mutation backwards to the original state. In all his models, Armitage uses a deterministic growth model but allows for stochastic mutation. Although this facilitates the derivation of analytic expressions for mean size of the two classes, it leads to difficulties with the boundary conditions (there is a non-zero probability that one population may have negative size). Armitage develops these 'mixed' models to allow for different growth rates in the two populations, and for forward mutation or forward and backward mutation. Because of the difficulties with the boundary conditions, these models are only suggested for situations in which the mutation rate is very small in comparison with the growth rate. Armitage also devotes a large part of the paper to estimation methods. These will be referred to in chapter 5.

The restriction of no death in the Armitage models is overcome in a paper by Nissen-Meyer (1966) which also deals with populations of bacteria subject to mutation. In this study, death becomes important as bacteria are assumed to be exposed to an antibiotic which inflicts a heavy death rate on the population of unmutated bacteria (susceptibles). Mutation may occur, however, to a resistant strain, which is largely unaffected by the antibiotic, though still subject to natural birth and death. Migration is taken to be unidirectional and extremely low in intensity. As drugs are administered in varying doses over time, birth and death in the susceptible population must have time dependent rates. The concern of these studies is with extinction probabilities, particularly of the resistent strain, under different drug administration programmes or to choose an optimum programme, under various constraints, to maximize these extinction probabilities. To get tractable expressions, Nissen-Meyer uses an interesting approximation. The susceptible population may be considered as a birth and death process with birth rate $\lambda_1(t)$ and death rate $\mu_1(t) + \nu$ where ν is the migration rate. The probability of a migration (mutation) in the time interval $[t,t+\delta t]$ is $X(t)\delta t$, but $X(t)$ can be approximated by:

$$E(X(t)) = X(0)\exp\int_0^t (\lambda(u) - \mu(u) - \nu)du$$

where $X(t)$ is the number of susceptibles at time t. Then the population of resistants can be approximated by a birth and death process with birth parameter λ_2, death parameter μ_2, and feeding function $E(X(t))$. The theory for such a process is given by

Bailey (1964) and an analytic expression for the extinction probability of the resistant population is easily derived.

Migratory systems occur frequently in queuing systems. Jackson (1957) describes one such system in which poisson arrivals occur to each of m queues. After being served, an individual may leave the system, or stay in the system by joining one of the other queues. As is often the case in queuing applications, analytic expressions are derivable only for the equilibrium distribution of the system. Whittle (1967) has generalized Jackson's model and in Whittle (1968) stated and proved several theorems on necessary and sufficient conditions for equilibrium. An interesting specific case of Whittle's generalisation is the Fermi-Dirac model.

Each queue is of, at most, unit length, and migration can occur from any occupied queue to any unoccupied queue. This is the simplest example of a class of migration models, sometimes referred to as flow models, in which the migration probabilities are a function of the state an individual is migrating to, as well as the state it is migrating from. Such models could have great relevance to animal movement, but no work seems to have been done to develop stochastic models for such situations. This will not be dealt with in this thesis.

Migration models have been used in a wide variety of other fields: to describe traffic flow, the spread of rumours, the movement of particles among different "compartments" (bone, blood, etc.) of the body. Until recently, analysis of compartmental systems has been based on deterministic models, and so has not been discussed in this review. However, in a very recent work, Matis and Hartley (1970) have considered the stochastic model. The model developed is equivalent to the model developed in section 4.3.3 of this thesis. The most novel contribution of their paper, however, is the estimation procedure which they have devised. Their results are therefore discussed in more detail at the end of Chapter 5.

The methods used in the studies described in this section, and the difficulties encountered are meant to be illustrative of the available techniques rather than serve as an exhaustive survey.

2.3 Migration Models Appropriate for Animal Ecology.

In section 2.2 it was pointed out that many migration models are derived from the birth and death process. A thorough study of the mathematical properties of the general birth and death process has been carried out by Karlin and McGregor (1957a and 1957b). In these papers they give the Kolmogorov equations for the transition probabilities

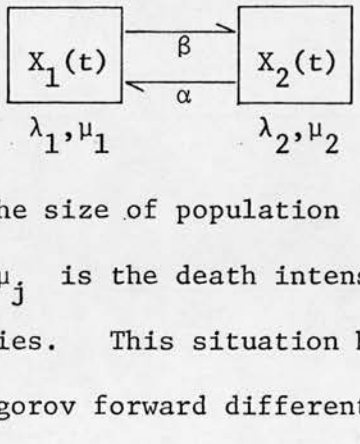
$$P_{ij}(t) = \text{Prob (process in state } i \text{ at time } 0 \text{ is in state } j \text{ at time } t\text{)}$$

of a general time homogeneous birth and death process and give sufficient conditions for a unique solution for the Kolmogorov equations. They give necessary and sufficient conditions for the classification of such processes as recurrent or transient, ergodic or recurrent null. In a later paper (Karlin and McGregor, 1961) they develop a powerful set of theorems for establishing the distribution for the time that the process occupies any specified set of states.

The immediate generalisation of the birth and death process to two dimensions has been called the Competition Process by Reuter (1961). There are two classes of individuals, and for each class, birth and death, possibly at different rates, may take place. Migration in either direction may take place. All instantaneous rates at time t are expressed as a general function of the numbers of individuals in both classes, at time t, but the process is taken to be time homogeneous (i.e. the rates are not a function of t). Thus migration rates may be determined by the number of individuals in the cell to which migration is to take place, thus allowing for 'flow' models. With the extension to include processes not homogeneous in time, this formulation covers the most useful models for describing animal migration between two different areas. Even without this extension, this model covers many systems of interest.

Reuter (1961) gives a sufficient condition for uniqueness of the solution of the Kolmogorov equations for the general time homogeneous competition process and proves several theorems on necessary and sufficient conditions for classifying the process. No further work appears to have been done on formulating properties of the general competition process.

The competition process in which all the individuals exhibit complete independence, and all individuals in the same class at any instant have identical, constant, instantaneous probabilities of birth, death or migration, might be called the linear competition process. This can be schematized by:



where $X_j(t)$ is the size of population j at time t, λ_j is the birth intensity, μ_j is the death intensity, and α and β are migration intensities. This situation has been studied by Puri (1968). He gives the Kolmogorov forward differential equations for the p.g.f. $G(w_1,w_2;t)$ of the probabilities $P_{x_1x_2}(t)$ where

$$P_{x_1x_2}(t) = \text{Prob}\ [X_1(t) = x_1,\ X_2(t) = x_2 \mid X_1(0) = 1,\ X_2(0) = 0] \ .$$

This initial condition seems rather restrictive, though necessary for tractability of the equations. Because of the independence conditions however, it will be seen, in section 3.4, that the effect of more general initial conditions can be accounted for.

The Kolmogorov equation is:

$$\frac{\partial G}{\partial t} + [(1-w_1)(\lambda_1 w_1-\mu_1) + \beta(w_1-w_2)]\frac{\partial G}{\partial w_1}$$
$$+ [(1-w_2)(\lambda_2 w_2-\mu_2) + \alpha(w_2-w_1)]\frac{\partial G}{\partial w_2} = 0$$

This equation is also valid if μ_j, λ_j, α and β are functions of time, but Puri is, in any case, unable to give a general solution for this equation. For the homogeneous case, he develops the backward Kolmogorov equations for the p.g.f. of the more general joint distribution $P_{x_1x_2y_1y_2}(t)$ where y_j is the value of the random variable $Y_j(t)$, the number of migrations out of state j during the interval $[0,t]$. This p.g.f. is referred to as $g_j(w_1w_2v_1v_2;t)$ where $j = 1$ or 2 depending on the initial conditions $g_1(w_1w_2v_1v_2;0) = w_1$ or $g_2(w_1w_2v_1v_2;0) = w_2$. Puri gives a complete solution for $G(w_1w_2;t)$ when $\lambda_1 = 0$ $\alpha = 0$. For $\lambda_2 = 0$, $\alpha = 0$, he gives a solution for $g_1(w_1w_2,v_1;t)$. With $\alpha = 0$ $Y_2(t)$ is always zero. The expression for these p.g.f.'s are complex and do not appear to be a form of any well known discrete multivariate distribution. Thus they are of little use for working out moments of the distribution; this can more easily be done using the defining differential equations directly. The expressions can be used for extinction probabilities such as:

$$\lim_{t \to \infty} P_{00}(t)$$

In both cases this turns out to be an exceedingly simple expression.

When analysing the linear competition process with birth and death occurring in both cells, Puri finds it necessary to resort to more involved techniques, even to deal with unidirectional migration.

The method is to define a sequence of stochastic processes identical to the original process except in the treatment of births. If a birth occurs, the newly created individual is defined to follow the n^{th} process, whereas the parent is defined to follow the $n-1^{th}$ process. The processes are thus defined recursively, with the process corresponding to $n = 0$ being the null process in which all individuals are dead from the start. The p.g.f. for the n^{th} such process, $f^{(n)}(w_1,w_2,v_1,v_2;t)$, is developed to correspond with the system defined by $g_1(w_1,w_2,v_1,v_2;t)$ and a recursion relation is derived defining $f^{(n)}$ in terms of $f^{(n-1)}$. Puri proves theorems to show that the sequence $\{f^{(n)}\}$ is monotone increasing, for fixed $(w_1w_2v_1v_2;t)$, and converges uniformly to $g_1(w_1,w_2,v_1,v_2;t)$. Further limit theorems are developed to derive the limiting extinction probability as for the more restrictive models mentioned above. The system used by Puri could be used to derive similar results for any of the marginal distributions of the process although he has not bothered to do this.

The first development of some of the competition models with explicit reference to animal populations is a study of unidirectional migration models by Pyke (1955). He considers two populations, A_1 and A_2 closed to birth and death, but with migration occurring from A_1 to A_2. He wishes to study the probability distribution

$$P_x(t) = \text{Prob}\ (\ x \text{ individuals migrate from } A_1 \text{ to } A_2 \text{ in time } [o,t))$$

using the general initial conditions: at $t = 0$ A_1 has x_{10} individuals, A_2 has x_{20} individuals. Pyke develops the differential-difference equations by defining a function:

$$p(x,t)\delta t = \text{Prob (an individual migrates from } A_1 \text{ to } A_2 \text{ in the interval } [t,t+\delta t) \mid x \text{ individuals have migrated in } [0,t))$$

He calls this the migration probability function. By assuming that the event of migrating is statistically independent of the activities of all other individuals in the population, it follows that

$$\begin{aligned} q_i(x,t)\delta t &= \text{Prob (} i \text{ of the } x_{10} - x \text{ individuals in } A_1 \text{ migrate to } A_2 \text{ in } [t,t+\delta t) \mid x \text{ have migrated in } [0,t)) \\ &\sim B\ (i;(x_{10} - x),\ [p(x,t)\delta t]) \end{aligned}$$

[The notation $B(i;N,p)$ will mean, throughout this thesis that i follows the binomial distribution with parameters N and p ; i.e.

$$Pr(i) = \binom{N}{i} p^i (1-p)^{N-i}$$]

Thus

$$\begin{aligned} q_0(x,t)\delta t &= (1 - p(x,t)\delta t)^{x_{10}-x} \\ &= 1 - (x_{10}-x)p(x,t)\delta t + 0(\delta t)^2 \\ q_1(x,t)\delta t &= (x_{10}-x)p(x,t)\delta t + 0(\delta t)^2 \\ q_i(x,t)\delta t &= 0(\delta t)^2 \qquad i > 1 \end{aligned}$$

These can then be used to define $P_x(t+\delta t)$ in terms of $P_x(t)$ and $P_{x-1}(t)$, and by the usual limiting process, one has the defining equations:

$$\frac{\partial P_0(t)}{\partial t} = -x_{10}p(0,t)\ P_0(t)$$

$$\frac{\partial P_{x+1}(t)}{\partial t} = (x_{10}-x)p(x,t)\ P_x(t) - (x_{10}-x-1)p(x+1,t)\ P_{x+1}(t)$$
$$\text{for } x = 0,1,\ldots,(x_{10}-1)$$

$$P_0(0) = 1 \qquad P_x(0) = 0 \qquad x > 0 \quad .$$

Thus the system is defined in terms of the migration probability function, and various forms of the function, may be examined:

(i) $p(x,t) = p$ where p is a constant independent of x and t

then $P_x(t) \sim B(x;x_{10},(1-e^{-pt}))$

(ii) $p(x,t) = p(t)$ where $p(t)$ is any function of time such that $p(t) \geq 0$ for all $t > 0$

then $P_x(t) \sim B(x;x_{10},(1-e^{-\gamma_t}))$

where $\gamma_t = \int_0^t p(u)du$.

He also considers two special cases:

(a) $p(t) = (t+a)^{-1}$ where a is a constant. $a > 0$ (this is a monotone decreasing function in t).

(b) $p(t) = (t-iT+a)^{-1}$ where $i = 1,2,\ldots,iT \leq t \leq (i+1)T$ (this is a cyclical function with period T).

(iii) $p(x,t) = f(x)$

Even the simplest forms for $f(x)$ lead to difficulties, since if $f(x)$ has terms in x^n it leads to a differential equation of order $n+1$ for $P_x(t)$. Even for $p(x,t) = ax+b$ he is unable to derive an expression for $P_x(t)$. The deterministic approximation yields an expression for $E(x)$, but he is unable to derive higher moments even using the cumulant generating function.

(iv) $$p(x,t) = \frac{p(t)}{x_{10}-x} \qquad x = 0,1,\ldots,x_1-1 \ .$$
$$= 0 \qquad x = x_{10} \ .$$

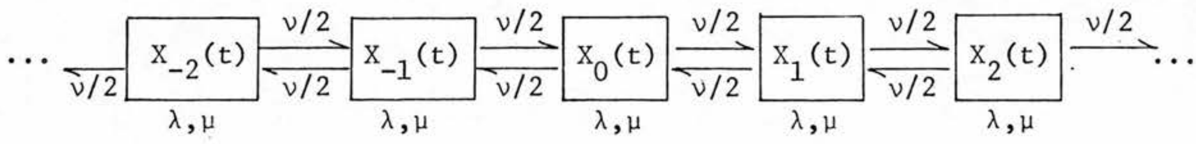
For this system, Pyke is able to solve the equations recursively and finds that

$$P_x(t) \sim \frac{1}{x!} \gamma_t^{\,x} \exp[-\gamma_t] \qquad x = 0,1,\ldots,x_1-1 \quad .$$

Thus x has a poisson distribution truncated on the right.

The main interest of Pyke's work is not so much the models (which are all special cases of Bartlett (1949)), but the general formulation in terms of $p(x,t)$, which will be extended to two way migration in chapter 3, and the large amount of work done on estimating the parameters of these models from the type of sampling experiments that are possible with animal populations. This work is discussed in section 5.2. The restriction of no birth or death in Pyke's model is a very restrictive one in working with animal populations.

A general extension to many dimensions of the migration model has been carried out by Bailey (1968). He develops equations for the M.G.F. of a migration system in which individuals have constant, independent probabilities of migrating to either of the adjacent cells in an unlimited string of cells in a linear array. The particular system discussed by Bailey is schemetized below:



Thus all cells are subject to the same birth and death rates and migration rates. If $K(\ldots\theta_i\ldots;t)$ is the cumulant generating function: i.e.

$$K(\ldots\theta_i\ldots;t) = \log\{E\{\exp(\sum_i \theta_i X_i(t))\}\}$$

where θ_i is the dummy variate corresponding to $X_i(t)$, then

Bailey develops the relation

$$\frac{\partial K}{\partial t} = \sum_{i=-\infty}^{\infty} \{\lambda(e^{\theta_i}-1) + \mu(e^{-\theta_i}-1) + \frac{\nu}{2}(e^{-\theta_i+\theta_{i+1}}-1) + \frac{\nu}{2}(e^{-\theta_i+\theta_{i-1}}-1)\} \frac{\partial K}{\partial \theta_i}$$

By expanding this equation and equating coefficients of θ_i on both sides, he gets a differential-difference equation for $E(X_i(t))$. By converting to the generating function

$$G(s,t) = \Sigma\, E(X_i(t))s^i$$

he derives a differential equation for G , which can be solved using Laplace transforms and the initial condition $X_j(0) = a_j$. The expression for the mean is

$$E(X_i(t)) = (\exp(\lambda-\mu-\nu)t) \sum_j a_j I_{i-j}(\nu t)$$

where $I_n(z)$ is the modified Bessel function of the first kind; that is, the coefficient of t^n in the expansion of $\exp\{(t + 1/t)/2\}$.

When all individuals start out at the origin (0th cell), this process describes a damped wave emanating from the origin. A similar development, using equations derived from coefficients of the c.g.f. of second order in θ , is given by Bailey to develop the expressions for the variances. These expressions are quite complex. Adke (1969) has developed the expressions for the covariances between any two cells, and has generalized the process to allow time dependent transition rates.

Bailey's model for cells in one dimension has immediate applicability to many animal migration studies; e.g., it might be used to describe fish movement between successive ponds in a stream. The model has two major failings however which limit its usefulness for such purposes: it is assumed that all migration rates are identical; also, in any actual experiment, one would examine only a small number of cells out of what is a much larger, but still finite, totality of cells. A method of adapting Bailey's model to describe this situation has not been developed.

Bailey also develops expressions for the means for a system of cells located at the modes of a two-dimensional or three-dimensional lattice. Migration is again assumed to be to adjacent cells and all migration rates are assumed identical. The two-dimensional model might be used to describe the migration of animals among a number of contiguous quadrats laid out on the ground, but the failings mentioned for the one-dimensional model still apply a fortiori. In particular, even if migration among quadrats is the result of random movement, it is not clear that the Bailey model with equal migration rates to adjacent cells is a good description unless quadrats are small and number of migrations is very large. In this situation, it might be more appropriate to consider migration to be over a continuous two-dimensional space (instead of between discrete quadrats) and then, given the distribution of individuals over the space at time t, impose an arbitrarily situated set of quadrats on the space and thus derive moments for the numbers in each quadrat etc. This, in fact, is the approach used by Dempster (1957) using models, for diffusion over the space, developed by Skellam (1951). This model assumes that individuals move out of areas of high density into areas of low density, the number migrating being dependent on this density difference and the mobility of the individuals. This is equivalent to models for electric current flow through a metal where current is equal to the conductivity, α, of the metal multiplied by the potential difference across it. If $f(x,y)$ is a function representing the frequency of individuals distributed over a two-dimension surface with axes x and y, then this system of flux is described by the differential equation:

$$\frac{\partial f}{\partial t} = \alpha\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) \qquad (2.3.1)$$

If death is simultaneously acting on the population, with constant rate μ, then the equation becomes:

$$\frac{\partial f}{\partial t} = \alpha\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) - \mu f \qquad (2.3.2)$$

The parameter α is thus a measure of the intensity of mobility. If x and y are considered to take the discrete values $\ldots -2,-1,0,+1,+2,\ldots$ thus indicating the quadrat at position x,y on a two-dimensional lattice, this model can be used to describe changes of numbers of individuals on each quadrat due to migration among the quadrats, or death (in the general sense, to include emigration beyond the quadrats examined). Dempster does this by assuming that $f(x,y)$ can be approximated by the general quadratic surface:

$$a + bx + cx^2 + ey + gy^2 + hxy \quad .$$

The fitting of this model will be referred to in chapter V.

This model of Skellam's was originally postulated as a description for the spacial diffusion of a population after several generations of individuals whose offspring are scattered according to some distribution about the parent, but it is evident from the above example, that these models can be very useful for describing net movement between specified areas. Equation 2.3.1 arises as a generalisation to continuous time of the well-known random walk in two dimensions with no barriers, with the added generalisation that the position of the next step is a random variable defined by a general distribution with mean zero. Skellam shows that, starting with a particle at the origin, the distribution quickly approaches

$$f(x,y;t) = \frac{1}{ta^2\pi} \exp\{-(x^2+y^2)/(ta^2)\}$$

the distribution defined by (2.3.1). This is, of course, the symmetrical bivariate normal distribution with variance increasing with t . This model, which describes the tendency of animals to move from areas of high density to areas of low density can easily be combined with models describing death, as in (2.3.2), or as Skellam does, with the malthusian law of growth and the logistic law of growth, which give, respectively:

$$\frac{\partial f}{\partial t} = \alpha(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}) + cf$$

$$\frac{\partial f}{\partial t} = \alpha(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}) + c_1 f - c_2 f^2$$

Moreover, the coefficients c, c_1, c_2 may be functions of x and y , so that the model describes different growth patterns over different areas. Skellam gives several specific examples for diffusion constrained to one dimension, and for the two-dimensional case.

These models of Skellam's would be most appropriate for describing animal movement that is continuous over a surface, but where the experimenter has examined an arbitrarily imposed system of quadrats (as in Dempster (1957)). This approach appears to be more promising than using the two-dimensional models of Bailey. Incidentally, the close analogy between Bailey's and Skellam's models suggests that the "damped wave" of Bailey's models may quickly tend to the shape of a normal distribution with increasing numbers of changes of state.

Before leaving the problem of describing migration among quadrats on a surface, mention should be made of an early series of studies on Tsetse fly carried out by Jackson (1947, but see his bibliography for four previous studies in the same series. The present discussion is summarised from Jackson (1940)). Jackson carried out a marking experiment to determine the number of flies on an area and their pattern of movement. He set out four quadrats, each 2 x 2 miles square in the following pattern:

1	2
3	4

Flies were marked in area 1 and capturing in all four quadrats was carried out in subsequent weeks. No explicit model for population interchange is assumed, except that flies are equally likely to migrate in any direction and that flies marked in 1 do not move more than half a mile beyond the border of quadrat 1 (an assumption he later made more explicit). These were the only assumptions, along with the assumptions involved in using the Petersen estimate of population size, required to estimate the emigration rate out of the area and the migration rate between quadrats.

The models of Bailey, Skellam, and Jackson, are all mainly descriptive of migration that is over a continuous two-dimensional space. Frequently, in animal ecology, movement may be among a number of distinct and separated areas, with migration possible between any pair of areas. This might be the situation when insects, say, migrate between different fields or plants, or when small mammals migrate among isolated areas of appropriate habitat such as small woodlots. The first model proposed to handle this situation appears to be the work of Chapman and Junge (1956). The same situation was examined in more detail by Darroch (1961). The models were developed in conjunction with marking experiments in which animals are captured in s areas at time 0, marked, returned, and recapturing is carried out at time t in r areas. Since estimation is the main concern of these studies, they will be described in detail in chapter V. However, mention should be made at this point of the underlying model for migration assumed in these studies. For simplicity, we will assume here that $r = s$. Then Chapman and Junge, and Darroch assume that the probability that an individual marked in cell i at time 0 is alive and in cell j at t is γ_{ij}, that this probability is the same for all (marked) individuals in cell i at time 0, and that animals migrate independently of one another. Darroch is in fact able to relax these assumptions somewhat. With only two cells being considered, this model would appear to bear a close relation to the linear competition process (with no birth). However, these models are effectively concerned with discrete points in time (the sampling times), and the markov process model is not developed by considering the limiting process as $t \to 0$. Nevertheless, both of these models are completely stochastic, and we would expect them to lead to very similar descriptions of the distribution of animals at time t. It will be shown in section 4.3 that the two types of model are in fact equivalent in many instances.

These completely stochastic models have been widely used in describing sampling experiments (Darroch 1958, 1959; Jolly 1965, Seber 1965) because they usually lead to probability distributions which are the product of multinomial distributions. As sampling is often taken to produce binomially distributed observations, the convenience of these models, in forming the joint distribution of the structure at time t

and the sample observations, is considerable. When migration is involved, we shall see (section 4.3) that these discrete time models lead to distributions for the structure of the population at time t which can be represented as the sums of multinomially distributed variables.

As mentioned in section 2.1, migration may occur not only between or among distinct physical areas, but also between different classes of individual. Where these classes represent age groups or developmental stages which must be passed through in order, matrix models have proved useful for describing the populations. (See Leslie (1948 and 1945) for the former, Lefkovitch (1965) for the latter.) Leslie's model is deterministic, and is given by:

$$\underline{n}_t = A\underline{n}_{t-1}$$

where $\underline{n}_t$ is a vector of length m whose j^{th} element is the size of age group j at time t $(j = 0,1,\ldots m-1)$, and A is an $m \times m$ matrix of the form

$$A = \begin{bmatrix} F_0 & F_1 & F_2 & \cdots & F_{m-2} & F_{m-1} \\ P_0 & & & & & \\ & P_1 & & & & \\ & & P_2 & \ddots & & \\ & & & & P_{m-2} & 0 \end{bmatrix} \qquad (2.3.3)$$

Thus P_j is the proportion of animals of age group j which survive to join the $(j+1)^{th}$ age group, and F_j is the number of offspring born into the 0^{th} age group by parents in the j^{th} group, expressed as a fixed proportion of the age group size. Pollard (1966) has considered the stochastic situation where these proportions are interpreted as probabilities, and has shown how to manipulate the matrices to derive means, variances and covariances. He also allows for immigration. Obviously, by introducing matrix differentiation, this model could be extended to continuous time. Matrix models could also be readily defined to describe logistic-type growth. If physical migration occurs between two areas, the Pollard model can easily be adapted to give

$$\begin{pmatrix} \underline{n}_1 \\ \underline{n}_2 \end{pmatrix}_t = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix} \cdot \begin{pmatrix} \underline{n}_1 \\ \underline{n}_2 \end{pmatrix}_{t-1}$$

where $\underline{n}_i$ is the age group vector in area i $(i=1,2)$ and the $m \times m$ matrix A_{xy} is of the same form as (2.3.3) but with the additional subscripts xy on each element so that:

P_{jxy} = prob (an individual of age group j alive in x survives to age group $j+1$ and is in area y at the next time period)

F_{jxy} = prob (an individual of age group j alive in x produces an offspring which survives and is in area y at the next time period)

Matrix models were also used by Usher and Williamson (1970), in a deterministic extension of Bailey's (1968) models in one and two dimensions. They considered edge effects when only a finite number of cells were involved, and considered the (matrix) equations for the size of each cell at time t, given that two classes of individuals existed: those subject to migration, and the other sedentary. The two classes were assumed to have different birth and death rates.

CHAPTER THREE

MIGRATION MODELS:

MIGRATION BETWEEN TWO AREAS CLOSED TO BIRTH AND DEATH

3.1 Introduction

As indicated in section 2.3, models used for describing animal migration have been based on

(i) stochastic models for describing movement between discrete areas (Pyke (1955), Puri (1968), Bailey (1968), etc.);

(ii) deterministic models for describing the distribution of animals over a surface and the changes of this distribution in time (Skellam (1951), Dempster (1957));

(iii) matrix models for describing movement between age classes. These can be taken as deterministic (Leslie (1945), (1948)), or stochastic (Pollard (1966), Lefkovitch (1965)) and are suitable for extension to migration between discrete areas as well.

In this chapter, the formulation of Pyke (1955) is extended to allow migration in both directions, and some of the properties of these models are developed for specific cases. These properties will prove useful in later work on the estimation of the parameters of the models.

3.2 A General Stochastic Formulation

Let A_1 and A_2 denote two areas, where, initially (at $t = 0$) A_1 has x_{10} individuals in it, and A_2 has x_{20} individuals. The two areas are closed to birth and death, but migration occurs between the two areas. Then if the number in A_1 at time t is $X_1(t)$, and in A_2, $X_2(t)$, for all $t \geq 0$

$$X_1(t) + X_2(t) = x_{10} + x_{20} = N \qquad (3.2.1)$$

Now suppose that migration of any individual to the opposite area is independent of the activity of all other individuals in the system, except that the probability of migration may be a function of the number of individuals in the area from which migration occurs. Then we may define

$$\alpha(x_2,t)\delta t = \text{Prob}(\text{an individual in } A_2 \text{ migrates to } A_1 \text{ in the interval } [t,t+\delta t)\,|\,X_2(t) = x_2)$$

$$\beta(x_1,t)\delta t = \text{Prob}(\text{an individual in } A_1 \text{ migrates to } A_2 \text{ in the interval } [t,t+\delta t)\,|\,X_1(t) = x_1)$$

where $x_1 + x_2 = N$.

Also, if

$$q_i(x_2,t) = \text{Prob}(\ i \text{ of the } x_2 \text{ individuals in } A_2 \text{ migrate in } [t,t+\delta t))$$

$$p_j(x_1,t) = \text{Prob}(\ j \text{ of the } x_1 \text{ individuals in } A_1 \text{ migrate in } [t,t+\delta t))$$

then, by the independence condition given above, i and j have the distributions

$$i \sim B[i;x_2,\alpha(x_2,t)\delta t]$$

$$j \sim B[j;x_1,\beta(x_1,t)\delta t]$$

Thus, the probability of no change in the system during the interval $[t,t+\delta t)$, given $X_1(t) = x_1$, is:

$$\begin{aligned}
&[p_0(x_1,t)\delta t][q_0(N-x_1,t)\delta t] \\
&= (1 - \beta(x_1,t)\delta t)^{x_1}(1 - \alpha(N-x_1,t)\delta t)^{N-x_1} \\
&= (1 - x_1\beta(x_1,t)\delta t + 0(\delta t)^2)(1 - (N-x_1)\alpha(N-x_1,t)\delta t + 0(\delta t)^2) \\
&= 1 - x_1\beta(x_1,t)\delta t - (N-x_1)\alpha(N-x_1,t)\delta t + 0(\delta t)^2 \quad . \qquad (3.2.2)
\end{aligned}$$

Also, the probability of a net migration of one individual from A_1 to A_2 in $[t,t+\delta t)$ is

$$\begin{aligned}
&[p_1(x_1,t)\delta t][q_0(N-x_1,t)\delta t] + [p_2(x_1,t)\delta t][q_1(N-x_1,t)\delta t] + \ldots \\
&= [x_1\beta(x_1,t)\delta t(1-\beta(x_1,t)\delta t)^{x_1-1}](1-\alpha(N-x_1,t)\delta t)^{N-x_1} \\
&= x_1\beta(x_1,t)\delta t + 0(\delta t)^2 \quad . \qquad (3.2.3)
\end{aligned}$$

And similarly, the probability of a net migration of one individual from A_2 to A_1 in $[t,t+\delta t)$ is:

$$(N-x_1)\alpha(N-x_1,t)\delta t + 0(\delta t)^2 \; . \qquad (3.2.4)$$

And the probability of a net migration of more than one individual in either direction is

$$0(\delta t)^2 \qquad (3.2.5)$$

Because the system is conservative (in the sense defined by (3.2.1)), the state of the system at any time t is defined by

$$P_x(t) = \text{Prob}(\ A_1 \text{ has } x \text{ individuals at time } t\ ) \; .$$

The four events described above prescribe all possible events that can occur in the interval $[t,t+\delta t)$, and as they are mutually exclusive events,

$$\begin{aligned} P_x(t+\delta t) = {} & P_x(t)[1-x\beta(x,t)\delta t - (N-x)\alpha(N-x,t)\delta t] \\ & + P_{x-1}(t)[(N-x+1)\alpha(N-x+1,t)\delta t] \\ & + P_{x+1}(t)[(x+1)\beta(x+1,t)\delta t] \\ & + 0(\delta t)^2 \ . \end{aligned}$$

By subtracting $P_x(t)$ from both sides of this equation, dividing by δt , and taking the limit as $\delta t \to 0$, we derive the differential difference equation:

$$\begin{aligned} \frac{\partial P_x(t)}{\partial t} = {} & - [x\beta(x,t) + (N-x)\alpha(N-x,t)]P_x(t) \\ & + [(N-x+1)\alpha(N-x+1,t)]P_{x-1}(t) \\ & + [(x+1)\beta(x+1,t)]P_{x+1}(t) \qquad (3.2.6) \end{aligned}$$

for $0 < x < N$.

The boundary conditions for $x = 0$ and $x = N$ may be established in the same way, and give

$$\frac{\partial P_0(t)}{\partial t} = - [N\alpha(N,t)]P_0(t) + [\beta(1,t)]P_1(t)$$

$$\frac{\partial P_N(t)}{\partial t} = - [N\beta(N,t)]P_N(t) + [\alpha(1,t)]P_{N-1}(t)$$

These boundary conditions will create no special difficulties as they can be considered to be of the form (3.2.6) with $P_x(t) \equiv 0$ for x negative or $x > N$.

As in Puri (1968), we can also develop the joint probability distribution for $X_1(t)$, $Y_1(t)$ and $Y_2(t)$ where $Y_j(t)$ is the number of individuals that have migrated <u>out of</u> A_j in $[0,t)$. We define

$$G(x,y_1,y_2;t) = \text{Prob}(X_1(t)=x,\ Y_1(t)=y_1,\ Y_2(t)=y_2)$$

The event $[x_1(t+\delta t)=x,\ Y_1(t+\delta t)=y_1,\ Y_2(t+\delta t)=y_2]$ can occur in exactly one of the four following ways:

(i) $[X_1(t)=x,\ Y_1(t)=y_1,\ Y_2(t)=y_2]$ and there is no change in $[t,t+\delta t)$

(ii) $[X_1(t)=x-1,\ Y_1(t)=y_1,\ Y_2(t)=y_2-1]$ and one individual migrates from A_2 to A_1 in $[t,t+\delta t)$

(iii) $[X_1(t)=x+1,\ Y_1(t)=y_1-1,\ Y_2(t)=y_2]$and one individual migrates from A_1 to A_2 in $[t,t+\delta t)$

(iv) More than one change of state occurs in $[t,t+\delta t)$.

The probabilities for the changes of state are given by (3.2.2) - (3.2.5) and so

$$\begin{aligned} G(x,y_1,y_2;t+\delta t) &= [1-(N-x)\alpha(N-x,t)\delta t - x\beta(x,t)\delta t]G(x,y_1,y_2;t) \\ &\quad + [N-(x-1)][\alpha(N-x+1,t)\delta t]G(x-1,y_1,y_2-1;t) \\ &\quad + (x+1)\beta(x+1,t)\delta t\ G(x+1,y_1-1,y_2;t) \\ &\quad + O(\delta t)^2 \ . \end{aligned}$$

By a similar limiting process, we have

$$\begin{aligned} \frac{\partial G}{\partial t}(x,y_1,y_2;t) = -[&(N-x)\alpha(N-x,t) + x\beta(x,y_1,y_2;t) \\ &+ (N-x+1)\alpha(N-x+1,t)\ G(x-1,y_1,y_2-1;t) \\ &+ (x+1)\beta(x+1,t)\ G(x+1,y_1-1,y_2;t) \qquad (3.2.7) \end{aligned}$$

By defining $G(x,y_1,y_2;t) = 0$ whenever $x < 0$ or $x > N$ or $y_1,y_2 < 0$, equation holds for all $0 \leqslant x \leqslant N$.

Finally, in instances where the equilibrium distribution is not immediately obvious, we will make use of a method due to Lederman and Reuter (1954) and outlined in Cox and Miller section 4.5. The theory applies to all markov processes with a finite state space, with all states accessible from any other state (irreducible processes). The theory requires that the process be represented in the form

$$P'(t) = P(t)Q$$

where $P(t)$ is the $(N+1) \times (N+1)$ matrix with i, jth element

$$Pij(t) = \text{prob}(X(t) = j \mid X(0) = i)$$

$$i,j = 0, 1 \ldots N$$

and Q is an $(N+1) \times (N+1)$ matrix giving the instantaneous transition probabilities. For the general form of the migration model, Q is of the form:

$$Q = \begin{bmatrix} q_0 & q_{01} & 0 & 0 & ..0 \\ q_{10} & q_1 & q_{12} & 0 & ..0 \\ 0 & q_{21} & q_2 & q_{23} & \\ 0 & 0 & q_{32} & q_3 & \\ \vdots & \vdots & & & \\ 0 & 0 & & & q_N \end{bmatrix}$$

where,

$$\begin{aligned} q_{i,i+1} &= (N-i)\alpha(N-i,t) && i = 0 \ldots N-1 \\ q_{i,i-1} &= i\beta(i,t) && i = 1 \ldots N \\ q_{ii} = q_i &= -(q_{i,i+1}+q_{i,i-1}) && i = 1 \ldots N-1 \\ q_0 &= -q_{01} && \\ q_N &= -q_{N,N-1} && \\ q_{ij} &= 0 \quad j \neq i,i-1,i+1 && \end{aligned}$$

If Q is not a function of t (i.e. the process is time homogeneous), then the equilibrium distribution as t tends to infinity exists, and it is determined by the eigenvalue of Q with largest real part. An application of the Perron Frobenius theorem indicates that this eigenvalue must have value 0. The equilibrium distribution can then be shown to be independent of the initial conditions, and the p.g.f. is given by

$$(1 \quad s \quad s^2 \quad \dots \quad s^N)\,(c_0, \quad c_1, \quad c_2 \quad \dots \quad c_N)'$$

where the vector of elements c_j is the left (row) eigenvector of Q corresponding to the eigenvalue $\lambda = 0$, and standardized so that

$$\sum_{j=0}^{N} c_j = 1.$$

3.3 The Simple Linear Model

Let $\alpha(x_2,t) = \alpha$

$\beta(x_1,t) = \beta$

where α and β are constant with respect to x_1, x_2 and t. This process is the continuous time analog of the Ehrenfest process. This model has been studied frequently (e.g. McQuarrie (1967), and is a particular case of Bartlett's (1949) general "evolutionary" model). The results given here for the initial conditions specified below do not appear in any well known text, but the methods are all standard and are given, for example, in Cox and Miller (1965), Chapter 4. For this case, (3.2.6) becomes

$$\frac{\partial P_x(t)}{\partial t} = -[x\beta+(N-x)\alpha]P_x(t)+(N-x+1)\alpha P_{x-1}(t)+(x+1)\beta P_{x+1}(t) \tag{3.3.1}$$

This equation can be solved by introducing the generating function:

$$P(s,t) = \sum_{x=0}^{\infty} P_x(t)s^x$$

Multiplying (3.3.1) by s^x and summing over x gives

$$\frac{\partial P(s,t)}{\partial t} = -\beta s\frac{\partial P}{\partial s} - N\alpha P + \alpha s\frac{\partial P}{\partial s} + N\alpha sP - \alpha s^2\frac{\partial P}{\partial s} + \beta\frac{\partial P}{\partial s}$$

$$\therefore \quad \frac{\partial P}{\partial t} + (\alpha s+\beta)(s-1)\,\frac{\partial P}{\partial s} - N\alpha(s-1)P=0$$

This partial differential equation is of the Lagrange form with auxiliary equations.

$$\frac{\partial t}{\partial 1} = \frac{\partial s}{(\alpha s+\beta)(s-1)} = \frac{\partial P}{PN\alpha(s-1)}$$

It is easily solved by standard methods and gives

$$\frac{P(s,t)}{(\alpha s+\beta)^N} = F\left(\frac{(s-1)}{(\alpha s+\beta)}\,e^{-(\alpha+\beta)t}\right) \tag{3.3.2}$$

where F is an arbitrary function. By using the initial condition

$$P(s,0) = s^{x_{10}}$$

in the equation (3.3.2), the form of the function F becomes

$$F(u) = \frac{(1+u\beta)^{x_{10}}(1-u\alpha)^{N-x_{10}}}{(\alpha+\beta)^N}$$

thus

$$P(s,t) = (\alpha s+\beta)^N\,F\left(\frac{(s-1)}{(\alpha s+\beta)}\,e^{-(\alpha+\beta)t}\right)$$

$$= \left[\frac{\beta(1-e^{-(\alpha+\beta)t})+(\alpha+\beta e^{-(\alpha+\beta)t})s}{\alpha+\beta}\right]^{x_{10}}\left[\frac{(\beta+\alpha e^{-(\alpha+\beta)t})+\alpha(1-e^{-(\alpha+\beta)t})s}{\alpha+\beta}\right]^{x_{10}}$$

$$= [(1-p_1)+p_1s]^{x_{10}}[(1-p_2)+p_2s]^{x_{20}} \tag{3.3.3}$$

Thus $P_x(t)$ has the distribution of the sum of two v binomial variates, r and q, where

$$r \sim B(r;x_{10},p_1) = B(r;x_{10}, \frac{\alpha+\beta e^{-(\alpha+\beta)t}}{\alpha+\beta})$$

$$q \sim B(q;x_{20},p_2) = B(q;x_{20}, \frac{\alpha(1-e^{-(\alpha+\beta)t})}{\alpha+\beta})$$

and so the following properties can immediately be derived.

(i) $$P_x(t) = \sum_{i=0}^{x} \binom{x_{10}}{i} p_1^{\,i}(1-p_1)^{x_{10}-i}\binom{x_{20}}{x-i}p_2^{\,x-i}(1-p_2)^{x_{20}-x+i}$$

(ii) $$E(X(t)) = x_{10}p_1 + x_{20}p_2$$

$$= \frac{N\alpha}{\alpha+\beta}(1-e^{-(\alpha+\beta)t}) + x_{10}e^{-(\alpha+\beta)t} \qquad (3.3.4)$$

(iii) $$\text{Var}\ (X(t)) = x_{10}p_1(1-p_1) + x_{20}p_2(1-p_2) \qquad (3.3.5)$$

(iv) $$\lim_{t\to\infty} P_x(t) = (\alpha+\beta)^{-N}\sum_i \binom{x_{10}}{i}\alpha^i\beta^{x_{10}-i}\binom{x_{20}}{x-i}\alpha^{x-i}\beta^{x_{20}-x+i}$$

$$= (\alpha+\beta)^{-N}\alpha^x\beta^{N-x}\sum_i \binom{x_{10}}{i}\binom{x_{20}}{x-i}$$

$$= \frac{\alpha^x\beta^{N-x}}{(\alpha+\beta)^N}\binom{N}{x}$$

thus the limiting distribution is binomial with

$$B(x_\infty;N;\frac{\alpha}{\alpha+\beta}) = B(x_\infty;N,\frac{r}{r+1})$$

$$\text{where } r = \frac{\alpha}{\beta}$$

Thus in the limit, the distribution is completely determined by the ratio of the two migration intensity parameters.

(v) If at $t = 0$, $x_{10} = 0$ then

$$x \sim B(x;x_{20},\frac{\alpha(1-e^{-(\alpha+\beta)t})}{\alpha+\beta}) \qquad (3.3.6)$$

(vi) If at $t = 0$ $x_{20} = 0$, then

$$x \sim B(x;x_{10}, \frac{\alpha+\beta e^{-(\alpha+\beta)t}}{\alpha+\beta}) \qquad (3.3.7)$$

(vii) For large N, the distribution of x will approach the normal distribution with mean and variance given by ii and iii above. x may be considered to be the number of "successes" out of N independent binomial trials, x_{10} of which have probability p_1 of success, x_{20} having probability p_2 of success. In this case, the Central Limit Theorem applies.

(viii) To derive an expression for $\text{cov}[X(t_2), X(t_1)]$ where $t_2 > t_1$, we require an expression for $E(x_2,x_1)$, where $x_2 = X(t_2)$ $x_1 = X(t_1)$.

$$E(x_2x_1) = \sum_{x_1=0}^{N} \sum_{x_2=0}^{N} x_1x_2 \text{ Prob}[X(t_2)=x_2|X(t_1)=x_1]\text{Prob}[X(t_1)=x_1| X(0)=x_{10}]$$

$$=\sum_{x_1}\sum_{x_2} x_1x_2 \text{ Prob}[X(t_2-t_1)=x_2|X(0)=x_1]\text{Prob}[X(t_1)=x_1|X(0)=x_{10}]$$

(since the process is time homogeneous)

$$=\sum_{x_1} x_1 \text{ Prob}[X(t_1)=x_1|X(0)=x_{10}]\sum_{x_2} x_2 \text{ Prob}[X(t_2-t_1)=x_2| X(0)=x_1]$$

$$=\sum_{x_1} x_1 \text{ Prob}[X(t_1)=x_1|X(0)=x_{10}]E(X(t_2-t_1)|X(0)=x_1)$$

now if we consider p_1 and p_2 to be functions of time,

$$=\sum_{x_1} x_1 \text{ Prob}[X(t_1)=x_1|X(0)=x_{10}](x_1p_1(t_2-t_1)+(N-x_1)p_2(t_2-t_1))$$

$$=[p_1(t_2-t_1)-p_2(t_2-t_1)]\ x_1^2 \text{ Prob}[X(t_1)=x_1|X(0)=x_{10}]$$

$$+Np_2(t_2-t_1)E(X(t_1)|X(0)=x_{10})$$

$$=[p_1(t_2-t_1)-p_2(t_2-t_1)][\mathrm{var}(X(t_1))+(E\ X(t_1))^2]$$

$$+Np_2(t_2-t_1)E(X(t_1))$$

thus $\mathrm{cov}(X(t_2),X(t_1)) = E(x_1x_2)-E(x_1)E(x_2)$

$$=[p_1(t_2-t_1)-p_2(t_2-t_1)][\mathrm{var}\ X(t_1)+(E(X(t_1)))^2]$$

$$+N[p_2(t_2-t_1)-p_2(t_2)]E(X(t_1))$$

$$-x_{10}[p_1(t_2)-p_2(t_2)]E(X(t_1))$$

If $t_2-t_1 = t_1 = t$ it may be shown that:

$$p_1(t_2) = [p_1(t)]^2+[1-p_1(t)][p_2(t)]$$

$$p_2(t_2) = [p_2(t)][p_1(t)]+[1-p_2(t)][p_2(t)]$$

and the expression for the covariance simplifies considerably to:

$$\mathrm{cov}[X(t_1)\ X(t_2)] = \mathrm{cov}[X(t),\ X(2t)]$$

$$= \mathrm{var}\ (X(t))$$

(ix) To derive the joint distribution for $X(t)$, $Y_1(t)$, $Y_2(t)$, substitute $\alpha(x,t) = \alpha$ $\beta(x,t) = \beta$ in (3.2.7). This gives

$$\frac{\partial G}{\partial t}(x,y_1,y_2;t) = -[(N-x)\alpha+x\beta]G+(N-x+1)\alpha\ G(x-1,y_1,y_2-1;t)$$

$$+(x+1)\beta\ G(x+1,y_1-1,y_2;t) \qquad (3.3.3)$$

Let $H(r,s_1,s_2;t)$ be the generating function

$$\sum_{r=0}\ \sum_{s_1=0}\ \sum_{s_2=0} r^{x_1}\ s_1^{y_1}\ s_2^{y_2}\ G(x,y_1,y_2;t)$$

Multiplying (3.3.3) by $r^{x_1}\ s_1^{y_1}\ s_2^{y_2}$ and summing gives

$$\frac{\partial H}{\partial t} = -N\alpha H+(\alpha-\beta)r\,\frac{\partial H}{\partial r} + N\alpha r s_2 H-\alpha r^2 s_2\,\frac{\partial H}{\partial r} + \beta s_1\frac{\partial H}{\partial r}$$

$$\frac{\partial H}{\partial t} = N\alpha[rs_2-1]H+[(\alpha-\beta)r-\alpha r^2 s_2+\beta s_1]\,\frac{\partial H}{\partial r} \qquad (3.3.4)$$

Some care must be exercised in performing the summations to assure that terms near the lower limit of the summation are included in the expressions for the partial derivatives. If $G(x,y_1,y_2;t)$ is defined to be zero for $x > N$ or $x < 0$ then (3.3.4) holds. This equation is again of the Lagrange form, but the auxiliary equations do not give a simple solution. Moreover, we will only want to use expressions for the cumulants to second order. (3.3.4) can be converted to the cumulant generating function $K(\theta_r,\theta_1,\theta_2;t) = \ln H(e^{\theta_r}, e^{\theta_1}, e^{\theta_2};t)$ by means of the relations:

$$\frac{1}{H}\,\frac{\partial H}{\partial r}\,e^{\theta_r} = \frac{\partial K}{\partial \theta_r}$$

and $$\frac{1}{H}\,\frac{\partial H}{\partial t} = \frac{\partial K}{\partial t}$$

(3.3.4) then becomes

$$\frac{\partial K}{\partial t} = N\alpha[e^{\theta_r+\theta_2}-1]+[(\alpha-\beta)-\alpha e^{\theta_r+\theta_2}+\beta e^{\theta_1-\theta_r}]\,\frac{\partial K}{\partial \theta_r} \qquad (3.3.5)$$

where $K(\theta_r, \theta_1, \theta_2;t)$

$$= (k_{100}\theta_r+k_{010}\theta_1+k_{001}\theta_2)+\frac{1}{2!}(k_{200}\theta_r^2+k_{020}\theta_1^2+k_{002}\theta_2^2$$

$$+2k_{110}\theta_r\theta_1+2k_{101}\theta_r\theta_2+2k_{011}\theta_1\theta_2) + \dots$$

and $k_{ijk} = k_{ijk}(t)$ is the mixed cumulant with i corresponding to $X(t)$, j to $Y_1(t)$, k to $Y_2(t)$. By substituting this expansion of K to second order terms in (3.3.5) and equating coefficients in θ, the following relations may be derived.

1. terms in θ_r:

$k'_{100}(t) = N\alpha-(\alpha+\beta)k_{100}$

This, on solution with $k_{100}(0) = x_{10}$ gives the expression for the mean of $X(t)$.

2. terms in θ_2:

$k'_{001}(t) = N\alpha-\alpha k_{100}$

$\therefore\ k_{001}(t) = \alpha\int_0^t N-k_{100}(u)du$

$$\therefore\ E(Y_2(t)) = \alpha\int_0^t E(X_2(u))du. \tag{3.3.6}$$

3. terms in θ_1:

$k'_{010}(t) = \beta k_{100}$

$$\therefore\ E(Y_1(t)) = \beta\int_0^t E(X_1(u))du \tag{3.3.7}$$

4. terms in $\theta r^2/2$

$k'_{200}(t) = -2(\alpha+\beta)k_{200}+N\alpha-(\alpha-\beta)k_{100}$

solution of this equation gives var $(X(t))$ as above.

5. the remaining terms were isolated but not solved: giving the following equations:

$k'_{020}(t) = 2\beta k_{110}+\beta E(X(t))$

$k'_{002}(t) = N\alpha-2\ k_{101}-\alpha E(X(t))$

$k'_{110}(t)+\beta k_{110}(t) = \beta[\text{var}(X(t))-E(X(t))]$

$k'_{101}(t)+\alpha k_{101}(t) = \alpha[N-\text{var}(X(t))-E(X(t))]$

$k'_{011}(t) = -\alpha k_{110}+\beta k_{101}$

3.4 The Quadratic Model (Density dependent migration)

As the density of either area increases, there may be increasing pressure on individuals in the area to migrate out. Density is a function of the area and the number of individuals in the area, but so long as the size of each area remains constant for all t, we may take the density to be proportional to the number of individuals in the area. Thus if the migration intensity is proportional to density,

$$\beta(x_1,t) = \beta x_1$$
$$\alpha(x_2,t) = \alpha x_2 = (N-x_1)\alpha$$

Equation (3.2.6) then gives:

$$P_x'(t) = -[x^2\beta+(N-x)^2\alpha]P_x(t)+[(N-x+1)^2\alpha]P_{x-1}(t)+[(x+1)^2\beta] P_{x+1}(t)$$

$$P_0'(t) = -[\alpha N^2]P_0(t)+\beta P_1(t)$$
$$P_N'(t) = -[\beta N^2]P_N(t)+\alpha P_{N-1}(t) \qquad (3.4.1)$$

if we again introduce the generating function

$$P(s,t) = \sum_k P_x(t)s^x$$

and apply it to the equations (3.4.1), we have

$$\begin{aligned}\frac{\partial P(s,t)}{\partial t} &= -(\alpha+\beta)\sum x^2P_x(t)s^x-\alpha N^2\sum P_x(t)s^x+2N\alpha\sum xP_x(t)s^x \\ &\quad +\alpha N^2\sum P_{x-1}(t)s^x-2\alpha N\sum(x-1)P_{x-1}(t)s^x+\alpha\sum(x-1)^2P_{x-1}(t)s^x \\ &\quad +\beta\sum(x+1)^2P_{x+1}(t)s^x \\ &= \frac{\partial^2P}{\partial s^2}[s(s-1)(\alpha s-\beta)] + \frac{\partial P}{\partial s}[(s-1)(s\alpha-2Ns\alpha-\beta)]+P[N^2\alpha(s-1)]\end{aligned} \qquad (3.4.2)$$

since

$$\sum x^2 P_x(t)s^x = s^2\sum x(x-1)P_x(t)s^{x-2}+s\sum xP_x(t)s^{x-1}$$

$$= s^2\frac{\partial^2 P}{\partial s^2}+s\frac{\partial P}{\partial s} \qquad \text{etc.}$$

This partial differential equation of second order appears to be insoluble even for $\alpha = \beta$. When $\alpha = \beta$, this model is a particular case of a quadratic birth and death process derived by Moran (1958) to describe a situation in genetics (for gene mutation). His model, in the notation of this thesis, is given by

$$\alpha(x_2,t) = a(N-x_2)x_2+b(N-x_2)$$

$$\beta(x_1,t) = a(N-x_1)x_1+c(N-x_1)$$

where a, b and c are constants, with b, c > 0 and $a > \max[-\frac{b}{N-1}, \frac{-c}{N-1}]$. Setting $a = -\alpha$ and $c = b = N\alpha$ gives the model discussed in this chapter with $\alpha = \beta$, and satisfies the above restrictions. Bather (1963) develops an expression for $P_x(t)$ and $\lim_{t\to\infty} P_x(t)$, but the expression is very complex, cannot be expressed in closed form, and would be unwieldy for computation. With $a = -\alpha$ $c = b = N\alpha$, some simplification results, but the expression for $P_x(t)$ could still not be achieved in closed form.

In order to derive the mean and variance for X(t), we will develop the cumulant generating function. Becuase of the enormous amount of algebra involved, a more restricted g.f. is developed than the p.g.f. defined by (3.2.7). We will develop a generating function for G(x,y;t) where

$$x = X(t) \qquad y = Y_1(t)+Y_2(t)$$

$$= \text{total number of migrations in } [0,t].$$

This will not represent any real restriction since

$$Y_1(t)+Y_2(t) = y$$
$$Y_1(t)-Y_2(t) = x_{10}=X(t)$$

whence it will be possible to isolate $Y_1(t)$ and $Y_2(t)$ separately. If $P_{x,y}(t) = \text{Prob}[X(t) = x,\ Y(t) = y]$, then we get a differential difference equation similar to (3.4.1) namely:

$$\begin{aligned} P'_x(t) = {} & -[(N-x)^2\alpha+x^2\beta]P_{xy}\ (t)[N-(x-1)]^2\alpha P_{x-1,y-1}(t) \\ & +(x+1)^2\beta P_{x+1,y-1} \end{aligned}$$

Converting to the joint p.g.f. $G(s,r;t) = \sum\limits_{x} \sum\limits_{y} s^x r^y P_{xy}(t)$

$$\begin{aligned} \frac{\partial G}{\partial t} = {} & \frac{\partial^2 G}{\partial s^2}[-(\alpha+\beta)s^2+\alpha s^3 r+\beta sr] \\ & + \frac{\partial G}{\partial s}[-(\alpha+\beta)s+2N\alpha s+\alpha s^2 r(1-2N)+\beta r] \\ & +G[\alpha N^2(rs-1)] \end{aligned}$$

this can be converted to the m.g.f. $\psi(\theta_s,\theta_r;t) = G(e^{\theta_s},e^{\theta_r};t)$ using the relations

$$\frac{\partial\psi}{\partial\theta_s} = \frac{\partial G}{\partial s}\, e^{\theta_s} \qquad \frac{\partial^2\psi}{\partial\theta_s^2} = \frac{\partial G}{\partial s} e^{\theta_s}+ \frac{\partial^2 G}{\partial s^2} e^{2\theta_s}$$

$$\begin{aligned} \therefore\ \frac{\partial\psi}{\partial t} = {} & \frac{\partial^2\psi}{\partial\theta_s^2}[-(\alpha+\beta)+\alpha e^{\theta_s+\theta_r}+\beta e^{\theta_r-\theta_s}] \\ & + \frac{\partial\psi}{\partial\theta_s}\,[2N\alpha(1-e^{\theta_s+\theta_r})]-\alpha N^2(1-e^{\theta_s+\theta_r})\psi \end{aligned}$$

and taking logs, gives the cumulant generating function $K(\theta_r,\theta_s;t) = \log\psi$

$$\frac{\partial^2 K}{\partial\theta_s^2} + \left(\frac{\partial K}{\partial\theta_s}\right)^2 = \frac{1}{\psi}\,\frac{\partial^2\psi}{\partial\theta_s^2}$$

$$\therefore\ \frac{\partial K}{\partial t} = \left[\frac{\partial^2 K}{\partial\theta_s^2} + \left(\frac{\partial K}{\partial\theta_s}\right)^2\right][-(\alpha+\beta)+\alpha e^{\theta_s+\theta_r}+\beta e^{\theta_r-\theta_s}] \ +$$

$$+\frac{\partial K}{\partial \theta_s}[2N\alpha(1-e^{\theta_s+\theta_r})]-\alpha N^2(1-e^{\theta_s+\theta_r})$$

As in section 3.3, equations for the means and variances may be derived by expanding K and identifying terms in θ. This gives the following equations, where

$$k_{10} = E(X(t)) \qquad k_{20} = var(X(t))$$

$$k_{01} = E(Y(t)) \qquad k_{02} = var(Y(t))$$

$$k_{11} = cov(X(t),Y(t))$$

$$k'_{10} = (\alpha-\beta)(k_{10}^2+k_{20})-2N\alpha k_{10}+\alpha N^2 \qquad (3.4.3)$$

$$k'_{01} = (\alpha+\beta)(k_{10}^2+k_{20})-2N\alpha k_{10}+\alpha N^2 \qquad (3.4.4)$$

$$k'_{20}(t) = 2(\alpha-\beta)(2k_{10}k_{20}+k_{30})+(\alpha+\beta)(k_{10}^2+k_{20})-4N\alpha k_{20}$$
$$-2N\alpha k_{10}+\alpha N^2 \qquad (3.4.5)$$

$$k'_{11}(t) = (\alpha-\beta)(2k_{10}k_{11}+k_{21})+(\alpha+\beta)(2k_{10}k_{20}+k_{30})$$
$$+(\alpha-\beta)(k_{10}^2+k_{20})-2N\alpha(k_{11}+k_{20}+k_{10})+\alpha N^2 \qquad (3.4.6)$$

$$k'_{02}(t) = 2(\alpha+\beta)(2k_{10}k_{11}k_{21})+(\alpha+\beta)(k_{10}^2+k_{20})-4N\alpha k_{11}$$
$$-2N\alpha k_{10}+\alpha N^2 \qquad (3.4.7)$$

(i) Mean and Variance when $\alpha = \beta$

Setting $\alpha = \beta$ in (3.4.3) gives, for E(X(t)):

$$2N\alpha k_{10}(t)+k'_{10}(t) = \alpha N^2$$

and if $k_{10}(0) = x_{10}$ $x_{20} = N-x_{10}$, the solution is:

$$E(X(t)) = \frac{1}{2}[N+(x_{10}-x_{20})\exp(-2N\alpha t)] \qquad (3.4.8)$$

Equation (3.4.5) for var(X(t)) becomes

$$2\alpha(2N-1)k_{20}(t)+k'_{20}(t) = 2\alpha k_{10}^2-2N\alpha k_{10}+\alpha N^2$$

and the solution, using the condition $\mathrm{var}(X(0)) = 0$ is:

$$\mathrm{var}(X(t)) = \frac{N^2}{4(2N-1)}(1-e^{-2\alpha(2N-1)t})+\frac{(x_{10}-x_{20})^2}{2}e^{-4N\alpha t}(e^{2\alpha t}-1) \tag{3.4.9}$$

(ii) The Equilibrium Distribution when $\alpha \neq \beta$.

The theory outlined in section 3.2 for the equilibrium process applies. For this case, the matrix Q is of the form:

$$\begin{bmatrix} -N^2\alpha & N^2\alpha & 0 & 0 & \cdots & 0 \\ \beta & -\beta-(N-1)^2\alpha & (N-1)^2\alpha & 0 & \cdots & 0 \\ 0 & 2^2\beta & -2^2\beta-(N-2)^2\beta & (N-2)^2\alpha\ldots & & 0 \\ 0 & 0 & 3^2\beta & & & 0 \\ \vdots & \vdots & \vdots & & & 1^2\ \alpha^2 \\ 0 & 0 & 0 & & N^2\beta & -N^2\beta \end{bmatrix}$$

Now if $\underline{c}' = (c_0,c_1\ldots c_N)$ is the left eigenvector corresponding to the eigenvalue $\lambda = 0$, then

$$\underline{c}'Q = (0,0,0,\ldots)$$

$$\therefore\ -N^2\alpha c_0+\beta c_1 = 0 \tag{3.4.11}$$

$$N^2\alpha c_0-[\beta+(N-1)\alpha]c_1+2^2\beta c_2=0 \tag{3.4.12}$$

$$\vdots$$

$$(N-k+2)^2\alpha c_{k-2}-[(k-1)^2\beta+(N-k+1)^2\alpha]c_{k-1}+k^2\beta c_k = 0$$

$$\begin{array}{l} \cdot \\ \cdot \qquad\qquad k = 2,3 \ldots N \\ \cdot \end{array}$$

$$1^2\alpha c_{N-1}-N^2\beta c_N = 0$$

solving recursively
$$\begin{aligned} c_1 &= (N^2\alpha)c_0/\beta \\ c_2 &= (N-1)^2\alpha c_1/(2^2\beta) \\ &\cdot \\ &\cdot \\ &\cdot \\ c_k &= (N-k+1)^2\alpha c_{k-1}/(k^2\beta) \\ &\qquad k = 1 \ldots N \end{aligned}$$

The solution of this difference equation, in terms of c_o is

$$\begin{aligned} c_k &= \frac{N^{(k)}}{k!}\frac{N^{(k)}}{k!}\left(\frac{\alpha}{\beta}\right)^k c_0 \\ &= [\tbinom{N}{k}]^2(\tfrac{\alpha}{\beta})^k c_0 \end{aligned}$$

and c_0 may be determined from the normalizing condition

$$\sum_{k=0}^{N} c_k = c_0\sum_{k=0}^{N}\binom{N}{k}^2\left(\frac{\alpha}{\beta}\right)^k = 1.$$

Thus the equilibrium distribution $P_x(\infty)$, say, is

$$P_x(\infty) = \binom{N}{x}^2\left(\frac{\alpha}{\beta}\right)^x / \left(\sum\binom{N}{x}^2\left(\frac{\alpha}{\beta}\right)^x\right) \qquad (3.4.13)$$

For the case $\alpha = \beta$, the denominator becomes

$$\sum_{x=0}^{N}\binom{N}{x}^2 = \binom{2N}{N} \qquad \text{(hypergeometric identity).}$$

and thus $\quad P_x(\infty) = \binom{N}{x}^2/\binom{2N}{N}.$

This is simply the hypergeometric distribution with mean

$$N.\ \frac{N}{2N} = \frac{N}{2} \qquad (3.4.14)$$

and variance

$$N\ \frac{N}{2N}.\frac{N}{2N}.\frac{2N-N}{2N-1} = \frac{N^2}{4(2N-1)} \qquad (3.4.15)$$

$$\doteq \frac{N}{8} \text{ for N large}$$

There does not appear to be a simple closed expression for the mean and variance when $\alpha \neq \beta$. When $\alpha = \beta$, the equilibrium distribution may be derived directly from the p.g.f. (3.4.2), since $\frac{\partial P(s,t)}{\partial t} = 0$, and the equation then reduces to

$$s(1-s)\frac{\partial^2 P}{\partial s^2} + [1-(-N-N+1)s]\ \frac{\partial P}{\partial s} + (-N)(-N)P = 0$$

which is the form of the hypergeometric differential equation for $F(\alpha,\beta,\gamma;x)$, the general hypergeometric function with $\alpha = \beta = -N \quad \gamma = 1 \quad x = s$ (see, for instance, Handbook of Mathematical Functions (1964), Chapter 15).

(iii) Approximations for the mean and variance when $\alpha \neq \beta$.

The deterministic analog of the general migration process is defined, by:

$$\frac{d\overline{X}(t)}{t} = -x\beta(x,t)+(N-x)\alpha(N-x,t)$$

For the density dependent model, this leads to the same expression for the mean as is given by equation (3.4.3) with $k_{20} = 0$, namely

$$\frac{\partial k_{10}(t)}{\partial t} = (\alpha-\beta)k_{10}^2-2N\alpha k_{10}+\alpha N^2$$

or $$\frac{dk_{10}}{(\beta-\alpha)k_{10}^2+2N\alpha k_{10}-N^2\alpha} = -dt$$

The discriminant of the quadratic in k_{10} is

$$q^2 = 4N^2\alpha\beta$$

and so, the solution of the differential equation is of the form

$$\frac{1}{q}\log\left(\frac{2(\beta-\alpha)k_{10}+2N\alpha-q}{2(\beta-\alpha)k_{10}+2N\alpha+q} = t+c\right.$$

The constant c is determined from $k_{10}(0) = x_{10}$ and this leads to the expression for $k_{10}(t) = \overline{X}(t)$

$$\tilde{\overline{X}}(t) = N\left\{\frac{\alpha x_{20}f_1(t)+\sqrt{\alpha\beta}x_{10}f_2(t)}{(\alpha x_{20}+\beta x_{10})f_1(t)+\sqrt{\alpha\beta}Nf_2(t)}\right\} \qquad (3.4.16)$$

where

$$f_1(t) = 1-\exp(-2N\sqrt{\alpha\beta}t)$$
$$f_2(t) = 1+\exp(-2N\sqrt{\alpha\beta}t)$$

in the limit as t becomes large the term $f_1(t)$ and $f_2(t)$, tend to 1 (in fact, very quickly due to the size of the exponent). This gives an approximation for the mean of the equilibrium distribution as

$$\tilde{\overline{X}}\ (\infty) = N\left\{\frac{x_{20}+\sqrt{\alpha\beta}x_{10}}{\alpha x_{20}+\beta x_{10}+\sqrt{\alpha\beta}N}\right\}$$

$$= \frac{N\sqrt{\alpha}}{\sqrt{\alpha}+\sqrt{\beta}}\,\frac{(\sqrt{\alpha}x_{20}+\sqrt{\beta}x_{10}}{(\sqrt{\alpha}x_{20}+\sqrt{\beta}x_{10}}$$

$$= \frac{N\sqrt{r}}{\sqrt{r+1}}\text{where } r = \frac{\alpha}{\beta}$$

An approximation for the variance may also be derived by using equation (3.4.5), with the left hand side set to zero, and assuming $k_{30} \approx 0$. Solving this for k_{20} in terms of k_{10} gives

$$k_{20} = \frac{-(\alpha+\beta)k_{10}^2+2N\alpha k_{10}-\alpha N^2}{4(\alpha-\beta)k_{10}+\alpha+\beta-4N\alpha}$$
$$= \frac{-(r+1)k_{10}^2+2Nrk_{10}-rN^2}{4(r-1)k_{10}+r+1-4Nr}$$

and substituting the approximation $k_{10} \simeq N\sqrt{r}/(\sqrt{r}+1)$ gives, after simplifying:

$$\tilde{V}(\infty) = k_{20} = \frac{2N^2 r}{(\sqrt{r}+1)^2(4N\sqrt{r}-r-1)} \qquad (3.4.18)$$

Note that this expression reduces to $\frac{N^2}{4(2N-1)}$ when $r = 1$ as in equation (3.4.15).

(iv) Accuracy of the Approximations for the Mean and Variance of the Equilibrium Distribution.

The approximate expressions for the mean and variance, denoted $\tilde{\bar{X}}(\infty)$ and $\tilde{V}(\infty)$, as given by equations (3.4.17) and (3.4.18) are compared with the exact expressions

$$\bar{X}(\infty) = \frac{1}{c_0}\sum x\binom{N}{x}^2 r^x \qquad (3.4.19)$$

and

$$V(\infty) = \frac{1}{c_0}\sum x^2\binom{N}{x}^2 r^x-[\bar{X}(\infty)]^2 \qquad (3.4.20)$$

(where c_o is the normalizing constant and $r = \frac{\alpha}{\beta}$). The results of this comparison are given in table 3.1,for various values of N and r. The results are given only for $r < 1$, since the error

$$E_x = \tilde{\bar{X}}(\infty)-\bar{X}(\infty)$$

is the same for r and for $\frac{1}{r}$. This follows from the fact that

$$\tilde{X}_r(\infty)+\tilde{X}_{1/r}(\infty) = \frac{N\sqrt{r}}{\sqrt{r}+1} + \frac{N\sqrt{1/r}}{\sqrt{1/r}+1}$$
$$= N$$

and (by a rather long, but routine combinatorial argument which will not be given here)

$$\bar{X}_r(\infty)+\bar{X}_{1/r}(\infty) = \frac{1}{c_0}\sum x\binom{N}{x}r^x + \frac{1}{c_0}\sum x\binom{N}{x}\left(\frac{1}{r}\right)^x$$
$$= N$$

thus

$$\tilde{X}_r-\bar{X}_r = -(\tilde{X}_{1/r}-\bar{X}_{1/r})$$

That is the errors of the approximation for the mean use of the same absolute size for r and 1/r, but are of opposite sign in the two cases.

It appears from table 3.1, that for all $r < .5$, the approximation <u>overestimates</u> the true mean. Since the approximation is exact for $r = 1$, it is probable that the approximation overestimates the true mean for all $r < 1$ and underestimates the true mean for all $r > 1$. From fig. 3.1, where the absolute size of the error is plotted against N, it is clear that a general pattern exists for all r. The error rises for the first few values of N and then begins a monotonic decline. Since the true mean increases with N, the relative error $(\tilde{X}-\bar{X})/\bar{X}$ approaches zero very rapidly.

No such clear pattern emerges for the approximation for the variance, but it is clear from table 3.1 that the approximation is excellent when α and β are of roughly the same magnitude. The approximation appears to be, consistantly, an overestimate of V when $r < 1$.

A better approximation for the mean and variance may be derived using equations (3.4.5) and (3.4.3) with $k'_{10}(t) = k'_{20} = k_{30} = 0$. This leads to two (nonlinear) equations in the mean and variance. Elimination of the variance V, leaves a cubic equation in $\overline{X}$;

$$4(1-r)^2\overline{X}^3+12Nr(1-r)\overline{X}^2-4Nr(1+N-3Nr)\overline{X}-2N^2r(2Nr-1) = 0$$

$$\text{provided } r \neq 1$$

This expression does not yield a convenient analytic expression for $\overline{X}$, and so is of little interest. Moreover for some values of r (e.g. $r < 1/2N$) it does not give a unique positive real root, so that one would have to examine the stability of the solution, or use other information about the process to select the correct solution. Nevertheless, in a number of particular cases tried, the solution given by this equation was always slightly better than the approximation (3.4.17).

Table 3.1

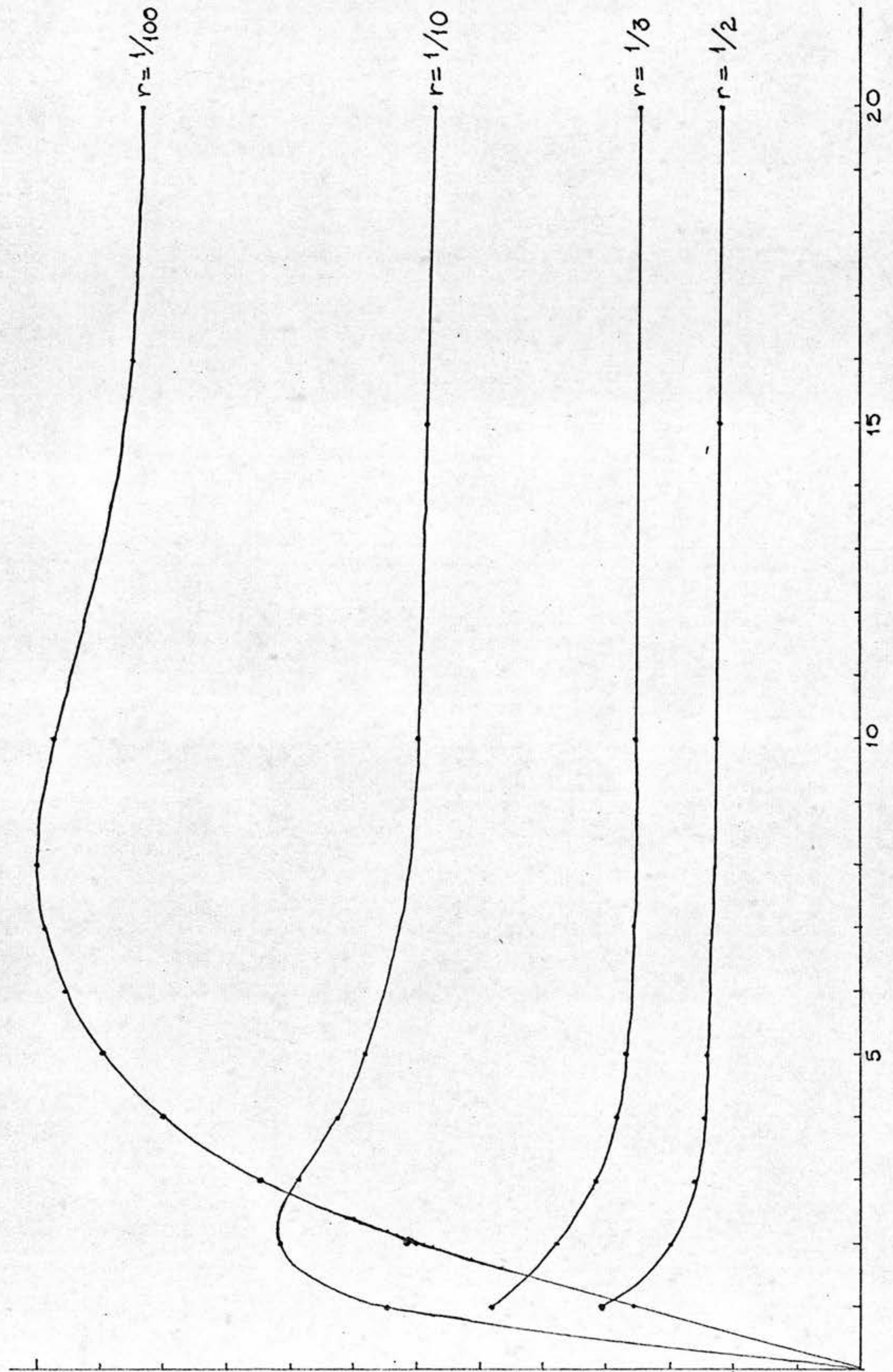
Comparison of the Approximate Means and Variances for the Equilibrium Distribution with the Exact Values.

The approximations for the mean and variance, $\tilde{\bar{X}}(\infty)$ and $\tilde{V}(\infty)$ are derived from equations (3.4.17) and (3.4.18) respectively; the exact values, $\tilde{\bar{X}}(\infty)$ and $V(\infty)$ from (3.4.19) and (3.4.20) respectively. $E_x = \tilde{\bar{X}}(\infty)-\bar{X}(\infty)$. $E_v = \tilde{V}(\infty)-V(\infty)$. $r = \alpha/\beta$

N	$\bar{X}(\infty)$	$\tilde{\bar{X}}(\infty)$	E_x	$V(\infty)$	$\tilde{V}$	E_v
			$r = 1/2$			
5	2.02	2.07	.048	.6751	.6786	.0035
10	4.10	4.14	.045	1.274	1.281	.007
20	8.24	8.28	.044	2.488	2.493	.005
			$r = 1/3$			
5	1.76	1.83	.075	.648	.656	.008
10	3.59	3.66	.071	1.222	1.231	.009
20	7.25	7.32	.069	2.369	2.389	.020
			$r = 1/10$			
5	1.045	1.201	.156	.525	.552	.027
10	2.260	2.400	.140	.981	1.000	.019
15	3.467	3.604	.137	1.423	1.453	.030
20	4.671	4.805	.135	1.866	1.908	.042
100	23.895	24.025	.131	9.140	9.207	.067
			$r = 1/100$			
5	.215	.455	.240	.185	.417	.232
10	.654	.909	.255	.392	.552	.160
20	1.592	1.818	.226	.864	.946	.082

fig. 3.1

Error in the Approximation for the Mean of the Quadratic Model.



(v) The Expected Number of Migrations in the Interval [0,t)

Equations (3.4.4) defines the expected number of migrations $k_{01}(t) = E(Y(t))$, in terms of $E(X(t))$ and $V(X(t))$. For $\alpha = \beta$, this equation gives:

$$k'_{01}(t) = 2\alpha(k_{10}+k_{20})-2N\alpha k_{10}+\alpha N^2.$$

Substituting the expressions (3.4.8) and (3.4.9), and integrating, using the initial condition $k_{01}(0) = 0$, gives:

$$k_{01}(t) = \frac{\alpha N^3 t}{(2N-1)} + (x_{10}-x_{20})^2\left[\frac{(e_1-1)}{8N} - \frac{(e_1e_2-1)}{2(2N-1)}\right]+ \frac{N^2(e_1e_2-1)}{4(2N-1)^2}$$

where $e_1 = \exp[-4N\alpha t]$ $e_2 = \exp[2\alpha t]$. The last term is less than 1 and can be ignored. 2N-1 may be replaced by N when N is large (except in the exponentials) to give the simpler expression:

$$k_{01}(t) \doteq \frac{N^2 t}{2} + \frac{(x_{10}-x_{20})^2}{8N}(1+e_1-2e_1e_2)$$

When $\alpha \neq \beta$:

Use equation (3.4.3) to express $(k^2_{10}+k_{20})$ in terms of $k_{10}(t)$, and substitute into (3.4.4). This gives:

$$k'_{10}(t) = [(\alpha+\beta)k'_{10}(t)+4N\alpha\beta k_{10}-2N^2\alpha\beta]/(\alpha+\beta)$$

and on integrating both sides:

$$k_{01}(t) = [(\alpha+\beta)k_{10}(t)+4N\alpha\beta\int k_{10}(t)dt-2N^2\alpha\beta t]/(\alpha+\beta)+c$$

where c is the constant of integration.

Now $k_{10}(t)$ can be approximated by $\tilde{X}(t)$ (equation 3.4.16), and it may be shown that

$$\int \tilde{X}(t)dt = \frac{N(\alpha-\sqrt{\alpha\beta})t}{(\alpha-\beta)} - \frac{1}{(\alpha-\beta)} \log[D]$$

where D is the denominator of the expression (3.4.16) for $\tilde{X}(t)$. Substituting this into the equation, and solving for c from the initial conditions $k_{01}(0) = 0$, $k_{10}(0) = x_{10}$ gives the following approximation for $k_{01}(t)$, say $\tilde{Y}(t)$, in terms of the approximation $\tilde{X}(t)$ for the mean:

$$\tilde{Y}(t) = \frac{\alpha+\beta}{\alpha-\beta} [\tilde{X}(t)-x_1] + \frac{2N^2\alpha\beta t}{(\sqrt{\alpha}+\sqrt{\beta})^2} - \frac{4N\alpha\beta}{(\alpha-\beta)^2} \log\left[\frac{(\alpha x_{20}+\beta x_{10})}{2N\sqrt{\alpha\beta}}(1-e)+\frac{(1+e)}{2}\right]$$

$$\text{where } e = \exp[-2N\sqrt{\alpha\beta}t]. \qquad (3.4.19)$$

(vi) The Tendency of the distribution of X(t) to normality and the Accuracy of the Approximations $\tilde{X}(t)$ and $\tilde{Y}(t)$.

With the linear model considered in section 3.3, the convergence of the distribution of X(t) to the normal distribution was an immediate result of the fact that, because of the independence properties involved, the random variable X(t) can be considered to be the outcome of the sum of N completely independent random variable. This situation does not appear to apply to the density dependent model. The behaviour of any one of the N individuals is determined by the location of the remaining N-1 individuals in the system. The equilibrium distribution with $\alpha = \beta$ is a hypergeometric distribution and it is well known that, for large N, this distribution approaches a binomial distribution and hence can be approximated by a normal distribution. But for the distribution at time t, it is not easy to develop a proof that the distribution of x(t) approaches the normal, especially as no analytic expression for the characteristic function, or the cumulant g.f. is available. Nevertheless, it appears that, under most conditions, the normal approximation is a good one; in fact the convergence appears to be even more rapid than for the linear model. The evidence for this is entirely from simulations. The approximation will be worst when E(X(t)) is close to N or 0. This occurs when x_{10} is close to N or 0, $r = \alpha/\beta$ is either very small or very large, and the elapsed time, t, is small. In such situations, the distribution is necessarily skewed due to the constraint $0 < Y(t) < N$. However, with increasing t, E(X(t)) very quickly approaches the mean of the equilibrium distribution, so that even if $x_{10} = 0$ or N, X(t) tends to "move away" from these boundaries very swiftly. For example, with

$$N = 20 \qquad \alpha = 1 \qquad \beta = 2 \qquad x_{10} = N = 20 \qquad t = .01$$

the expected number of migrations that occur in [0,.01] is only 5.876 (result based on 500 simulations of the process), yet E(X(.01)) = 14.448 = 20-5.552. The equilibrium mean and variance, given by the approximation formulae, are 8.284 and $(1.579)^2$. Thus in most of the realizations of the process, the first five or six migrations were all out of A_1, with X(t) moving monotonically towards the equilibrium mean. The approximation for E(X(.01)) given by (3.4.16) is 14.392. The frequency distribution of X(.01) over the 500 simulations may be compared with the expected frequency for the normal approximation with the same mean and variance. It can be seen that the distribution is skewed to the right, but the normal approximation is not so far different, despite the fact that the conditions for a good fit, mentioned above, are not met:

$x(.01) =$	≤ 10	11	12	13	14	15	16	17	≤ 18
$f(x(.01))$ from 500 simulations	10	21	42	73	85	129	79	46	15
$f(x(.01))$ from the normal approximation	6.9	18.0	44.3	79.9	106.6	105.4	72.9	44.4	21.9

$\chi^2_6 = 14.03$

Another assurance of the closeness of the normal approximation is given by the accuracy of the approximation $\tilde{Y}(t)$ (equation 3.4.19) and $\tilde{X}(t)$ (3.4.16) for $E(Y(t))$. This is demonstrated, for a number of simulated realizations of the process, in table 3.2. The approximations were both derived using the assumption that the third cumulant is zero. If the distribution for $X(t)$ were markedly skewed, we would not expect these approximations to be quite so good.

Explanation of Table 3.2:

The parameters of the process are given in the first line of each section, along with the values for the mean and the variance of the equilibrium distribution (equations 3.4.17, 3.4.18). For each set of parameters, the state after a number of equal time intervals is given by the following expressions:

$X_s(t) \pm 2S_x$: the mean of $X(t)$ over 300 simulations along with twice the s.e. of the mean.

$\tilde{\overline{X}}(t)$: the approximation for $E(X(t))$ given by equation (3.4.16).

$\overline{Y}_s(t)\pm 2S_y$: the mean of $Y(t)$, the total number of migratations in $[0,t)$, over 300 simulations, along with twice the s.e. of the mean.

$\tilde{Y}(t)$: the approximation for $E(Y(t))$ calculated from equation (3.4.18).

Table 3.2

Comparison of Approximate expressions for Mean state and Mean Number of Migrations with Simulated Values. [For explanation see text].

$N = 2{,}000$	$x_{10} = 500$	$\alpha = 1 \quad \beta = 2$	$\tilde{\bar{X}}(\infty) = 828.43$	$\tilde{V}(\infty) = (15.58)^2$
t	$\bar{X}_s(t) \pm 2s_x$	$\tilde{\bar{X}}(t)$	$\bar{Y}_s(t) \pm 2s_y$	$\tilde{\bar{Y}}(t)$
$\frac{2}{3} \times 10^{-4}$	598.70 ± 1.28	599.00	180.80 ± 1.36	180.56
$\frac{4}{3} \times 10^{-4}$	668.43 ± 1.65	669.05	359.69 ± 1.85	358.60
2×10^{-4}	718.04 ± 1.72	718.14	537.96 ± 2.05	536.58
$N = 20$	$x_{10} = 20$	$\alpha = 1 \quad \beta = 2$	$\tilde{\bar{X}}(\infty) = 8.284$	$\tilde{V}(\infty) = (1.579)^2$
t	$\bar{X}(t) \pm 2s_x$	$\tilde{\bar{X}}(t)$	$\bar{Y}_s(t) \pm 2s_y$	$\tilde{Y}(t)$
3/400	$15.41 \pm .19$	15.44	$4.74 \pm .19$	4.68
6/400	$12.63 \pm .18$	12.77	$8.14 \pm .20$	7.90
9/400	$11.05 \pm .20$	11.14	$10.75 \pm .21$	10.52
12/400	$9.98 \pm .20$	10.12	$13.29 \pm .23$	12.87
$N = 200$	$x_{10} = 200$	$\alpha = 1 \quad \beta = 10$	$\tilde{\bar{X}}(\infty) = 48.05$	$\tilde{V}(\infty) = (13.56)^2$
t	$\bar{X}_s(t) \pm 2s_x$	$\tilde{\bar{X}}(t)$	$\bar{Y}_s(t) \pm 2s_y$	$\tilde{Y}(t)$
2×10^{-4}	$142.75 \pm .65$	143.07	$57.85 \pm .66$	57.44
4×10^{-4}	$111.31 \pm .59$	112.15	$91.47 \pm .61$	90.57
6×10^{-4}	$92.48 \pm .56$	93.22	$114.13 \pm .59$	113.36
8×10^{-4}	$79.86 \pm .53$	80.77	$131.53 \pm .58$	130.95

3.5 The General Non-homogeneous Model (Time Dependent Migration)

The general time dependent migration process is specified by

$$\alpha(x_2,t) = \alpha(t)$$

$$\beta(x_1,t) = \beta(t)$$

where $\alpha(t),\ \beta(t) \geqslant 0$ for $t\varepsilon[0,T]$

This leads to the following partial differential equation for the p.g.f. $P(s,t)$:

$$\frac{\partial P}{\partial t} + [s\alpha(t)+\beta(t)](s-1)\frac{\partial P}{\partial s} - N(s-1)\alpha(t)P = 0$$

As in the linear model, this equation is of the Lagrange form, and is solved using the auxiliary system:

$$\frac{ds}{[s\alpha(t) + \beta(t)](s-1)} = \frac{dt}{1} = \frac{dP}{PN(s-1)\alpha(t)}$$

From the first equation

$$\frac{ds}{dt} = (\alpha s+\beta)(s-1)$$

or, setting $s-1 = 1/z$,

$$\frac{ds}{dt} = -\frac{1}{z^2}\frac{dz}{dt}$$

and

$$\frac{dz}{dt} + [\alpha(t) + \beta(t)]z = -\alpha(t)$$

hence

$$z = \frac{1}{s-1} = [c_1 - A_t]/E_t \qquad (3.5.2)$$

where c_1 is an arbitrary constant

$$A_t = \int_0^t \alpha(u)[\exp\{\int_0^u [\alpha(w) + \beta(w)]dw\}]du$$

$$E_t = \exp\{\int_0^t [\alpha(u) + \beta(u)]du\}$$

From the second equation

$$\frac{dP}{dt} = N\alpha(t)(s-1)P$$

$$\therefore \quad \log P = c_2 + N\int_0^t \alpha(w)(s-1)dw \tag{3.5.3}$$

where c_2 is an arbitrary constant.

Substituting for (s-1) using equation (3.5.2), equation (3.5.3) becomes:

$$\log P = c_2 + N\int_0^t \{\frac{\alpha(w).E_w}{[c_1 - A_w]}\}dw$$

but

$$\frac{\partial A_w}{\partial w} = \alpha(w).E_w$$

hence $$\log P = \log c_3 - N \log[c_1 - A_t] \tag{3.5.4}$$

where $\log c_3$ is an arbitrary constant. Thus,

$$c_3 = P[c_1 - A_t]^N$$

From (3.5.2), we have

$$c_1 = [E_t/(s-1)] + A_t$$

and thus $$c_3 = P[E_t/(s-1)]^N$$

The solution of the partial differential equation is of the form $c_3 = F(c_1)$ where F is an arbitrary function, and so

$$P(s,t)[E_t/(s-1)]^N = F\{[E_t/(s-1)] + A_t\} \tag{3.5.5}$$

At $t = 0$,
$$P(s,0) = s^{x_{10}}$$
$$E_0 = 1$$
$$A_0 = 0$$

and so the function F is of the form

$$F(u) = (1 + u)^{x_{10}} u^{N-x_{10}}$$

Setting $u = [E_t/(s-1) + A_t]$ in $F(u)$ and substituting into 3.5.5 gives the following solution for $P(s,t)$ after some re-arrangement.

$$P(s,\ t) = [(1-p_1) + p_1 s]^{x_{10}}[(1-p_2) + p_2 s]^{N-x_{10}} \tag{3.5.6}$$

where

$$p_1 = (A_t+1)/E_t$$
$$p_2 = A_t/E_t$$

Note that, so long as $0 \leq p_1, p_2 \leq 1$, this distribution is the distribution of the sum of two binomial variates. The form of the p.g.f. is identical to that of the simple linear model (see equation 3.3.3) although p_1 and p_2 are defined differently.

Minimal conditions on the functions $\alpha(t)$, $\beta(t)$ such that the constraint $0 \leqslant p_1, p_2 \leqslant 1$ applies were not determined. However it would seem reasonable that, so long as these two functions are non-negative, and the integrations required by the definitions of A_t and E_t can be performed, the distribution should be well defined and the constraint hold. Some particular forms of this distribution may be of practical importance:

1. When $\alpha(t) = \alpha\rho(t) \quad \beta(t) = \beta\rho(t)$.

In this situation α and β are constants (independent of t) and $\rho(t)$ is an arbitrary non-negative function of time (presumed to be integrable over any time interval. α and β may be chosen such that $\rho(0) = 1$). Thus the system describes a population in which the tendency to migrate from A_1 to A_2 is always proportional to the tendency to migrate in the other direction, but the intensity of migration may fluctuate over time. This may be due to some pattern of behaviour (e.g. high activity during the day, low activity at night) which occurs over time.

In this case we find: $E_t = \exp[f(t)]$

$$A_t = \frac{\alpha}{\alpha+\beta}[\exp[f(t)] - 1]$$

where $f(t) = (\alpha+\beta)\int_0^t \rho(u)du$

hence $p_2 = \alpha\{1 - \exp[-f(t)]\}/(\alpha+\beta)$

$$p_1 = \{\alpha + \beta\exp[-f(t)]\}/(\alpha+\beta)$$

Note that this gives a p.g.f. for this model which is identical to the p.g.f. of the simple linear model (equation 3.3.3) expect that all terms of the form $\exp[-(\alpha+\beta)t]$ are replaced by $\exp[-f(t)]$. Thus all the results of section 3.3 apply to this model by making this substitution for the exponential terms.

2. When $\alpha(t) = \alpha\rho(t) \quad \beta(t) = \alpha[1 - \rho(t)]\ \rho(t_1) \geqslant \rho(t_2) \geqslant 0$

when $t_1 < t_2$

In this situation, $\rho(t)$ is non-negative and monotone decreasing in time. Such a model might describe a population in which the general mobility is constant over time, but shifts from a tendency to migrate in one direction to a tendency to migrate the other. This occurs, for instance, when A_2 represents a feeding area which becomes depleted or unsuitable in time, but at the same time, area A_1 begins to become increasingly attractive, because of a relatively undepleted food source, or because the food 'comes into season' later on A_1 than on A_2. In this case we have $E_t = \exp(\alpha t)$, and $A_t = \int_0^t \alpha\rho(u)\exp(\alpha u)du$.

Thus, it is not possible to proceed further without adopting an explicit functional form for $\rho(u)$. The forms for $\rho(u)$ that are most likely to occur are:

1. exponential: e.g. $\alpha\rho(t) = \alpha\exp[-ct]$
2. sigmoidal: e.g. $\alpha\rho(t) = k[1 - \frac{(k-\alpha)}{(k-\alpha) + \alpha\exp[-ct]}]$

   $k > \alpha > 0 \quad c > 0 \quad \alpha, k, c$ constant

Both these forms can be integrated to form E_t and A_t, and hence analytic solutions for the p.g.f. are obtainable. For the sigmoidal form, $\rho(t)$ approaches 1 asymptotically as t approaches 0 and approaches 0 asymptotically as t becomes large; and $\rho(t)$ is monotone decreasing in t. However many other functions could be given with these same properties.

In section 4.3.4 we describe how time dependent processes may be simulated using the exponential form of $\rho(t)$ above as an example.

CHAPTER FOUR

STOCHASTIC MIGRATION:
MODELS WITH BIRTH AND DEATH OCCURRING

4.1 Introduction

In dealing with actual populations, the restriction of no birth, assumed for the models in the previous section, is not a severely limiting constraint. Even if births are occurring, they are often recognizably different from the older individuals in the population; or births may occur only over a short period, or at a low rate, relative to the period of study. In the first case new births can be recognized and their effect eliminated from any analysis of migration. In the latter case, the experimental period may be chosen so that the effect of births on the population is negligible. Moreover, as is often done in experimental work, one may establish a sub population of marked animals, study their behaviour, and assume it to be representative of the entire population. Additions to the marked population may be made, but these are considered, usually, as a distinct sub-population. Thus "birth" in the natural sense cannot occur into the marked population.

The restriction of no death, on the other hand, is a severely limiting constraint. Even if the study period is very short compared with the life span of individuals, we cannot eliminate the effects of death. It must be remembered that we are dealing with mobile populations, so that death must include those individuals who migrate permanently out of the areas under study. Thus the marked sub populations will also be subject to death.

In this chapter we will develop some methods of dealing with populations which are free to migrate between two areas and are also subject to birth and death.

Birth and death will be taken to be stochastic, as in a birth and death process. As mentioned in Chapter 2, the most general work done for this model is the work of Puri (1968), but he does not develop analytic expressions for the p.g.f. when birth occurs in both populations, and does not give expressions for the means. His work is mainly of use in developing extinction probabilities. In this thesis, interest is mainly on estimating the parameters of the models, so that moments of the distribution assume some importance. Analytic expressions for these are developed in the next section. The model assumed is the linear model of complete independence: that is each individual has a constant probability of undergoing any of the possible changes of state in any interval $[t,t+\delta t]$ and this probability is not a function of t nor of the distribution of individuals between the two areas. This is a bit artificial as it allows for unlimited growth if the birth rate exceeds the death rate. A more realistic model would probably be of the logistic or density dependent type, with the rates changing as density increases. The algebraic difficulties of such a model would be immense; the linear model, however, may be a reasonable description of animal populations with low densities, or over short periods of time. Limiting properties as t becomes large are unlikely to be applicable to real populations except when the populations are declining: i.e. no birth, or death rate exceeding birth rate.

4.2 The Linear Model with Birth Death and Immigration

Because of the introduction of stochastic birth and death and immigration, this process is no longer conservative in the sense of the linear model of section 3.2. That is, we do not have

$$X_1(t)+X_2(t) = x_{10}+x_{20} = N \qquad \text{for all } t > 0$$

It will therefore be necessary to consider the two dimensional process defined by

$$P_{ij}(t) = \mathrm{Prob}[X_1(t) = i,\ X_2(t) = j] \qquad i,\ j = 0,\ 1,\ 2\ \ldots$$

We will assume that birth and death occur in the stochastic fashion defined by the simple birth and death process. That is

Pr (an individual in area A_x gives birth to another individual in area A_x in $[0,\delta t)) = \lambda_x\delta t+0(\delta t)^2$ $\quad x = 1,2.$

Pr (an individual in area A_x dies in $[0,\delta t)) = \mu_x\delta t+0(\delta t)^2$

We will also assume that immigration into either A_1 or A_2 is possible from outside the system, and that arrivals are poisson distributed with mean $\nu_x t$. That is:

Pr (a new individual enters area A_x from outside the system in $[0,\delta t)) = \nu_x\delta t+0(\delta t)^2$

As before, we will also have two way migration within the system defined by:

Pr (an individual in A_2 migrates to A_1 in $[0,\delta t)) = \alpha\delta t+O(\delta t)^2$

Pr (an individual in A_1 migrates to A_2 in $[0,\delta t)) = \beta\delta t+O(\delta t)^2$

Thus the event $[X_1(t+\delta t) = i,\ X_2(t+\delta t) = j]$ can occur in exactly one of the following ways:

$[X_1(t) = i,\ X_2(t) = j]$ and no change occurs in $[t,t+\delta t)$

$[X_1(t) = i,\ X_2(t) = j-1]$ and there is an increase of 1 in A_2 in $[t,t+\delta t)$, by birth or by immigration.

$[X_1(t) = i-1,\ X_2(t) = j+1]$ and there is a migration of 1 individual.

$[X_1(t) = i,\ X_2(t) = j+1]$ and there is one death in A_2 in $[t,t+\delta t)$

$[X_1(t) = i+1,\ X_2(t) = j]$ and there is one death in A_1 in $[t,t+\delta t)$

$[X_1(t) = i+1,\ X_2(t) = j-1]$ and there is a migration of 1 individual from A_1 to A_2 in $[t,t+\delta t)$.

$[X_1(t) = i-1,\ X_2(t) = j]$ and there is an increase of 1 in A_1 in $[t,t+\delta t)$, by birth or by immigration.

More than one change occurs.

The probabilities of each of these changes are given above, so that a difference equation for $P_{ij}(t+\delta t)$ is defined in the usual way (as in section 3.2) and by the usual limiting process, gives the differential difference equations for $P_{ij}(t)$:

$$\frac{\partial P_{ij}(t)}{\partial t} = [-(\nu_1+\nu_2)-j(\lambda_2+\mu_2+\alpha)-i(\lambda_1+\mu_1+\beta)]P_{ij}(t)$$

$$+[\nu_2+(j-1)\lambda_2]P_{i,j-1}(t)+(j+1)\mu_2 P_{i,j+1}(t)$$
$$+[\nu_1+(i-1)\lambda_1]P_{i-1,j}(t)+(i+1)\mu_1 P_{i+1,j}(t)$$
$$+\alpha(j+1)P_{i-1,j+1}(t)+\beta(i+1)P_{i+1,j-1}(t) \qquad (4.2.1)$$

The boundary condition for $P_{00}(t)$ follows from (4.2.1) by considering $P_{i,j}(t) = 0$ for $i,j < 0$.

If the p.g.f. is defined by

$$G(r,s;t) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} r^i s^j P_{ij}(t)$$

(4.2.1) may be transformed to give

$$\frac{\partial G}{\partial t} = G[\nu_2(s-1)+\nu_1(r-1)]+\frac{\partial G}{\partial s}[(s\lambda_2-\mu_2)(s-1)+(r-s)\alpha]$$
$$+\frac{\partial G}{\partial r}[(r\lambda_1-\mu_1)(r-1)+(s-r)\beta] \qquad (4.2.2)$$

It is not possible to integrate this equation analytically. Converting to the cumulant generating function:

$$K(\theta_r,\theta_s;t) = k_{10}(t)\theta_r+k_{01}(t)\theta_s+k_{20}(t)\frac{\theta_r^2}{2!} + \dots$$

where
$$k_{10}(t) = E(X_1(t))$$
$$k_{01}(t) = E(X_2(t))$$
$$k_{20}(t) = \mathrm{var}(X_1(t))$$ etc. gives

$$\frac{\partial K}{\partial t} = \nu_2(e^{\theta_s}-1)+\nu_1(e^{\theta_r}-1)$$
$$+\frac{\partial K}{\partial \theta_s}[(e^{\theta_s}\lambda_2-\mu_2)(1-e^{-\theta_s})+(e^{\theta_r-\theta_s}-1)\alpha]$$
$$+\frac{\partial K}{\partial \theta_r}[(e^{\theta_r}\lambda_1-\mu_1)(1-e^{-\theta_r})+(e^{\theta_s-\theta_r}-1)\beta]$$

Expanding both sides of this equation and identifying coefficients of first and second order terms in θ, gives the following system of differential equations for the means, variances, and covariance.

$$\begin{bmatrix} k'_{10}(t) \\ k'_{01}(t) \\ k'_{20}(t) \\ k'_{11}(t) \\ k'_{02}(t) \end{bmatrix} = \left[\begin{array}{ccc|ccc} \nu_1 & (\lambda_1-\mu_1-\beta) & \alpha & 0 & 0 & 0 \\ \nu_2 & \beta & (\lambda_2-\mu_2-\alpha) & 0 & 0 & 0 \\ \hline \nu_1 & +(\lambda_1+\mu_1+\beta) & \alpha & 2(\lambda_1-\mu_1-\beta) & 2\alpha & 0 \\ 0 & -\beta & -\alpha & \beta & 2(\lambda_1+\lambda_2-\mu_1-\mu_2-\alpha-\beta) & \alpha \\ \nu_2 & \beta & (\lambda_2+\mu_2+\alpha) & 0 & 2\beta & 2(\lambda_2-\mu_2-\alpha) \end{array}\right] \begin{bmatrix} k_{10}(t) \\ k_{01}(t) \\ k_{20}(t) \\ k_{11}(t) \\ k_{01}(t) \end{bmatrix} \tag{4.2.3}$$

This system of equations can be partitioned as indicated and will be denoted by

$$\left[\begin{array}{c} \underline{k}'_1 \\ \hline \underline{k}'_2 \end{array}\right] = \left[\begin{array}{c} \underline{V}_1 \\ \hline \underline{V}_2 \end{array}\right] + \left[\begin{array}{c|c} C_{2x2} & 0 \\ \hline B_{3x3} & K_{3x3} \end{array}\right] \cdot \left[\begin{array}{c} \underline{k}_1 \\ \hline \underline{k}_2 \end{array}\right] \tag{4.2.4}$$

Evidently the equation for the vector of means $\underline{k}_1(t)$ can be given directly as

$$\frac{\partial}{\partial t} \underline{k}_1(t) = \underline{V}_1+[C]\underline{k}_1(t) \tag{4.2.5}$$

The solution of this system of equations will first be given for the situation when $\nu_1 = \nu_2 = 0$.

1. Means when no immigration occurs.

Equation (4.2.4) becomes

$$\left[\begin{array}{c} k'_1 \\ \hline k'_2 \end{array}\right] = \left[\begin{array}{c|c} C_{2x2} & 0 \\ \hline B_{3x2} & K_{3x3} \end{array}\right] \left[\begin{array}{c} k_1 \\ \hline k_2 \end{array}\right] \tag{4.2.6}$$

which is of the form $\frac{\partial}{\partial t} \underline{k}(t) = [A]\underline{k}(t)$ where $\underline{k}(t)$ is a column vector of length n each of whose elements is a function of t, and A is an nxn matrix of elements not involving t.

The solution of this system is

$$\underline{k}(t) = \exp\{[A]t\}\underline{k}(0) \qquad (4.2.7)$$

where $\exp\{[A]t\} = t(I+A + \frac{A.A}{2!} + \frac{A.A.A}{3!} + \ldots)$

This result is given by Finkbeiner (1960), Chapter 10. Moreover, he shows that any power series of matrices converges provided the eigenvalues are bounded in modulus by the scalar r and the scalar power series converges for the value r. Thus exp{[A]t} converges so long as the eigenvalues of A are bounded.

The method of calculating exp{[A]t} will be by means of the method used by Turner et.al. (1963). If ξ_i i = 1, ... n are the eigenvalues of the matrix A, then

$$\exp\{[A]t\} = \sum_{j=1}^{n} \frac{\prod\limits_{i\neq j}^{n} (\xi_i I-A)}{\prod\limits_{i\neq j} (\xi_i-\xi_j)} e^{\xi_j t} \qquad (4.2.8)$$

where I is the n x n identity matrix. We thus require the eigenvalues of the matrix.

$$A = \left|\begin{array}{c|c} C_{2x2} & 0 \\ \hline B_{3x2} & K_{3-3} \end{array}\right|$$

It is a simple matter to verify that

$$|A-\xi I| = |C-\xi I|.|K-\xi I|$$

so that the eigenvalues of [A] are the eigenvalues of [C] along with the eigenvalues of [K]. The equation $|C-\xi I| = 0$ has roots:

$$\xi_{\substack{1\\2}} = \frac{1}{2}[(\lambda_1+\lambda_2-\mu_1=\mu_2-\alpha-\beta)\pm\sqrt{(\lambda_1-\mu_1-\lambda_2+\mu_2-\beta+\alpha)^2+4\alpha\beta}\ ]$$

And the equation $|K-\xi I| = 0$ gives the cubic equation:

$$(\lambda_1-\mu_1+\lambda_2-\mu_2-\alpha-\beta-\xi)[(2\lambda_1-2\mu_1-2\beta-\xi)(2\lambda_2-2\mu_2-2\alpha-\xi)-4\alpha\beta] = 0$$

which has roots:

$$\xi_3 = \lambda_1-\mu_1+\lambda_2-\mu_2-\alpha-\beta$$

$$\xi_{\substack{4\\5}} = \xi_3\pm\sqrt{(\lambda_1-\mu_1-\lambda_2+\mu_2+\beta-\alpha)^2+4\alpha\beta}$$

These eigenvalues are more simply expressed in terms of

$$\gamma_1 = \lambda_1-\mu_1$$

$$\gamma_2 = \lambda_2-\mu_2$$

and are:

$$\gamma_1+\gamma_2-\alpha-\beta$$

$$\frac{1}{2}(\gamma_1+\gamma_2-\alpha-\beta\pm\sqrt{(\gamma_1-\gamma_2+\alpha-\beta)^2+4\alpha\beta})$$

$$\gamma_1+\gamma_2-\alpha-\beta\pm\sqrt{(\gamma_1-\gamma_2-\alpha+\beta)^2+4\alpha\beta} \qquad (4.2.9)$$

when $\gamma_1 = \gamma_2 = \gamma$ these simplify to:

$$\xi_3 = 2\gamma-\alpha-\beta$$

$$\xi_1 = \gamma-\alpha-\beta$$

$$\xi_2 = \gamma$$

$$\xi_4 = 2\alpha$$

$$\xi_5 = 2(\gamma-\alpha-\beta) \qquad (4.2.10)$$

Whether or not $\gamma_1 = \gamma_2$, the roots are all evidently of finite modulus, so that the solution (4.2.5) for the case of no immigration is:

$$\underline{k}_1'(t) = [C]\underline{k}_1(t)$$

and has solution

$$\underline{k}_1(t) = \exp\{[C]t\}\underline{k}_1(0)$$

where $\underline{k}_1(0) = (x_{10}, x_{20})'$.

Using (4.2.8) with $r = \xi_1 - \xi_2 = \sqrt{(\gamma_1-\gamma_2+\alpha-\beta)^2+4\alpha\beta}$ gives the expression for $\exp\{[C]t\}$.

$$= \frac{e^{\xi_2 t}}{r}\begin{bmatrix} (\gamma_2-\gamma_1-\alpha+\beta+r)/2 & -\alpha \\ -\beta & (\gamma_1-\gamma_2+\alpha-\beta+r/2 \end{bmatrix}$$

$$- \frac{e^{\xi_1 t}}{r}\begin{bmatrix} (\gamma_2-\gamma_1-\alpha+\beta-r)/2 & -\alpha \\ -\beta & (\gamma_1-\gamma_2+\alpha-\beta-r)/2 \end{bmatrix}$$

Thus, since $e^{\xi_1 t} = e^{r} e^{\xi_2 t}$

$$E(X_1(t)) = \frac{e^{\xi_2 t}}{r}[\frac{x_{10}}{2}[(\gamma_2-\gamma_1-\alpha+\beta)(1-e^{rt})+r(1+e^{rt})]-x_{20}\alpha(1-e^{-rt})]$$

$$E(X_2(t)) = \frac{e^{\xi_2 t}}{r}[\ \frac{x_{20}}{2}[(\gamma_1-\gamma_2-\beta+\alpha)(1-e^{rt})+r(1+e^{rt})]-x_{10}\beta(1-e^{-rt}] \tag{4.2.11}$$

It should be noted that $E(X_2(t))$ can be derived from $E(X_1(t))$ by replacing α by β, β by α, and changing the subscripts 1 and 2 to 2 and 1 respectively. This is as we would expect, since switching the parameters relating to area A_1 to those relating to A_2 and vice versa we would expect that this "mirror image" system should end up with the same properties only with the role of A_1 and A_2 switched. Since this property must always hold for both the means and the variance, we will henceforward quote be expression for A_1 only.

In the particular case $\lambda_1 = \lambda_2 \quad \mu_1 = \mu_2$, the expression (4.2.11) simplifies considerably to:

$$E(X_1(t)) = \frac{e^{(\lambda-\mu)t}}{(\alpha+\beta)}[x_{10}(\alpha+\beta e^{-(\alpha+\beta)t})+x_{20}\alpha(1-e^{-(\alpha+\beta)t})]$$

The similarity to the mean of the linear process of section 3.2 is obvious.

2. The Variances and Covariance when no Immigration occurs

We will only consider the case:

$$\lambda_1 = \lambda_2 \quad \mu_1 = \mu_2$$
$$\gamma = \lambda-\mu_1$$

because of the algebraic difficulties encountered in the more general case. Since we already have the expression for $\underline{k}_1(t)$, the vector of means, the equation for the vector $\underline{k}_2(t)$ of variances and the covariance between $x_1(t)$ and $x_2(t)$, may be expressed as

$$\underline{k}_2'(t) = [B]\underline{k}_1(t)+[K]\underline{k}_2(t) \qquad (4.2.13)$$

where B and K are the matrices of expression (4.2.6). Pre-multiplying both sides of (4.2.13) by the matrix $\exp\{-t[K]\}$ gives

$$\exp\{-t[K]\}\underline{k}_2'(t)-\exp\{-t[K]\}[K]\underline{k}_2(t) = \exp\{-t[K]\}[B]\underline{k}_1(t) \qquad (4.2.14)$$

It may be shown by using the theorems of Finkbeiner, Chapter 10, that the left hand side represents the differential:

$$\frac{d}{dt}\exp\{-t[K]\}\underline{k}_2(t)$$

It is obvious then, that the elements of the vector represented by the product of the terms on the right hand side of this equation may be integrated term by term. This process can be denoted by

$$\int_0^t \exp\{-u[K]\}[B]\underline{k}_1(u)du. \quad +\underline{c}$$

where c is the vector of constants (0,0,0)' determined from the fact that $\underline{k}_2(0) = (0,0,0)'$.

Thus integrating both sides of (4.2.12) gives:

$$\underline{k}_2(t) = \exp\{[K]t\}\int_0^t \exp\{-t[K]\}[B]\underline{k}_1(u)du \qquad (4.2.15)$$

Thus we require the matrix $\exp\{[K]t\}$ for the matrix

$$[K] = \begin{bmatrix} 2(\alpha-\beta) & 2\alpha & 0 \\ \beta & 2\gamma-\alpha-\beta & \alpha \\ 0 & 2\beta & 2(\gamma-\alpha) \end{bmatrix}$$

whose eigenvalues are $2\gamma-\alpha-\beta$, 2γ and $2(\gamma-\alpha-\beta)$. The desired matrix is easily derived following Turner et al's method (e.g. 4.2.8). Now if we set:

$$e_1 = \exp[2(\gamma-\alpha-\beta)t]$$

$$e_2 = \exp[(2\gamma-\alpha-\beta)t]$$

$$e_3 = \exp[2\gamma t]$$

then $\exp\{[K]t\}$

$$= \frac{1}{(\alpha+\beta)^2}\begin{bmatrix} \beta^2e_1+2\alpha\beta e_2+\alpha^2e_3 & -2\alpha\beta e_1-2\alpha(\alpha-\beta)e_2+2\alpha^2e_3 & \alpha^2e_1-2\alpha^2e_2+\alpha^2e_3 \\ -\beta^2e_1+\beta(\beta-\alpha)e_2+\alpha\beta e_3 & 2\alpha\beta e_1+(\alpha-\beta)^2e_2+2\alpha\beta e_3 & -\alpha^2e_1+\alpha(\alpha-\beta)e_2+\alpha\beta e_3 \\ \beta^2e_1-2\beta^2e_2+\beta^2e_3 & -2\alpha\beta e_1-2\beta(\beta-\alpha)e_2+2\beta^2e_3 & \alpha^2e_1+2\alpha\beta e_2+\beta^2e_3 \end{bmatrix} \tag{4.2.16}$$

To form the integral, we also require the matrix

$$\exp\{-t[K]\}$$

Since the eigenvalues of K are distinct and non zero, the matrix [K] may be expressed in the spectral form

$$[K] = [Q^{-1}D_\xi Q]$$

where D_ξ is the diagonal matrix of eigenvalues, and Q is the row modal matrix of [K]. Manipulation of these matrices will then show that

$$\begin{aligned} \exp\{[K]t\} &= \exp[Q^{-1}D_\xi Q]t \\ &= \exp[Q^{-1}D_{\xi t}Q] \\ &= Q^{-1}D_{e\xi t}Q \end{aligned}$$

and

$$\begin{aligned} \exp\{-t[K]\} &= [\exp\{[K]t\}]^{-1} \\ &= [Q^{-1}D_{e\xi t}Q]^{-1} \\ &= Q^{-1}D_{e-\xi t}Q \end{aligned}$$

Thus it is evident that $\exp\{-t[K]\}$ is identical to (4.2.16) except that e_i is replaced by $e_{-i} = [e_i]^{-1}$. The integral on the right of (4.2.15) may now be developed. The development is facilitated by expressing the matrix [B] as the sum of the two matrices :

$$[B] = (\lambda+\mu)\begin{bmatrix}1 & 0\\ 0 & 0\\ 0 & 1\end{bmatrix}+\begin{bmatrix}\beta & \alpha\\ -\beta & -\alpha\\ \beta & \alpha\end{bmatrix}$$

$$= [B_1]+[B_2] \quad \text{(say)}$$

It happens that vectors with elements in the proportions $(1 \ \ -1 \ \ 1)'$ are eigenvectors of the $\exp\{-t[K]\}$ corresponding to the eigenvalue $\exp[2(\gamma-\alpha-\beta)t]$, and thus considerable simplification results by expressing $[B]$ in this way.

The integration $\int_0^t \exp\{-t[K]\}[B_1]\underline{k}_1(u)du$
$+\int_0^t \exp\{-t[K]\}[B_2]\underline{k}_2(u)du$

can be carried out and gives:

$$\frac{\lambda+\mu}{(\alpha+\beta)^2}\begin{bmatrix} c_1e_1+c_2e_2+(x_{10}\alpha^2-x_{20}\alpha^2+2x_{10}\alpha\beta)e_3/\gamma \\ -c_1e_1-c_2e_2+(x_{10}\beta^2+x_{20}\alpha^2)e_3/\gamma \\ c_1e_1+c_2e_2-(x_{20}\beta^2-x_{10}\beta^2+2x_{20}\alpha\beta)e_3/\gamma \end{bmatrix}$$

where

$$c_1 = \frac{(x_{10}+x_{20})\alpha\beta(\lambda+\mu+2\alpha+2\beta)}{(\lambda+\mu)(-\gamma+2\alpha+2\beta)} \qquad e_1 = (e^{(-\gamma+2\alpha+2\beta)t}-1)$$

$$c_2 = \frac{(x_{10}\beta-x_{20}\alpha)(\beta-\alpha)(\lambda+\mu+\alpha+\beta)}{(\lambda+\mu)(-\gamma+\alpha+\beta)} \qquad e_2 = (e^{(-\gamma+\alpha+\beta)t}-1)$$

$$e_3 = (e^{-\gamma t}-1)$$

The final expression for $\underline{k}_2(t)$ is now derived by pre-multiplying by $\exp\{[K]t\}$. Again this is facilitated by the fact that the vector $(1 \ \ -1 \ \ 1)'$ is the eigenvector corresponding to the eigenvalue $\exp[2(\gamma-\alpha-\beta)t]$. The final result is

$$\mathrm{var}(X_1(t)) = \frac{\lambda+\mu}{(\alpha+\beta)^2}\{c_1(e^{\gamma t}-e^{2(\gamma-\alpha-\beta)t})+c_2(e^{(\gamma-\alpha-\beta)t}-e^{2(\gamma-\alpha-\beta)t})$$

$$+\frac{2\alpha(x_{10}\beta-x_{20}\alpha)}{\gamma}(e^{(2\gamma-\alpha-\beta)t}-e^{(\gamma-\alpha-\beta)t})+\frac{N\alpha^2}{\gamma}(e^{2\gamma t}-e^{\gamma t})\} \qquad (4.2.17)$$

$$\mathrm{cov}(X_1(t),X_2(t)) = -\frac{(\lambda+\mu)}{(\alpha+\beta)^2}\{c_1(e^{\gamma t}-e^{2(\gamma-\alpha-\beta)t})+c_2(e^{(\gamma-\alpha-\beta)t}-e^{2(\gamma-\alpha-\beta}$$

$$\frac{(\alpha-\beta)(x_{10}\beta-x_{20}\alpha)}{\gamma}(e^{(2\gamma-\alpha-\beta)t}-e^{(\gamma-\alpha-\beta)t})+\frac{N\alpha\beta}{\gamma}(e^{2\gamma t}-e^{\gamma t})\} \qquad (4.2.18)$$

The result for $\mathrm{var}(X_2(t))$ was, as expected, the "mirror image" of $\mathrm{var}(X_1(t))$. The expressions given above were verified by substituting back into the original equation (4.2.13) and were seen to satisfy the identity.

As mentioned in section 4.1, the situation when death occurs, but no birth, may be of some importance in marking experiments. In this case $\lambda = 0 \quad \gamma = -\mu$ and the expression for $\mathrm{var}(X_1(t))$ simplifies (after rearranging terms) to:

$$\mathrm{var}(X_1(t)) = \frac{x_{10}e_1}{(\alpha+\beta)^2}[\alpha+\beta e_2][\alpha+\beta-e_1(\alpha+\beta e_2)]$$

$$+\frac{x_{20}e_1}{(\alpha+\beta)^2}[\alpha-\alpha e_2][\alpha+\beta-e_1(\alpha-\alpha e_2)]$$

where $e_1 = \exp[-\mu t] \quad e_2 = \exp[-(\alpha+\beta)t]$. Note that this is the form of the variance of the sum of two binomially distributed variables. The covariance also simplifies to the form of the covariance of the sum of two binomial variables:

$$\mathrm{cov}(X_1(t)X_2(t))$$

$$= -\frac{x_{10}e_1^2}{(\alpha+\beta)^2}[\alpha+\beta e_2][\beta(1-e_2)]$$

$$-\frac{x_{20}e_1^2}{(\alpha+\beta)^2}[\alpha(1-e_2)][\beta+\alpha e_2].$$

3. Means, Variances and Covariances when Immigration Occurs

The full equation for the means $\underline{k}(t) = (k_1(t), k_2(t))'$ is given by equation (4.2.5):

$$\frac{\partial}{\partial t}\,\underline{k}(t) = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} \gamma_1-\beta & \alpha \\ \beta & \gamma_2-\alpha \end{bmatrix}\underline{k}(t)$$

$$= \underline{v}+[C]\underline{k}(t)$$

We shall consider only the situation where $\gamma_1 = \gamma_2 = \gamma$. The above equation is of the same form as the equation for the variances (4.2.13), so its solution is

$$\underline{k}(t) = \exp\{[C]t\}(\int_0^t \exp\{-u[C]\}\underline{v}\,du+\underline{k}(0))$$

This yields the expression for the mean

$$X_1(t) = \frac{1}{\alpha+\beta}[\frac{\alpha(\nu_1+\nu_2)}{\gamma}(e^{\gamma t}-1)+\frac{(\beta\nu_1-\alpha\nu_2)}{(\gamma-\alpha-\beta)}(e^{(\gamma-\alpha-\beta)t}-1)$$

$$+(x_{10}\beta-x_{20}\alpha)e^{(\gamma-\alpha-\beta)t}+(x_{10}+x_{20})\alpha e^{\gamma t}] \qquad (4.2.20)$$

This expression evidently reduces to (4.2.12) when $\nu_1 = \nu_2 = 0$, as we would expect.

The equation for the variances, in the notation of (4.2.4) is

$$\frac{\partial}{\partial t}\,\underline{k}_2'(t) = [K]\underline{k}_2(t) = \underline{v}_2+[B]\underline{k}_1(t)$$

Again, this is of the form of the equation for the variances when no immigration occurs. It was solved in a similar manner, but the expression for the variances and covariances is so exceedingly cumbersome that the expression will not be given here.

4. Moments for the Number of Births, Deaths, etc.

If we define the multidimensional process

$$P_{\{x_i\}}(t) = P_{x_1x_2\ldots x_8}(t)$$

$$= \mathrm{Prob}[X_1(t) = x_1;\ x_2(t) = x_2;$$

number of births plus immigrants into A_1 in $[0,t) = x_3$

number of births plus immigrants into A_2 in $[0,t) = x_4$

number of deaths in A_1 in $[0,t) = x_5$

number of deaths in A_2 in $[0,t) = x_6$

number of transfers from A_1 to A_2 in $[0,t) = x_7$

number of transfers from A_2 to A_1 in $[0,t) = x_8]$

Then it is a simple matter to form the p.g.f. and hence the cumulant generating function $K(\{\theta^{x_i}\};t)$. The differential equation for K is:

$$\begin{aligned}\frac{\partial K}{\partial t} &= [\nu_1(e^{\theta_{x1}+\theta_{x3}}-1)+\nu_2(e^{\theta_{x2}+\theta_{x4}}-1)] \\ &\quad + \frac{\partial K}{\partial \theta_{x_1}}[\lambda_1(e^{\theta_{x1}+\theta_{x3}}-1)+\mu_1(e^{\theta_{x5}-\theta_{x1}}-1)+\beta(e^{\theta_{x2}+\theta_{x7}-\theta_{x1}}-1)] \\ &\quad + \frac{\partial K}{\partial \theta_{x_2}}[\lambda_2(e^{\theta_{x2}+\theta_{x4}}-1)+\mu_2(e^{\theta_{x6}-\theta_{x2}}-1)+\alpha(e^{\theta_{x1}+\theta_{x8}-\theta_{x2}}-1)]\end{aligned}$$

If this equation is expanded and linear terms identified on both sides, the following results for the means are proved.

Let $\quad M_x(t) = \int_0^t X_x(u)du \qquad x = 1,\ 2$

The expected number of new entries in A_x (by birth or immigration) in $[0,t)$

$$= \nu_x t + \lambda_x M_x(t)$$

Since immigration is by poisson distributed arrivals with mean $\nu_x t$, the expected numbers of births in $[0,t)$

$$= \lambda_x M_x(t)$$

Expected number of deaths in A_x in $[0,t)$

$$= \mu_x M_x(t).$$

Expected number of transfers from A_1 to A_2 in $[0,t)$

$$= \beta M_1(t)$$

Expected number of transfers from A_2 to A_1 in $[0,t)$

$$= \alpha M_2(t)$$

Terms of second order in θ lead to equations for the variances of these variables that are in the matrix differential equation form of (4.2.13). Their solution will not be reproduced here.

4.3 A General Formulation for Models with Independent Migration

When migration of the individuals is independent, (that is, the probability that any individual migrates in $[t,t+\delta t)$ is independent of the location of the remaining individuals in the population), a probability distribution for the individuals at time t, given initial conditions at time 0, can be developed without resorting to the markov process models. The approach is to define this probability distribution in terms of the probabilities:

$$Q_{ij} = \text{Prob [an individual in area } A_i \text{ at time 0 is in area } A_j \text{ at time t].}$$

If the population is closed, then

$$\sum_j Q_{ij} = 1$$

Death is easily accounted for by the generalisation

$$Q_{ij} = \text{Prob [an individual alive in area } A_i \text{ at time 0 is alive and in area } A_j \text{ at time t].}$$

$$1-\sum_j Q_{ij} = \text{Prob [an individual alive in area } A_i \text{ at time 0 has died by time t].}$$

The models are completely stochastic, so we would expect them to lead to distributions which are related in some way to the distributions defined by the markov processes for independent systems (as developed in sections 3.3, 3.5, 4.2). We will demonstrate this relation. The advantage of the models of this section is that their simplicity will allow us to write down the P.G.F. for many distributions which are not amenable to development using the theory of markov processes. The main disadvantage is that the Q_{ij} are not explicit functions of t, and so can only be used to describe the distribution at a single point in time, t, and are not useful for predicting a future observations, say $t+\tau$, unless a new set of Q_{ij} related to the interval $[t,t+\tau]$ are estimated. In other words, in estimation problems, the number of parameters increases with the number of observation times, whereas with the markov process models, this is not the case.

1. Migration between two closed areas:

$$\begin{aligned} Q_{11}+Q_{12} &= 1 \\ Q_{22}+Q_{21} &= 1 \end{aligned} \qquad (4.3.1)$$

If at time 0 there are x_{10} individuals in A_1, then the number of these that are in A_1 at time t, say x_{1t} has distribution:

$$B(x_{1t};x_{10},Q_{11})$$

because of the independence properties. Also if there are x_{20} individuals in A_2 at time 0, then the number of these individuals in A, at time t, say x_{2t}, has distribution

$$B(x_{2t};x_{20},Q_{21})$$

Therefore $x = x_{1t}+x_{2t}$ is the sum of two independent binomially distributed variates and so has p.g.f.

$$G(r) = (Q_{12}+Q_{11}r)^{x_{10}}(Q_{22}+Q_{21}r)^{x_{20}} \qquad (4.3.2)$$

The similarity with the p.g.f. for the linear closed model (equation 3.3.3) is immediately evident, and the relation is made explicit by noting that

$$Q_{11} = \text{prob}[x_{1t} = 1|x_{10} = 1,x_{20} = 0]$$

and the distribution of x_{1t} is given by equation (3.3.7) with $x_{10} = 1$

$$\therefore\ Q_{11} = \binom{1}{1} \frac{\alpha+\beta e^{-(\alpha+\beta)t}}{\alpha+\beta}$$

also $Q_{21} = \text{prob}[x_{2t} = 1 | x_{20} = 1, x_{10} = 0]$ and the distribution of x_{2t} is given by equation (3.3.6) with $x_{20} = 1$

$$\text{i.e.}\qquad Q_{21} = \binom{1}{1} \frac{\alpha(1-e^{-(\alpha+\beta)t})}{(\alpha+\beta)}$$

Substitution of these values for Q_{21}, Q_{11}, and using (4.3.1) to define Q_{12} and Q_{22} in (4.3.2) gives the p.g.f. of the markov process (3.3.3) exactly.

2. Migration between two areas with death occurring

Let $G(r_1 r_2 r_3)$ be the generating function

$$G(r_1 r_2 r_3) = \sum_{x_1 x_2 x_3} P_{x_1 x_2 x_3} r_1^{x_1} r_2^{x_2} r_3^{x_3}$$

where $P_{x_1 x_2 x_3}$ = Prob [A_1 has x_1 individuals at time t, A_2 has x_2 individuals and $x_3 = x_{10}+x_{20}-x_1-x_2$ individuals have died in [0,t)]

then it is evident that the vector $(x_1 x_2 x_3)$ is the sum of two vectors representing the distribution of the x_{10} individuals at time t, and the x_{20} individuals at time t. As each of these has the trinomial distribution, we can immediately write down:

$$\begin{aligned} G(r_1 r_2 r_3) = {} & (Q_{11}r_1+Q_{12}r_2+(1-Q_{11}-Q_{12})r_3)^{x_{10}} \\ & (Q_{21}r_1+Q_{22}r_2+(1-Q_{21}-Q_{22})r_3)^{x_{20}} \qquad (4.3.3.) \end{aligned}$$

An immediate consequence of the form of G [as the p.g.f. of the sum of two independent trinomial distributions] is that:

$$E(x_1) = x_{10}Q_{11}+x_{20}Q_{21}$$
$$E(x_2) = x_{10}Q_{12}+x_{20}Q_{22}$$
$$\text{cov}(x_1x_2) = -x_{10}Q_{11}Q_{12}-x_{20}Q_{22}Q_{21}$$
$$\text{var}(x_1) = x_{10}Q_{11}(1-Q_{11})+x_{20}Q_{21}(1-Q_{21})$$
$$\text{var}(x_2) = x_{10}Q_{12}(1-Q_{12})+x_{20}Q_{22}(1-Q_{22})$$

These expressions may be compared with the expressions for mean and variance of the model of the previous section, with $\lambda_1 = \lambda_2 = 0$ and $\mu_1 = \mu_2$. (Equation (4.2.12 with $\lambda = 0$, and 4.2.19, respectively). This suggests that the equivalence with the markov process model is demonstrated if

$$Q_{11} = \frac{\alpha e^{-\mu t}}{(\alpha+\beta)}(\alpha+\beta e^{-(\alpha+\beta)t})$$

$$Q_{21} = \frac{\alpha e^{-\mu t}}{(\alpha+\beta)}(1-e^{-(\alpha+\beta)t})$$

Comparison of the expressions for $E(x_2)$ further suggests

$$Q_{22} = \frac{\alpha e^{-\mu t}}{\alpha+\beta}(\beta+\alpha e^{-(\alpha+\beta)t})$$

$$Q_{12} = \frac{\alpha e^{-\mu t}}{\alpha+\beta}(1-e^{-(\alpha+\beta)t})$$

Substitution of these into (4.3.3) gives the p.g.f. of the model in terms of the parameters α, β and the time t. If this is the p.g.f. of the markov process of section 4.2 (with no birth or immigration and equal death rates in A_1 and A_2), then this p.g.f. (with $r_3 = 0$) must satisfy the differential equation for the markov process p.g.f. (4.2.2

with $\nu_1 = \nu_2 = \lambda_1 = \lambda_2 = 0$ and $\mu = \mu_1 = \mu_2$). It was verified that this is so.

Note that in developing (4.3.3) there was no assumption made that death rates were equal. Thus the generating function defined by 4.3.3 also describes the markov process defined by equation 4.2.2 with $\nu_1 = \nu_2 = \lambda_1 = \lambda_2 = 0$ and $\mu_1 \neq \mu_2$. The expressions for the Q_{ij} in terms of the parameters α, β, μ_1, μ_2 are complex, however, as will be seen below. Nevertheless, for certain types of experiments, the Q_{ij} can be estimated (as in Chapman and Junge (1956), and Darroch (1961))for given t even though it is difficult to obtain estimates of the parameters α, β, μ_1, μ_2.

3. Migration Among Many Areas:

Let there be n areas, $A_1 \ldots A_n$. Initially, assume area A_i contains x_{i0} individuals. Q_{ij} i, j = 1 ... n are defined as at the beginning of this section, and migration is independent. Death may be accounted for by considering an additional area A_{n+1} such that

$$\left.\begin{aligned} Q_{i\ n+1} &\geqslant 0 \\ Q_{n+1,i} &= 0 \end{aligned}\right\} \text{ for } i = 1,\ldots n$$

Then at time t, each of the groups of individuals that were in x_{i0} at time 0, will have a multinomial distribution over the areas $A_1 \ldots A_n$ with corresponding probabilities Q_{ij} j = 1 ... n. The p.g.f. for each of these multinomial distributions is thus

$$(Q_{i1}r_1+Q_{i2}r_2+Q_{i3}r_3\ \dots\ Q_{in}r_n)^{x_{i0}} \qquad i = 1 \dots n$$

And each of these n multinomial (vector) variates is independent. The state of the process at time t is given by the sum of these n variates and so the p.g.f. of the system at time t given the x_{i0} is

$$G(r_1 \dots r_n) = \prod_{j=1}^{n} (Q_{i1}r_1+Q_{i2}r_2+\dots Q_{in}r_n)^{x_{i0}} \qquad (4.3.4)$$

The equivalent markov process model, as mentioned in section 2.2, was developed by Bartlett (1949). It is defined in terms of the parameters

$\gamma_{ij}\delta t$ = Prob (an individual in area A_i at time t migrates to area A_j in the interval $[t,t+\delta t)$).

Bartlett gives the p.g.f. $G(r_1r_2 \dots r_n;t)$ as

$$\frac{\partial G}{\partial t} = \sum_i \sum_{j\neq i} \gamma_{ij}(r_j-r_i)\frac{\partial G}{\partial r_i} \qquad (4.3.5)$$

To obtain the relation between the Q_{ij} and the γ_{ij}, we convert this p.g.f. to the m.g.f. $M(s_1 \dots s_n;t) = G(e^{s_1} \dots e^{s_n};t)$, using the identities

$$\frac{\partial M}{\partial t} = \frac{\partial G}{\partial t} \qquad \frac{\partial M}{\partial S_i}\, e^{S_i} = \frac{\partial G}{\partial r_i}$$

Therefore

$$\frac{\partial M}{\partial t} = \sum_i \sum_{j\neq i} \gamma_{ij}(e^{s_j-s_i}-1)\frac{\partial M}{\partial s_i}$$

Now, following the example of Bharucha-Reid mentioned in section 2.2, the probability Q_{ij} is given by the expected number of individuals in A_j given the initial condition; at time 0, there is one individual in A_i, and no individuals in all other areas A_j $j \neq i$. We therefore require an expression for $m_i(t)$, the mean in area A_i at time t. By expanding the above partial differential equation and identifying coefficients of s_i, we find

$$\frac{\partial m_i(t)}{\partial t} = -m_i(t) \sum_{j \neq i} \gamma_{ij} + \sum_{j \neq i} \gamma_{ji} m_j(t)$$

If $\underline{m}(t)$ is the column vector of means, then

$$\frac{\partial}{\partial t}\underline{m}(t) = [\Gamma]\underline{m}(t) \qquad (4.3.5)$$

where $[\Gamma]$ is the matrix

$$\begin{bmatrix} -\sum\limits_{j \neq 1} \gamma_{1j} & \gamma_{21} & \gamma_{31} & \cdots & \gamma_{n1} \\ \gamma_{12} & -\sum\limits_{j \neq 2} \gamma_{2j} & \gamma_{32} & \cdots & \gamma_{n2} \\ \gamma_{13} & \gamma_{23} & -\sum\limits_{j \neq 3} \gamma_{3j} & \cdots & \gamma_{n3} \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{1n} & \gamma_{2n} & \gamma_{3n} & \cdots & -\sum\limits_{j \neq n} \gamma_{nj} \end{bmatrix}$$

It was mentioned in the previous section, that equation (4.3.5) has solution

$$\underline{m}(t) = \exp\{-[\Gamma]t\}\underline{m}(0)$$

Thus, if $\underline{m}(0) = \underline{1}i$ where $\underline{1}i$ is the column vector of length n, with a 1 in row i and zeros elsewhere, and if $\underline{Q}_i = (Q_{i1}, Q_{i2} \cdots Q_{in})'$

$$\therefore \underline{Q}_i = \exp\{-[\Gamma]t\}\underline{1}_i \qquad (4.3.6)$$

Now it was also shown in the previous section, that

$$\exp\{-[\Gamma]t\} = Q^{-1}D_{e^{-\xi t}}Q \qquad (4.3.7)$$

where Q is the row modal matrix, and D is the diagonal matrix with diagonal elements the negative exponentials of the eigenvalues of the matrix $t[\Gamma]$. If the i, jth element of Q is q_{ij} and of Q^{-1} is q^{ij}, then expanding (4.3.7), substituting in (4.3.6) and identifying terms, it follows that

$$\mathbb{Q}_{ij} = \sum_{s=1}^{n} q^{js} q_{si} e^{-\xi_s t} \qquad (4.3.8)$$

Substitution of these values in (4.3.4) gives the general solution for the markov process defined by Bartlett. This result however is both more explicit, and for more general initial conditions than the result given by Bartlett (1949).

For the closed model for migration between two areas,

$$\Gamma = \begin{bmatrix} -\beta & \alpha \\ \beta & -\alpha \end{bmatrix}$$

and the spectral decomposition

$$Q\Gamma Q^{-1} = D_{\xi} \qquad \text{is}$$

$$\begin{bmatrix} 1 & 1 \\ -\beta & \alpha \end{bmatrix} \begin{bmatrix} -\beta & \alpha \\ \beta & -\alpha \end{bmatrix} \begin{bmatrix} \alpha/(\alpha+\beta) & -1/(\alpha+\beta) \\ \beta/(\alpha+\beta) & 1/(\alpha+\beta) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -(\alpha+\beta) \end{bmatrix}$$

thus

$$\begin{aligned} \mathbb{Q}_{11} &= \frac{\alpha}{\alpha+\beta}.1.e^{\circ} + \frac{-1}{(\alpha+\beta)}.\beta.e^{-(\alpha+\beta)t} \\ &= \frac{\alpha-\beta e^{-(\alpha+\beta)t}}{(\alpha+\beta)} \qquad \text{as before} \end{aligned}$$

The other values for Q_{ij} also are found to be the same as before.

The extension to three areas in general causes analytic difficulties as the eigenvalues of Γ are not simple expressions. The particular case where $\gamma_{12} = \beta$ $\gamma_{21} = \alpha$ $\gamma_{13} = \mu_1$ $\gamma_{23} = \mu_2$, $\gamma_{31} = \gamma_{32} = 0$ represents the migration process between two areas, A_1 and A_2, and unequal death rates in the two areas. In this case, the matrix Γ is

$$\begin{bmatrix} -(\beta+\mu_1) & \alpha & 0 \\ \beta & -(\alpha+\mu_2) & 0 \\ \mu_1 & \mu_2 & 0 \end{bmatrix}$$

and has eigenvalues

$$\xi_1 = 0$$

$$\xi_{\substack{2\\3}} = -\frac{1}{2}\left(\alpha+\beta+\mu_1+\mu_2 \mp \sqrt{(\alpha+\beta+\mu_2-\mu_1)^2-4\beta(\mu_2-\mu_1)}\right)$$

The spectral representation of Γ is

$$\begin{bmatrix} 1 & 1 & 1 \\ c-\mu_1-\sqrt{d} & c-\mu_2-\sqrt{d} & 0 \\ c-\mu_1+\sqrt{d} & c-\mu_2+\sqrt{d} & 0 \end{bmatrix} [\Gamma] \quad \frac{1}{2\sqrt{d}(\mu_2-\mu_1)} \begin{bmatrix} 0 & c-\mu_2+\sqrt{d} & -c+\mu_2+\sqrt{d} \\ 0 & -c+\mu_1-\sqrt{d} & c-\mu_1-\sqrt{d} \\ 2\sqrt{d}(\mu_2-\mu_1) & \mu_2-\mu_1 & \mu_1-\mu_2 \end{bmatrix} = [D_\xi].$$

where $c = \frac{1}{2}(\alpha-3\beta-\mu_1-\mu_2)$

$$\sqrt{d} = \sqrt{(\alpha+\beta-\mu_1+\mu_2)^2/4-\beta(\mu_2-\mu_1)}$$

Expressions for the Q_{ij} can thus be derived in terms of α, β, μ_1 and μ_2, but they are very cumbersome. In general, it will not be possible to get convenient expressions for the $\mathbb{Q}_{ij}$ in terms of the γ_{ij} when three or more areas are considered. As will be mentioned in the next section, however, it is possible to get estimates for the $\mathbb{Q}_{ij}$ for a given experiment involving marking at time zero and recapturing at time t. Thus, it is possible to derive a numerical estimate for the matrix $[\mathbb{Q}]$ where this is the matrix with ijth element $\mathbb{Q}_{ij}$. Now from (4.3.6) we have

$$[\mathbb{Q}] = \exp\{-[\Gamma]t\}$$

$$\therefore \ [\Gamma] = -\frac{1}{t}\log[\mathbb{Q}]$$

where $\log[\mathbb{Q}]$ is defined by a suitable power series in $[\mathbb{Q}]$ chosen such that the scalar power series converges for the modulus of all eigenvalues of $[\mathbb{Q}]$. It is thus possible to derive a numerical estimate of $[\Gamma]$. By calculating the eigenvalues, and eigenvectors of $[\Gamma]$, one can then use (4.3.8) to derive the expressions for the $\mathbb{Q}_{ij}$ as explicit functions of t. We have thus used the information from observations at only two points in time to derive the p.g.f. for general t, and can use this p.g.f. to make statements about the probability distribution for the system at any time in the interval $[0,t)$, or to make predictions about the system at a later time $t+\tau$. Uncertainty about the $\mathbb{Q}_{ij}$ due to the estimation procedure may be incorporated in such probability statements by repeating the process on a computer over and over, using values of the $\mathbb{Q}_{ij}$ to which a suitable error term has been added.

4. Time Dependent Migration Between Two Areas.

The only restriction imposed in developing the p.g.f. for the distribution of $X_1(t)$ as given in equation (4.3.2) is that individuals migrate independently. This condition holds for the time dependent models of section 3.5, and the form of the parameters $\mathbb{Q}_{ij}$ is easily identified (in terms of $\alpha(t)$, $\beta(t)$) by comparing equation (4.3.2) with equation (3.5.6), namely

$$\mathbb{Q}_{11} = (A_t + 1)/E_t \qquad \mathbb{Q}_{12} = 1 - \mathbb{Q}_{11}$$
$$\mathbb{Q}_{21} = A_t/E_t \qquad \mathbb{Q}_{22} = 1 - \mathbb{Q}_{21}$$

where $E_t = \exp\{\int_0^t [\alpha(u) + \beta(u)]du\} \qquad A_t = \int_0^t \alpha(u)E_u \, du.$

In this section, we also present a simulation method of deriving the p.g.f. of a time dependent migration model. This method was developed before discovering the analytic solution given above, but is nevertheless included as it illustrates a number of interesting features of these time dependent models. First, it shows that for some time dependent models, there may be a non zero (and, indeed, large) probability that an individual never migrates again after a specified point in time, say t. Thus time dependent models can represent situations where migration tends, with increasing time, to depopulate one area, until it is finally completely depleted. There is a non zero probability that one population can become depleted and stay depleted. Second, the p.g.f. can be derived by this method without evaluating A_t; only values of E_t are required. Thus, this method might be useful when the integral E_t can be formed, but A_t cannot. This would save one from resorting to numerical integration.

The simulation gives expressions for the p.g.f. at specified times $t_1 \; t_2 \; \ldots \; t_n$. The method is to simulate several hundred realizations of the markov process using the initial condition $X_1(t_0) = 1 \;\; X_2(t_0) = 0$.

The proportion of simulations having $X(t_i) = 1$ then gives an estimate for $\mathbb{Q}_{11}$ at time t_i. The variance of the estimate is $\mathbb{Q}_{11}(1 - \mathbb{Q}_{11})/n_s$ where n_s is the number of simulations. Thus, by increasing n_s, any desired precision may be achieved. Similarly one may derive estimates for $\mathbb{Q}_{21}$ by the same process using the initial condition $X_1(t_0) = 0$, $X_2(t_0) = 1$. The method described below can easily be adapted to simulate the model for general initial conditions, $X_1(t_0) = x_{10}$ $X_2(t_0) = x_{20}$ and this might be useful for investigating other properties of the model which are not amenable to analytic investigation. As an example of the simulation, we have used the exponential model of section 3.5.2. That is

$$\alpha(x_1,t) = \alpha(1 - e^{-ct})$$
$$\beta(x_2,t) = \alpha e^{-ct}$$

By a suitable choice of time scale we can arrange that $c = 1$. The analytic expressions for $\mathbb{Q}_{11}(t)$, $\mathbb{Q}_{21}(t)$ are then

$$\mathbb{Q}_{11}(t) = (\alpha e^{-t} - e^{-\alpha t})/(\alpha-1)$$
$$\mathbb{Q}_{21}(t) = (\alpha e^{-t} - \alpha e^{-\alpha t})/(\alpha-1)$$

It will be seen that the results of an actual simulation reported below (with $\alpha = 25$) conform closely to these expressions.

If $q(x_1, t) = x_1\alpha(x_1,t) + x_2\beta(x_2,t)$, then we will show (in section 6.5) that the time ρ_t (the time lapse until the next migration occurs given that $X_1(t) = x_1$ $X_2(t) = x_2 = N-x_1$) has distribution function:

$$F(\tau) = \text{Prob}(\rho_t \leq \tau) = 1 - \exp[-\int_t^{t+\tau} q(x_1,u)du]$$
$$= 1 - \exp[-f(\tau)]$$

In this example, with $N = 1$ we have

$$q(x,t) = x\alpha + \alpha(1-2x)\exp[-t]$$
$$f(\tau) = x\alpha\tau - \alpha(1-2x)e^{-t}(e^{-\tau}-1)$$

Now if $F(\tau) = 1 - \exp[-f(\tau)]$ is to be a true distribution function, we require $F(0) = 0$, $\lim_{\tau\to\infty} F(\tau) = 1$ and $F(\tau)$ is monotone increasing in τ.

To simulate realizations of the process, we will need to select a random value of τ, say τ_r. This can be done using the fact that $F(\tau)$ is uniformly distributed on [0,1]. Since there is only one individual in the system and it is either in A_1 or A_2 at time t, we need to consider the two cases:

(i) Select a random τ_r when $X(t) = 1$.

$$f(\tau) = \alpha\tau-\alpha e^{-t}(1-e^{-\tau})$$

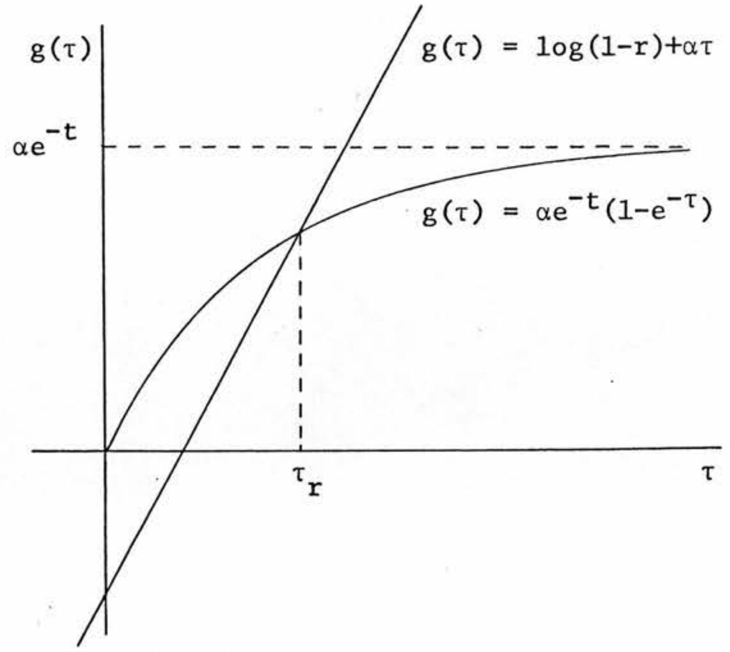
$$f(0) = 0$$

$$\lim_{\tau\to\infty} f(\tau) = \infty$$

It is also easily shown, graphically, that $f(\tau)$ is monotone increasing in τ since $\alpha\tau \geqslant \alpha e^{-t}\tau \geqslant \alpha e^{-t}(1-e^{-\tau})$ for all τ. Thus $F(\tau) = 1-\exp[\alpha e^{-t}(1-e^{-\tau})-\alpha\tau]$ is a distribution function. Let r be a random number from the uniform distribution on [0,1] and let $F(\tau) = r$. This then leads to the equation for τ_r

$$\log(1-r)+\alpha\tau_r = e^{-t}(1-e^{-\tau_r})$$

Now the function on the left represents a line with positive slope (α) and negative intercept. The function in τ on the right is always positive and monotone increasing for $\tau > 0$. Thus the two functions have a unique intersection for positive τ, and this r is easily found using Newton's iterative procedure. (e.g. see Noble (1964) section 2.3). This situation is illustrated graphically below.



(ii) Select a random τ_r when $X(t) = 0$.

$$f(\tau) = \alpha e^{-t}(1-e^{-\tau})$$

$$f(0) = 0$$

$$\lim_{\tau\to\infty} f(\tau) = \alpha e^{-t}$$

thus $\lim_{\tau\to\infty} F(\tau) = 1-\exp[-\alpha e^{-t}]$

$= \text{Prob [a migration occurs in } [t,\infty)|X(t) = 0)$

Thus, given $X(t) = 0$, there is a probability $\exp[-\alpha e^{-t}]$ that the one individual in the system never returns to A_1. Thus to select a random value of τ, we choose a random number r from the uniform distribution on [0,1]. Then if $r\varepsilon[1-\exp(-\alpha e^{-t}),1]$ $\tau_r = \infty$. If $r\varepsilon[0,1-\exp(-\alpha e^{-t})]$, setting $r = F(\tau)$ and solving for τ gives

$$\tau_r = -\log(1+\frac{e^t}{\alpha} \log r)$$

We are now able to simulate the process to derive values for $Q_{11}(t_i)$ and $Q_{21}(t_i)$ for any specified set of t_i, where t_0 is the initial time and

$$t_0 < t_1 < t_2 \dots < t_n.$$

The method of simulating is best indicated in the flow diagram, fig. 4.1. This method of simulation was carried out with $\alpha = 25.0$, and using the method described above for selecting the τ_r. The values of $Q_{11}(t_j)$ and $Q_{21}(t_j)$, derived from 1,000 simulations, are given in table 4.1, along with the standard errors of the Q_{i1}, for $t_j = .05, .10, .15 \dots 2.5$; $i = 1,2$. In order to see if these discrete values could be described as lying along some curve that is a function of continuous time, these points were plotted (see figure 4.2). The points in Fig. 4.2 are the values of $Q_{i1}(t_j)$ against t_j as given by table 4.1. The smooth curve is the function $Q_{i1}(t) = \exp[-.975t]$, and can be seen to be an excellent approximation for $Q_{11}(t)$ and $Q_{21}(t)$ for the range $0.2 \leqslant t < 2.5$. The approximating curve was chosen in a purely empirical manner, so there is no theoretical reason for this to be the correct expression for $Q_{i1}(t)$. Indeed, it is evident that this cannot be the exact expression, since by definition $Q_{21}(0) = 0$. However, the approximation is evidently quite adequate for $0.2 < t < 2.5$, and so, for the initial conditions $X_1(0) = x_{10}$, and $X_2(0) = x_{20}$, the p.g.f. of $X_1(t)$ may be approximated by:

$$\begin{aligned} G(r) &= (1-e(t)+e(t)r)^{x_{10}}(1-e(t)+e(t)r)^{x_{20}} \\ &= (1-e(t)+e(t)r)^N \end{aligned}$$

where $e(t) = \exp[-.975t]$, $\qquad 0.25 \leqslant t \leqslant 2.5$

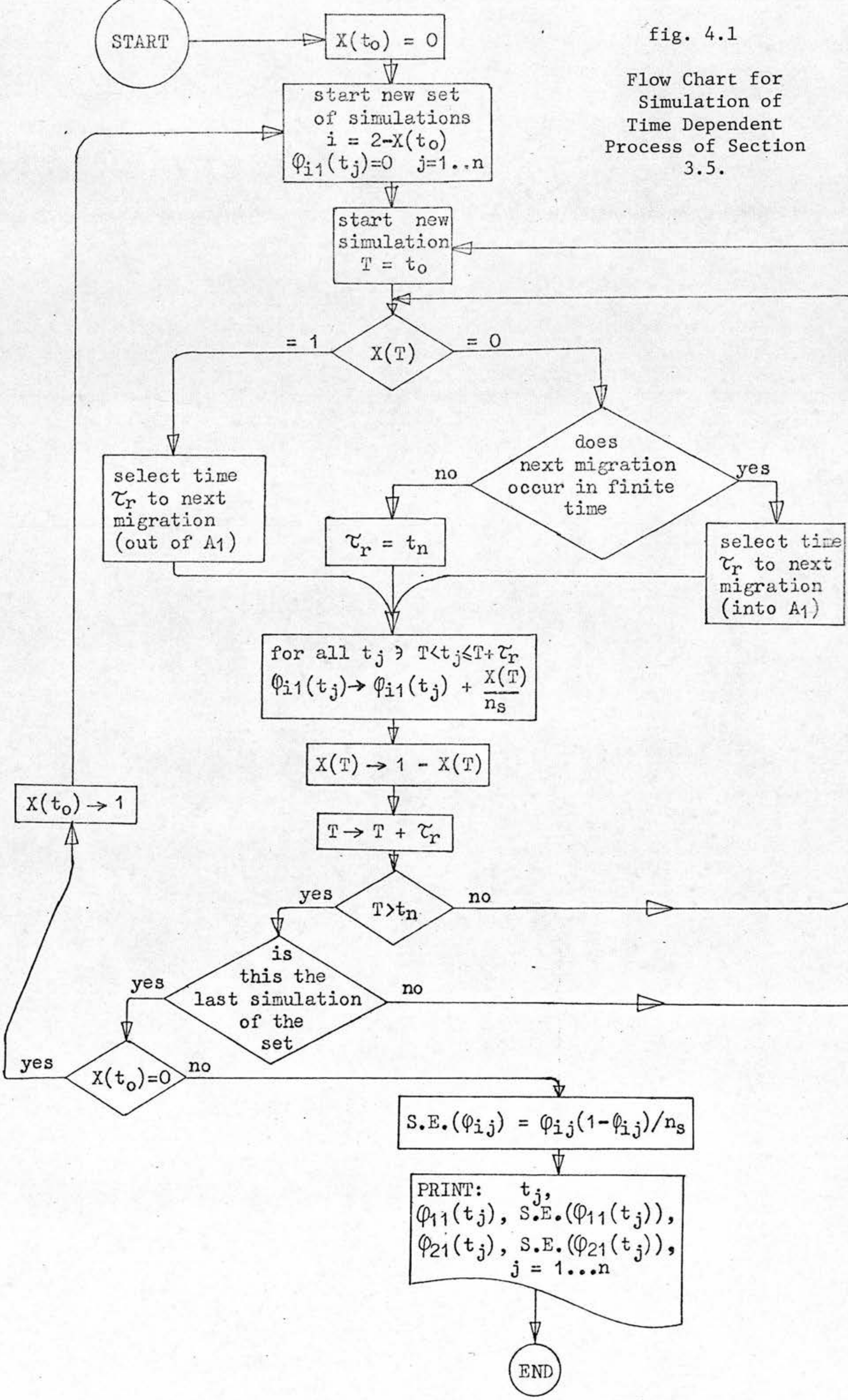


fig. 4.1

Flow Chart for Simulation of Time Dependent Process of Section 3.5.

Table 4.1

Data for figure 4.1

$Q_{ij}(t)$: true value of parameter (see page 99)

$\hat{Q}_{ij}(t)$: estimates value of $Q_{ij}(t)$, from mean over 1,000 simulations

$\widehat{s.e.}(\hat{Q}_{ij}(t))$: estimated standard error = $\sqrt{(\hat{Q}_{ij})(1 - \hat{Q}_{ij})/1000}$

t	$Q_{11}(t)$	$\hat{Q}_{11}(t)$	$\widehat{s.e.}(\hat{Q}_{11}(t))$	$Q_{21}(t)$	$\hat{Q}_{21}(t)$	$\widehat{s.e.}(\hat{Q}_{21}(t))$
.05	.979	.980	.0044	.692	.680	.0148
.10	.939	.923	.0084	.857	.842	.0115
.15	.896	.888	.0100	.872	.851	.0113
.20	.853	.864	.0108	.845	.838	.0117
.25	.811	.813	.0123	.809	.793	.0128
.30	.771	.779	.0131	.771	.785	.0130
.35	.734	.713	.0143	.734	.737	.0139
.40	.698	.679	.0148	.698	.713	.0143
.45	.664	.664	.0149	.664	.693	.0146
.50	.632	.642	.0152	.632	.621	.0153
.60	.572	.561	.0157	.572	.564	.0157
.70	.517	.515	.0158	.517	.509	.0158
.80	.468	.459	.0158	.468	.459	.0158
.90	.424	.443	.0157	.424	.424	.0156
1.0	.383	.402	.0155	.383	.354	.0151
1.1	.347	.335	.0149	.347	.328	.0148
1.2	.314	.309	.0146	.313	.340	.0149
1.3	.284	.278	.0142	.284	.300	.0145
1.4	.257	.237	.0134	.257	.289	.0143
1.5	.232	.228	.0133	.232	.230	.0133
1.6	.210	.201	.0127	.210	.204	.0127
1.7	.190	.180	.0121	.190	.195	.0125
1.8	.172	.169	.0119	.172	.178	.0121
1.9	.156	.157	.0115	.156	.166	.0118
2.0	.141	.141	.0110	.141	.140	.0110
2.1	.128	.125	.0105	.128	.140	.0110
2.2	.115	.100	.0095	.115	.120	.0103
2.3	.104	.099	.0094	.104	.100	.0095
2.4	.095	.102	.0096	.095	.102	.0096
2.5	.086	.105	.0097	.086	.096	.0093

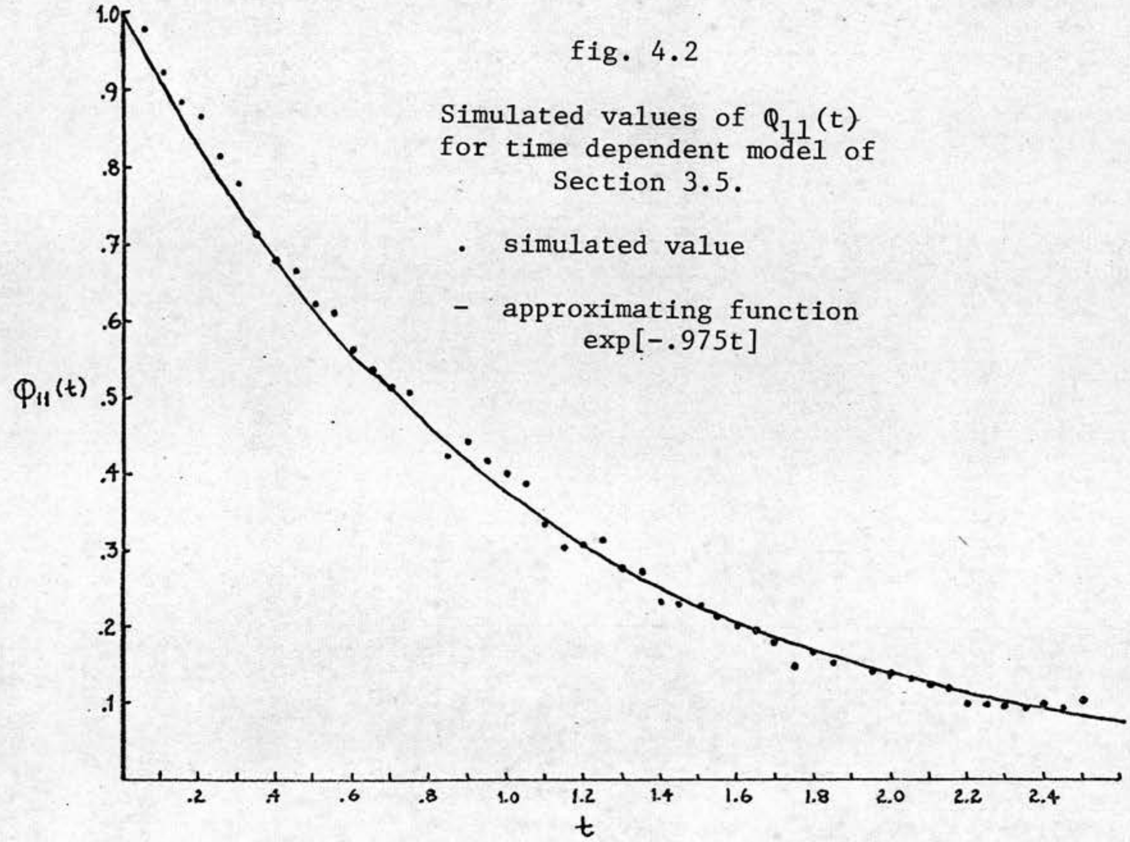
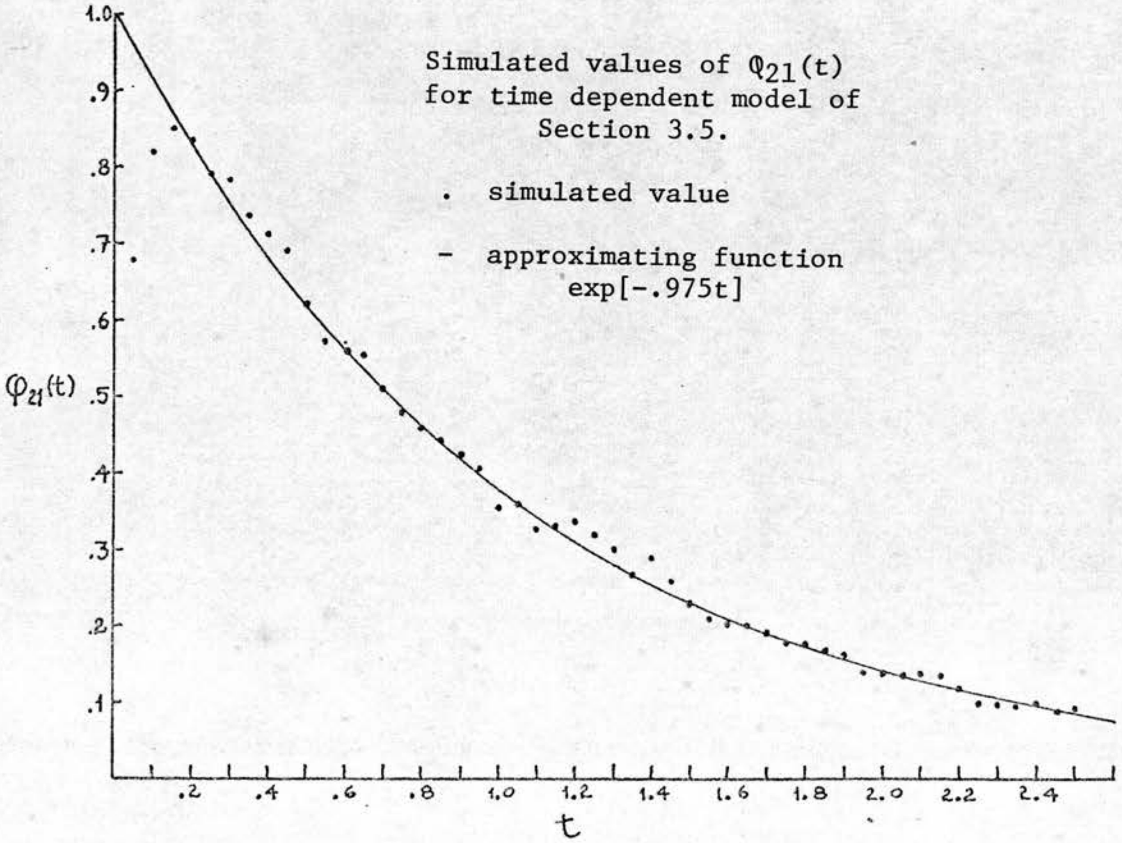


fig. 4.2

Simulated values of $\varphi_{11}(t)$ for time dependent model of Section 3.5.



Simulated values of $\varphi_{21}(t)$ for time dependent model of Section 3.5.

4.4 Some Graphical Illustrations of the Models:

In this section we give a very few examples of the expected changes in $X_1(t)$ over time when different migration models are operating on a two dimensional system. In so far as is possible, the parameters of each different model have been taken to be the same, so that differences in the graphs of $E(X_1(t))$ are due to the different mechanisms of migration, rather than simply due to different migration rates. Four models were considered:

1. X1: Simple Linear Model.

This is the simple linear model of section 3.3. The parameter values were:

$$N = 20 \qquad \alpha = \beta = 0.5$$

Table 4.2 and Table 4.3 list the expected value $X1 = E(X_1(t))$ for various values of t, and also list the corresponding standard deviation of $X_1(t)$, S1. The values in table 4.2 are for $x_{10} = 0$, and in table 4.3 for $x_{10} = 20$. The two corresponding graphs of X1 are shown in Fig. 4.3 and Fig. 4.4.

2. X2: Linear Model with Death.

This is the model of section 4.2 with no immigration or birth, but with stochastic death. The parameter values were:

$$N = 20 \qquad \alpha = \beta = 0.5 \qquad \mu_1 = \mu_2 = 0.5$$

The equations for the mean and variance are 4.2.12 (with $\lambda = 0$) and 4.2.19 respectively. This system could also be taken to represent the system in which permanent migration out of A_1 and A_2 is occurring at the same rate as migration between the two areas. The means X2 and corresponding standard deviations S2 are listed for $x_{10} = 0$ (table 4.2) and for $x_{20} = 20$ (table 4.3), and these means are graphed against time in fig. 4.3 and fig. 4.4, respectively.

3. X3: The Quadratic Model

This is the model of section 3.4 (density dependent migration). The parameter values were:

$$N = 20 \qquad \alpha = \beta = 0.5$$

The means X3 and corresponding standard deviations S3 are tabulated and graphed for $x_{10} = 0$ and $x_{10} = 20$ as in 1. and 2. above. However, the quadratic model causes equilibrium to be approached with such rapidity, that the means and standard deviations are also given (in Table 4.5) for a more appropriate time scale. The graph of X3 for this finer time scale is given in fig. 4.5. Fig. 4.5 also gives a number of further graphs of X3 for different initial values x_{10}.

4. X4: The Time Dependent Model.

This is the model of section 4.3.4, where the migration intensity functions are time dependent and are given by

$$\alpha(x_1,t) = x_1\alpha(1-\exp(-ct))$$
$$\beta(x_2,t) = (N-x_1)\alpha\ \exp(-ct)$$

The parameter values were:

$$N = 20 \quad \alpha = 25 \quad c = 1.0$$

The Q_{11}, Q_{21} were calculated by the method of simulation described in section 4.3.4, and the means, tabulated in table 4.2 (for $x_{10} = 0$) and table 4.3 ($x_{10} = 20$) are derived from

$$X4 = x_{10}Q_{11}+(N-x_{10})Q_{21}$$
$$(S4)^2 = x_{10}Q_{11}(1-Q_{11})+(N-x_{10})Q_{21}(1-Q_{21})$$

Since there is some slight error of estimation in Q_{11} and Q_{21}, the tabulated values are not exact. The corresponding curves in fig. 4.3 and fig. 4.4 represent the smooth curve that best fitted these points. As in 3. above, further tables and graphs are given for this model (table 4.5, fig. 4.6) to represent the mean with a finer time scale and for several different values of x_{10}.

X1 = 0.

Table 4.2. Data for fig. 4.3.

	linear		linear with death		quadratic		time dependent	
T	X1	S1	X2	S2	X3	S3	X4	S4
0.2000	1.8127	1.283899	1.6402	1.227062	9.8168	1.605585		
0.4000	3.2968	1.659323	2.6992	1.528040	9.9966	1.601284	13.5411	2.091175
0.6000	4.5119	1.869231	3.3425	1.668494	9.9999	1.601281	11.1421	2.221436
0.8000	5.5067	1.997627	3.6913	1.734932	10.0000	1.601281	9.1681	2.228317
1.0000	6.3212	2.079260	3.8340	1.760403	10.0000	1.601281	7.5438	2.167571
1.2000	6.9881	2.132231	3.8351	1.760599	10.0000	1.601281	6.2073	2.069006
1.4000	7.5340	2.167013	3.7413	1.743968	10.0000	1.601281	5.1076	1.950187
1.6000	7.9810	2.190020	3.5861	1.715547	10.0000	1.601281	4.2027	1.821972
1.8000	8.3470	2.205307	3.3936	1.678631	10.0000	1.601281	3.4581	1.691215
2.0000	8.6466	2.215495	3.1809	1.635546	10.0000	1.601281	2.8455	1.562256
2.2000	8.8920	2.222298	2.9599	1.588028	10.0000	1.601281	2.3414	1.437798
2.4000	9.0928	2.226848	2.7387	1.537426	10.0000	1.601281	1.9266	1.319459
2.6000	9.2573	2.229891	2.5229	1.484805	10.0000	1.581139		
2.8000	9.3919	2.231929	2.3160	1.431019	10.0000	1.581139		
3.0000	9.5021	2.233294	2.1202	1.376752	10.0000	1.581139		
3.2000	9.5924	2.234209	1.9367	1.322548	10.0000	1.581139		
3.4000	9.6663	2.234821	1.7659	1.268839	10.0000	1.581139		
3.6000	9.7268	2.235232	1.6078	1.215963	10.0000	1.581139		
3.8000	9.7763	2.235508	1.4622	1.164182	10.0000	1.581139		
4.0000	9.8168	2.235692	1.3286	1.113692	10.0000	1.581139		
4.2000	9.8500	2.235816	1.2062	1.064637	10.0000	1.581139		
4.4000	9.8772	2.235899	1.0944	1.017122	10.0000	1.581139		
4.6000	9.8995	2.235954	0.9925	0.971214	10.0000	1.581139		
4.8000	9.9177	2.235991	0.8997	0.926952	10.0000	1.581139		
5.0000	9.9326	2.236016	0.8153	0.884354	10.0000	1.581139		

X1 = 20.

Table 4.3. Data for fig. 4.4.

	linear		linear with death		quadratic		time dependent	
T	X1	S1	X2	S2	X3	S3	X4	S4
0.2000	18.1873	1.283899	16.4565	4.219729	10.1832	1.605585		
0.4000	16.7032	1.659323	13.6754	3.939675	10.0033	1.601284	13.5411	2.091175
0.6000	15.4881	1.869231	11.4739	3.659431	10.0001	1.601281	11.1421	2.221436
0.8000	14.4933	1.997627	9.7151	3.392372	10.0000	1.601281	9.1681	2.228317
1.0000	13.6788	2.079260	8.2966	3.144370	10.0000	1.601281	7.5438	2.167571
1.2000	13.0119	2.132231	7.1411	2.917267	10.0000	1.601281	6.2073	2.069006
1.4000	12.4660	2.167013	6.1904	2.710797	10.0000	1.601281	5.1076	1.950187
1.6000	12.0190	2.190020	5.4005	2.523648	10.0000	1.601281	4.2027	1.821972
1.8000	11.6530	2.205307	4.7378	2.354074	10.0000	1.601281	3.4581	1.691215
2.0000	11.3533	2.215495	4.1767	2.200215	10.0000	1.601281	2.8455	1.562256
2.2000	11.1080	2.222298	3.6975	2.060281	10.0000	1.601281	2.3414	1.437798
2.4000	10.9072	2.226848	3.2852	1.932623	10.0000	1.601281	1.9266	1.319459
2.6000	10.7427	2.229891	2.9277	1.815780	10.0000	1.581139		
2.8000	10.6081	2.231929	2.6159	1.708464	10.0000	1.581139		
3.0000	10.4979	2.233294	2.3424	1.609569	10.0000	1.581139		
3.2000	10.4076	2.234209	2.1013	1.518135	10.0000	1.581139		
3.4000	10.3337	2.234821	1.8878	1.433346	10.0000	1.581139		
3.6000	10.2732	2.235232	1.6982	1.354500	10.0000	1.581139		
3.8000	10.2237	2.235508	1.5291	1.280993	10.0000	1.581139		
4.0000	10.1832	2.235692	1.3781	1.212306	10.0000	1.581139		
4.2000	10.1500	2.235816	1.2429	1.147993	10.0000	1.581139		
4.4000	10.1228	2.235899	1.1216	1.087663	10.0000	1.581139		
4.6000	10.1005	2.235954	1.0127	1.030980	10.0000	1.581139		
4.8000	10.0823	2.235991	0.9146	0.977646	10.0000	1.581139		
5.0000	10.0674	2.236016	0.8264	0.927399	10.0000	1.581139		

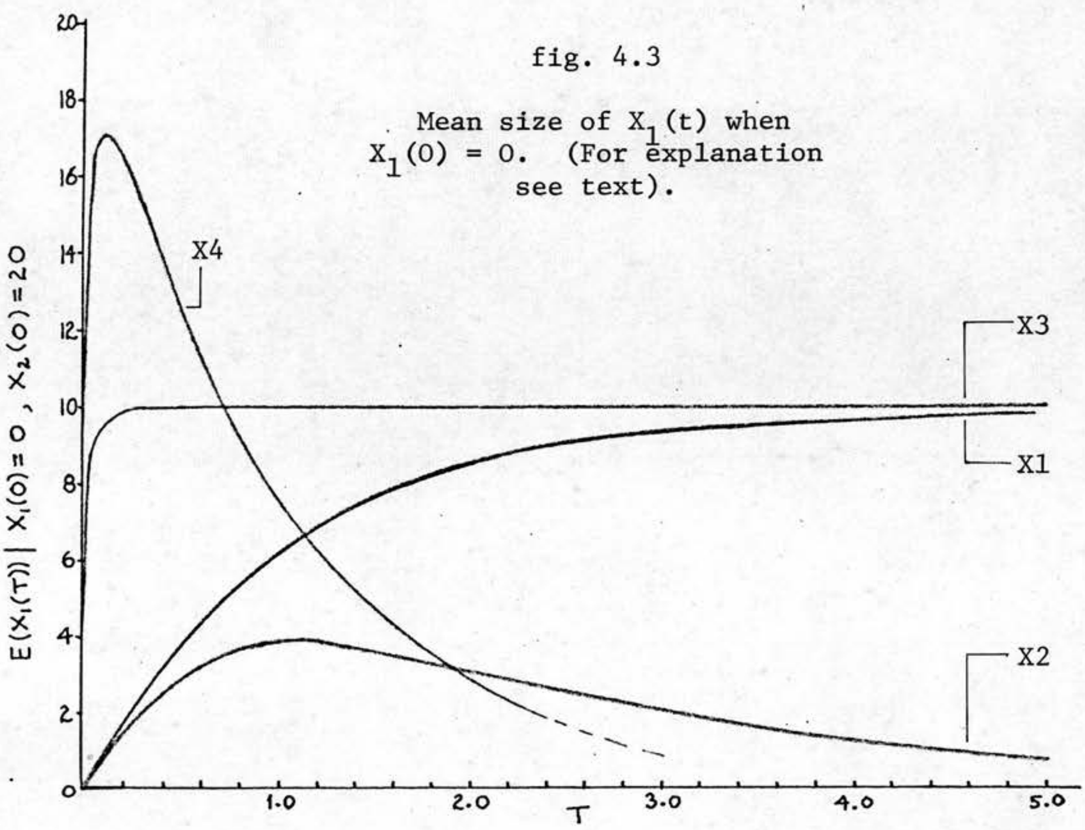


fig. 4.3

Mean size of $X_1(t)$ when $X_1(0) = 0$. (For explanation see text).

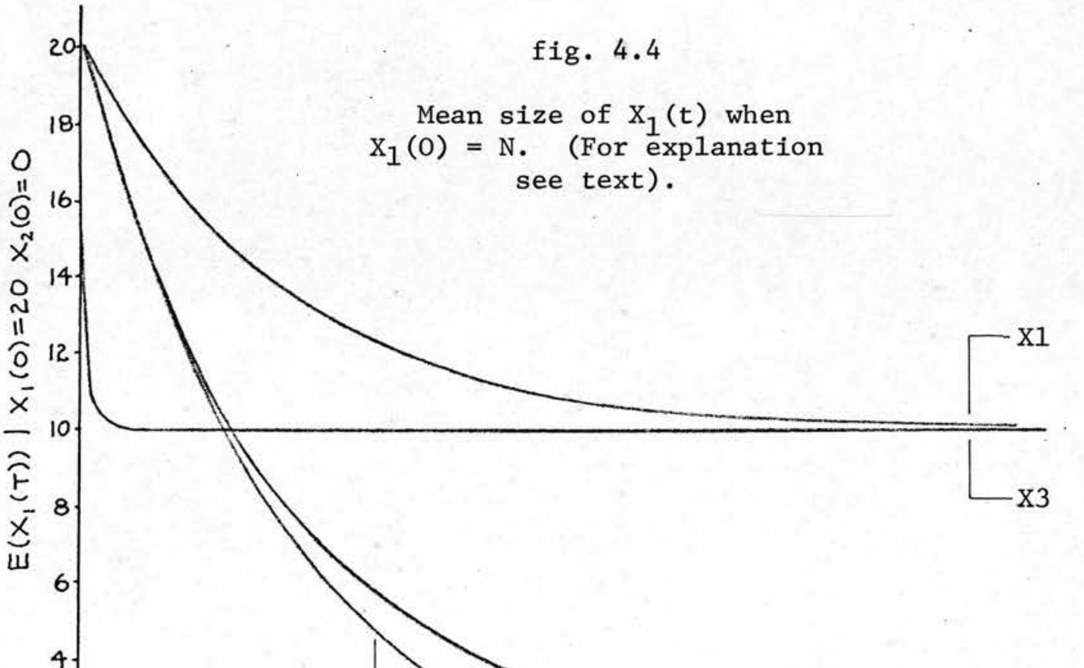


fig. 4.4

Mean size of $X_1(t)$ when $X_1(0) = N$. (For explanation see text).

Table 4.4 Data for Fig. 4.5

Quadratic Model

X1 = 0 X1 = 20

T	X3	S3	X3	S3
102	1.8127	1.4749	18.1873	1.4749
.02	3.2968	1.7900	16.7032	1.7900
.03	4.5119	1.8981	15.4881	1.8981
.04	5.5067	1.9166	14.4933	1.9165
.05	6.3212	1.8940	13.6788	1.8940
.06	6.9881	1.8454	13.0119	1.8545
.07	7.5340	1.8107	12.4600	1.8107
.08	7.9810	1.7691	12.0190	1.7691
.09	8.3470	1.7327	11.6530	1.7327
.10	8.6466	1.7022	11.3533	1.7022
.11	8.8920	1.6776	11.1080	1.6776
.12	9.0928	1.6584	10.9072	1.6584
.13	9.2573	1.6435	10.7427	1.6435
.14	9.3919	1.6323	10.6081	1.6323
.15	9.5021	1.6239	10.4979	1.6239
.16	9.5924	1.6176	10.4076	1.6176
.17	9.6663	1.6131	10.3337	1.6131
.18	9.7268	1.6097	10.2732	1.6097
.19	9.7763	1.6073	10.2237	1.6073
.20	9.8168	1.6056	10.1832	1.6056

Table 4.5 Data for Fig. 4.6

Time Dependent Model

X1 = 0 X1 = 20

T	X4	S4	X4	S4
.01	4.46	1.8615		
.02	7.76	2.1793	19.86	.3722
.03	10.28	2.2350		
.04	12.24	2.1793	19.60	.6261
.05	13.76	2.0720		
.06	14.90	1.9492	19.28	.8331
.07	16.00	1.7787		
.08	16.60	1.6799	19.00	.9749
.09	16.88	1.6227		
.10	17.26	1.5377	18.84	1.0454
.11				
.12	17.62	1.4480	18.42	1.2063
.13				
.14	17.40	1.5040	18.18	1.2862
.15				
.16	17.30	1.5282	17.46	1.4892
.17				
.18	17.12	1.5701	17.40	1.5040
.19				
.20	17.02	1.5925	17.26	1.5377

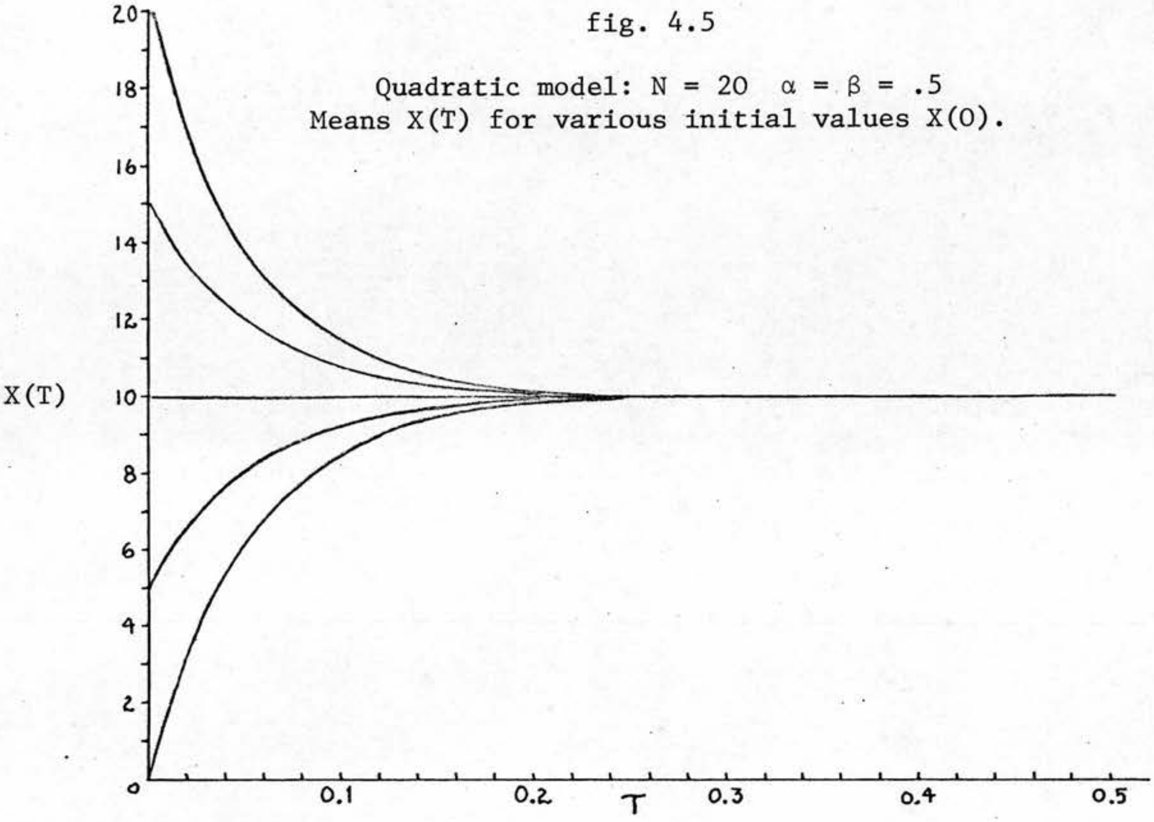


fig. 4.5

Quadratic model: N = 20 $\alpha = \beta = .5$
Means X(T) for various initial values X(0).

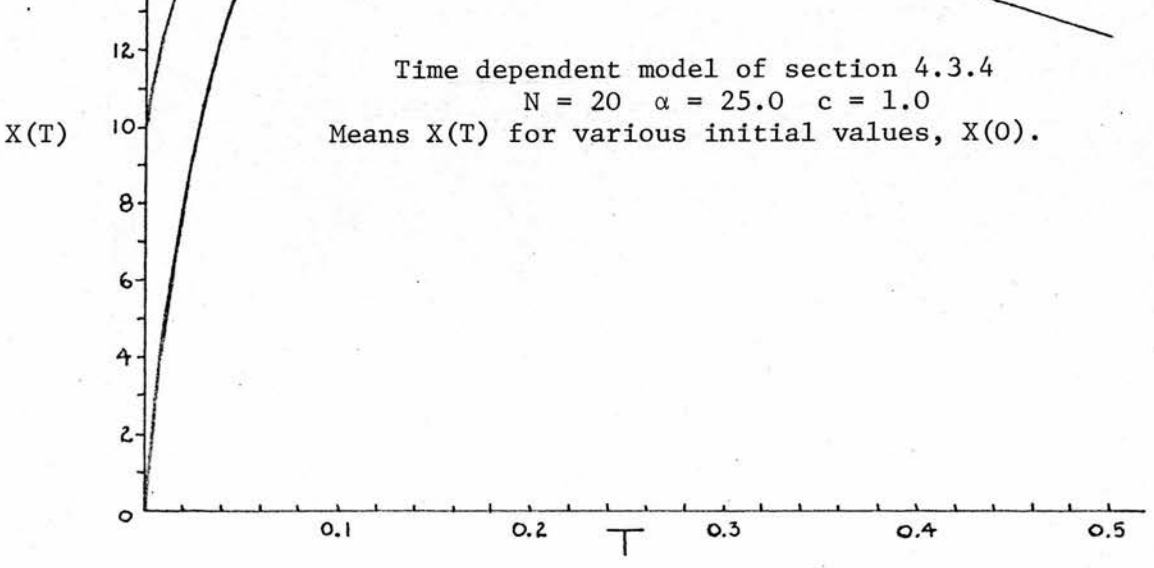


fig. 4.6

Time dependent model of section 4.3.4
N = 20 $\alpha = 25.0$ c = 1.0
Means X(T) for various initial values, X(0).

CHAPTER FIVE

ESTIMATION PROCEDURES:
SURVEY OF PREVIOUS WORK

5.1 Introduction

Despite the very large volume of literature concerning models to describe migration processes, suprisingly little attention has been paid to the problem of estimating the parameters of these models. The standard textbooks dealing with stochastic processes are concerned almost exclusively with the development of the p.g.f. of processes, and its properties. There is no treatment of general techniques for estimation of parameters of that class of stochastic process which most concerns us here: Markov processes. An exception is Bartlett (1955) who devotes a short section to such problems. Bartlett's work will be described briefly in the next section. A more general treatment of estimation techniques for markov processes is developed in a monograph by Billingsley (1961), and this will be outlined in some detail in the next section. This is virtually all that exists in the way of general estimation theory for markov processes. The remaining literature that is relevant to this thesis deals with estimation procedures developed for particular models, using standard estimation techniques: method of moments, maximum likelihood, minimum chi-square etc. The only fields where much attention has been paid to estimation of the parameters of markov processes are epidemiology and immunology. Developments in these fields will be reviewed briefly in section 5.3. However, the type of data available for estimation purposes in these fields will often not be the sort of information that can be collected on animal populations. The types of experiments which animal ecologists can carry out are of two main types:

1. Complete counts of, or samples from, the population at fixed points in time. In the latter case, experimenters may also determine the relative effort expended at each point in time in capturing the sample. Such effort models are described in Beverton and Holt (1957), but since we have not investigated such methods in this thesis, effort models will not be given further mention.

2. Marking Experiments: A subpopulation of marked animals is introduced into the population, either by adding marked animals to the population, or by capturing, marking, and releasing animals. At subsequent times, samples are drawn from the population, and animals may be removed, or returned. In the latter case, unmarked animals may be marked, or marked ones remarked. Such experiments thus yield data consisting of the capture histories of various classes of marks (including the numbers of unmarked individuals in each sample). The number of variations of this basic process is immense, and there is an extensive literature on the models assumed for specific experiments of this type, and the estimation techniques developed to deal with them. A comprehensive review of the literature is given by Cormack (1968). In this chapter, we shall review only those applications of marking experiments which deal with

populations subject to immigration. Very little literature exists for this restricted subject, and it will be possible to review it (section 5.4) in some detail. Section 5.4 will also deal with estimation techniques for animal populations subject to migration using data of the first type mentioned above. A more extensive review of estimation techniques for animal populations will not be undertaken in this thesis. Where such techniques are used in Chapters 6 to 8, a brief outline of how the method was developed by previous researchers will be given in those chapters, prior to application of their methods.

5.2 Estimation Techniques for Markov Processes

Bartlett (1955, Chapter 8) devotes a chapter of his book to problems of inference with reference to stochastic processes. Much of this is devoted to a review of standard theory (e.g. the properties of the likelihood function), and to developing tests of goodness of fit. This latter problem, while important, will not be considered in this thesis. Bartlett then uses the likelihood principle to derive estimates when observations on the state of the process are made at discrete times (i.e. observe $X(t_1)$, $X(t_2)$ ... $X(t_n)$ for given (t_i), and when the process is observed continuously (i.e. observe $X(t) = z_i$ for $T_i \leqslant t < T_{i+1}$, where $T_0 = 0$, T_i is the time of occurrance of the ith change of state). The problem of forming estimates using this latter type of data is dealt with more rigorously and in more detail by Billingsley (1961), and so will be discussed later.

Bartlett notes that, for any markov process, the likelihood of the observations at discrete times (as described above) is

$$L = \Pr(X(t_0)) \prod_{r=1}^{n} \Pr(X(t_r) | X(t_{r-1})) \qquad (5.2.1)$$

This suggests that many stationary (i.e. time homogeneous) processes may be approximated by the autoregressive system

$$X(t_r) = \beta X(t_{r-1})+Y_r$$

where the Y_r are normally and independently distributed with mean m and variance v. Bartlett develops the maximum likelihood estimates (m.l.e.) for β and v for such a system when $m = 0$. He also shows how the emigration-immigration process (immigration by poisson arrivals at rate ν, independent stochastic emigration with intensity μ) can be approximated by just such an autoregressive model ($\beta = \exp[-\mu]$, $v = m = \nu/\mu$ for m large). Nevertheless, the estimate for β is a complicated expression, even for this simplest case. Bartlett suggests another method of estimation, again derived from methods for dealing with autoregressive schemes; estimates may be derived by comparing the observed covariance between consecutive observations, and the variance of the observations, with their analytical expressions. A wide range of techniques have been developed, mainly using correlation or regression analysis, for dealing with discrete observations from time series models (many of these are developed in Bartlett (1955) Chapter 9), but the literature is vast, and it was not possible to investigate if these techniques might also be applied to discrete observations from markov process models.

The first part of Billingsley's monograph, (Billingsley, 1961) is devoted to statistical inferences about markov chains. When observations of a markov process are made at discrete, equally spaced, times, these observations represent a realization of an embedded process which is a markov chain defined by the transition probabilities

$$p_{ij}(\theta) = \Pr(X(t_r) = j | X(t_{r-1}) = i)$$

(where θ is the set of parameters of the model for which we wish to form estimates). For some of the models derived in the first half of this thesis, we have been able to give analytic expressions for the $p_{ij}(\theta)$, so this work is relevant. Billingsley develops the expression for the likelihood (which is just the expression (5.2.1) given by Bartlett), and states sufficient conditions for the m.l.e. to be consistent, and have the usual large sample properties. These conditions and properties are developed in measure-set-theoretic terms which would be difficult to summarize here. The condition simplifies when the state space is finite, and is:

The set D of (i,j) such that $p_{ij}(\theta) > 0$ is independent of θ and each $p_{ij}(\theta)$ has continuous partial derivatives of third order for all θ.

The dxr matrix: $[\partial p_{ij}(\theta)/\partial\theta_u]$ has rank r for all θ where r is the number of parameters $u = 1, \dots r$ and d is the number of elements in D).

For all θ, the system is ergodic.

This condition obviously holds for the time homogeneous closed models of chapter 3 of this thesis. It may not hold for some time dependent models, or for models that incorporate death (the ergodicity assumption may be violated). In these cases, one can still derive m.l.e. for the parameters but the large sample properties are not assured.

Much of the first part of Billingsley's monograph is devoted to developing large sample distribution theory for hypothesis testing and goodness of fit tests. He also gives several examples of specific markov chains, for which estimates are derived. As we have seen in the first part of this thesis, it is not always possible to derive expressions for the $p_{ij}(\theta)$, and even when this is possible, the expression may be so complicated that the m.l. equations are difficult to solve.

The second part of Billingsley's thesis deals with estimation techniques for markov processes which are observed continuously for a predetermined length of time T. During this time n(T) changes of state occur (n(T) is thus a random variable). If the system starts off with $X(0) = z_1$, remains in that state a length of time r_1, then jumps to z_2, stays in that state a length of time r_2 etc., then the set of observations is

$$\{(z_k, r_k) \mid 1 \leq k \leq n(T)+1\}$$

where

$$r_{n(T)+1} = T - \sum_{k=1}^{n(T)} r_k \tag{5.2.2}$$

If

$$p_\theta(\delta t, \xi, A) = \text{Prob}\{X(t+\delta t) \varepsilon A \mid X(t) = \xi\}$$

where A is any set of states of the process, then Billingsley defines two functions

$$q(\xi;\theta) = \lim_{\delta t \to 0} [1-p_\theta(\delta t,\xi,\{\xi\})]/\delta t \tag{5.2.3}$$

and

$$q(\xi,A;\theta) = \lim_{\delta t \to 0} p_\theta(\delta t,\xi,A)/\delta t \tag{5.2.4}$$

which may be interpreted, respectively, as the instantaneous probability of a change of state at time t, given that $X(t) = \xi$; and, as the instantaneous probability of a change of state from ξ to $\eta \varepsilon A$. Then, using theorems due to Doob (1953, page 266), he shows that

$$\begin{aligned} &\text{Prob}(r_k > a | r_1 \ldots r_{k-1}, z_1 \ldots z_k) \\ &= \text{Prob}(r_k > a | z_k) = \exp[-aq(z_k;\theta)] \end{aligned} \tag{5.2.5}$$

and

$$\begin{aligned} &\text{Prob}(z_{k+1} \varepsilon A | r_1 \ldots r_k, z_1 \ldots z_k) \\ &= \text{Prob}(z_{k+1} \varepsilon A | z_k) = q(z_k,A;\theta)/q(z_k;\theta) \\ &= \Pi_\theta(\xi,A) \qquad \text{(say)} \end{aligned} \tag{5.2.6}$$

In order to give a rigorous development of the likelihood, it is necessary to introduce the density $f(\xi,\eta;\theta)$ for $\Pi_\theta(\xi,A)$

$$\Pi_\theta(\xi,A) = \int_A f(\xi,\eta,\theta)\lambda\ \partial\eta$$

where λ is a meausre on the state space and the $\Pi_\theta(\xi,A)$ are absolutely continuous with respect to this measure and have density f. The log likelihood of the observations $\{z_n,r_n\}$ is then shown to be:

$$\log L = \sum_{k=1}^{n(T)-1} \{\log f(z_k,z_{k+1};\theta) + \log q(z_{k+1};\theta) - r_{k+1}q(z_{k+1};\theta)\} \tag{5.2.7}$$

This expression is not quite exact, as it ignores the information from the observation $r_{n(T)+1}$ (equation 5.2.2) so that the situation is identical to an experiment in which the number n, of changes of state to be observed is predetermined ($n = n(T)$). The loss of information, relative to the whole sample, will be negligible when $n(T)$ is large. Billingsley also develops sufficient conditions for the development of the above likelihood expression and for proving asymptotic properties for the estimates, and asymptotic distributions for hypothesis testing. These simplify considerably if the state space is finite. Then $p_\theta(t,\xi,A)$ becomes the transition probability

$$p_{ij}(\delta t) = P_r(X(t+\delta t) = j|X(t) = i)$$

and the functions $q(z_n,A;\theta)$ and $q(z_n;\theta)$ become

$$q_{ij}(\theta) = \lim_{\delta t\to 0} (p_{ij}(\delta t))/\delta t \qquad (5.2.8)$$

$$q_i(\theta) = \lim_{\delta t\to 0} (1-p_{ij}(\delta t))/\delta t \qquad (5.2.9)$$

and thus we have

$$\Pi_{ij}(\theta) = \text{Prob}\{z_{n+1} = j|z_n=i\}$$

$$q_{ij}(\theta) = q_i(\theta)\Pi_{ij}(\theta)$$

and $$q_i(\theta) = \sum_{j\neq 1} q_{ij}(\theta)$$

The log likelihood is now easily expressed in terms of these functions and is given by Billingsley as:

$$\log L = \sum \log \Pi_{ij}(\theta)+\log q_i(\theta) - r_k\, q_i(\theta)$$

$$= \sum \log q_{ij}(\theta) - r_k\, q_i(\theta) \qquad (5.2.10)$$

where the summation is over $k = 1, \ldots n(T)$, and for each k, i takes the value z_k, j, the value z_{k+1}. Obviously, this is more conveniently grouped in terms of the states $i, j = 1 \ldots s$ (s finite) which the system can occupy. In this thesis, for models without birth or immigration we have $s = N$. Also, only one transition occurs at a time, so $j = i-1$ or $i+1$. Therefore, let:

$n_{i\ i+1}$ be the number of transitions (out of the $n(T)$) from state i to i+1.

$n_{i\ i-1}$ be the number of transitions from state i to i-1.

g_i be the amount of time spent in state i.

$= \sum_k r_k$ where $k = \{k | z_k = i\}$.

$$\therefore \log L = \sum_{i=0}^{N} [n_{i,i+1} \log q_{i\ i+1}(\theta) + n_{i\ i-1} \log q_{i\ i-1}(\theta) - g_i q_i(\theta)] \tag{5.2.11}$$

Billingsley also establishes the conditions under which this expression was developed and conditions that assure the consistency of the m.l.e. and the asymptotic distribution theory which he goes on to develop. These are:

1. The sample function, $X(t)$, $0 \leq t < T$, is a right continuous step function.

2. The $\lim_{t \to 0} p_{ii}(t;\theta) = 1$ holds for all i, and the limits (5.2.8) and (5.2.9) hold, so that $q_i(\theta)$, $q_{ij}(\theta)$ exist $\forall\, i,j$.

3. For all θ and i, $q_i(\theta) > 0$.

4. The set D of (i,j) such that $q_{ij}(\theta) > 0$ is independent of θ, and the $q_{ij}(\theta)$ have continuous third order partial derivatives $\forall \theta$. The dxr matrix $[\partial q_{ij}(\theta)/\partial \theta_u]$ $(u = 1 \dots r)$ has rank r. (Thus $r \leqslant d$).

5. For all θ, the markov chain defined by $\{z_k\}$ $k = 1 \dots n(T)$ is ergodic.

These estimation methods of Billingsley's will be applied to the models of this thesis in Chapter 6. Where necessary, extensions of the theory will also be outlined in Chapter 6.

5.3 Estimation of Parameters of Migration Models: Fields other than Animal Ecology.

Most of the estimation methods used in fields where migration models are applied to real situations, consist of manipulating the model to derive analytic expressions for some properties of the model which have observable counterparts in the real world. The analytic expressions are equated to the observations and the equations solved for the parameters of the model. Complex models may involve many parameters, and finding sufficient properties to identify all the parameters may tax both experimenter and statistician.

An example of this process is described by Bartlett (1949) who describes the work of Soper (1929) for estimating the parameters of a (deterministic) migration model for measles epidemics. The model accounts for observed damped oscillatory behaviour of outbreaks. Equations for the amplitude, period, and damping factor can be solved for the three parameters of the model. Bartlett (1949) has criticized this model as over simplified, and emphasizes the necessity for incorporating stochastic factors. The fitting of a complex model of this type, is described in detail in Bartlett (1956); it is very much an empirical method, combining direct knowledge (e.g. incubation period), indirect estimates (by the process described above), and an enlightened combination of guesswork and adjustment of parameter estimates until simulated realisations of the model reflected the important characteristics of actual data.

In most applications of migration models, one is restricted by the fact that it is not possible to observe repeated realisations of the process. Thus one is unable to make use of the second or higher moments of the model at a given time for forming estimates, nor can one make use of the central limit theorem properties of means to improve precision of estimates. An exception is the field of bacterial genetics, where one can set up large numbers of replicate populations. Armitage (1952) describes standard experiments with mutating bacteria where this is the case. Growth can be directly observed and it is required to estimate the mutation (migration rate; in one direction only) rate. Armitage reviews the methods suggested by previous workers, and compares them for efficiency against the m.l. estimate. These include

1. setting the proportion of cultures where no mutation occurs equal to its theoretical expectation.

2. solving $\overline{y}$ (theoretical) = $\overline{y}$ (observed) where $\overline{y}$ is the average size of mutated colonies.

3. solving E(m) (theoretical) = E(m) (observed) where m is the number of mutations that occur.

4. solving median (m) theoretical = median (m) observed.

It is surprising that Armitage does not consider moment estimates based on variances and covariances, as much of the deductive part of his paper is given over to the development of expressions for these statistics. He does examine the form of the analytic expressions for $\bar{y}$, and suggests that parameters of the model may be estimated from a single realisation by comparing the form of the approach to the equilibrium state as time elapses with this property for observed $\bar{y}$.

In animal ecology, one will rarely be able to establish replicate populations. One might achieve this by means of marking individuals so that there are a number of distinct subpopulations, all of which can be assumed to act in the same manner. Another situation where replication effectively occurs, is in the situation described by Bailey (1968) for migration among cells in a linear array (see section 2.3). If, at $t = 0$, each individual is given a mark to indicate which cell it is in at $t = 0$, then, because all birth/death/migratio rates are identical for all cells, the distribution, at time $\tau > 0$, of individuals marked initially in cell i represents one replication of the process if the origin is taken to be cell i.

There appears to be only a single experiment in animal ecology literature which makes use of replication for estimating migration intensities without resorting to marking, and this can most conveniently be described here. This is the work of Shiyomi (1967, 1968), studying the movement of aphids among individual blades of barley planted in a square grid. He assumes that, initially, insects have a distribution (with mean μ, variance σ^2 known) over plants; that an insect leaves a plant in $[0,t)$ with constant independent probability q, to join a pool of insects on the ground, and that the number of insects climbing from the ground back onto the plants has the poisson distribution with intensity λ. This model doesn't take account of the (non zero) probability that the pool on the ground may become depleted, but the model seems adequate of q is reasonably large and λ is not too large. Then so long as $\mu \neq \sigma^2$, it is possible to get moment estimates for λ and q using the observed mean and variance of the number of individuals per plant at time t. The proportion r of individuals which leave a plant in $[0,t)$ and have climbed onto a plant again by time t is then estimated from the equation

$$\lambda = \mu . q . r$$

In the second paper (Shiyomi (1968)), he deduced the expression for the likelihood. The equations for the m.l. estimates were solved iteratively using the moment estimates as initial values. Shiyomi does not give expressions for the variance of either type of estimate, and there is little difference between the two types of point estimates, both appearing to give similar and excellent fits of the model to the data.

5.4 Estimation Techniques for Animal Populations Subject to Migration

Interesting theoretical work was done by Pyke (1955) in estimation of migration rates. The first part of the thesis is devoted to three models for estimating the number of migrants, while making minimal assumptions about the manner in which migrations take place. In all three models he considers migration between two otherwise closed areas, A_1 and A_2, and in each area, there are x_i individuals of one type, y_i individuals of another type, at $t = 0$. Such a situation would occur if the number of females y_i, and males x_i were known for two populations at $t = 0$.

The three models and the estimation procedures applied to them are as follows:

1. M_{x1} of the x_1 and M_{y1} of the y_1 individuals, migrate from A_1 to A_2 in $[0,t)$. A sample of size n is taken (binomial sampling) in A_2 at time t, and x are of the type x_i (the x_1 individuals being indistinguishable from the x_2 individuals). It is assumed that M_{x1} has the distribution

$$B(M_{x1};\ M,\ \ x_1/(x_1+y_1)) \qquad (5.4.1)$$

$M = M_{x1}+M_{y1}$ is the parameter which he wishes to estimate.

The conditional probability $\Pr(x|M_{x1};M)$ is easily formed, as is the joint probability $\Pr(x,M_{x1};M)$. The true likelihood is the latter probability summed over $M_{x1} = 0 \ldots M$. Pyke is unable to find the M.L.E. for M, but gets a moment estimate by equating x to its (unconditional) expectation. He shows that this gives the same estimate as is obtained by selecting M to maximize $\Pr(x|M_{x1};M)$ with M_{x1} replaced by $E(M_{x1})$. If ν $G^*(\theta)$ is the m.g.f. associated with this expression (i.e. the conditional m.g.f. of x given M_{x1}, with M_{x1} replaced by $E(M_{x1})$) and $G(\theta)$ is the true expression for the unconditional m.g.f. of x, then Pyke proves that $G^*(\theta)$ converges to $G(\theta)$ for large M. Thus for large M, Pyke's estimate converges to the m.l.e. Pyke investigates the asymptotic variance and asymptotic bias of the estimate $\hat{M}$ by expanding $\hat{M}^2$ and $\hat{M}$ in Taylor series about the point $E(x/n)$. The estimate is $\hat{M} = (x_1+y_1)[nx_2-x(x_2+y_2)]/[x(x_1+y_1)-nx_1]$ but the expressions for the bias and variance are too complex to be given here. If samples are taken in both A_1 and A_2 it is possible to take M_{x1} and M_{y1} as parameters and thus avoid making any assumptions about the distribution of M_{x1}. In this case, m.l.e. are easily derived for M and M_{x1}

2. If the x/y dichotomy is arranged by marking one group (the x_i) in each area, then it is easy to arrange that the x_1 are distinguishable from the x_2 individuals. In his second model, Pyke makes use of this additional information. He is able to derive the m.l.e. for M amd M_x. The asymptotic variance for M is easily derived by the Taylor series method, but for M_x he again finds it necessary to assume that M_x has the binomial distribution (5.4.1).

3. In the third model migration is allowed in both directions and the x_i are assumed to be indistinguishable. Sampling is carried out in both areas. Pyke gives m.l.e. for M amd M_x, where these are now net migration figures, and can thus

take negative or positive values. He derives the likelihood when the x_i and y_i are distinguishable and shows that this cannot yield estimates of all four migration parameters (the actual migration in each direction, as opposed to net migration).

These estimation methods are based on rather unrealistic assumptions (no death, all initial sizes known etc.) and do not incorporate the sort of assumptions that can usually be safely made (e.g. marked and unmarked animals are subject to the same migration rates).

In the second part of his thesis, Pyke developed the stochastic migration models already described in section 2.3, and developed estimates for the parameters. For model one (independent migration, one way only), he develops the likelihood of the form (5.2.1) for complete observations at $t_1 \ldots t_k$. This easily yields a m.l.e. of the migration intensity, and its variance is also derived. Pyke also develops the likelihood for a sampling experiment like that of model 2 above. He derives an estimate by the method described for his first model; i.e. by replacing M_x by $E(M_x)$ in the expression for $Pr(x|M_x)$. He proves that the estimate approaches the m.l.e. when the population sizes at $t = 0$ are large.

For the time dependent model, with migration intensity $\alpha(t)$, Pyke develops the likelihood for complete counts at times $t_1 \ldots t_k$, and derives estimates for $v_i = \int_{t_i}^{t_{i+1}} \alpha(u)du$, (and variances) and suggests that these estimates may be used to determine the form of the function $\alpha(t)$, by noting that for $t_{i+1}-t_i$ small:

$$v_i \doteq (t_{i+1}-t_i)\alpha[(t_{i+1}+t_i)/2]$$

This gives estimates $\hat{\alpha}(t)$ at the midpoint of each time interval, and as he also develops the variance covariance matrix for these values of $\hat{\alpha}(t)$, it would be possible to fit any functional form for $\alpha(t)$ by the method of least squares.

The methods of Pyke were quickly superseded by a much more general method of Chapman and Junge (1956). Marked animals are released in each of r populations at t_1. It is not necessary that initial population sizes be known. As with Pyke's first three models, estimates are formed without making any assumptions about the manner in which individuals migrate, nor will the estimates be invalidated if tagging affects migration. To derive variances for these estimates, it was necessary to adopt additional conditions. Chapman and Junge state these assumptions in terms of distributional properties but they may be provided for by the following conditions:

1. random (binomial) sampling is carried out in each population at recapturing time t_2; all animals alive in a population at time t_2 have equal and independent probabilities of being taken in the sample.

This condition is sufficient to derive estimates. To derive variances, they also require:

2. all individuals alive in population i have equal and independent probabilities of migrating to population j between $t = 0$ and $t = \tau$.

The following notation is required. The notation is consistent with the unified notation recommended by Cormack (1968), and is used for work in this thesis involving marking:

N_{12ij} : number of animals in A_i at $t_1 (t = 0)$ and in A_j at $t_2 (t = \tau)$.

M_{12ij} : number of animals, that were marked in A_i at t_1, and are alive and in A_j at t_2.

m_{12ij} : number of animals, that were marked in A_i at t_1, and are captured in A_j at t_2.

$M_{12(i)j}$: number of animals, alive and in A_i at t_1 (though not necessarily marked) that are captured in A_j at t_2.

n_{1i} : number sampled in A_i at t_1 $= \sum_j M_{12ij} = M_{12i\cdot}$

n_{2j} : number sampled in A_j at t_2 $= \sum_i M_{12(i)j}$

Capital letters thus represent unobservable random variables, and small letters, observations. A dot in the subscript list indicates summation over that subscript. Since there are only two sampling periods, we need not specify t_1 and t_2 by the first two subscripts, so for the present discussion, these will be dropped. Condition 1 above then leads to the following expectation:

$$E(m_{ij} | M_{ij}) = n_{2j} \frac{M_{ij}}{N_{\cdot j}}$$

Summing this over j, and setting expectations equal to observations gives the equation for the moment estimate of the $\hat{N}_{\cdot j}$, namely

$$\sum_j m_{ij}\hat{N}_{.j}/n_{2j} = n_{1i} \qquad (5.4.2)$$

If $[M]$ is the matrix $[m_{ij}]$ and $\underline{n}_1' = (n_{11} \ldots n_{1r})$ $\underline{n}_2' = (n_{21} \ldots n_{2r})$ $\underline{N}' = (N_{.1} \ldots N_{.r})$, then the total size is estimated by

$$\hat{N}_{..} = \underline{n}_2'[M]^{-1}\underline{n}_1$$

(5.4.2) is easily solved for the $\hat{N}_{.j}$. If condition 2 is also assumed, then

$$E(m_{ij}) = n_{1i}N_{ij}/N_{i.}$$

which, in the same way, leads to equations for the moment estimates $\hat{N}_{i.}$, namely

$$\sum_j m_{ij}\hat{N}_{i.}/n_{1i} = n_{2j}$$

which leads to estimates for $\hat{N}_{ij} = \dfrac{m_{ij}\ \hat{N}_{i.}\ \hat{N}_{.j}}{n_{1i}\ n_{2j}}$

Chapman shows that the estimate $\hat{N}_{..}$ is consistent, and shows that other estimates (e.g. the Petersen estimate, ignoring migration, and some estimates suggested by previous workers, notably Schaeffer (1951)) can be decidedly misleading. If death occurs, the Chapman and Junge estimates of $\hat{N}_{i.}$ are still valid, and so an estimate of $\hat{N}_{..}$ (at t_1) is available (assuming equal and independent probabilities of death for all animals in the same population at t_1). The $N_{.j}$ are not estimable, however, nor are the N_{ij}.

By assuming that the m_{ij}, given the M_{ij} are multinomially distributed (with mean given by (5.4.2), and that the M_{ij} are multinomially distributed with $E(M_{ij}) = (n_{1i}\ N_{ij}/N_{i.})$ it is possible to derive the unconditional variances and covariances of the m_{ij}, and hence, using the Taylor expansion of $\hat{N}_{..}$ to linearize the expression, it is possible to derive an expression for the asymptotic variance of $\hat{N}_{..}$. Variances for the $\hat{N}_{ij}$ were not derived but this can easily be done by the same method. Chapman and Junge's paper should be referred to for references to earlier work where two sample marking experiments are applied to migrating populations. The theory of these eariler papers is obviated by their paper however, so they will not be reviewed here.

Darroch (1961) considered the same situation as Chapman and Junge, but investigated the likelihood derived using the conditions mentioned above. He shows that the estimates for $N_{.j}$ defined by (5.4.2) are m.l.e. He also derives the variance-covariance matrix for the $\hat{N}_{.j}$, and shows that the bias is negligible for large $N_{.j}$. Darroch proceeds by developing m.l.e. for the $\rho_j = 1/p_j$, p_j = Prob (an animal, alive in stratum j at t_2 is taken in the sample) and θ_{ij} = Prob (an animal, marked in population i at t_1 is in population j at t_2, given that it survives to t_2). The estimates of the $\hat{N}_{.j}$ are single valued functions of these m.l.e., and so are also m.l.e., but in addition, this formulation allows him to examine other questions. He shows that if the matrix $[\hat{\theta}_{ij}]$ is close to singularity, the estimates $\hat{N}_{.j}$ will be very inaccurate. This leads him to recommend ways of grouping data to rectify this. Also, if the independence assumptions for movement and recapturing fail to hold, this will affect the variance covariance matrices of the $\hat{\rho}_i$ and the $\hat{\theta}_{ij}$. He is thus able to examine the effect of relaxing these assumptions and shows that the estimates are still consistent, and develops modified expressions for the asymptotic variances. Darroch also gives a detailed explanation of how estimates must be modified if the number of populations involved at t_1 differs from the number sampled at t_2.

One of the main disadvantages of the two sample marking experiment, is that, unless dead individuals are actually sampled at t_2 (as happens in the Schaeffer experiment with salmon), one can say very little about death rates. Death may include true death or may be taken in the extended sense to mean permanent migration out of the populations under study. Two methods of overcoming this are

(i) adopt a model (and experimental procedure) that allows for explicit estimation of the "death" rate

(ii) extend sampling to three or more points in time to allow for the estimation of additional parameters.

The first approach is taken by Dempster (1957), and the second by Iwao (1963).

The model used by Dempster, for migration of grasshoppers among contiguous quadrats laid out on the ground, is based on the diffusion models of Skellam (1951) and has already been described in section 2.3. The distribution of individuals over the surface is assumed to be defined by

$$\frac{\partial f}{\partial t} = \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) - \mu f$$

Now if one considers a set of nine quadrats in a 3 x 3 grid, the number of animals in each quadrat may be denoted by f_{xy} $x, y = 1 \ldots 3$ and this may be approximated, at t_1, by the general quadratic surface

$$f_{xy}(t_1) = a+bx+cx^2+ey+gy^2+hxy$$

Then Dempster finds estimates for the coefficients by least squares, using the observations at t_1, and then estimates $c_1 = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$ by taking the second derivatives of the quadratic equation to give

$$\hat{c}_1 = 2(\hat{c}+\hat{g})$$

The change in numbers in the central square between t_1 and t_2, say $f_{22}(t_1)-f_{22}(t_2)$ is thus the result of diffusion of individuals from adjacent quadrats (αc_1) minus losses due to death ($\mu f_{22}(t_1)$). This gives an equation for the two parameters α and μ:

$$f_{22}(t_1)-f_{22}(t_2) = \alpha\hat{c}_1 - f_{22}(t_1)$$

Dempster had a 6 x 3 array of quadrats, and so developed four sets of equations in the two unknowns, by considering the four possible 3 x 3 sets of quadrats, centred on quadrats f_{22}, f_{32}, f_{42}, f_{52}. These four equations were then solved for the two unknowns. In Dempsters experiment he took $f_{xy}(t_1)$ to be the number of first in star locusts, and $f_{xy}(t_2)$ to be the number of second instar locusts, so that μ was in fact an estimate of first instar mortality.

In many ways this method of estimation is rather inadequate (for instance, there are no estimates of the variances of $\hat{\alpha}$ and $\hat{\mu}$, and these variances appear to be considerable), and this example by no means exploits all possible applications of Skellam's models. These models could probably be more successfully combined with marking experiments.

Iwao's method, for estimating the survival rates of individuals that are capable of migration between two areas, makes use of three sampling periods. It is an extension of an earlier method developed by Richards and Waloff (1954), which used data from two sampling periods. They thus needed to make the assumption that the fraction surviving on each population was the same, and that the ratio of the sampling fractions at the two sampling times was known for each population, in order to reduce the number of parameters. Iwao's method of development is essentially the same as that of Richards and Waloff, but, being more general, only Iwao's estimates will be described.

Iwao develops a deterministic model, but there is no essential difference if we state the model in stochastic terms and then proceed by using moment estimates. As this is more consistent with work developed in this thesis, we shall do this, and, so, require the following notation (which is consistent with that defined above in describing the work of Chapman and Junge, and will be used again in Chapter 8).

$m_{t_i t_j ij}$	The number of individuals captured in A_j at sampling time t_j that were last seen (and marked) at t_i in population A_i [previous capture history, prior to the most recent marking does not enter into Iwao's estimates].
$M_{t_i t_j ij}$	The number of individuals, alive in A_j at t_j that were last marked in A_i at time t_i.
$p_{t_j j}$	Prob (an individual, alive in A_j at time t_i is taken in the t_jth sample).
$Q_{t_i ij}$	Prob (an individual, alive in A_i at time t_i is alive and in A_j at time t_i+1).
$N_{t_j j}$	The number of individuals alive in A_j at t_j.

$n_{t_j j}$ The size of the sample taken in A_j at t_j.

This notation becomes more convenient if sampling times are denoted by numbers (1, 2 ...) and the populations by A or B, to denote populations A_1 and A_2 respectively. Thus

$(1-\mathbb{Q}_{1AA}-\mathbb{Q}_{1AB})$ = Prob (an individual, alive in population A at the time the first sample is taken, does not survive until the time the second sample is taken).

Now if the conditions of the Chapman Junge experiment are fulfilled (including independent and equal survival probabilities, between t = 1 and t = 2 of all individuals alive in the same population at t = 1), then we have the following expectations:

$$E(m_{12AA}) = N_{1A}p_{1A}p_{2A}\mathbb{Q}_{1AA}$$
$$E(m_{12BA}) = N_{1B}p_{1B}p_{2A}\mathbb{Q}_{1BA}$$
$$E(m_{12AB}) = N_{1B}p_{1A}p_{23}\mathbb{Q}_{1AB}$$
$$E(m_{12BB}) = N_{1B}p_{1B}p_{2B}\mathbb{Q}_{1BB}$$

By removing the expectation signs, and manipulating these equations, Iwao develops the following relations for the moment estimates:

$$\hat{\mathbb{Q}}_{1AB} = \hat{\mathbb{Q}}_{1BB} \cdot n_{1B} \cdot m_{12AB}/(n_{1A} \cdot m_{12AA})$$
$$\hat{\mathbb{Q}}_{1BA} = \hat{\mathbb{Q}}_{1AA} \cdot n_{1A} \cdot m_{12BA}/(n_{1B} \cdot m_{12BB})$$

Iwao then notes that several methods exist for estimating the survival on a single population, all based on (at least) three sampling periods [e.g. Leslie (1952), Bailey's triple catch method (Bailey (1951)), Jolly (1965)]. He therefore requires that all animals, taken at the second sampling time also be marked (or remarked if necessary) and returned to the population in which they were sampled. A third sample is then taken at a later time, providing the additional observations necessary to estimate $\hat{Q}_{1AA}$ and $\hat{Q}_{1BB}$. Iwao uses Leslie's three point sampling method (Leslie (1952)), to estimate Q_{1AA}, Q_{1BB}, but, Leslie's model is deterministic. Since we have stated the model in stochastic terms here, we can instead use the method of Jolly (1965), with three sampling periods and no loss on capture but immigration must be allowed for, as animals may enter from the opposite population. It may be shown that the Leslie and the Jolly estimates of survival in this particular case are identical. Then ignoring all observations in population B, we have

$$\hat{Q}_{1AA} = \hat{M}_{21AA}/m_{1A}$$

$$= \frac{\text{(estimated number of animals, marked in A at t = 1, that are alive and in A at t = 2)}}{\text{(size of sample in A at t = 1)}}$$

and the Jolly estimate for $\hat{M}_{21AA}$ is derived from the equation:

$$\frac{m_{13AA}}{\hat{M}_{12AA}-m_{12AA}} = \frac{m_{23AA}}{n_{2A}} \qquad (5.4.3)$$

and hence

$$\hat{Q}_{1AA} = (n_{2A}m_{13AA}+m_{12AA}m_{23AA})/(m_{23AA}n_{1A})$$

and similarly

$$\hat{Q}_{1BB} = (n_{2B}m_{13BB}+m_{12BB}m_{23BB})/(m_{23BB}n_{1B})$$

However, as Jolly points out (Jolly (1965), p. 238-9), equation 5.4.3 is reasonable, under the assumptions used in constructing his model for behaviour in a single population, because the following argument holds (adapted here to the three sample situation). At $t = 2$ two classes of marked animals may be distinguished: the m_{12} marked animals that were captured at $t=2$, (and could only have been marked at $t = 1$), and the $M_{12}-m_{12}$ animals that were marked at $t = 1$ and were not captured at $t = 2$. Just after $t = 2$, there are the two groups: the $M_{12}-m_{12}$, and the n_2 ($= m_{12}+u_2$ where u_2 are the unmarked animals in the second sample) animals just released. Of the former, m_{13} are caught at $t = 3$. Under the assumptions of Jolly's model, the proportion of subsequent recaptures (at $t = 3$) from each of these groups at $t = 2$ is expected to be the same, so that the equation

$$\frac{m_{13}}{M_{12}-m_{12}} = \frac{m_{23}}{n_2}$$

holds and can be used to estimate M_{12}. However, one of the explicit assumptions is that, if animals leave the population, they do so permanently; they do not leave between $t = 1$ and $t = 2$ and re-enter between $t = 2$ and $t = 3$. If they do, and are in B at $t = 2$, the m_{13AA} may not only be members of $M_{12AA}-m_{12AA}$, but may also be members of the $M_{12AB}-m_{12AB}$, and the equation (5.4.3) is inappropriate for estimating $\hat{M}_{12AA}$. Iwao appears to be unaware of this restriction.

If, however, all individuals migrate at most once between $t = 1$ and $t = 3$ as may happen if this time interval is relatively short, the m_{13AA} will all necessarily have been members of $M_{12AA}-m_{12AA}$ and Iwao's estimate will be valid. The presence of any appreciable numbers of individuals with the capture histories m_{123ABA} or m_{123BAB} will indicate that one should not use Iwao's estimates. Also the experimenter should make the interval between $t = 2$ and $t = 3$ as short as possible so that any individuals that may have migrated between t = 1 and t = 2 do not have a chance to return before the last sample is taken.

If these assumptions hold, Jolly's method also provides an estimate of N_{2A} and N_{2B}

$$\hat{N}_{2A} = \frac{\hat{M}_{21AA}}{\alpha_{2A}}$$

where $$\alpha_{2A} = \text{(number of marked animals in second sample, ignoring those marked in B)} / n_{2A}$$

$$= m_{12AA}/n_{2A}$$

and similarly

$$\hat{N}_{2B} = \hat{M}_{21BB} . n_{2B}/m_{12AA}$$

Finally N_{1A} and N_{1B} may be derived from the equations

$$N_{2A} = N_{1A}Q_{1AA}+N_{1B}Q_{1BA}$$

$$N_{2B} = N_{1A}Q_{1AB}+N_{1B}Q_{1BB}$$

In a later paper (Iwao et al (1966)), these estimates are applied to a mark recapture experiment with fifth instar insects in the field. Sampling took place every three days for a month and estimates were calculated for successive sets of three sampling periods. Complete capture histories are not given so it is not possible to tell whether the assumption, that insects migrate at most once, is fulfilled. Iwao has not yet developed expressions for the variances of the survival estimates.

In a recent work, Matis and Hartley (1970) develop the general model for linear migration (with death) for the n-dimensional system. Their deduction of the properties of this model leads to the same conclusions as were developed (independently) in section 4.3.3 of this thesis, namely, they demonstrate that the distribution is the convolution of multinomial distributions, and identify the form of the parameters Q_{ij}. The nth cell is, in fact, taken to be the "death" state, from which there is no return, so there are $(n-1)^2$ linearly independent $Q_{ij}(t)$, defining the distribution at time t.

$$Q_{ij}(t) \quad i, j = 1 \dots n-1$$

$$Q_{in}(t) \quad i = 1 \dots n$$

$$= \text{prob. (an individual, alive in cell } i \text{ at time } 0 \text{ is dead at time } t).$$

$$= 1 - \sum_{j=1}^{n-1} Q_{ij}(t)$$

Their estimation procedure is based on observation of the total number in the system at times $t_1, t_2 \dots t_s$: $N_T(t_i) = \sum_{j=1}^{n-1} N_j(t_i)$. They show that

$$\text{cov}[N_T(t_i)\ N_T(t_j)] = \sum_{s=1}^{n-1} N_s(0)(1 - Q_{sn}(t_j))Q_{sn}(t_i)$$

$$i, j = 1 \dots n-1$$

and so have an expression for $\sum$, the variance-covariance matrix of the vector of observations $\underline{N}_T$ in terms of the parameters Q_{ij} and hence, in terms of the γ_{ij}. The γ_{ij} are then estimated using the following iterative technique.

1. Use the Gauss-Newton Linearization technique, and a set of initial values for the parameters γ_{ij} to iterate to the best linear unbiased estimates of γ_{ij} under the assumption that $\sum = \sigma^2 I$.

2. Use the "improved" parameter estimates to valuate $\sum$.

3. Use Aitken's generalized least squares theorem iteratively to derive the best linear unbiased estimate of the γ_{ij} using this value of $\sum$.

4. Repeat steps two and three until successive results of step 3 agree to any desired accuracy. Variances for the estimates are then calculated using the last estimate of $\sum$.

Matis and Hartley encounter no difficulties in having this method converge to admissible estimates when the method was applied to both simulated and experimental data. However they were unable to demonstrate theoretically that this would always be the case.

The estimation procedure is similar to iterative schemes suggested in this thesis (section 7) for the more restricted case of observations at only two times. In this thesis, we also consider the situation where the data are the set of $N_i(t_j)$ $i = 1 \ldots n-1$, $j = 1 \ldots s$, and not merely the set of totals $N_T(t_j)$ $j = 1 \ldots s$. The above iterative scheme might also be applied in this case, but it would involve developing an expression for the $s(n-1) \times s(n-1)$ variance-covariance matrix for the $N_i(t_j)$. The results of section 3.3 viii suggest that this is practicable only when the observations are at equally spaced times ($t_i = it$), but this method could certainly bear investigation in the case of such data.

CHAPTER SIX

ESTIMATION FROM STEP FUNCTION OBSERVATIONS

6.1 Introduction

The most complete information available for estimating the parameters of a markov process, is complete knowledge of the step function; that is, the experimenter knows the initial state, and then observes each subsequent change of state and the time lapse between each change. The estimation theory has been developed by Billingsley (1961) and is described in section 5.2. In this chapter the theory is adapted to the models of Chapters 3 and 4, and estimates derived. It is rare, in animal ecology, to have such detailed information, and no such data seems to appear in recent ecological journals. It should be possible, however, to carry out such experiments in certain instances. For example, with populations of small rodents, one may be able to surround an area by a trip wire, or detector strip, such that every time an animal crosses the boundary, a mark is recorded which indicates the time and direction of crossing. If no such automatic monitoring is possible, direct observation may be feasible, and will yield useful information if migration rates are high. In many situations, it may be unreasonable to insist that the initial state is known. In this chapter, we will also determine if it is possible to estimate the initial state from step function data.

Although it may frequently be impossible to obtain step function observations, the sort of estimates that are possible with such data may suggest ways of forming estimates from less complete data.

6.2 Estimates for Closed Populations

To implement the theory of section 5.2, we require expressions for the function $q_{ij}(\theta)$ and $q_i(\theta)$ as defined by equations (5.2.8) and (5.2.9) respectively. Comparison of these definitions with the development of expressions for the transition probabilities in the general formulation of section 3.2 gives immediately:

$$q_{i,i-1}(\theta) = \lim_{\delta t\to 0} \frac{i\beta(i,t)\delta t+O(\delta t)^2}{\delta t} \quad \text{(from equation 3.2.3)}$$

$$= i\beta(i,t) \qquad (6.2.1)$$

$$q_{i,i+1}(\theta) = \lim_{\delta t\to 0} \frac{(N-i)\alpha(N-i,t)\delta t+O(\delta t)^2}{\delta t} \quad \text{(from equation 3.2.4)}$$

$$= (n-i)\alpha(N-i,t) \qquad (6.2.2)$$

$$q_{i,j}(\theta) = \lim_{\delta t\to 0} \frac{O(\delta t)^2}{\delta t} \quad \text{(from equation 3.2.5)}$$

$$= 0 \qquad j \neq i+1,\ i-1$$

and

$$q_i(\theta) = \sum_{j\neq i} q_{ij}(\theta) = i\beta(i,t)+(N-i)\alpha(N-i,t) \qquad (6.2.3)$$

6.3 The Linear Model

$$q_{i,i+1} = (N-i)\alpha$$
$$q_{i,i-1} = i\beta$$

and the likelihood of observations $\{z_k, r_k | k = 1 \ldots n(T)\}$ is given by (5.2.11)

$$\log L = \sum_{i=0}^{N} n_{i,i+1} \log (N-i)\alpha - g_i(N-i)\alpha + \sum_{i=0}^{N} n_{i,i-1} \log \beta i - g_i i\beta \qquad (6.3.1)$$

Differentiating w.r.t. α and β and setting the result equal to zero gives:

$$\frac{\partial \log L}{\partial \alpha} : (\textstyle\sum n_{i,i+1})/\alpha - \sum g_i(N-i) = 0$$

$$\frac{\partial \log L}{\partial \beta} : (\textstyle\sum n_{i,i-1})/\beta - \sum i g_i = 0$$

hence

$$\hat{\alpha} = \frac{\sum n_{i,i+1}}{\sum (N-i) g_i}$$

$$= \frac{u}{NT - \sum i g_i} \qquad (6.3.2)$$

where u is the number of upward transitions (i.e. changes of state that result in a unit increase in the size of the population of A_1), and T is the time to the last change of state. Also, if there are d downward transitions:

$$\hat{\beta} = \frac{d}{\sum i g_i} \qquad (6.3.3)$$

Note that $\sum i g_i = \int_0^T X_1(t)dt$, and $(NT - \sum i g_i) = \int_0^T X_2(t)dt$, so that the m.l.e. $\hat{\alpha}$ and $\hat{\beta}$ are similar to moment estimates derived from equations [(3.3.6) and (3.3.7)] for the expected number of upward and downward transitions in [0,T].

To derive the asymptotic variances, we have by differentiating again and taking expectations

$$-E\,\frac{\partial^2 \log L}{\partial \beta^2} = \frac{E(d)}{\beta^2}$$

$$\therefore \text{ as.var } (\hat{\beta}) = \beta^2/E(d)$$

and, similarly as.var $(\hat{\alpha}) = \alpha^2/E(u)$.

Integrating the expressions (3.3.7) for E(d), E(u), gives

$$E(d) = \frac{N\alpha\beta T}{(\alpha+\beta)} - \frac{\beta(x_{20}\alpha - x_{10}\beta)}{(\alpha+\beta)^2}[1-\exp(-(\alpha+\beta)T)] \qquad (6.3.4)$$

and $$E(u) = \frac{N\alpha\beta T}{(\alpha+\beta)} + \frac{\alpha(x_{20}\alpha - x_{10}\beta)}{(\alpha+\beta)^2}[1-\exp(-(\alpha+\beta)T)] \qquad (6.3.5)$$

Also $\frac{\partial^2 \log L}{\partial\alpha\partial\beta} = 0$ so that $\hat{\alpha}$ and $\hat{\beta}$ are asymptotically independent (as T becomes large).

When $\alpha = \beta$, the log likelihood simplifies to

$$\log L = \sum(n_{i,i+1}+n_{i,i-1})\log\alpha - g_i N\alpha$$

giving $$\hat{\alpha} = (u+d)/NT$$

$$\text{as. var}(\hat{\alpha}) = \alpha^2/(E(u)+E(d))$$

It proved impossible to derive the exact distribution for $\hat{\alpha}$ and $\hat{\beta}$, so one must rely on the asymptotic properties. Since the conditions given by Billingsley are fulfilled, the estimates are consistent, and asymptotically normally distributed. However it is of some interest to know how well the estimates serve in "small sample" situations. In particular, we can examine the estimates for unbiasedness, and the adequacy of the normal approximation. If these properties hold, then one can use $\hat{\alpha}$ and the variance of $\hat{\alpha}$ to form confidence intervals for α, and similarly for β. In any real situation, the variance of $\hat{\alpha}$ can be estimated by

$$\hat{\text{var}}(\hat{\alpha}) = (\hat{\alpha})^2/\hat{E}(u)$$

where $\hat{E}(u)$ is (6.3.5) evaluated at $\hat{\alpha}$, $\hat{\beta}$. It is also of interest to see if this expression, and the similar expression for $\hat{\text{var}}(\hat{\beta})$ are adequate in small samples. To investigate these questions, a number of simulations were carried out, the results of which are summarized in

Table 6.1.

1. Explanation of Table 6.1

Five sets of simulations are reported. For each set, 300 simulations were carried out, and the results are in the following form for each set.

line 1: the parameters of the model

line 2: (theoretical) properties of the model

line 3: distribution of $\hat{\alpha}$: mean (m), variance (v), skewness (s) over 300 simulations.
distribution of $\hat{var}(\hat{\alpha})$: mean (m), standard deviation ($\sqrt{v}$)

line 4: as for line 3, but for $\hat{\beta}$ and $\hat{var}(\hat{\beta})$

line 5: $\hat{p}_A$, the porportion of the realisations for which the nominal 95% confidence interval ($\hat{\alpha} \pm 1.96\sqrt{\hat{var}(\hat{\alpha})}$) includes the true value α, with the s.e. of $\hat{p}_A$. Similarly $\hat{p}_B$ and s.e. ($\hat{p}_B$).

The sampling variance of the estimate over the 300 simulations may be used as a basis for comparing the asymptotic variance (line 2) and the estimate of the variance. The precision of the estimate of the variance is indicated by its s.d. ($\sqrt{v}$). The skewness coefficient (s) is the third sample moment, divided by the cube of the s.d. If the sample comes from a normal distribution, s is approximately

Table 6.1

Simulation of the Simple Linear Model

with

M.L.E. for α and β from Step Function Observations

(for explanation, see text)

$N = 2000$ $x_{10} = 500$ $t = 0.3$ $\alpha = 1$ $\beta = 2$

$E(u) = 44.78$ $E(d) = 30.44$ as. $var(\hat{\alpha}) = .0233$ as. $var(\hat{\beta}) = .1314$

$\hat{\alpha}$: $m = .9994$ $v = .0222$ $s = -.06$ $\hat{var}(\hat{\alpha})$: $m = .0223$ $\sqrt{v} = .0034$

$\hat{\beta}$: $m = 2.0609$* $v = .1360$ $s = +.24$ $\hat{var}(\hat{\beta})$: $m = .1357$ $\sqrt{v} = .0251$

$\hat{p}_A = .9467$ S.E. $= .0130$ $\hat{p}_B = .9400$ S.E. $= .0137$

$N = 100$ $x_{10} = 25$ $t = .03$ $\alpha = 20$ $\beta = 40$

$E(u) = 42.32$ $E(d) = 35.36$ as. $var(\hat{\alpha}) = 9.452$ as. $var(\hat{\beta}) = 42.245$

$\hat{\alpha}$: $m = 20.313$ $v = 9.760$ $s = +.22$ $\hat{var}(\hat{\alpha})$: $m = 9.467$ $\sqrt{v} = 1.780$

$\hat{\beta}$: $m = 41.222$* $v = 47.697$ $s = +.03$ $\hat{var}(\hat{\beta})$: $m = 47.637$ $\sqrt{v} = 11.717$

$\hat{p}_A = .9400$ S.E. $= .0137$ $\hat{p}_B = .9433$ S.E. $= .0134$

$N = 20$ $x_{10} = 10$ $t = 1.0$ $\alpha = 1$ $\beta = 5$

$E(u) = 15.56$ $E(d) = 22.21$ as. $var(\hat{\alpha}) = .0643$ as. $var(\hat{\beta}) = 1.1257$

$\hat{\alpha}$: $m = 1.0017$ $v = .0596$ $s = .24$ $\hat{var}(\hat{\alpha})$: $m = .0646$ $\sqrt{v} = .0182$

$\hat{\beta}$: $m = 5.3502$* $v = 1.3688$ $s = .68$ $\hat{var}(\hat{\beta})$: $m = 1.3337$ $\sqrt{v} = .5637$

$\hat{p}_A = .9533$ S.E. $= .0122$ $\hat{p}_B = .9367$ S.E. $= .0141$

$N = 20$ $x_{10} = 20$ $t = 1.0$ $\alpha = 1$ $\beta = 5$

$E(u) = 13.90$ $E(d) = 30.52$ as. $var(\hat{\alpha}) = .0720$ as. $var(\hat{\beta}) = .8191$

$\hat{\alpha}$: $m = 1.0061$ $v = .0725$ $s = .19$ $\hat{var}(\hat{\alpha})$: $m = .0731$ $\sqrt{v} = .0226$

$\hat{\beta}$: $m = 5.2002$* $v = 1.0219$ $s = .69$ $\hat{var}(\hat{\beta})$: $m = .9113$ $\sqrt{v} = .3452$

$\hat{p}A = .9233$ S.E. $= .0154$ $\hat{p}_B = .9367$ S.E. $= .0141$

$N = 20$ $x_{10} = 20$ $t = .25$ $\alpha = 1$ $\beta = 5$

$E(u) = 2.01$ $E(d) = 14.96$ as. $var(\hat{\alpha}) = .4978$ as. $var(\hat{\beta}) = 1.6715$

$\hat{\alpha}$: $m = 1.0009$ $v = .5516$ $s = .88$ $\hat{var}(\hat{\alpha})$: $m = .5226$ $\sqrt{v} = .4532$

$\hat{\beta}$: $m = 5.3751$* $v = 1.4707$ $s = .78$ $\hat{var}(\hat{\beta})$: $m = 1.9388$ $\sqrt{v} = .8476$

* 99% confidence interval for $E(\hat{\beta})$ excludes true value β.

normally distributed as $N(0, 6/n)$, or in this case, with $n = 300$, $s \sim N(0,(.1414)^2)$. Values of s that are larger than about .3 or .4 in absolute value indicate marked asymmetry in the distribution. (This test is described in Snedecor and Cochran (1967).

The bias of the estimate may be examined by deriving a confidence interval for the expected value of the estimate. Thus, for example, if the mean value of $\hat{\beta}$ over 300 realisations is m_β, and its sampling variance v_β, and if these 300 values come from some distribution with mean $E(\hat{\beta})$ (say), then if $E(\hat{\beta}) = \beta$, the estimate is unbiased. By the central limit theorem, $m_\beta \sim N(E(\hat{\beta}), v_\beta/300)$ and so, a 99% confidence interval for $E(\hat{\beta})$ is

$$m_\beta \pm 2.57\sqrt{v_\beta}/\sqrt{300}$$

If this confidence interval excludes the true value β, it is fairly certain that the estimate is in fact biased, and where this is the case, we shall refer to the estimate as being "detectably biased".

For all models, $\alpha < \beta$, since each model, "viewed" from A_2 instead of A_1 gives the same result with α and β(and A_1 and A_2) interchanged. Thus the first set can also be considered as the model $N = 2000$ $x_{10} = 1500$ $\alpha = 2$ $\beta = 1$.

2. Discussion of the Results in Table 6.1

As one might expect, the asymptotic normal properties for the estimates are more nearly fulfilled the larger u and d are. It also appears that these properties are least satisfactorily fulfilled by the larger of the two estimates. The larger of the two estimates tended to have a more skewed distribution, and to exhibit greater bias. (For all sets in table 6.1 $\hat{\beta}$ as "detectably biased", although in no case does the bias appear to be greater than 10%. There was no "detectable bias" in $\hat{\alpha}$ in any of the sets.). For d or u less than 20, the estimates show noticeable skewness. As expected, the skewness is positive due to the constraint; $\alpha,\beta > 0$. Nevertheless, confidence statements based on the normal approximation do not appear to be seriously misleading (at least in the long run!). The estimates $\hat{\text{var}}(\hat{\alpha})$, $\hat{\text{var}}(\hat{\beta})$ appear to be virtually unbiased, though somewhat unprecise if u or $d < 20$.

Interpretation of the last set of results is made difficult by the very low value of E(u). In over 12% of the simulations u was observed to be zero. In such a situation, the experimenter would presumably fit a model allowing migration from A_1 to A_2 only. In this case, the estimate of β (6.3.3) is still the m.l.e. for the stochastic migration model with migration from A_1 to A_2 only. Also, the asymptotic variance $\beta^2/E(d)$ is correct (with E(d) calculated from 6.3.4 with $\alpha = 0$: i.e.

$$E(d) = \beta[1-\exp(-\beta T)] \qquad).$$

Thus, the estimate of β is the m.l.e., regardless of whether the model allows one way migration only, or two way migration. The experimenter will, however, be wrong in assuming that migration is unidirectional. For this set of simulations, it was also found that the confidence intervals for α were quite unsatisfactory, due to the extreme skewness in the distribution of $\hat{\alpha}$. In fact, if the 95% confidence interval for α is calculated, $(\hat{\alpha}\underline{+}1.96\sqrt{\hat{\text{var}}(\hat{\alpha})})$, it includes inadmissible values for α $(\alpha < 0)$ in over half the simulations.

To summarize, it seems that $\hat{\alpha}$ and $\hat{\beta}$ are quite good estimates so long as u and d are reasonably large (>20) regardless of N, x_{10}, x_{20}, but some bias can be expected, particularly in the larger parameter, though not exceeding 10%. For u, d > 20 the asymptotic normal theory may safely be used for making confidence statements about α and β.

3. Estimation for the Linear Model when N, x_{10} are Unknown

In this case, the theory of Billingsley (section 5.2) no longer holds, as the size of the state space for the observations is a function of the unknowns x_{10} and x_{20}. Instead of the observations $\{z_k, r_k | k = 1 \ldots n(T)\}$, we now have only the observations $\{w_k, r_k | k = 1 \ldots n(T)\}$ where

$$w_1 = 0$$

and

$$z_k = X_1(t) = x_{10}+w_k \quad \text{for} \quad \sum_{j=0}^{k-1} r_j \leqslant t < \sum_{j=0}^{k} r_j \qquad (r_0 = 0)$$

and since z_k can take values 0 ... N, the w_k may take values $-x_{10}$ ... -1, 0, 1, ... x_{20}. Now if we define $\tilde{n}_{i,i+1}$ as the number of times $w_k = i$, $w_{k+1} = i+1$ $(i = -x_{10} \dots x_{20}-1)$ and $\tilde{n}_{i,i-1}$ as the number of times $w_k = i$, $w_{k+1} = i-1$ $(i = -x_{10}+1, \dots x_{20})$, and $\tilde{g}_i = \sum r_k$ where the summation is over all r_k such that $w_k=i$, then the log likelihood is:

$$\begin{aligned}\log L = &\sum_{i=-x_{10}}^{x_{20}-1} \tilde{n}_{i,i+1} \log[N-(x_{10}+i)]\alpha \\ &+ \sum_{i=-x_{10}+1}^{x_{20}} \tilde{n}_{i,i-1} \log(x_{10}+i)\beta \\ &- \sum_{i=-x_{10}}^{x_{20}} \tilde{g}_i[(N-x_{10}-i) +(x_{10}+i)\beta] \qquad (6.3.6)\end{aligned}$$

$$\text{if } -x_{10} \leq w_k \leq x_{20} \quad k = 1 \dots n(T)$$

$$= 0 \quad \text{otherwise}$$

It would be difficult to deal with this likelihood in a rigorous fashion. There is obviously some information about x_{10} and $x_{20} = N-x_{10}$ from the fact that

$$\begin{aligned}x_{20} &> \max\{w_k | k = 1 \dots n(T)\} \\ -x_{10} &< \min\{w_k | k = 1 \dots n(T)\} \qquad (6.3.7)\end{aligned}$$

but if N is large, and x_{10} is not close to 0 or N, and if T is not large, then $|X_1(t)-x_{10}|$ may be quite small relative to x_{10} and x_{20}, $\forall\, 0 \leq t \leq T$ and the information from (6.3.7) will be negligible. In this case, we might attempt to maximize (6.3.6), for the four parameters $(\alpha, \beta, N, x_{10})$, ignoring the information of (6.3.7). It is convenient to redefine the likelihood in terms of the four parameters (α, β, z, y) where $z = \beta x_{10}$ $y = \alpha x_{20}$. Thus the four parameters are effectively continuous, so that differentiation is possible.

$$\begin{aligned}\log L = {} & \sum \tilde{n}_{i,i+1} \log(y - i\alpha) - yT + \alpha \sum i\tilde{g}_i \\ & + \sum \tilde{n}_{i,i-1} \log(z + i\beta) - zT - \beta \sum i\tilde{g}_i \qquad (6.3.8)\end{aligned}$$

The log likelihood is thus the sum of two terms, the first in α and y only, the second in β and z only. They can thus be maximized separately. Differentiating the log likelihood w.r.t. α and y, and setting equal to zero, does not lead to equations that are easily solved. To see if maximization of the likelihood would lead to useful estimates in a numerical case, one simulation of the model with parameters

$$N = 2{,}000 \quad x_{10} = 500 \quad \alpha = 1 \quad \beta = 2 \text{ and } T = .03$$

was generated to derive a set of the $\tilde{n}_{i,i+1}$, $\tilde{n}_{i,i-1}$ and $\tilde{g}_i$ assuming N, x_{10} unknown. These values were substituted into the first term of the likelihood, and values of α and y were sought to maximize this function, using a computer program for the simplex method of Nelder and Mead (see, for example, Box et al (1969)). This did not converge to admissible values of α and y. Plots of the likelihood surface revealed a long ridge, almost parallel to the line $y = 1500$, and having a maximum height in the positive quadrant at $\alpha = 0$. Similar results occurred if the second term was maximized in z and β. Subsequent simulations lead to the same result.

It thus appears that z and y are estimable from such experiments, but not α and β. This means that virtually identical observations $\{z_k, w_k\}$ can occur either with large numbers and low migration intensity, or with low numbers and high migration intensity. That this is the case is born out by comparing the first set of simulations in table 6.1 with the second set. In both cases we have $y = 1500 \quad z = 1{,}000$, but in the first set $N = 2{,}000$ whereas in the second $N = 100$. Nevertheless, there is such a close similarity between the observations from these two models that, for a given realization with x_{10} and x_{20} unknown, it would be impossible to discriminate between the two. A summary of the observations for these two models, over 300 simulations is given below:

N	u (upward transitions)		d (downward transitions)		$\sum i\tilde{g}_i$	
	mean	s.d.	mean	s.d.	mean	s.d.
2000	44.78	6.59	30.44	5.31	.011	.253
100	42.32	5.53	35.36	4.77	.124	.085

Thus for "small samples" at least [i.e. when T is sufficiently small that u and d are not large, and the equilibrium state of the process is not reached], it appears that step function observations will not yield estimates of α, β, x_{10} and x_{20} for the linear model. Even if it is known that $\alpha = \beta$ this will apply, since large values of x_{10} and x_{20} can compensate for small values of α and vice versa.

However, for large T, the "ridge" in the likelihood surface begins to drop away for very small and very large values of $\hat{\alpha}$ or $\hat{\beta}$ and useful estimates can be obtained. In practice, both to avoid the "small sample" non-identifiability problems mentioned above, and to get adequate initial estimates so that iterative maximization procedures will converge, it is necessary to have T sufficiently large that the distribution achieves equilibrium, and is observed in the equilibrium state for a short period. This can be determined by plotting the step function $\{w_k\}$ against time. The function will tend either to rise (if $x_{10} <$ the equilibrium mean m_e) or to fall ($x_{10} > m_e$), and then level off and oscillate about m_e. After the step function has levelled off, one can form a visual estimate of the average value of w_k at equilibrium, say $\overline{w}_e$, and this gives an estimate of $m_e - x_{10}$. As a first estimate, we may consider the process to be in the equilibrium state from the time the step function first crosses the line $w_k = \overline{w}_e$. Let this time be T_e. Then the mean squared deviation:

$$[\sum (z_k - m_e)^2 r_k]/(T-T_e)$$
$$\doteq [\sum (w_k - \overline{w}_e)^2 r_k]/(T-T_e) = \hat{v}_e \text{ (say)}$$

where summation is over all k such that $t_k = \sum_{i=0}^{k-1} r_i \varepsilon [T_e, T]$. This $\hat{v}_e$ can thus be used as an estimate of the equilibrium variance of the process, v_e. By using the w_k, $\overline{w}_e$ and $\hat{v}_e$, initial estimates of α, β, z and y can be derived, and these may be used as starting values for an iterative maximization of the likelihood (6.3.6). The process of finding initial estimates in some particular cases is described below. Once m.l.e. are obtained in this way, estimates of their variances and covariances are available from the inverse of the matrix of negative second differential coefficients of the log likelihood. Most computer packages for maximizing likelihoods also give this numerical variance covariance matrix as a standard feature.

(i) Initial estimates when $\alpha = \beta$.

The equilibrium distribution is $B(z_k; N, \frac{1}{2})$ (section 3.3.iv with $\alpha = \beta$) which leads to the moment estimates:

$$\hat{N} = 4\hat{v}_e$$

$$\frac{\hat{N}}{2} - \hat{x}_{10} = \overline{w}_e$$

$$\therefore \qquad \hat{x}_{10} = 2\hat{v}_e - \overline{w}_e$$

The m.l.e. for α when x_{10} and x_{20} are known is a function of N only, and so may be estimated in this situation, using $\hat{N}$:

$$\hat{\alpha} = (u+d)/\hat{N}T$$

It is interesting to note that α can be estimated even if observations begin when the system is in equilibrium despite the fact that the equilibrium distribution is not a function of α. This follows, of course, from the fact that the distribution of u and d, even at equilibrium <u>is</u> a function of α.

These initial estimates for α, N and x_{10} can then be used to derive starting values for $z = x_{20}\alpha$ and $y = x_{10}\alpha$ for maximizing the log likelihood:

$$\log L = \sum \tilde{n}_{i,i+1} \log(y-i\alpha) + \sum n_{i,i-1} \log(z+i\alpha) - (z+y)T$$

One is least likely to lose appreciable information by ignoring (6.3.7) when N is large. If x_{10} is close to N/2 when observations begin, then x_{10} will not be estimable and the iterative procedure for maximizing log L will not converge. On the other hand, if x_{10} is close to 0, $\min\{w_k\}$ will yield useful information on x_{10}, though $\max\{w_k\}$ will yield little information on x_{20} if N is large, and conversely if x_{10} is close to N.

(ii) Initial estimates when $\alpha \neq \beta$.

It was found to be impossible to derive simple initial estimates for the general case with x_{10} and x_{20} unknown. If x_{10} is known (but not N), then initial estimates are easily derived. This includes the important case of $x_{10} = 0$. Adding x_{10} to each w_k then gives the z_k, and therefore, we can also determine $n_{i,i+1}$, $n_{i,i-1}$ and $\sum ig_i$ ($\sum ig_i = \sum i\tilde{g}_i + x_{10}T$). From observation of the process at equilibrium, we can form the estimates

$$\hat{m}_e = \bar{w}_e + x_{10}$$

$$\hat{v}_e = [\sum (z_k - \hat{m}_e)^2 r_k]/(T-T_e)$$

The equilibrium distribution is $B(z_k;N,p)$ where $p = \alpha/(\alpha+\beta)$

and setting

$$\hat{N}\hat{p} = \hat{m}_e$$
$$\hat{N}\hat{p}(1-\hat{p}) = \hat{v}_e$$

we may determine $\hat{N}$ and $\hat{p}$. The m.l.e. for β is immediately available from the equation

$$\hat{\beta} = (\sum n_{i,i-1})/\sum ig_i \qquad (6.3.8)$$

Thus estimates of $z = x_{20}\alpha$ and α can be formed and used for maximizing the log likelihood:

$$\log L = \sum n_{i,i+1}\log(z-i\alpha)-g_i(z-i\alpha)$$
$$+\sum n_{i,i-1}\log i\beta-g_i i\beta$$

It can be seen from this equation that the m.l.e. of β is in fact as given by (6.2.8) and so only the first term, in α and z, need be maximized.

If $\alpha >> \beta$ or $\beta >> \alpha$ the information provided by 6.3.7 is likely to be considerable and should not be neglected.

These methods are suitable for forming initial estimates when $|x_{10}-m_e|$ is of moderate size. If $|x_{10}-m_e|$ is at all large, especially if N, α and β are also large, then the number of transitions required for the system to reach equilibrium may be extremely large. [e.g. for N = 2,000 $x_{10} = 500$ $\alpha = 1$ $\beta = 2$ $m_e = 666.6$ and over 2,000 migrations are required before the system approaches equilibrium.] In such instances, it may be possible to estimate (by extrapolation if necessary) the point to which the step function is approaching and levelling off (i.e. $\hat{m}_e$ if x_{10} is known, $\overline{w}_e$ if not), although it is not possible to form a reliable estimate of v_e. In this case, the form of the approach to equilibrium may be used for deriving initial estimates, since:

$$\text{if } x_{10} > m_e,\ z_k - m_e = w_k - \overline{w}_e > 0 \ \forall k \ni t_k \varepsilon [0, T_e]$$
$$\overline{w}_e \text{ is negative}$$

$$\text{and } E(z_k - m_e) = (x_{10} - \frac{N\alpha}{\alpha+\beta})\ e^{-(\alpha+\beta)t_k} \qquad \text{(from 3.3.4)}$$
$$= -\overline{w}_e\ e^{-(\alpha+\beta)t_k}$$

$$\text{if } x_{10} < m_e \ \ \overline{w}_e \text{ is positive and } E(m_e - z_k) = \overline{w}_e\ \ e^{-(\alpha+\beta)t_k}$$

Thus, the regression of $\log|z_k - m_e| = \log|w_k - \overline{w}_e|$, on t_k, constrained to pass through $\log|w_e|$ at $t = 0$, is approximately linear with slope $-(\alpha+\beta)$. Thus $(\alpha+\beta)$ can be estimated using standard linear regression. The estimate will be approximately unbiased but very inefficient, due to the high correlations between the z_k and non-homogeneity of variance.

When $\alpha = \beta$, the regression coefficient r is an estimate of -2α, leading to the initial estimates

$$\hat{\alpha} = r/(-2) \qquad \hat{N} = (u+d)/(\hat{\alpha}T) \qquad \hat{x}_{10} = \overline{w}_e - \hat{N}/2.$$

When $\alpha \neq \beta$, x_{10} known, r is an estimate of $-(\alpha+\beta)$, leading to the initial estimates:

$$\hat{\beta} = d/(\sum ig_i) \qquad \hat{\alpha} = |r| - \hat{\beta} \qquad \hat{N} = \hat{m}_e r/\hat{\alpha}$$

When $\alpha \neq \beta$, x_{10} unknown, the following four equations in

α, β, x_{20} and x_{10} may be solved

$$\alpha = u/(x_{20}T - \sum i\tilde{g}_i)$$

$$\beta = d/(\sum i\tilde{g}_i - x_{10}T)$$

$$(\alpha+\beta) = -r$$

$$(x_{20}\alpha - x_{10}\beta)/(\alpha+\beta) = \overline{w}_e$$

for initial estimates for maximizing the log likelihood (6.3.6).

6.4 The Quadratic Model

$$q_{i\ i+1} = (N-i)^2\alpha$$

$$q_{i\ i-1} = i^2\beta$$

and the likelihood of the observations $\{z_k, r_k | k = 1 \ldots n(T)\}$ is:

$$\log L = \sum_{i=0}^{N} n_{i,i+1} \log(N-i)^2\alpha - g_i(N-i)^2\alpha$$

$$+ \sum n_{i,i-1} \log i^2\beta - g_i i^2\beta$$

Differentiating and setting equal to zero gives:

$$\frac{\partial \log L}{\partial \alpha}: \frac{\sum n_{i,i+1}}{\alpha} - \sum g_i (N-i)^2 = 0$$

$$\frac{\partial \log L}{\partial \beta}: \frac{\sum n_{i,i-1}}{\beta} - \sum g_i i^2 = 0$$

$$\therefore \qquad \hat{\alpha} = \frac{u}{\sum (N-i)^2 g_i}$$

$$\hat{\beta} = \frac{d}{\sum i^2 g_i}$$

as. $\mathrm{var}(\hat{\alpha}) = \alpha^2/E(u)$

as. $\mathrm{var}(\hat{\beta}) = \beta^2/E(d)$

as. $\mathrm{cov}(\hat{\alpha}\hat{\beta}) = 0$

As with the linear model, the asymptotic variances cannot be calculated in any real situation and must instead be estimated by substituting $\hat{\alpha}$ and $\hat{\beta}$ for α and β, and using approximations for E(u) and E(d). Now we have expressions for X, the expected number in A, at time T (equation 3.4.16), and for Y the total expected number of migrations (in both directions) in the interval [0,T) (equation 3.4.19). Then, as noted at the beginning of section 3.4, E(d) and E(u) can be calculated, since

$$E(d) = (Y-X+x_{10})/2$$
$$E(u) = (Y+X-x_{10})/2$$

and so the estimate $\hat{E}(d)$ and $\hat{E}(u)$ can be derived by substituting $\hat{\alpha}$, and $\hat{\beta}$ into these expressions. This then leads to the estimates for the asymptotic variances

$$\hat{\mathrm{var}}(\hat{\alpha}) = \hat{\alpha}^2/\hat{E}(u)$$
$$\hat{\mathrm{var}}(\hat{\beta}) = \hat{\beta}^2/\hat{E}(d).$$

As with the linear model, an investigation of the distribution of the estimates $\hat{\alpha}$, $\hat{\beta}$, and the variance estimates $\hat{\mathrm{var}}(\hat{\alpha})$, $\hat{\mathrm{var}}(\hat{\beta})$, was carried out by means of simulation. The results of these simulations are given in Table 6.2. The form of the table is exactly as described for Table 6.1.

1. Discussion of the Results in Table 6.2

It has already been remarked (section 3.2.6) that, when migration is density dependent, there is a pronounced tendency for the first several changes of state to be in one direction.

This is increasingly true with larger $|x_{10}-\overline{X}(\infty)|$ (where $\overline{X}(\infty)$ is the equilibrium mean). For example, with $N = 100$ $\alpha = 1$ $\beta = 2$ $x_{10} = 0$, $\overline{X}(\infty) = 41.4$ and it was found (by repeated simulation) that in over 50% of realizations of the process, the first twenty migrations were <u>all</u> from A_2 to A_1.

There is thus a considerable danger, if estimates are formed after observing just a few transitions, that one may assume that migration is unidirectional. It was found also, that if u (or d) was very small (<5), the estimate of α (or β) was extremely untrustworthy, usually being very much larger than the true value. With small populations ($N < 50$), to avoid these two problems, observations must be continued until the process has virtually approached equilibrium. This is the case with the last three sets of simulations in table 6.2 ($E(X(t))$ is within one s.d. of $\overline{X}(\infty)$ in all three cases). In this case, the estimates of α and β are quite good even though u is sometimes quite small.

In general there is no appreciable bias in the estimates except when $\beta >> \alpha$. As with the linear model, it is the larger of the two estimates which shows the most bias and skewness. The asymptotic normal properties of the estimates are more nearly fulfilled with increasing u, d, and for u, $d > 15$ the asymptotic normal theory may safely be applied.

Table 6.2

Simulation of the Quadratic (Density Dependent) Model with M.L.E. for α and β from Step Function Observations (for explanation, see description of table 6.1 in text)

$N = 100 \quad x_{10} = 25 \quad t = .016 \quad \alpha = 1 \quad \beta = 2$

$E(u) = 62.33 \quad E(d) = 46.10 \quad \text{as. var}(\hat{\alpha}) = .0160 \quad \text{as. var}(\hat{\beta}) = .0868$

$\hat{\alpha}$: $m = 1.0068 \quad v = .0154 \quad s = .11 \qquad \widehat{\text{var}}(\hat{\alpha})$: $m = .0163 \quad \sqrt{v} = .0028$

$\hat{\beta}$: $m = 2.0147 \quad v = .0880 \quad s = .25 \qquad \widehat{\text{var}}(\hat{\beta})$: $m = .0888 \quad \sqrt{v} = .0204$

$\hat{p}_A = .9433 \quad \text{S.E.} = .0134 \quad \hat{p}_B = .9433 \quad \text{S.E.} = .0134$

$N = 100 \quad x_{10} = 25 \quad t = .0025 \quad \alpha = 1 \quad \beta = 2$

$E(u) = 12.45 \quad E(d) = 4.35 \quad \text{as. var}(\hat{\alpha}) = .0803 \quad \text{as. var}(\hat{\beta}) = .9153$

$\hat{\alpha}$: $m = .9964 \quad v = .0804 \quad s = .33 \qquad \widehat{\text{var}}(\hat{\alpha})$: $m = .0809 \quad \sqrt{v} = .0266$

$\hat{\beta}$: $m = 1.9233 \quad v = .8653 \quad s = .58 \qquad \widehat{\text{var}}(\hat{\beta})$: $m = .9074 \quad \sqrt{v} = .5039$

$\hat{p}_A = .9333 \quad \text{S.E.} = .0144 \quad \hat{p}_B = .8800 \quad \text{S.E.} = .0188$

$N = 20 \quad x_{10} = 20 \quad t = .05 \quad \alpha = 1 \quad \beta = 10$

$E(u) = 9.62 \quad E(d) = 24.80 \quad \text{as. var}(\hat{\alpha}) = .1040 \quad \text{as. var}(\hat{\beta}) = 4.0325$

$\hat{\alpha}$: $m = 1.0319 \quad v = .1124 \quad s = .68 \qquad \widehat{\text{var}}(\hat{\alpha})$: $m = .1095 \quad \sqrt{v} = .0424$

$\hat{\beta}$: $m = 10.3460^{*} \quad v = 4.7831 \quad s = .49 \qquad \widehat{\text{var}}(\hat{\beta})$: $m = 4.4169 \quad \sqrt{v} = 1.7413$

$\hat{p}_A = .9233 \quad \text{S.E.} = .0154 \quad \hat{p}_B = .9333 \quad \text{S.E.} = .0144$

$N = 20 \quad x_{10} = 20 \quad t = .05 \quad \alpha = 1 \quad \beta = 2$

$E(u) = 3.76 \quad E(d) = 14.89 \quad \text{as. var}(\hat{\alpha}) = .2662 \quad \text{as. var}(\hat{\beta}) = .2686$

$\hat{\alpha}$: $m = 1.0089 \quad v = .3159 \quad s = .94 \qquad \widehat{\text{var}}(\hat{\alpha})$: $m = .2992 \quad \sqrt{v} = .2526$

$\hat{\beta}$: $m = 2.0398 \quad v = .2833 \quad s = .48 \qquad \widehat{\text{var}}(\hat{\beta})$: $m = .2864 \quad \sqrt{v} = .1216$

$\hat{p}_A = .8900 \quad \text{S.E.} = .0181 \quad \hat{p}_B = .9467 \quad \text{S.E.} = .0130$

see continuation on next page

Table 6.2 continued

Simulation of the Quadratic (Density Dependent) Model with M.L.E. for α and β from Step Function Observations (for explanation, see description of table 6.1 in text)

$N = 20$ $x_{10} = 20$ $t = .10$ $\alpha = 1$ $\beta = 2$

$E(u) = 10.40$ $E(d) = 22.08$ as. $\mathrm{var}(\hat{\alpha}) = .0962$ as. $\mathrm{var}(\hat{\beta}) = .1812$

$\hat{\alpha}$: $m = 1.0303$ $v = .0907$ $s = .34$ $\widehat{\mathrm{var}}(\hat{\alpha})$: $m = .1042$ $\sqrt{v} = .0443$

$\hat{\beta}$: $m = 2.0533$ $v = .2212$ $s = .58$ $\widehat{\mathrm{var}}(\hat{\beta})$: $m = .1944$ $\sqrt{v} = .0734$

$\hat{p}_A = .9500$ S.E. $= .0126$ $\hat{p}_B = .9333$ S.E. $= .0144$

2. Estimation for the Quadratic Model when N, x_{10} are Unknown

As in subsection 2 above (for the linear model), the state space for the observations is defined in terms of the unknown parameters. We will again give a non rigorous treatment of the likelihood, assuming that $|X_1(t)-x_{10}|$ is small compared with x_{10} and x_{20} so that we can reasonably ignore the information about the range of the observations. Using the notation of section 6.3.3, the log likelihood is:

$$\begin{aligned}\log L = & \sum_{i=-x_{10}}^{x_{20}-1} \tilde{n}_{i,i+1} \log(x_{20}-i)^2\alpha \\ & + \sum_{i=-x_{10}+1}^{x_{20}} \tilde{n}_{i,i-1} \log(x_{10}+i)^2\beta \\ & - \sum_{i=-x_{10}}^{x_{20}} \tilde{g}_i[(x_{20}-i)^2\alpha+(x_{10}+i)^2\beta] \qquad (6.4.1)\end{aligned}$$

setting $y = x_{20}\alpha \quad z = x_{10}\beta$

$$\begin{aligned}\log L = & \sum\tilde{n}_{i,i+1} \log(y-i\alpha)^2/\alpha - \tilde{g}_i(y-i\alpha)^2/\alpha \\ & + \sum\tilde{n}_{i,i-1} \log(z+i\beta)^2/\beta - \tilde{g}_i(z+i\beta)^2 \qquad (6.4.2)\end{aligned}$$

The first derivatives of 6.4.1 do not lead to easily solved equations, so that maximization must be carried out iteratively. The problems of identifiability encountered with the linear model are not as severe with data from the quadratic model, but estimates are very poor unless u and d are greater than 100-150. For small or moderate sized populations (N < 200) this virtually implies that the system must be observed until equilibrium is achieved, or longer. Even with small samples, the likelihood surface appears to be unimodal with a maximum occurring within the positive quadrant. This latter property was not always found to be the case with the linear model.

The variance-covariance matrix of $\hat{\alpha}$, $\hat{y}$, and of $\hat{z}$, $\hat{\beta}$, may be estimated by inverting the matrix of negative second differentials, evaluated at the m.l.e. These differentials are:

$$-\frac{\partial^2 \log L}{\partial\alpha^2} = 2\sum\frac{i^2 n_{i,i+1}}{(y-i\alpha)^2} + \frac{2Ty^2}{\alpha^3} - \frac{u}{\alpha^2}$$

$$-\frac{\partial^2 \log L}{\partial\alpha\partial y} = -2\left(\sum\frac{i n_{i,i+1}}{(y-i\alpha)^2} + \frac{Ty}{\alpha^2}\right)$$

$$-\frac{\partial^2 \log L}{\partial y^2} = 2\left(\sum\frac{n_{i,i+1}}{(y-i\alpha)^2} + \frac{T}{\alpha}\right)$$

and

$$-\frac{\partial^2 \log L}{\partial\beta^2} = 2\sum\frac{i^2 n_{i,i-1}}{(z+i\beta)^2} + \frac{2Tz^2}{\beta^3} - \frac{d}{\beta^2}$$

$$-\frac{\partial^2 \log L}{\partial\beta\partial z} = -2\left(\sum\frac{i n_{i,i-1}}{(z+i\beta)^2} + \frac{Tz}{\beta^2}\right)$$

$$-\frac{\partial^2 \log L}{\partial z^2} = 2(\sum\frac{n_{i,i-1}}{(z+i\beta)^2} + \frac{T}{\beta}$$

Once the variances and covariances for y, α and z, β have been calculated, approximate variances for x_{10}, and x_{20} can be derived from the asymptotic relations

$$\frac{\text{var } x_{20}}{x_{20}} \doteqdot \frac{\text{var } y}{y^2} + \frac{\text{var } \alpha}{\alpha^2} - \frac{\text{cov } (\alpha y)}{\alpha y}$$

and

$$\frac{\text{var } x_{10}}{x_{10}} \doteqdot \frac{\text{var } z}{z^2} + \frac{\text{var } \beta}{\beta^2} - \frac{\text{cov } (\beta z)}{\beta z}$$

Two typical examples of estimation for the quadratic model when x_{10}, x_{20} are unknown will illustrate the adequacy (or inadequacy) of the estimates. Table 6.3 shows data from two realizations of the model: $N = 100 \quad x_{10} = 25 \quad \alpha = 1 \quad \beta = 2$. For the first set of data, $T = .016$ which gave observations $u = 62 \quad d = 45$. For this type of estimation, this is a small sample. The second set of data in table 6.3 is for a moderate sized sample: $T = .045 \quad u = 164 \quad d = 142$. Properties of the estimates $\hat{y}$, $\hat{\alpha}$ (from maximisation of the first term of 6.4.2) and x_{10}, $\hat{\alpha}$ (using 6.4.1) will be illustrated. The results for z, β were not calculated, but could be expected to be similar.

For the small sample, the maximum of (6.4.2) occurred at $\hat{\alpha} = 8.2 \quad \hat{y} = 287.1 \quad (\hat{x}_{20} = 35)$. The variance-covariance matrix for $\hat{\alpha}$, $\hat{y}$ was estimated as:

$$\begin{bmatrix} .0232 & .5649 \\ .5649 & 142.926 \end{bmatrix}$$

Clearly, the estimates are quite inadequate and the variance estimates are misleading in their precision. In fig. 6.1 we give a plot of the contours of constant relative likelihood for both 6.4.1 ($L(x_{20},\alpha)$) and 6.4.2 ($L(y,\alpha)$). The relative likelihood is the value of the likelihood, for some set of parameter values expressed as a fraction of the likelihood evaluated at its maximum. Thus $R(x_{20},\alpha) = L(x_{20},\alpha)/L(\hat{x}_{20},\hat{\alpha})$ for 6.4.1, and $R(y,\alpha) = L(y,\alpha)/L(\hat{y},\hat{\alpha})$ for 6.4.2. Thus all values of the parameters within the contour $R = .1$ are parameter sets which lead to a sampling probability for the observed data that is at least one tenth of the most likely sampling probability. Thus $R(\theta)$ is a measure of relative plausibility of sets of parameter values θ. Interpretation of likelihoods in this way has been discussed by Sprott and Kalbfleish (1965, 1969) and Barnard, Jenkins and Winston (1962). $R(\theta)$ may be shown to have, asymptotically, the shape of a (multivariate) normal distribution with mean $\hat{\theta}$ and variance-covariance matrix $V(\hat{\theta})$, where $V(\hat{\theta})$ is the asymptotic variance-covariance matrix of the m.l. estimates. (Sprott and Kalbfleish 1969). While a normal shaped relative likelihood does not necessarily mean that the asymptotic normal properties of the m.l. estimates themselves hold the converse (a relative likelihood which does not have the normal shape) may be taken as an indication that, either the asymptotic normal properties of the estimates do not hold, or one has observed a pathological (outlying) set of data. In either case, one would not wish to apply the asymptotic normal theory for forming confidence regions.

It can be seen from fig. 6.1 that $R(x_{20},\alpha)$ is far from normal in shape, the contours having a distinct crescent shape. However, the shape is perceptibly improved by the transformation to the variates y,α, whose contours are much more nearly elliptical. This "improvement" was found to be quite generally the case for several other data sets, and was not a property peculiar to this one set of data. However, $R(y,\alpha)$ shows considerable skewness, the contours being closer together towards the axes $y = 0$ $\alpha = 0$. It may be noted that the true parameter values ($\alpha = 1$ $y = x_{20} = 75$)

fall within the contour $R = 1/10$. Thus, even for this small sample, the true values turn out to be relatively plausible, whereas they would not be deemed reasonable values if the asymptotic normal properties of $\hat{\alpha}$ and $\hat{y}$ were used for deriving confidence intervals for α, y.

For the larger sample, the maximum of (6.4.2) occurred at $\hat{\alpha} = .88$ $\hat{y} = 72.2$ $(\hat{x}_{20} = 82)$.

Despite this considerable improvement in the point estimates with the larger sample, the likelihood surface did not display the asymptotic properties of the relative likelihood. Fig. 6.2 shows the contours of constant likelihood for $L(x_{20},\alpha)$ (equation 6.4.1) and for $L(y,\alpha)$ (6.4.2). The characteristic crescent shape of $R(x_{20},\alpha)$ is still prominent, but again, $R(y,\alpha)$ has more elliptical contours, although there is considerable skewness. The range of plausible values for α is much smaller than the range of plausible values indicated by the relative likelihood from the small sample. ($0 < \alpha \leqslant 5$ as opposed to $0 < \alpha \leqslant 40$, approximately). However, the range of plausible values for x_{20} is larger ($30 < x_{20} < 250$ as opposed to $20 < x_{20} < 150$ approximately). Moreover, the matrix of negative second differentials of the log likelihood was ill conditioned and gave negative estimates for the variances of the estimates. The true parameter values ($\alpha = 1.0$, $y = x_{20} = 75$) were within the contour $R = .6$

Table 6.3

Data from Single Realizations of the

Quadratic Model with $N = 100$ $x_{10} = 25$ $\alpha = 1$ $\beta = 2$

(N and x_{10} assumed unknown)

Small sample: $T = .016$

$u = 62$ $\sum i\tilde{g}_i = .2192$

$d = 45$ $\sum i^2\tilde{g}_i = 3.3356$

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$\tilde{n}_{i,i+1}$	1	2	2	2	1	1	1	1	1	1	7	4	8	3	4	6	7	6	4	0
$n_{i,i-1}$	0	0	1	1	1	0	0	0	0	0	0	6	3	7	2	3	5	6	6	4

Large sample: $T = .045$

$u = 164$ $\sum i\tilde{g}_i = .8028$

$d = 142$ $\sum i^2\tilde{g}_i = 15.3256$

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
$n_{i,i+1}$	1	1	2	1	1	1	1	1	1	1	1	3	8	7	9	17	13	11	14	16	16	9	6	8	7	5	2	1	0
$n_{i,i-1}$	0	0	0	1	0	0	0	0	0	0	0	0	2	7	6	8	16	12	10	13	15	15	8	6	8	7	5	2	1

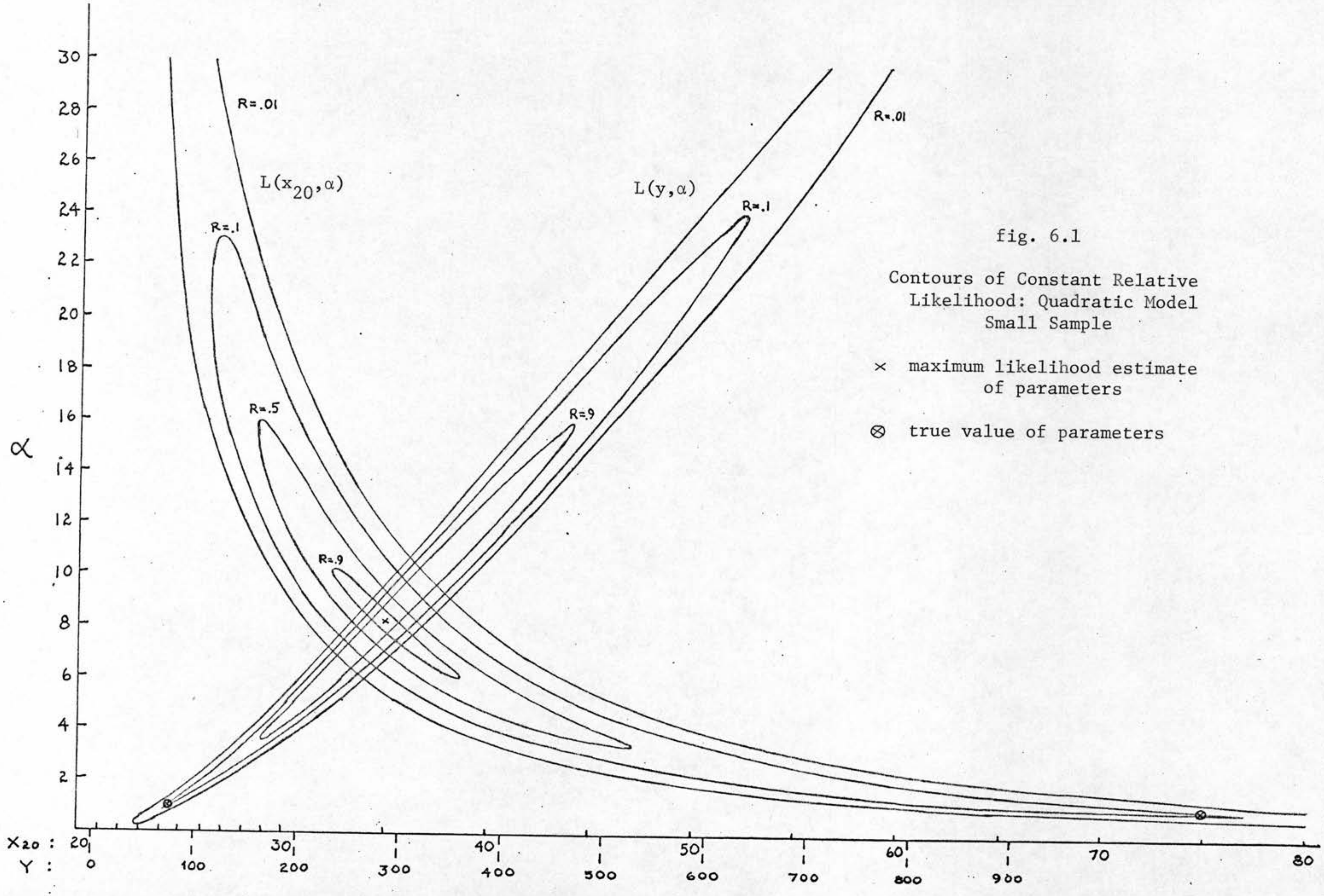


fig. 6.1

Contours of Constant Relative Likelihood: Quadratic Model Small Sample

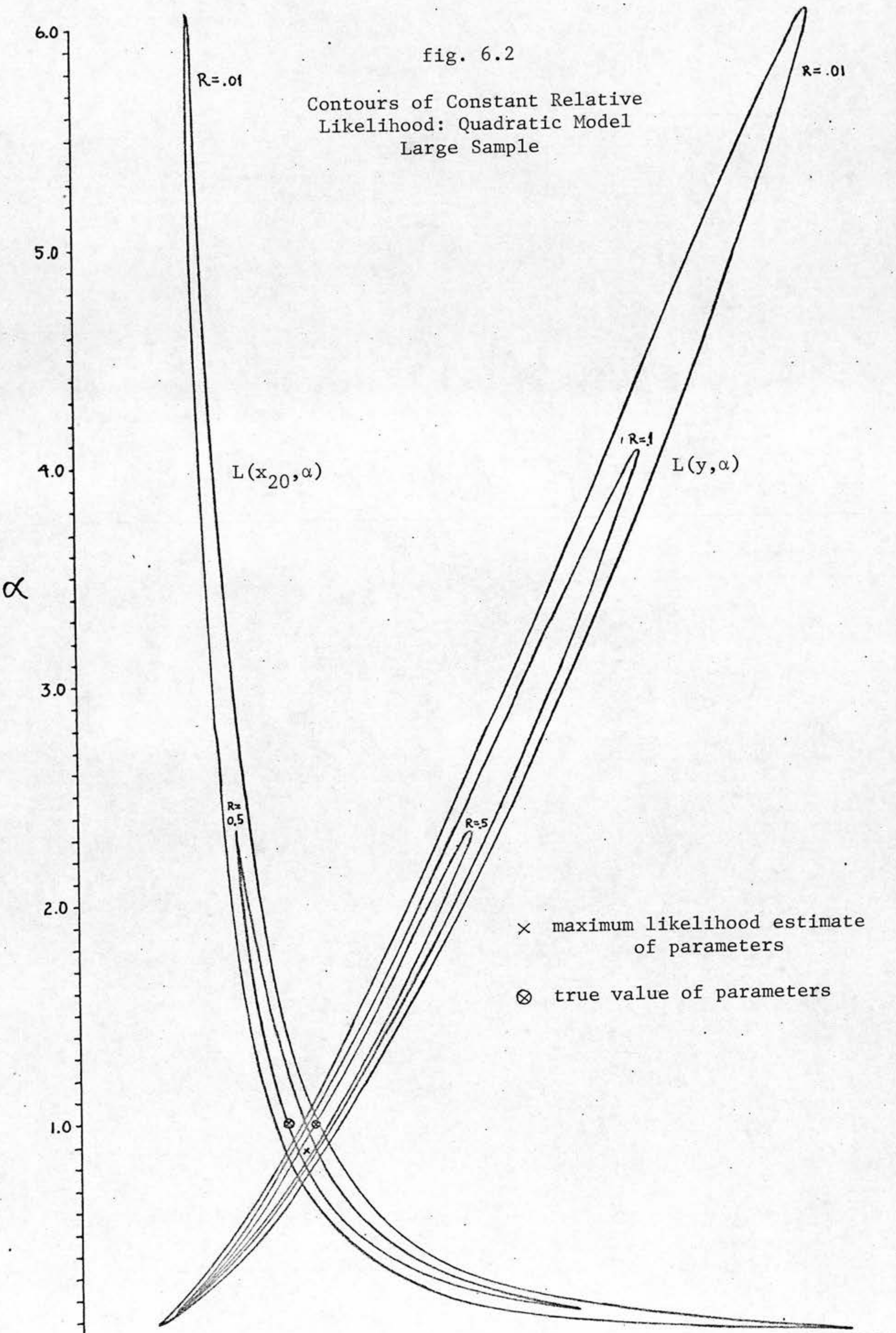


fig. 6.2

Contours of Constant Relative Likelihood: Quadratic Model Large Sample

6.5 Models Not Homogeneous in Time (Time Dependent Migration)

The methods of Billingsley are developed for time homogeneous markov processes, but there is no difficulty in adapting them to handle markov processes whose transition rates are time dependent. We now require the observations

$$\{(z_k, r_k, t_{k+1}) | k = 1, \ldots n(T)\}$$

where $t_{k+1} = \sum_{i=0}^{k} r_i$ is the time at which the kth change of state occurs. Now the functions defined in section (5.2) will be explicit functions of time, so we now define

$$p_\theta(t, \delta t, \xi, A) = \text{Prob}\{X(t+\delta t) \varepsilon A | X(t) = \xi\}$$

and definitions of $q(\xi,\theta,t)$ and $q(\xi,A,\theta,t)$ follow from (5.2.3) and (5.2.4) using $p_\theta(t,\delta t,\xi,A)$ instead of $p_\theta(\delta t,\xi,A)$. The theorems of Doob, for the time dependent process (equations 5.2.5 and 5.2.6), become:

$$\text{Prob}\{r_k > a | r_1 \ldots r_{k-1}, z_1 \ldots z_k, t_1 \ldots t_k\}$$

$$= \text{Prob}\{r_k > a | z_k, t_k\}$$

$$= \exp[-\int_{t_k}^{t_k+a} q(\xi,\theta,t)dt$$

and

$$\text{Prob}\{z_{k+1} \varepsilon A | r_1 \ldots r_k, z_1 \ldots z_k, t_k \ldots t_{k+1}\}$$

$$= \text{Prob}\{z_{k+1} \varepsilon A | z_k, t_{k+1}\}$$

$$= q(z_k, A, \theta, t_{k+1}) / q(z_k, \theta, t_{k+1})$$

The development of the likelihood, and the conditions required of the functions q follow in an exactly analogous fashion to that described in section 5.2. The only additional condition required is that $q(\xi,\theta,t)$ be integrable over all $t\varepsilon[0,T]$ and that the other conditions on the functions q hold for all $t\varepsilon[0,T]$.

If the state space is finite, the log likelihood, analogous to (5.2.10), ignoring $r_{n(T)+1}$, is:

$$\log L = \sum_{k=1}^{n(T)} \{\log q(z_k,z_{k+1},\theta,t_{k+1}) - \int_{t_k}^{t_{k+1}} q(z_k,\theta,t)dt\} \tag{6.5.1}$$

Because of the integral in this expression, it is impossible, in general, to group terms according to the state occupied (value of z_k), as was done in section 5.2. In most cases, this makes the likelihood very difficult to deal with analytically. The first term can be grouped conveniently in terms of the states $i = 0,1 \dots N$, using the following notation:

u_{ij} time at which the jth change of state from i to i+1 occurs. $j = 1 \dots n_{i,i+1}$

d_{ij} time at which the jth change of state from i to i-1 occurs. $j = 1 \dots n_{i,i-1}$

$q_{i,i+1}(\theta,t) = q(z_k,z_{k+1},\theta,t_{k+1})$ when $z_k = i$, $z_{k+1} = i+1$, $t_{k+1} = t$.

$q_{i,i-1}(\theta,t) = q(z_k,z_{k+1},\theta,t_{k+1})$ when $z_k = i$, $z_{k+1} = i-1$, $t_{k+1} = t$.

Then (6.5.1) becomes

$$\log L = \sum_{i=0}^{N} \{ \sum_{j=1}^{n_{i,i+1}} \log q_{i,i+1}(\theta,u_{ij}) + \sum_{j=1}^{n_{i,i-1}} \log q_{i,i-1}(\theta,d_{ij})\} - \sum_{k=1}^{n(T)} \int_{t_k}^{t_{k+1}} q(z_k,\theta,t)dt \tag{6.5.2}$$

We will now apply these methods to some of the models mentioned in section 3.5.

1. Model of section 3.5.1

For the first model in section 3.5, it was assumed that the migration intensity functions were:

$$\alpha(x_2,t) = \alpha\rho(t)$$
$$\beta(x_1,t) = \beta\rho(t)$$

where $\rho(t)$ is any positive, integrable function in t. We then have, using equations 6.2.1, 6.2.2, 6.2.3 respectively

$$\begin{aligned} q_{i,i+1}(\theta,t) &= (N-i)\alpha\rho(t) \\ q_{i,i-1}(\theta,t) &= i\beta\rho(t) \\ q(z_k,\theta,t) &= [(N-z_k)\alpha+z_k\beta]\rho(t) \\ \theta &= \{\alpha,\beta,\{\rho\}\} \text{ where } \{\rho\} \text{ is the set of parameters necessary for defining } \rho(t). \end{aligned}$$

and the log likelihood, (6.5.2), after simplification, is:

$$\log L = u\log\alpha + d\log\beta + \sum_{i=0}^{N}\{\sum_{j=1}^{n_{i,i+1}}\log\rho(u_{ij}) + \sum_{j=1}^{n_{i,i-1}}\log\rho(d_{ij})$$

$$-N\alpha\int_0^T\rho(t)dt - (\beta-\alpha)\sum_{k=1}^{n(T)} z_k\int_{t_k}^{t_{k+1}}\rho(t)dt.$$

This cannot be dealt with further unless an explicit functional form for $\rho(t)$ is postulated. However, even for the simplest functional forms, the likelihood proves analytically intractable for deriving estimates. Moreover, in any real situation, one is probably unaware of the functional form of $\rho(t)$ and interest centres on a method of estimation that would reveal the form of this function. Altogether, then, it seems that step function estimates are not very useful in this situation. More appropriate estimation procedures to deal with this model are developed in the next chapter.

2. Model of section 3.5.2

$$\alpha(x_1,t) = \alpha\rho(t)$$
$$\beta(x_2,t) = \alpha(1-\rho(t))$$

$$\therefore \quad q_{i,i+1}(\theta,t) = (N-i)\alpha\rho(t)$$
$$q_{i,i-1}(\theta,t) = i\alpha(1-\rho(t))$$

The log likelihood is, after simplification:

$$\log L = (u+d)\log\alpha + \sum_{i=0}^{N}\{\sum_{j=1}^{n_{i,i+1}}\log\rho(u_{ij}) + \sum_{j=1}^{n_{i,i-1}}\log(1-\rho(d_{ij}))\}$$

$$-N\alpha\int_0^T\rho(t)dt - \alpha\sum_{i=0}^{N} ig_i + 2\alpha\sum_k z_k\int_{t_k}^{t_{k+1}}\rho(t)dt$$

For the system considered in section 3.5 and 4.3 $\rho(t)$ was taken to be e^{-ct}. If c is known, the likelihood is easily maximized analytically, but if c is also unknown, the likelihood could be maximized using standard iterative procedures in any numerical instance.

6.6 Multi-dimensional models

The methods of deriving estimates from step function observations can easily be extended for use with multi-dimensional models. For the non-closed two dimensional models of section 4.2, this would require that observations be made of the times of occurrence of each birth, death and immigration, as well as the times of migrations. This is not a very realistic requirement. However, certain types of transitions may be observable (say, migrations) while the others are not. The m.l.e. are thus of some interest in determining how such information might be used.

The methods of section 5.2 were thus extended to the general two dimensional system of section 4.2. The development is routine, although it involves a more cumbersome notation, and will not be given here. For the particular case with migration and death only ($\mu_1 = \mu_2 = \mu$) the m.l.e. are:

$$\hat{\alpha} = u/\sum_i \sum_j j g_{ij}$$

$$\hat{\beta} = d/\sum_i \sum_j i g_{ij}$$

where g_{ij} is the time the system spends in the state $X_1(t) = i$, $x_2(t) = j$. Thus

$$\hat{\alpha} = u/\int_0^T X_2(t)dt$$

$$\hat{\beta} = d/\int_0^T X_1(t)dt$$

as in section 6.3. The m.l.e. of μ is

$$\hat{\mu} = D/\sum_i \sum_j (i+j)g_{ij}$$

where D is the total number of deaths occurring in [0,T]. Thus if complete counts are made to supplement the observations u and d, say at times $t_0 = 0, t_1, \dots t_n = T$, then a rough estimate of the integral is given by:

$$\int_0^T X_1(t)dt \simeq \sum_{j=1}^{n-1} (t_{j+1}-t_j)X_1(t_j)$$

and similarly for $\int_0^T X_2(t)dt$. D is then known as well since

$$\begin{aligned} D &= X_1(0)+X_2(0)-X_1(T)-X_2(T) \\ &= N\ (0)-N(T) \end{aligned}$$

If n is large, then an improved estimate of μ can be obtained using the fact that N(t) follows the death process (see for example, Bharucha-Reid p. 85). Thus $N(t_j)$ has the distribution:

$$B\{N(t_j);\ N(0),\ \exp[-\mu t_j])\}$$

and hence the regression of $\log (N(t_j)/N(0))$ on t_j passes through the origin and is linear with slope $-\mu$. The $N(t_j)$ however are highly correlated if the $t_{j+1}-t_j$ are small, and the assumption of homogeneity of variance is violated as well. The regression estimate for μ is thus very inefficient, though approximately unbiased.

However, the distribution of $N(t_j)|N(t_{j-1})$ is independent of $N(t_k)$ $k = 1 \dots j-2$ and has the distribution:

$$B\{N(t_j)|N(t_{j-1});\ N(t_{j-1}),\ \exp[-\mu(t_j-t_{j-1})]\}$$

Thus the likelihood of the observations $N(t_0) \dots N(t_n)$ is:

$$L = \prod_{i=1}^{n} \binom{N(t_{i-1})}{N(t_i)} (\exp[-\mu\Delta t_i])^{N(t_i)} (1-\exp[-\mu\Delta t_i])^{N(t_{i-1})-N(t_i)}$$

$$\text{where } \Delta t_i = (t_i - t_{i-1})$$

$\partial \log L/\partial\mu = 0$ gives the following equation for $\hat{\mu}$:

$$\sum_{i=1}^{n} \{[N(t_{i-1})-N(t_i)] \frac{\Delta t_i \exp[-\mu\Delta t_i]}{(1-\exp[-\mu\Delta t_i])}\} = \sum_{i=1}^{n} \Delta t_i N(t_i)$$

This may be solved iteratively using Newton's Iterative procedure, or the method of Maximum Likelihood Scoring (Bailey 1961, A ppendix 1), using the regression estimate for μ as the initial value. From the expected value of the negative second differential of log L, and using $E(N(t_j)) = N(t_0)\exp[-\mu t_j]$, we find:

$$\text{as. var}(\hat{\mu}) = [N_0 \sum_{i=1}^{n} \frac{(\Delta t_i)^2 \exp[-\mu\Delta t_i]\exp[-\mu t_{i-1}]}{(1-\exp[-\mu\Delta t_i])}]^{-1}$$

If $\Delta t_i = c$, say, $\forall i$, that is, observations are made at equal time intervals, these expressions simplify, and we find

$$\hat{\mu} = -\frac{1}{c}\log \hat{\mathbb{Q}}$$

$$\text{where} \qquad \hat{\mathbb{Q}} = 1-[N(t_0)-N(t_n)]/\sum_{i=0}^{n-1} N(t_i)$$

$\hat{\mathbb{Q}}$ may also be shown to be the maximum likelihood estimate of $\mathbb{Q} = \exp[-c\mu]$.

$$\text{as. var}(\hat{\mu}) = \frac{(1-\exp[-c\mu])^2}{N(t_0)c^2\exp[-c\mu](1-\exp[-cn\mu])}$$

$$\text{as. var}(\hat{\mathbb{Q}}) = \frac{\mathbb{Q}(1-\mathbb{Q})^2}{N(t_0)(1-\mathbb{Q}^n)}$$

CHAPTER SEVEN

ESTIMATION FROM COMPLETE COUNTS AT FIXED TIMES

7.1 Introduction.

In this chapter, we will produce methods of estimation which are based on observations of the total numbers in each area at specified times $t_0, t_1 \ldots t_n$. In effect, this implies knowledge of the net migration out of each area in each interval $[t_{i-1}, t_i)$, since $X(t_i) = X(t_{i-1}) + u(t_{i-1}) - d(t_{i-1})$, where $d(t_{i-1})$ is the total number of migrations out of the area in $[t_{i-1}, t_i)$, and $u(t_{i-1})$ is the number into the area. It is evident that, as equilibrium is approached, such observations will not be very informative with regard to the parameters of the model, as the net migration may frequently be zero, or close to zero, and the distribution of $X(t)$ may be independent of some of the parameters altogether, or may be a function of some function of the parameters, leading to non-identifiability.

We will consider two types of experiments in this chapter. In the first type (complete counts without marking), all individuals are assumed to be indistinguishable, so that at time t_i, the experimenter knows only the current location of the individual, but nothing of its previous history. Thus, as time advances, unless the experimenter deliberately causes disequilibrium in the system(by removing all animals in one area, for instance, or by transferring individuals from one area to another), the system inevitably moves towards equilibrium and the observations become less and less informative. These estimates then, can best be used in situations where the system begins in a state of extreme disequilibrium (say, with all individuals in a single area at t_0), and complete counts are made until the system begins to approach equilibrium.

In the second type of experiment (complete counts with marking), it is assumed that all individuals are distinguishable. Usually, this means that the experimenter must mark or otherwise identify each individual in the population, and keep a record of its location at each t_i. In fact, at each t_i it is only necessary to know the location of each individual at the previous time t_{i-1}, and not its entire capture history. Such experiments will be much more informative than the first type and do not become less informative as time progresses, since at each time t_i, after the counts have been made, there are r (where r is the number of areas) subpopulations, each distinguishable at t_i and t_{i+1}, and each in the state of extreme disequilibrium described above. For the independence models of chapters 3 and 4, these subpopulations act independently and so can effectively be considered as replications of the migration system (with different initial conditions, if necessary). The quadratic (density dependent) model of Chapter 3 cannot be considered in this way, and will not be analysed for this type of experiment.

Often, it will be more convenient to estimate the Q_{ij} (in the notation of section 4.3) rather than the parameters of the markov process. We have noted in section 4.3 that it is always possible, given a numerical set of the Q_{ij} with variances and covariances, to estimate the parameters of the markov process. Where estimates of the Q_{ij} are given, we will thus not be too concerned to give analytic means of deriving estimates for the parameters of the markov process. As we pointed out earlier, one of the difficulties in estimating the Q_{ij} is that, if r and n are large, the number of parameters proliferates. If, however, $t_i - t_{i-1}$ = constant $i = 1 \ldots n$, then the number of Q_{ij} will not increase with n (unless migration is time dependent). Nevertheless, if r is large, it may be more convenient to estimate the parameters of the markov process directly.

7.2 The Linear Model:

7.2.1 Estimates for a closed population using complete counts without marking.

The model is defined by $Q_{11}(t)$ and $Q_{21}(t)$ where

$$Q_{11}(t) = (\alpha+\beta \exp[-(\alpha+\beta)t])/(\alpha+\beta)$$
$$Q_{21}(t) = \alpha(1-\exp[-(\alpha+\beta)t])/(\alpha+\beta) \qquad (7.2.1)$$

The simplest estimates of Q_{11} and Q_{21} are given by the moment estimates using complete counts at $t_1 = t_0+t$ and $t_2 = t_0+2t$. Then if we observe $X_1(t_1) = n_1$ and $X_1(t_2) = n_2$, and if $X_1(t_0) = x_{10}$

$$E(n_1) = x_{10}Q_{11}(t)+(N-x_{10})Q_{21}(t)$$
$$E(n_2|n_1) = n_1Q_{11}(t)+(N-n_1)Q_{21}(t)$$

Removing expectations and solving, gives

$$\hat{Q}_{11} = \frac{(N-x_{10})n_2-(N-n_1)n_1}{N(n_1-x_{10})}$$

$$\hat{Q}_{21} = \frac{n_1^2-x_{10}n_2}{N(n_1-x_{10})}$$

These estimates were found to be extremely poor, in most circumstances giving highly inaccurate estimates, or inadmissible values ($Q < 0$ or $Q > 1$). As moment estimates are widely, and sometimes indiscriminately, used by biologists, it is of some interest to investigate what conditions, if any, lead to acceptable estimates, using the moment equations.

For the estimates to be reasonable, we will also insist that admissible estimates of α and β should be obtained. These moment estimates are:

$$\hat{\beta} = \frac{(\hat{Q}_{11}-1)t}{(1-\hat{Q}_{11}+\hat{Q}_{21})} \log(\hat{Q}_{11}-\hat{Q}_{21})$$

$$\hat{\alpha} = \hat{Q}_{21}\hat{\beta}/(1-\hat{Q}_{11})$$

Now, $$\hat{Q}_{11}-\hat{Q}_{21} = (n_2-n_1)/(n_1-x_{10})$$

and this quantity must be positive if estimates of α and β are to be obtained at all. This will be the case iff x_{10}, n_1 and n_2 are in a strictly monotone sequence (either increasing or decreasing). Evidently, there is a high probability of this condition failing if x_{10} or n_1 is close to equilibrium. Moreover, if x_{10} or x_{20} is large, the variances of n_1 and n_2 are also large, and if t is small, so that $E(|n_2-n_1|)$ or $E(|n_1-x_{10}|)$ is of the same order of magnitude as $\text{var}(n_1)$, then, again, there is a high probability that the condition will fail.

Even if this condition does hold, there is no assurrance that the estimates will be usefully precise, or even give admissible values. The asymptotic variance of Q_{11} and Q_{21} can be derived using the standard δ-technique, but the expression is too complicated to be of any use in examining the precision analytically. We will not give the expression here, but only note that, like the estimates themselves, the variance "blows up" if $n_1 = x_1$. An example will make clear the inadequacy of moment estimates even when the condition of monotonicity holds. For $x_{10} = 500$ $x_{20} = 1500$ $\alpha = 1$ $\beta = 2$ $t = .05$ a single simulation gave $n_1 = 520$ $n_2 = 561$. This gave estimates $\hat{Q}_{11} = 1.8$ $\hat{Q}_{21} = -.25$ (the true values are $Q_{11} = .9071$, $Q_{21} = .0464$), and $\hat{\beta} = -10.9$ $\hat{\alpha} = -3.4$. The asymptotic variances (calculated using the true values for Q_{11} and Q_{21}) were:

$$\text{as. var } \hat{Q}_1 = (.475)^2$$
$$\text{as. var } \hat{Q}_2 = (.15)^2$$

It is evident that the best conditions for using moment estimates are when x_{10} is 0 (or N) and n_2 is observed just as the system begins to approach equilibrium. This will, in fact, produce a good estimate for Q_{21} (or Q_{11} if $x_{10} = N$), but the other estimate will still be quite imprecise. For the case $x_{10} = 0$, we then have:

$$\hat{Q}_{11} = [Nn_2-(N-n_1)n_1]/Nn_1$$
$$\hat{Q}_{21} = n_1/N$$
$$\therefore \quad \text{var } \hat{Q}_{21} = \frac{1}{N^2} \text{var}(n_1) = \frac{1}{N} Q_{21}(1-Q_{21})$$
$$(Nn_1)^2 \text{ var } \hat{Q}_{11} \doteqdot (2n_1-N-NQ_{11})^2 \text{ var } n_1 + N^2 \text{ var } n_2$$
$$+2N(2n_1-N-NQ_{11}) \text{ cov}(n_1 n_2)$$

but from section 3.3, we have var $n_1 = \text{cov}(n_1 n_2)$; hence

$$\text{var } \hat{Q}_{11} \doteqdot \left[\frac{(2n_1-NQ_{11})^2-N^2}{N^2 n_1^2}\right] \text{var } n_1 + \frac{1}{n_1^2} \text{ var } n_2$$

An example where the conditions for a good moment estimate should be close to optimal, is the model $N = 250$ $x_{10} = 0$ $\alpha = 1$ $\beta = 1$ and $t = .5$. Here, $E(n_2) = 108.1$, so the system is close to equilibrium by the time the second observation is made. $\mathbb{Q}_{11} = .684$ $\mathbb{Q}_{21} = .316$ and the estimates, from observations $n_1 = 76$ $n_2 = 118$ are:

$$\begin{aligned} \hat{\mathbb{Q}}_{11} &= .857 \\ \operatorname{var}(\hat{\mathbb{Q}}_{11}) &\doteqdot (.10)^2 \\ \hat{\mathbb{Q}}_{21} &= .304 \\ \operatorname{var}(\hat{\mathbb{Q}}_{21}) &= (.05)^2 \\ \hat{\beta} &= .095 \\ \hat{\alpha} &= .202 \end{aligned}$$

Even in this situation, only the estimate for $\mathbb{Q}_{21}$ was adequate. If the experimenter knows that $\alpha = \beta$, then $\mathbb{Q}_{11} = 1-\mathbb{Q}_{21}$ and only one observation is necessary to form the moment estimate. The estimate is then

$$\begin{aligned} \hat{\mathbb{Q}}_{11} &= (N-n_1-x_{10})/(N-2x_{10}) \\ \operatorname{var} \hat{\mathbb{Q}}_{11} &= N\mathbb{Q}_{11}(1-\mathbb{Q}_{11})/(N-2x_{10})^2 \end{aligned}$$

This estimate is unbiased and will be quite precise if N is large and observations begin when the system is not close to equilibrium. This appears to be the only situation in which the moment estimates can be used with reasonable assurance.

We may hope to improve the precision of estimates for $\mathbb{Q}_{11}$, $\mathbb{Q}_{21}$ by taking multiple observations, say at times $t_0, t_1 \ldots t_n$. The likelihood of the observation $\{X_1(t_r); r = 1 \ldots n\}$ is then

$$L = \prod_{r=1}^{n} \Pr(X(t_r) | X(t_{r-1}))$$

where $\Pr(X(t_r)|X(t_{r-1}))$ is the convolution of the two binomial distributions (as shown in section 3.3);

$$B(x;N-X(t_{r-1}),\ \mathbb{Q}_{21}(\Delta t_r))$$
$$B(x;X(t_{r-1}),\ \mathbb{Q}_{11}(\Delta t_r))$$

where $\Delta t_r = t_r - t_{r-1}$ and $\mathbb{Q}_{11}(t)$, $\mathbb{Q}_{21}(t)$ are as in equation 7.2.1. These probabilities cannot be expressed in closed form, and consequently, the likelihood proved to be entirely intractible for estimating α and β. Even if $\Delta t_r = c$ (constant) $\forall\, r = 1 \ldots n$, in which case the likelihood is a function of the parameters $\mathbb{Q}_{11}(c)$ and $\mathbb{Q}_{21}(c)$, it still proved impossible to derive estimates (for $\mathbb{Q}_{11}$ and $\mathbb{Q}_{21}$) from the likelihood. An approximation to the likelihood was derived by using the fact (pointed out in section 3.3) that $X(t_r)$ is approximately normally distributed. $\Pr(X(t_r)|X(t_{r-1}))$ was replaced by the probability of $X(t_r)$ from the normal distribution with mean $E(X(t_r)|X(t_{r-1}))$ (equation 3.3.4) and variance $\operatorname{var}(X(t_r)|X(t_{r-1}))$ (equation 3.3.5). This expression, too, resisted all efforts to derive estimates for α and β, or $\mathbb{Q}_{11}(c)$, $\mathbb{Q}_{21}(c)$.

We may, however, proceed using regression methods. The distribution of $X(t_r)|X(t_{r-1})$ is independent of $X(t_i)$ $i = 1 \ldots r-2$, and we have already noted that $X(t_r)$ is approximately normally distributed. (If Δt_r is very small, the normal approximation will not be very good. However if Δt_r is sufficiently large that $X(t_r) - X(t_{r-1}) > 10$ or so, the approximation is quite good). The only assumption of standard regression theory that is not met is the requirement of homogeneity of variance. The variances of the $X(t_r)|X(t_{r-1})$ differ for different r, and moreover, the variances are a function of the parameters. This can be overcome in two ways: by transforming the observations in some way that stabilizes the variance, or by using weighted least squares. Standard techniques are available for selecting a transformation to stabilize the variance (e.g. see Anscombe, 1948), but these require that the variance be expressible as a function of the mean. As this is not possible for the linear model, or the other models considered in this chapter, we will not use this technique.

The weighted least squares technique for estimating a set of parameters $\underline{\theta}$, given $E(X(t_r)) = \mu_r(\theta)$ and $\mathrm{var}(X(t_r)) = v_r(\theta)$, is to minimize the weighted sum of squares

$$\sum_r \{(X(t_r)-\mu_r(\theta))^2/v_r(\theta)\}$$

This may be done iteratively, by first minimizing the unweighted sum of squares:

$$\sum_r \{(X(t_r)-\mu_r(\theta))^2\}$$

to give estimates of θ. First estimates of the variances may then be obtained, and the weighted sum of squares may be minimized to give improved estimates of θ. The process may be continued until successive estimates of θ converge. For most of the models in this chapter $\mu(\theta)$ is a linear function of the parameters. When this is the case, and the variances are known, except for a constant, it is well known that the weighted least squares estimates of the parameters are minimum variance linear unbiased estimates (by the Markov theorem on least squares; the result holds even if the assumption of normality is violated: see David and Neyman (1938)). However, the variances, here, are not known except for a constant, but are functions of the parameters. It would seem reasonable that the above iterative scheme should converge to estimates that are also m.v.l.u., but no general proof of this is given in any of the standard texts on regression or least squares theory. A proof of this would be beyond the scope of this thesis (writer). Even less is known of the properties of estimates obtained by this iterative procedure when $\mu(\theta)$ is not a linear function of the estimates. However, if the properties of normality and independence hold, then the likelihood of the observations is

$$L = \prod_{r=1}^{n} \frac{1}{\sqrt{2\pi v_r(\theta)}} \exp\{-(X(t_r)-\mu_r(\theta))^2/v_r(\theta)\}$$

or

$$\log L = -\tfrac{1}{2}\sum \log v_r(\theta) - \sum\{(X(t_r)-\mu_r(\theta))^2/v_r(\theta)\}$$

Minimizing the weighted sum of squares is thus equivalent to maximizing the log likelihood, but ignoring the first term, in $\log v_r(\theta)$. If this term does not make a large contribution to the likelihood, relative to the second term, then the estimates from weighted least squares will be approximately the m.l.e. and have the asymptotic properties of m.l. estimates. However, the effect of omitting this term from the likelihood for these models was not investigated.

1. Estimates by weighted least squares when $\Delta t_r = c$.

$$\text{let } p_i = X(t_i)/N$$
$$\text{and let } \mathbb{Q}_{11}(c) = \mathbb{Q}_{11}, \quad \mathbb{Q}_{21}(c) = \mathbb{Q}_{21}$$

then, using the results of section 3.3, we have

$$\mu(\theta) = E(p_r|p_{r-1}) = \mathbb{Q}_{21}+(\mathbb{Q}_{11}-\mathbb{Q}_{21})p_{r-1}$$

$$\begin{aligned} v(\theta) = \operatorname{var}(p_r|p_{r-1}) &= \frac{1}{N^2}\operatorname{var}(X(t_r)|X(t_{r-1})) \\ &= \frac{p_{r-1}}{N}\mathbb{Q}_{11}(1-\mathbb{Q}_{11})+\frac{(1-p_{r-1})}{N}\mathbb{Q}_{21}(1-\mathbb{Q}_{21}) \end{aligned}$$

Hence, the regression of p_r on p_{r-1} is linear. The least squares analysis may be carried out iteratively as described above. Each stage of the iteration is a linear weighted regression, and this technique is described in many standard texts (e.g. Draper and Smith (1966) p. 77). Once final estimates are known, final values of the weights may also be estimated. Derivation of the variances and covariances for the regression coefficients for known weights is also a standard technique. Use of the final estimates of the weights in this derivation should give estimates for these variances and covariances which are quite good. Then if the estimate of the slope of the linear regression is b, and of the intercept is a,

$$\begin{aligned} \hat{\mathbb{Q}}_{21} &= a \\ \hat{\mathbb{Q}}_{11} &= b+a \\ \operatorname{var}\mathbb{Q}_{21} &= \operatorname{var} a \\ \operatorname{cov}(\mathbb{Q}_{11}\mathbb{Q}_{21}) &= \operatorname{var} a + \operatorname{cov}(ab) \\ \operatorname{var}\mathbb{Q}_{11} &= \operatorname{var} a + \operatorname{var} b+2\operatorname{cov}(ab) \end{aligned}$$

If $c = \Delta t_r$ is small, $\mathbb{Q}_{21}$, the probability that an individual in A_2 at time t is in A_1 at time t+c, may be very small, and it may be necessary to constrain the regression to have a non-negative estimate for a.

If observations begin when the system is near equilibrium, the slope of the regression will be close to zero. Also $\mathbb{Q}_{11} = \mathbb{Q}_{21} = \alpha/(\alpha+\beta)$ and this quantity may be estimated from the value of the intercept, a, of the regression line.

(ii) When $\alpha = \beta$ and $\Delta t_r = c$

If $\alpha = \beta$, $\mathbb{Q}_{21} = 1-\mathbb{Q}_{11}$ and

$$E(p_r | p_{r-1}) = (1-\mathbb{Q}_{11})+(2\mathbb{Q}_{11}-1)p_{r-1}$$

$$\text{var}(p_r | p_{r-1}) = \frac{1}{N}\mathbb{Q}_{11}(1-\mathbb{Q}_{11}) = \text{constant } \forall t_r$$

Thus the linear regression of $\frac{X(t_r)}{N}$ on $\frac{X(t_{r-1})}{N}$ gives the m.v.l.u. estimate of $1-\mathbb{Q}_{11}$ (the intercept) and hence of $\mathbb{Q}_{11}$.

(iii) When the Δt_r differ

If observations are made at irregular intervals, it is preferable to estimate α and β rather than the set of $\mathbb{Q}_{ij}(\Delta t_r)$. However, the $\mu_i(\theta)$ are no longer linear functions of the parameters $\theta = \{\alpha,\beta\}$ and so each stage of the iterative procedure described above, will itself require an iterative minimization of the weighted sum of squares. Thus a great deal of simplicity is lost when counts are at irregular intervals. The mean and variance of p_i are

$$\mu_i(\theta) = \alpha/(\alpha+\beta)+(p_{i-1}-\alpha/(\alpha+\beta))e_i$$

$$v_i(\theta) = \frac{(1-e_i)}{N(\alpha+\beta)^2}\{[\alpha^2+p_{i-1}(\beta^2-\alpha^2)]e_i+\alpha\beta]$$

where $e_i = \exp[-(\alpha+\beta)\Delta t_i]$.

(iv) When $\alpha = \beta$ and the Δt_r differ

$$\text{Then} \quad E\,\frac{(p_i - 1/2)}{(p_{i-1} - 1/2)} = \exp[-2\alpha\Delta t_i]$$

$$\text{var}\left(\frac{p_i - 1/2}{p_{i-1} - 1/2}\right) = \frac{1 - \exp[-4\alpha\Delta t_i]}{4N(p_{i-1} - 1/2)^2}$$

Thus an initial estimate of α may be formed by taking $\hat{\alpha} = -b/2$ where b is the slope of the linear regression of $\log[(p_i - 1/2)/(p_{i-1} - 1/2)]$ on Δt_i constrained to pass through the origin.

If the Δt_i do not differ greatly, the changes in variance, with increasing i, will be due chiefly to changes in $(p_{i-1} - 1/2)^2$, which approaches 0 with increasing elapsed time t_i. Thus if the p_i have a wide range (i.e. p_1 close to 1 or 0 and p_n close to 1/2), the observations give decreasing information about α as i increases. The weighted least squares procedure outlined above should thus be used to give more weight to the earlier observations.

7.2.2 Estimates for a Closed Population using Complete Counts with Marking.

1. When Δt_r = constant, r = 1 ... n.

Let $n_{i,i+1\ AB}$ be the number of individuals seen in $A(A_1)$ at t_i and in $B(A_2)$ at t_{i+1}. Similar definitions of $n_{i,i+1\ AA}$, etc. are required. Let n_{iA} be the number in A at t_i, n_{iB} the number in B. Then the likelihood of the observations is

$$L = \prod_{i=0}^{n-1} P_r(n_{i,i+1AA}, n_{i,i+1AB}, n_{i,i+1BA}, n_{i,i+1BB} | n_{iA}, n_{iB}) \qquad (7.2.5)$$

but

$$n_{i,i+1AA} + n_{i,i+1AB} = n_{iA}$$

$$n_{i,i+1BB} + n_{i,i+1BA} = n_{iB}$$

$$\therefore\ L = \prod_{i=0}^{n-1} P_r(n_{i,i+1BA}, n_{i,i+1AA} | n_{iA}, n_{iB})$$

$$= \prod_{i=0}^{n-1} P_r(n_{i,i+1BA} | n_{iB}) P_r(n_{i,i+1AA} | n_{iA}) \qquad (7.2.6)$$

From 3.3.6 and 3.3.7 we have $B(n_{i,i+1BA}; n_{iB}, Q_{21})$

and $B(n_{i,i+1AA}; n_{iA}, Q_{11})$

Thus, log L is, omitting terms not including Q_{11}, Q_{21}

$$\log L = \sum_{i=0}^{n-1} [n_{i,i+1BA} \log(Q_{21}) + (n_{iB} - n_{i,i+1BA}) \log(1-Q_{21})$$

$$+ n_{i,i+1AA} \log(Q_{11}) + (n_{iA} - n_{i,i+1AA}) \log(1-Q_{11})] \qquad (7.2.7)$$

The m.l.e. are, therefore:

$$\hat{Q}_{11} = \sum_{i=0}^{n-1} n_{i,i+1AA} / (\sum_{i=0}^{n-1} n_{iA})$$

$$\hat{Q}_{21} = \sum n_{i,i+1BA} / (\sum n_{iB})$$

The negative second differential of the log likelihood gives

$$-\frac{\partial^2 \log L}{\partial Q_{11}^2} = \frac{-\sum n_{i,i+1AA}}{Q_{11}^2} - \frac{\sum (n_{iA} - n_{i,i+1AA})}{(1-Q_{11})^2}$$

Taking expectations, using the conditional expectation

$$E(n_{i,i+1AA} | n_{iA}) = Q_{11} n_{iA}$$

gives

$$-E\frac{\partial^2 \log L}{Q_{11}^2} = E(\sum_{i=0}^{n-1} n_{iA})(\frac{1}{Q_{11}(1-Q_{11})})$$

and since $E(n_{0A}) = x_{10}$

$$E(n_{jA}) = NQ_{21}+(Q_{11}-Q_{21})n_{j-1A}$$

the expectation of $\sum n_{iA}$ may be derived recursively and is:

$$x_{10}(\frac{1-c^n}{1-c})+\frac{NQ_{21}}{(1-c)}[n+ \frac{c^{n-1}}{(1-c)}]$$

where $c = Q_{11}-Q_{21}$.

Since c is less than 1, and may be very close to zero, for n even moderately large, c^n approaches zero, and

$$-E\ \frac{\partial^2 \log L}{\partial Q_{11}^2} = \frac{x_{10}+NnQ_{21}}{Q_{11}(1-Q_{11})(1-Q_{11}+Q_{21})}$$

$$\therefore \text{as. } \operatorname{var}(\hat{Q}_{11}) = \frac{Q_{11}(1-Q_{11})(1-Q_{11}+Q_{21})}{x_{10}+NnQ_{21}}$$

Similarly, since $\sum n_{iB} = Nn-\sum n_{iA}$

$$\text{as. } \operatorname{var}(\hat{Q}_{21}) = \frac{Q_{21}(1-Q_{21})(1-Q_{11}+Q_{21})}{Nn(1-Q_{11})-x_{10}}$$

$$\text{as. } \operatorname{cov}(\hat{Q}_{11}, \hat{Q}_{21}) = 0$$

Note that if Δt_r is very small, Q_{21} and $(1-Q_{11})$ will be close to zero, and the variances of the estimates may be large even though N and n are large.

(ii) When Δt_r differ.

As in 7.2.1, we again estimate α and β, and again, we find that a great deal of simplicity is lost when counts are at irregular intervals. The likelihood cannot be maximized analytically, and maximazation must be carried out by some iterative scheme. The log likelihood is

$$\log L = \sum_{i=0}^{n-1}\{n_{i,i+1BA}\log[\alpha+\beta e(t_i)]+n_{i,i+1AB}\log[\beta+\alpha e(t_i)]$$

$$+n_{i,i+1BB}\log \beta[1-e(t_i)]+n_{i,i+1AA}\log \alpha[1-e(t_i)]\}$$

$$-Nn \log(\alpha+\beta)$$

where $e(t_i) = \exp-[(\alpha+\beta)(t_{i+1}-t_i)]$

If $\alpha = \beta$ this simplifies somewhat to

$$\log L = \sum\{(n_{i,i+1BA}+n_{i,i+1AB})\log(1+e(t_i))$$

$$+(n_{i,i+1BB}+n_{i,i+1AA})\log(1-e(t_i))\}$$

where $e(t_i) = \exp-[2\alpha(t_{i+1}-t_i)]$

7.3 The Linear Model with Death

7.3.1 Estimates using Complete Counts Without Marking

The model of section 4.2 with no birth or immigration and equal death rates has p.g.f.

$$(Q_{11}r_1+Q_{12}r_2+(1-Q_3r_3)^{x_{10}}(Q_{21}r_1+Q_{22}r_2+(1-Q_3)r_3)^{x_{20}}$$

(equation 4.3.3 with $Q_{13} = Q_{23} = 1-Q_3$). The probability of $X_1(t_i)$, $X_2(t_i)|X_1(t_{i-1})$, $X_2(t_{i-1})$ is thus the convolution of two trinomial distributions, and cannot be expressed in closed form. The likelihood of the observations $\{X_1(t_i), X_2(t_i)|i = 0 \dots n\}$, being the product of n such probabilities, proved to be entirely intractible.

However, it is possible to obtain regression estimates by making use of the properties of this model given in section 4.3.2. As in section 7.2, the estimation procedures are very much simpler if counts are made at equal time intervals.

1. When Δt_r = constant, $r = 1 \ldots n$

We may deal entirely with the Q_{ij}, since $Q_{ij}(\Delta t_r) = Q_{ij}$ $r = 1 \ldots n$. The relations between the Q_{ij} and the parameters μ, α, β of the markov process, are given in section 4.3.3. These lead to the following results:

$$Q_3 = \exp[-\mu t] \qquad \text{where } t = \Delta t_r$$
$$Q_{12} = Q_3 - Q_{11}$$
$$Q_{22} = Q_3 - Q_{21}$$

We also have expressions for the mean, variance and covariance of $X_1(t)$ and $X_2(t)$ in section 4.3.2, hence

$$\begin{aligned} E(X_1(t_i))+E(X_2(t_i)) &= E(N(t_i)) \\ &= X_1(t_{i-1})(Q_{11}+Q_{12})+X_2(t_{i-1})(Q_{21}+Q_{22}) \\ &= N(t_{i-1})Q_3 \end{aligned}$$

$$\begin{aligned} \text{var}(N(t_i)) &= \text{var}(X_1(t_i))+\text{var}(X_2(t_i))+2\ \text{cov}(X_1(t_i)X_2(t_i)) \\ &= N(t_{i-1})Q_3(1-Q_3) \end{aligned}$$

Thus the $N(t_i)$ follow the death process. The m.l.e. for Q_3 was developed in section 6.6 for observations from the death process, and so

$$\hat{Q}_3 = 1-[N(t_0)-N(t_n)]/\sum_{i=0}^{n-1} N(t_i)$$

$$\text{as.var } \hat{Q}_3 = Q_3(1-Q_3)^2/[N(t_0)(1-Q_3^n)].$$

Moreover, we have

$$
\begin{aligned}
D(t_i) &= X_1(t_i)-X_2(t_i)\\
\therefore E(D(t_i)) &= X_1(t_{i-1})(2Q_{11}-Q_3)+X_2(t_{i-1})(2Q_{21}-Q_3)\\
\mathrm{var}(D(t_i)) &= \mathrm{var}\ X_1(t_i)+\mathrm{var}\ X_2(t_i)-2\ \mathrm{cov}(X_1(t_i)X_2(t_i))\\
&= N(t_{i-1})Q_3(1-Q_3)+4(Q_{11}Q_{12}+Q_{21}Q_{22})\\
&= N(t_{i-1})Q_3(1-Q_3)+4Q_{11}(Q_3-Q_{11})+4Q_{21}(Q_3-Q_{21})
\end{aligned}
$$

Thus the iterative least squares procedure outlined in section 7.2 may be carried out using $\mu_i(\theta) = E(D(t_i))$ and $v_i(\theta) = \mathrm{var}(D(t_i))$. Each stage of the iterative procedure is a weighted multiple linear regression of the $D(t_i)$ on $X(t_{i-1})$ and $X_2(t_{i-1})$, and the method of deriving the estimates of the coefficients $b_1 = \widehat{(2Q_{11}-Q_3)}$ and $b_2 = \widehat{(2Q_{21}-Q_3)}$ is a standard technique (e.g. see Draper and Smith (1966)). Once final estimates of b_1 and b_2 are obtained, the variances and covariance of b_1 and b_2 can also be obtained, and then

$$
\begin{aligned}
\hat{Q}_{11} &= (b_1+\hat{Q}_3)/2\\
\hat{Q}_{21} &= (b_2+\hat{Q}_3)/2
\end{aligned}
$$

Approximate variances of Q_{11} and Q_{21} can be derived by assuming Q_3 to be constant. Then

$$
\begin{aligned}
\mathrm{var}\ \hat{Q}_{11} &\doteqdot (\mathrm{var}\ b_1)/4\\
\mathrm{var}\ \hat{Q}_{21} &\doteqdot (\mathrm{var}\ b_2)/4
\end{aligned}
$$

7.3.2 Estimates Using Complete Counts with Marking

1. When Δt_r = constant $r = 1 \ldots n$

As in 7.3.1, the model may be parameterized in terms of $\mathbb{Q}_{11}$, $\mathbb{Q}_{21}$ and $\mathbb{Q}_3$. The notation for the observations is as in section 7.2. The likelihood is, again, of the form 7.2.5, and so

$$L = \prod_{i=0}^{n-1} \binom{n_{iA}}{n_{i,i+1AA},\ n_{i,i+1AB},\ \delta n_{iA}} \mathbb{Q}_{11}^{\,n_{i,i+1AA}} (\mathbb{Q}_3-\mathbb{Q}_{11})^{n_{i,i+1AB}} (1-\mathbb{Q}_3)^{\delta n_{iA}}$$

$$\binom{n_{iB}}{n_{i,i+1BB},\ n_{i,i+1BA},\ \delta n_{iB}} (\mathbb{Q}_3-\mathbb{Q}_{21})^{n_{i,i+1BB}} \mathbb{Q}_{21}^{\,n_{i,i+1BA}} (1-\mathbb{Q}_3)^{\delta n_{iB}}$$

where $\delta n_{iA} = n_{iA}-n_{i,i+1AA}-n_{i,i+1AB}$ and similarly for δn_{iB}. Maximization of the log likelihood gives estimates

$$\hat{\mathbb{Q}}_3 = 1-(N_0-N_n)/(\sum_{i=0}^{n-1} N_i)$$

where $N_i = n_{iA}+n_{iB}$

$$\hat{\mathbb{Q}}_{11} = \hat{\mathbb{Q}}_3(\sum n_{i,i+1AA})/(\sum n_{i,i+1AA}+n_{i,i+1AB})$$

$$= \hat{\mathbb{Q}}_3(\sum n_{i,i+1AA})/(\sum n_{iA})$$

$$\hat{\mathbb{Q}}_{21} = \hat{\mathbb{Q}}_3(\sum n_{i,i+1BA})/(\sum n_{iB})$$

The estimate of $\hat{\mathbb{Q}}_3$ is the same estimate as that derived in 7.3.1 when counts are made without marking. Thus, as we would expect, marking the animals contributes no information for estimating the parameter of the death process. We then have, as before:

$$\text{as. var } \hat{\mathbb{Q}}_3 = \hat{\mathbb{Q}}_3(1-\mathbb{Q}_3)^2/[N_0(1-\mathbb{Q}_3)^n]$$

From the expectations of the negative second derivatives of the log likelihood, we find (after simplifying using the conditional expectations $E(n_{i,i+1AA}|n_{iA}) = n_{iA}\mathbb{Q}_{11}$ etc.):

$$\text{as. var } \hat{\mathbb{Q}}_{11} = \mathbb{Q}_{11}(\mathbb{Q}_3-\mathbb{Q}_{11})/\mathbb{Q}_3 E(\textstyle\sum n_{iA})$$

$$\text{as. var } \hat{\mathbb{Q}}_{21} = \mathbb{Q}_{21}(\mathbb{Q}_3-\mathbb{Q}_{21})/\mathbb{Q}_3 E(\textstyle\sum n_{iB})$$

$$\text{as. cov } (\hat{\mathbb{Q}}_{11}\hat{\mathbb{Q}}_{21}) = 0$$

$$\text{as. cov } (\hat{\mathbb{Q}}_{11}\hat{\mathbb{Q}}_{3}) = (\mathbb{Q}_3-\mathbb{Q}_{21})/E(\textstyle\sum n_{iB})$$

The expectations $E(\sum n_{iA})$, $E(\sum n_{iB})$ can be derived by finding the expression for $E(n_{iA})$ using the recursion relations:

$$E(n_{0A}) = x_{10}$$

$$E(n_{iA}) = E(n_{i-1A})\mathbb{Q}_{11}+E(n_{i-1B})\mathbb{Q}_{21}$$

$$E(n_{iB}) = E(n_{i-1A})\mathbb{Q}_{12}+E(n_{i-1B})\mathbb{Q}_{22}$$

By routine finite difference methods, this gives

$$E(n_{iB}) = c_1\mathbb{Q}_3^i-c_2(\mathbb{Q}_{11}-\mathbb{Q}_{21})^i$$

$$E(n_{iA}) = \frac{c_1}{(\mathbb{Q}_3-\mathbb{Q}_{11})}\,\mathbb{Q}_3^i-c_2(\mathbb{Q}_{11}-\mathbb{Q}_{21})^i$$

where $c_2 = [x_{10}+x_{20}(1-\mathbb{Q}_3)]/(\mathbb{Q}_{11}-\mathbb{Q}_{21}-\mathbb{Q}_3)$

$c_1 = x_{20}-c_2$

And summing, $E(\sum n_{iB}) = \dfrac{c_1}{1-\mathbb{Q}_3}(1-\mathbb{Q}_3^n)+\dfrac{c_2}{(1-\mathbb{Q}_{11}+\mathbb{Q}_{21})}[1-(\mathbb{Q}_{11}-\mathbb{Q}_{21})^n]$

$$E(\textstyle\sum n_{iA}) = \dfrac{c_1(1-\mathbb{Q}_3^n)}{(1-\mathbb{Q}_3)(\mathbb{Q}_3-\mathbb{Q}_{11})} - \dfrac{c_s}{(1-\mathbb{Q}_{11}+\mathbb{Q}_{21})}[1-(\mathbb{Q}_{11}-\mathbb{Q}_{21})^n]$$

For n large, these expectations converge to values independent of n, as one would expect, since the probability of all individuals eventually dying off is 1.

2. When the Δt_r differ:

The log likelihood, in α, β, and μ, may be maximized iteratively. It did not prove tractable for deriving analytic estimates. The log likelihood is:

$$\log L = \sum_{i=0}^{n-1} \{ -\mu t_i N(t_{i+1}) + (N(t_i) - N(t_{i+1})) \log(1-e^{-\mu\Delta t_i})$$

$$+ n_{i,i+1AA} \log(\alpha+\beta e^{-(\alpha+\beta)\Delta t_i}) + n_{i,i+1BB} \log(\beta+\alpha e^{-(\alpha+\beta)\Delta t_i})$$

$$+ (n_{i,i+1BA} + n_{i,i+1AB}) \log(1-e^{-(\alpha+\beta)\Delta t_i})$$

$$- N(t_{i+1}) \log(\alpha+\beta) + n_{i,i+1BA} \log \alpha + n_{i,i+1} \log \beta \}$$

7.4 The Time Dependent Model

We will consider methods of estimation for the non-homogeneous model of section 3.5 where the migration intensities are functions of time:

$$\alpha(x_2,t) = \alpha\rho(t)$$

$$\beta(x_1,t) = \beta\rho(t)$$

where $\rho(t)$ is some arbitrary function of time and $\rho(t) > 0 \quad t \in [t_0, t_n]$.

We will not assume any explicit form for $\rho(t)$ but will try to ascertain the form of the function by estimation. It was noted, in section 3.5 that the p.d.f. of $X_1(t)$ for this model is identical to that of the simple linear model, except that $\exp[-(\alpha+\beta)\delta t]$ was replaced by $\exp[-(\alpha+\beta)\int_t^{t+\delta t} \rho(t)dt = \tau(\delta t)$ (say). If δt is small, the integral is approximately equal to $\delta t\ \rho(\delta t/2)$ and hence, if we can obtain estimates of $\tau(\Delta t_r)$ $\Delta t_r = t_{r+1} - t_r$, $r = 0 \ldots n-1$, by making complete counts at times $t_0, t_1 \ldots t_n$, then by plotting $-\frac{1}{\Delta t_r} \log \hat{\tau}(\Delta t_r)$ against $t = t_r + \Delta t_r/2$, we will have a series of points that conform to the function $(\alpha+\beta)\rho(t)$ except for errors due to estimation.

If we also estimate these variances due to estimation, then least squares may be used to fit some suitable function for $\rho(t)$, or to test any hypothesis concerning $\rho(t)$ (such as $\rho(t)$ = constant). This idea, of estimating the $\tau(\Delta t_r)$ was first suggested by Pyke (1956) and was applied to models involving one way migration only.

We shall estimate the $\tau(\Delta t_r)$ and its variance rather than the $\log(\tau(\Delta t_r))$, as this proves to be more convenient. If the Δt_r are all equal (say $\Delta t_r = c \;\forall r$) then no generality is lost. The estimates $\hat{\tau}(\Delta t_r)$ then conform to the function $\exp[-c(\alpha+\beta)\rho(t)]$ $t = t_r + c/2$.

7.4.1 Estimates from Complete Counts Without Marking.

The likelihood of the observations $n_0 = X_1(t_0) \ldots n_n = X_1(t_n)$ is:

$$L = \prod_{i=1}^{n} \Pr(n_i | n_{i-1})$$

$$= \prod_{i=1}^{n} \sum_{j=0}^{n_i} \binom{n_{i-1}}{j} \binom{N-n_{i-1}}{n_i-j} (\alpha+\beta\tau_i)^j (\alpha-\alpha\tau_i)^{n_{i-1}-j} (\beta-\beta\tau_i)^{n_i-j} (\beta+\alpha\tau_i)^{N-n_i-n_{i-1}+j} (\alpha+\beta)^{-N}$$

Evidently, this likelihood involves n+2 parameters (τ_i, $i = 1 \ldots n$, α, β) and we have only n+1 observations. Clearly, estimation is possible only in the case $\alpha = \beta$ when the likelihood reduces to:

$$L = \prod_{i=0}^{n} \sum_{j=0}^{n_i} \binom{n_{i-1}}{j} \binom{N-n_{i-1}}{n_{i-1}} (1+\tau_i)^{n_{i-1}+n_i-2j} (1-\tau_i)^{N-n_{i-1}-n_i+2j}$$

$$\text{where } \tau_i = \exp[-2\alpha \int_{t_{i-1}}^{t_i} \rho(t)dt]$$

$$\doteqdot \exp[-2\alpha c \rho(t_{i-1}+c/2)]$$

$$\therefore \quad L = \prod_{i=1}^{n} K[(1-\tau_i)/(1+\tau_i)]^{n_i+n_{i-1}} (1-\tau_i)^N S$$

$$\text{where } S = \sum_{j=0}^{n_i} \frac{\binom{n_{i-1}}{j}\binom{N-n_{i-1}}{n_{i-j}}}{\binom{N}{n_i}} \left[\frac{(1-\tau_i)^2}{(1+\tau_i)^2}\right]^j$$

It proved impossible to maximize this likelihood for the $\hat{\tau}_i$. An approximation for the likelihood was developed by noting that the term S is in the form of the p.g.f. of a hypergeometric distribution with mean $n_i(n_{i-1}/N)$, and hence, if n_{i-1}/N is small, S could be replaced by the p.g.f. of a binomial distribution with the same mean, namely

$$S \simeq [(1-n_{i-1}/N) + \frac{n_{i-1}}{N}\frac{(1-\tau_i)^2}{(1+\tau_i)^2}]^{n_i}$$

This was thought to be a reasonable procedure, since the n_{i-1}/N must be small [or $1-n_{i-1}/N$ must be small, in which case we may consider $n_i = X_2(t_i)$]. If n_{i-1}/N is not small, the distribution will be near equilibrium, and as the equilibrium distribution is not a function of τ_i, this type of experiment is not suitable for estimating the τ_i. Maximization of the approximate expression for the log likelihood leads to a cubic equation for $\hat{\tau}_i$. However, the equation permits negative roots, and multiple positive roots, and, in several cases tried, was shown to give completely inadequate estimates.

Useful estimates may be derived, however, by using the moment equation for $p_i = n_i/N$, provided, again, that p_i or $1-p_i$ is small for all i. We then have

$$E(p_i|p_{i-1}) = \frac{1}{2} + (p_{i-1} - \frac{1}{2})\tau_i \qquad (7.4.1)$$

$$\begin{aligned} \mathrm{var}(p_i|p_{i-1}) &= \mathrm{var}(n_i)/N^2 \\ &= (1-\tau^2)/4N \end{aligned}$$

Removing the expectation sign from (7.4.1) and solving for τ_i gives the estimate

$$\hat{\tau}_i = (p_i-1/2)/(p_{i-1}-1/2)$$

Thus, so long as $p_i < 1/2 \ \forall i$ <u>or</u> $p_i > 1/2 \ \forall i$ the estimates $\hat{\tau}_i$ will all be positive. This will be the case so long as the system is not close to equilibrium. Also

$$\begin{aligned} E(\hat{\tau}_i) &= (E(p_i|p_{i-1})-1/2)/(p_{i-1}-1/2) \\ &= \tau_i \end{aligned}$$

so the estimate is unbiased, and

$$\begin{aligned} \mathrm{var}(\hat{\tau}_i) &= \mathrm{var}(p_i|p_{i-1})/(p_{i-1}-1/2)^2 \\ &= (1-\tau_i^2)/4N(p_{i-1}-1/2)^2 \end{aligned}$$

Again, notice that as the system tends to equilibrium, p_i approaches 1/2 and the variances of the $\hat{\tau}_i$ will tend to become large.

Table 7.1

Counts without Marking from the model with $\rho(t)dt = dt$

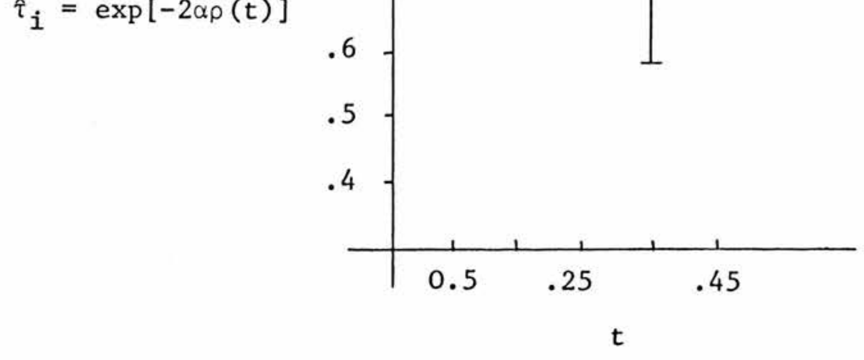
$N = 250 \quad \alpha = \beta = 1 \quad \Delta t = .1$

$\exp[-2\alpha \int_0^{.1} dt] = .8187$

i	0	1	2	3	4	5
n_i	0	16	32	52	71	76
$\hat{\tau}_i$	-	.8720	.8257	.8111	.7397	.9074
$\widehat{s.e.}(\hat{\tau}_i)$	-	.031	.041	.051	.073	.062

Fig. 7.1

Plot of $\hat{\tau}_i$ (with $\pm 2\ \widehat{s.e.}\ (\hat{\tau}_i)$) for data of table 7.1.



In table 7.1 we present an example of the estimates applied to data which, in fact, came from the time homogeneous model (simple linear model) with $N = 250$ $\alpha = \beta = 1$. Counts of $X_1(t_i)$ were made at equal time intervals $t = 0, .1, .2 \ldots .5$. Hence the $\hat{\tau}_i$ are estimates of the function $\exp[.2\alpha]$ for $t = .05, .15, .25 \ldots .45$. In Fig. 7.1 the estimated functional values are plotted against time. The parentheses above and below each point denote the points two standard deviations above and below each point. It is evident that the hypothesis $\rho(t) =$ constant is not inconsistent with the data, but the variances of the estimates are large (and obviously tend to increase with time as the system moves towards equilibrium).

7.4.2 Estimates from complete counts with marking

The likelihood of the observations $n_{i,i+1AA}$, $n_{i,i+1AB}$, $n_{i,i+1BA}$, $n_{i,i+1BB}$ $i = 0 \ldots n-1$ is given by 7.2.6 and from section 3.3 we have the distributions

$$B(n_{i,i+1BA};\ n_{iB}, r(1-\tau_i))$$
$$B(n_{i,i+1AA};\ n_{iA}, (r+(1-r)\tau_i))$$

where $r = \alpha/(\alpha+\beta)$. Since we have $2n+2$ observations and $n+2$ parameters it is now possible to estimate α, β, and τ_i $i = 0 \ldots n-1$ where

$$\tau_i = \exp[-(\alpha+\beta)\int_{t_i}^{t_{i+1}} \rho(t)dt]$$
$$\doteqdot \exp[-c(\alpha+\beta)\rho(t_i+c/2)]$$

when $t_{i+1}-t_i = c \ \forall i$.

It is easier to estimate only the $n+1$ parameters r and τ_i

$i = 0 \ldots n-1$. This will not prevent us from investigating the form of the function $\rho(t)$.

The log likelihood of the observations is:

$$\log L = \sum_{i=0}^{n-1} n_{i,i+1BA}\log[r(1-\tau_i)]+n_{i,i+1BB}\log[1-r+r\tau_i]$$

$$+n_{i,i+1AB}\log[(1-r)(1-\tau_i)]+n_{i,i+1A_A}\log[r+(1-r)\tau_i]$$

Differentiating w.r.t. r and setting equal to zero gives:

$$\sum_{i=0}^{n-1}\{\frac{n_{i,i+1BA}}{r}-\frac{n_{i,i+1BB}\tau_i}{(1-r+r\tau_i)}-\frac{n_{i,i+1AB}}{(1-r)}+\frac{n_{i,i+1AA}(1-\tau_i)}{r+(1-r)\tau_i}\}=0 \quad (7.4.2)$$

Differentiating w.r.t. τ_i and setting equal to zero gives, after some rearrangement:

$$\tau_i^2[r^2(n_{i,i+1AA}+n_{i,i+1BB}-n_{i,i+1AB}-n_{i,i+1BA})+r(n_{i,i+1AB}+n_{i,i+1BA})]$$

$$+\tau_i[2r^2-r+1)(n_{i,i+1AB}+n_{i,i+1BA})-Nr]$$

$$+[r(n_{i+1B}+n_{iB})-Nr^2-n_{BB}] = 0 \qquad (7.4.3)$$

These equations may be solved iteratively as follows: If the system is in equilibrium, an initial estimate for r is $r_{(1)} = \sum_{i=0}^{n} n_{iA}/Nn$. This value may then be used in (7.4.3) and the set of quadratic equations solved for the set of first estimates of τ_i. These in turn may be used in 7.4.2, and Newton's iterative method may be used to find a new value for r, the root of the equation 7.4.2. This process can be repeated until successive iterations are in sufficient agreement. It proved impossible to demonstrate analytically, that 7.4.3 and 7.4.2 always gave unique positive roots at each iteration, or that the above process would always converge. No examples were attempted.

If $\alpha = \beta$, the likelihood simplifies considerably. Here $r = 1/2$ and we have:

$$\log L = \sum_{i=0}^{n-1} \{(n_{i,i+1BA}+n_{i,i+1AB})\log(1-\tau_i)$$

$$+(n_{i,i+1BB}+n_{i,i+1AA})\log(1+\tau_i)\}$$

Hence

$$\hat{\tau}_i = (n_{i,i+1_{AA}}+n_{i,i+1BB}-n_{i,i+1AB}-n_{i,i+1BA})/N$$

$$\text{as. var } \hat{\tau}_i = \left[\frac{E(n_{i,i+1BA}+n_{i,i+1AB})}{(1-\tau_i)^2} + \frac{E(n_{i,i+1AA}+n_{i,i+1BB})}{(1+\tau_i)^2}\right]^{-1}$$

but

$$E(n_{i,i+1AB}|n_{iA}) = n_{iA}(1-\tau_i)/2 \qquad \text{etc.}$$

hence

$$\text{as. var}(\tau_i) = \frac{1}{2}\left[\frac{E(n_{iA}+n_{iB})}{(1-\tau_i)} + \frac{E(n_{iA}+n_{iB})}{(1+\tau_i)}\right]^{-1}$$

$$= \frac{1-\tau_i^2}{N}$$

In table 7.2 we again give an example of the estimation procedure for the time homogeneous model. The data of table 7.1 is, in fact, taken from table 7.2, but ignores the information due to the marking. The two sets of estimates are thus strictly comparable. It can be seen immediately by comparing figure 7.1 with figure 7.2 that there is no appreciable difference in the point estimates, but the standard errors of the estimates, when individuals are marked, are smaller and do not tend to increase with time. Thus, as the system approaches equilibrium, one may still derive usefully precise estimates of the τ_i if individuals are distinguishable.

Table 7.2

Counts with Marking from the Model with

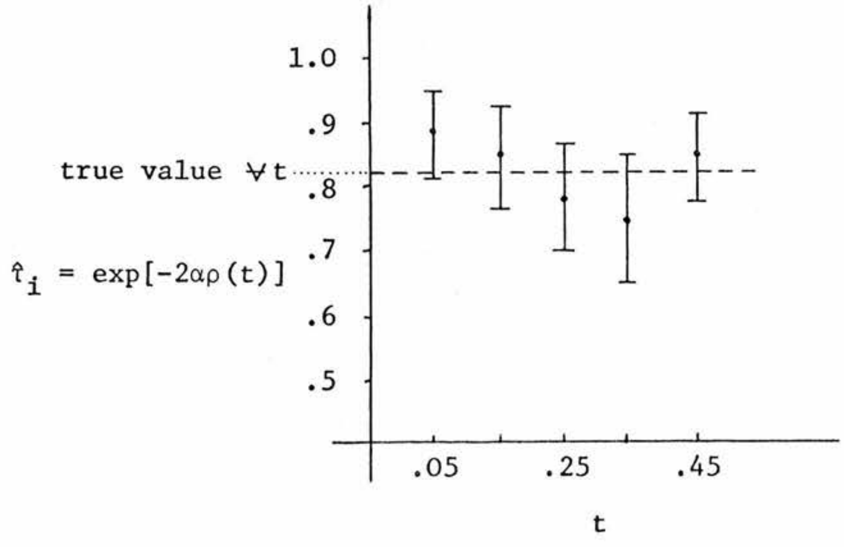
$\rho(t)dt = dt$

$N = 250 \quad \alpha = \beta = 1 \quad \Delta t = .1$

i	0	1	2	3	4	5
n_{iA}	0	16	35	52	71	76
$n_{i,i+1AA}$	0	16	30	47	64	
$n_{i,i+1AB}$	0	0	5	5	7	
$n_{i,i+1BA}$	16	19	22	24	12	
$n_{i,i+1BB}$	234	215	193	174	167	
$\hat{\tau}_i$	.872	.848	.784	.768	.848	
s.e.$(\hat{\tau}_i)$	.031	.034	.039	.041	.034	

Fig. 7.2.

Plot of $\hat{\tau}_i$ (with $\pm 2\ \widehat{s.e.}\ (\hat{\tau}_i)$) for data of table 7.2.



CHAPTER EIGHT

ESTIMATION FROM SAMPLING EXPERIMENTS WITH MARKING

8.1 Introduction

In this chapter we consider the problem of estimating population sizes, survival rates and migration rates using multiple mark-recapture samplings in two populations. We have already remarked, in section 4.3, that the representation of migration models as convolutions of multinomial distributions facilitates the determination of sampling distributions when binomial sampling is assumed, and when animals survive, migrate etc., independently of one another. Only this approach has been taken in this cahpter. It should be noted however that there are certain shortcomings of the method. First, populations subject to density dependent growth or migration cannot be treated in this way. More seriously, the number of parameters increases with the number of sampling occasions. Thus, to a certain degree, one of the purposes of multiple sample experiments is vitiated; the gain in precision of the estimates is not commensurate with the additional effort expended in taking more samples. However, the experimenter is still able to "monitor" the population over an extended period of time, and to note changes in population size or shifts in its distribution (over the two areas).

In section 5.4, the only extant, widely applicable methods of estimation from sampling data have been described. The work of Chapman and Junge (1956) and Darroch (1961) gives a thorough description of what can be derived from a two sample experiment. Such experiments cannot yield estimates of survival. To obtain information on survival rates, one must sample on at least three occasions. The only available estimates for a triple catch experiment were given by Iwao (1963), but, as pointed out in section 5.4, these estimates are invalid if animals migrate more than once. Moreover, since Iwao does not explicitly develop a model or sampling distribution, it would be difficult to derive variances for his estimates.

It seems worthwhile, therefore, to develop a quite general expression for the sampling distribution of the multiple mark recapture experiment where samples are drawn simultaneously from two populations on n occasions. As it turned out, this expression proves intractable for deriving estimates in the general case, but it provides a complete description of the experiment which is useful in determining the properties of any estimates which can be derived by other means.

8.2 The n-sample Experiment on Two Non-closed Populations

This experiment, and the model used to describe it, is an extension of the work of Jolly (1965). In that study, he developed a likelihood-type expression for the following experiment: samples are drawn from a (single) population at n consecutive sampling occasions. The population need not be homogeneous (in the sense that all individuals are subject to the same survival rates, capture rates, etc.), but if the population is stratified (i.e. non-homogeneous), it is assumed that individuals belong to one and only one stratum throughout the course of the experiment. In fact, Jolly develops estimates only for the case of a homogeneous population. In this chapter we shall also assume homogeneity. At each sampling occasion each captured individual is given a unique identifying mark (if it has not already been given one at a previous sampling occasion) and returned to the population. Thus, by recording which individuals were captured on each sampling occasion, the experimenter has a complete record of the capture history of all the animals taken in the course of sampling. It turns out that it is not necessary to know complete capture histories to estimate survival rates, population sizes, etc., but simply to know, for each animal observed at time j, the most recent time i ($i < j$) that the animal was captured. It will be shown that the same is true when migration occurs, with the added proviso that a record is made of where the individual was last captured (either on population A or population B) as well as when.

In Jolly's experiment, the sampling times are designated by $t = 1, 2, \ldots, n$ but there is no restriction that these times be equally spaced. It is assumed that sampling is with replacement or that the population is sufficiently large that the binomial approximation for the hypergeometric distribution is applicable. Losses on capture are allowed for, to take account of the probability that an animal dies before it can be returned to the population. These features will also be incorporated in this chapter. With the added assumptions of homogeneity and independence with respect to survival and capture probabilities, Jolly proceeds to develop a likelihood expression for this experiment as follows: the sample observations at sampling time $t = j$ are conditional upon a certain structure of the population at the time the jth sample is taken. This structure is defined by a number of unobservable random variables; the population size, the number of unmarked individuals in the population, the number of individuals having each possible class of capture histories. The probability of the observed sample, given this structure can be written down, and, given the sample observations and the structure at time j, the probability distribution of the structure at time j+1 can be determined. The product of these two expressions, multiplied over j = 1, ..., n then gives the joint distribution of the sample observations and the unobservable random variables. The true likelihood of the observations, then, is this product summed over all possible values of the unobservable random variables. As this procedure of summation makes the likelihood expression entirely intractable for estimation purposes, Jolly proposes that the unobservable random variables be considered as parameters and that estimates be formed by treating the joint distribution as a likelihood. By doing so he is able to derive a number of relationships among the estimates of the parameters and unobservable random variables at times j and j+1 which are functions of the observations. This set of relations can then be solved recursively to give estimates in terms of the observations only.

We shall now extend this method to the similar situation where sampling is carried out simultaneously on two populations A and B at times t = 1, 2, ..., n. Animals are subject to stochastic death. The probability that animal survives to t = j+1 given that it is alive at t = j in population A may differ from the probability that it survives to t = j+1 given that it is alive at t = j in population B. An individual may migrate in either direction between the two populations, and may do so more than once between j and j+1. Animals may be "lost on capture" and sampling is assumed to conform to the assumptions of the binomial distribution. The assumptions of independence and homogeneity apply to all these stochastic events. As in the work of Jolly, births are most readily accounted for by incorporating them in a deterministic fashion: a number of new individuals, of fixed but unknown size, enters each of the populations between j and j+1, and these new individuals are assumed to be indistinguishable from the unmarked individuals already present.

It might be pointed out that a much more parsimonious parameterization of this model is possible. Seber (1965) developed a model similar to that of Jolly without introducing the unobservable random variables into the likelihood expression. This nevertheless yielded identical estimates. The same might be done here, but the introduction of the extra parameters elucidates the structure of the system and the resulting sampling distribution.

8.2.1 Notation: The complete notation required for this chapter is given below, and is consistent with the unified notation suggested by Cormack (1968).

Let $j = 1 \dots n \quad i = 1 \dots j-1 \quad x = A$ or $B \quad y = A$ or B

where n is the number of sampling occasions.

Parameters:

N_{1x} : number of individuals in population x at time of first sample.

B_{jx} : number of new individuals, entering the populations for the first time between $t = j-1$ and $t = j$ and which are in population x at $t = j$.

$\mathbb{Q}_{jxy}$: prob (an animal, alive in x at $t = j$ is alive and in y at $t = j+1$).

$\mathbb{Q}_{jx}$: $= \mathbb{Q}_{jxA} + \mathbb{Q}_{jxB}$.
= prob (an animal, alive in x at $t = j$ is alive (in A or B) at $t = j+1$).

p_{jx} : prob (an animal, alive at $t = j$ is captured in the jth sample).

η_{jx} : prob (an animal, captured from x at t = j is released; i.e. does not become a "loss on capture").

Unobservable Random Variables

M_{ijxy}: number of individuals, alive in y at t = j that were last captured in x at t = i.

U_{jy} : number of unmarked animals alive in y at the time the jth sample is taken.

N_{jy} : $= U_{jy} + M_{\cdot j \cdot y} = U_{jy} + \sum_{i=1}^{j-1} \sum_{x=A,B} M_{ijxy}$

= total number of individuals alive in y at the time the jth sample is taken.

Observations

m_{ijxy}: number of individuals, caught in y at t = j that were last captured in x at t = i.

= number of the M_{ijxy} taken in the jth sample.

u_{jy} : number of unmarked animals taken in the jth sample from population y.

n_{jy} : total number of animals taken in the jth sample from population y.

s_{jy} : number of animals out of the n_{jy} that are released back into population y after the jth sample is taken.

8.2.2 Development of the Joint Distribution

The joint distribution of the sample observations and the unobservable random variables can now be derived. The structure at time j is given by the M_{ijxy} and the U_{jy} (i = 1, ..., j-1
x, y = A or B) and we need to develop the joint distribution of the

observations at $t = j$ and the structure at $t = j+1$ given the M_{ijxy}, U_{jy} and the parameters.

Consider first what may happen to a member of M_{ijxy} between j and $j+1$. It may:

1. be caught in population y in the jth sample in which case it next appears, if at all, in M_{jkyx} $(k > j)$.
2. not be caught and survive to become a member of M_{ij+1xA} or M_{ij+1xB}.
3. not be caught and die or emigrate permanently between j and $j+1$.

If the individual is caught (1.), then he becomes a member of m_{ijxy}. If 2. or 3. occur, the individual is one of $M_{ijxy} - m_{ijxy}$ individuals who will subsequently become a member of M_{ij+1xA} or M_{ij+1xB} (2.) or disappear from the system altogether (3.). The distribution of these uncaught individuals causes some difficulty. The M_{ij+1AA} and the M_{ij+1AB} groups can both be made up of individuals who were members of either the $M_{ijAA} - m_{ijAA}$ or the $M_{ijAB} - m_{ijAB}$ individuals that were not caught at j but were last seen in A at $t = i$. Similarly both M_{ij+1BA} and M_{ij+1BB} may be made up of individuals from either the $M_{ijBB} - m_{ijBB}$ or the $M_{ijBA} - m_{ijBA}$.

Consider the case of the M_{ij+1AA} and the M_{ij+1AB} individuals. Each of the $M_{ijAA} - m_{ijAA}$ individuals may fall into exactly one of the following classes, with probabilities as indicated:

x_1	survive in A to join M_{ij+1AA} (prob. = $\mathbb{Q}_{jAA}$).
y_1	survive in B to join M_{ij+1AB} (prob. = $\mathbb{Q}_{jAB}$).
z_1	do not survive to $t = j+1$ (prob. = $1 - \mathbb{Q}_{jA}$).

Thus (x_1, y_1, z_1) has the trinomial distribution which may be indicated by

$$(x_1, y_1, z_1) \sim T(x_1 y_1; M_{ijAA} - m_{ijAA}, \mathbb{Q}_{jAA}, \mathbb{Q}_{jAB}).$$

Similarly, the $M_{ijAB} - m_{ijAB}$ may be partitioned into (x_2, y_2, z_2) with probabilities conditional on M_{ijAB} and m_{ijAB}.

x_2 : survive in A to join M_{ij+1AA} (prob. $\mathbb{Q}_{jBA}$).
y_2 : survive in B to join M_{ij+1AB} (prob. $\mathbb{Q}_{jBB}$).
z_2 : don't survive to $t = j+1$ (prob. $1 - \mathbb{Q}_{jBA} - \mathbb{Q}_{jBB}$),

and so $(x_2, y_2, z_2) \sim T(x_2 y_2; M_{ijAB} - m_{ijAB}, \mathbb{Q}_{jBA}, \mathbb{Q}_{jBB})$. Thus the joint distribution of the M_{ij+1AA} individuals with the M_{ij+1AB} individuals (and the remaining $M_{ijAA} - m_{ijAA} + M_{ijAB} - m_{ijAB} - M_{ij+1AA} - M_{ij+1AB}$ that don't survive) is distributed as (x, y, z) where $x = x_1 + x_2$, $y = y_1 + y_2$, $z = z_1 + z_2$; that is (x, y, z) is the sum of two independent trinomially distributed (vector) variates: this will be written

$$PT(M_{ij+1AA}, M_{ij+1AB}; M_{ijAA} - m_{ijAA}, \mathbb{Q}_{jAA}, \mathbb{Q}_{jAB}, M_{ijAB} - m_{ijAB}, \mathbb{Q}_{jBA}, \mathbb{Q}_{jBB})$$

where $PT(a, b; N_1\ p_1\ p_2, N_2\ r_1\ r_2)$ indicates the probability of $(a, b, N_1 + N_2 - a - b)$

$$= \sum_{m=0}^{a} \sum_{n=0}^{b} \{\binom{N_1}{m, n, N_1-m-n} p_1^m p_2^n (1-p_1-p_2)^{N_1-m-n}\}\{\binom{N_2}{a-m, b-n, N_2-a-b+m+n} r_1^{a-m} r_2^{b-n} (1-r_1-r_2)^{N_2-a-b+m+n}\} \tag{8.2.0}$$

which can readily be seen to be the distribution of the sum of two independent trinomial vector variates.

Similarly, the joint distribution of M_{ij+1BB} and the M_{ij+1BA} individuals is:

$$PT(M_{ij+1BB}, M_{ij+1BA}; M_{ijBB} - m_{ijBB}, \mathbb{Q}_{BB}, \mathbb{Q}_{BA}, M_{ijBA} - m_{ijBA}, \mathbb{Q}_{AB}, \mathbb{Q}_{AA})$$

If we then write $B(r; N, p)$ to indicate that r has the binomial distribution:

$$\text{prob}(r) = \binom{N}{r} p^r (1 - p)^{N-r}$$

we may then write down the joint distribution of numbers in each group of marked individuals in the jth sample, and the structure of the populations at time j+1 with regard to marked individuals; this distribution is:

$$\begin{aligned}
&\prod_{i=1}^{j-1} P(m_{ijxy}, M_{ij+1xy} | p_{jx}, \mathbb{Q}_{jxy}, M_{ijxy}) \\
= &\prod_{i=1}^{j-1} B(m_{ijAA}; M_{ijAA}, p_{jA}).B(m_{ijBB}; M_{ijBB}, p_{jB}) \\
&.B(m_{ijBA}; M_{ijBA}, p_{jA}).B(m_{ijAB}; M_{ijAB}, p_{jB}) \\
&.PT(M_{ij+1AA}, M_{ij+1AB}; M_{ijAA} - m_{ijAA}, \mathbb{Q}_{jAA}, \mathbb{Q}_{jAB}, \\
&\qquad\qquad M_{ijAB} - m_{ijAB}, \mathbb{Q}_{jBA}, \mathbb{Q}_{jBB}) \\
&.PT(M_{ij+1BB}, M_{ij+1BA}; M_{ijBB} - m_{ijBB}, \mathbb{Q}_{jBB}, \mathbb{Q}_{jBA}, \\
&\qquad\qquad M_{ijBA} - m_{ijBA}, \mathbb{Q}_{jAB}, \mathbb{Q}_{jAA})
\end{aligned} \tag{8.2.1}$$

The joint distribution of the unmarked animals in the jth sample and in the two populations at time j+1 is similar. However, the new entries B_{j+1x} that enter the population between t = j and t = j+1 must be accounted for. We thus have

$$P(u_{jx}, U_{j+1x} | U_{jx}, B_{j+1x}, p_{jx}, \mathbb{Q}_{jxy})$$

$$= B(u_{jA}; U_{jA}, p_{jA}).B(u_{jB}; U_{jB}, p_{jB})$$

$$.PT(U_{j+1A} - B_{j+1A}, U_{j+1B} - B_{j+1B}; U_{jA} - u_{jA}, \mathbb{Q}_{jAA}, \mathbb{Q}_{jAB}, U_{jB} - u_{jB}, \mathbb{Q}_{jBA}, \mathbb{Q}_{jBB}) \qquad (8.2.2)$$

Finally, the distribution of the n_{jx} animals captured in x is:

$$P(s_{jx}, M_{jj+1xy} | \eta_{jx}, \mathbb{Q}_{jxy}, n_{jx})$$

$$= B(s_{jA}; n_{jA}, \eta_{jA}).T(M_{jj+1AA}, M_{jj+1AB}; s_{jA}, \mathbb{Q}_{jAA}, \mathbb{Q}_{jAB})$$

$$.B(s_{jB}; n_{jB}, \eta_{jB}).T(M_{jj+1BB}, M_{jj+1BA}; s_{jB}, \mathbb{Q}_{jBB}, \mathbb{Q}_{jBA}) \qquad (8.2.3)$$

The product of the three terms (8.2.1), (8.2.2), (8.2.3) completely defines the joint distribution of the sample observations at time j and the structure at time j+1, given the structure at time j, for j = 2, ..., n. For j = 1, the distribution is:

$$
\begin{aligned}
P_1 = {} & B(n_{1A}; N_{1A}, p_{1A}).B(n_{1B}, N_{1B}, p_{1B}) \\
& .PT(U_{2A} - B_{2A}, U_{2B} - B_{2B}; N_{1A} - n_{1A}, \phi_{1AA}, \phi_{1AB}, N_{1B} - n_{1B}, \phi_{1BA}\ \phi_{1BB}) \\
& .B(s_{1A}; n_{1A}, \eta_{1A}).T(M_{12AA}, M_{12AB}; s_{1A}, \phi_{1AA}, \phi_{1AB}) \\
& .B(s_{1B}; n_{1B}, \eta_{1B}).T(M_{12BB}, M_{12BA}; s_{1B}, \phi_{1BB}, \phi_{1BA})
\end{aligned}
$$

Thus the entire joint distribution is given by:

$$
\begin{aligned}
L = P_1 . \prod_{j=2}^{n} \{ [\prod_{i=1}^{j-1} P(m_{ijxy}, M_{ij+1xy}\ p_{1x}, \phi_{jxy}, M_{ijxy})] \\
.P(u_{jx}, U_{j+1x}\ U_{jx}, B_{j+1x}, p_{jx}, \phi_{jxy}) \\
.P(s_{jx}, M_{jj+1xy} | \eta_{jx}, \phi_{jxy}, n_{jx}) \}
\end{aligned}
\qquad (8.2.4)
$$

The formation of estimates is accomplished by finding the values of the unobservable random variables and parameters which maximize L. For continuous parameters, log L is differentiated with respect to the parameter and set equal to zero, and for discrete parameters (including the unobservable random variables), say N, the first difference of log L, log L(N) - log L(N-1) is set equal to zero. This system of equations must then be solved to give parameter estimates. At first glance, this might appear to be an impossible task, especially as L contains terms of the form PT (a, b) (see equation 8.2.0) which are not expressible in closed form. If births are assumed to occur, then the parameters B_{jA}, B_{jB} enter into the limits of the summation of the expression PT(a, b), and this does seem to make the expression L entirely intractable for forming estimates. If no births occur, however, some progress can be made, although the task of deriving estimates for the general case involves extremely torturous algebra and is perhaps impossible.

The method used was as follows:

If equation (8.2.0) is written

$$PT(a, b; N_1, p_1, p_2, N_2, r_1, r_2)$$

$$= \sum_{m=0}^{a} \sum_{n=0}^{b} T(m,n; N_1, p_1, p_2).T(a-m, b-n; N_2, r_1, r_2)$$

and if

$$R(a, b) = \sum_{m=0}^{a} \sum_{n=0}^{b} mT(m,n).T(a-m, b-n)$$

$$S(a, b) = \sum_{m=0}^{a} \sum_{n=0}^{b} nT(m,n).T(a-m, b-n),$$

then it is easy to show, by differentiation and a bit of combinatorial juggling that:

$$\frac{\partial \log PT}{\partial p_1} = \frac{-N_1}{(1-p_1-p_2)} + \frac{(1-p_2)}{(1-p_1-p_2)p_1} R(a,b) + \frac{p_1}{(1-p_1-p_2)p_1} S(a,b)$$

Similarly, expressions (which also include terms in R(a,b) and S(a,b)) were derived for the derivative of log PT with respect to p_2, r_1 and r_2 and for the difference of log PT with respect to N_1, N_2, a and b. Expressions for the differentials and differences of a general binomial or trionmial distribution are also easily derived. These formulae were then applied in forming the differentials and differences of log L. Setting these expressions equal to zero, gives a series of equations from which the difficult terms of the form R(a,b) and S(a,b) can be eliminated. This leaves a reduced set of equations with no terms involving summations, but it was not clear if enough independent equations remained to derive estimates. It would be pointless to give a detailed description of this process, but some results can be reported. The equations resulting from differentiation of log L by $\mathbb{Q}_{jBB}$, $\mathbb{Q}_{jAA}$, $\mathbb{Q}_{jAB}$, $\mathbb{Q}_{jBA}$ lead to the following relations among the estimates:

$$\hat{\mathbb{Q}}_{jAB}(\hat{N}_{jA} - n_{jA} + s_{jA}) + \hat{\mathbb{Q}}_{jBB}(\hat{N}_{jB} - n_{jB} + s_{jB}) = \hat{N}_{j+1B} \qquad (8.2.5)$$

$$\hat{\mathbb{Q}}_{jAA}(\hat{N}_{jA} - n_{jA} + s_{jA}) + \hat{\mathbb{Q}}_{jBA}(\hat{N}_{jB} - n_{jB} + s_{jB}) = \hat{N}_{j+1A} \qquad (8.2.6)$$

$$\frac{\partial \log L}{\partial p_{jx}} \text{ gives } \hat{p}_{jx} = n_{jx}/\hat{N}_{jx} \qquad (8.2.7)$$

$$\frac{\partial \log L}{\partial \eta_{jx}} \text{ gives } \hat{\eta}_{jx} = s_{jx}/n_{jx} \qquad (8.2.8)$$

The equations resulting by differencing log L have proved to be more difficult to deal with and, as yet, I have not been able to eliminate the terms involving summations from these equations.

Before leaving the problem of the general "likelihood" expression, it should be noted that the expression L is a function of the m_{ijxy} and u_{jx} only. That is, the joint distribution is given in terms of the observation of when and where each individual in the sample was last sighted. The observations do not make use of capture history previous to the most recent sighting. However, it should be noted that if an individual is captured r times, it enters into exactly r of the observations $\{m_{ijx,y}, u_{jx} | j = 1 \ldots n \quad i = 1 \ldots j-1 \quad x, y = A \text{ or } B\}$. Thus, in fact, the entire capture history of each individual is accounted for, and the $\{m_{ijx,y}, u_{jx}\}$ (or equivalently, the $\{m_{ijxy}, n_{jx}\}$), are jointly sufficient for estimating the parameters of the model.

8.3 Estimates from a Triple Catch Experiment

Although it was not possible to derive estimates for the general n-sample experiment of section 8.2, it was noted that the relations for the estimates given by equations 8.2.5 - 8.2.8 could be deduced directly from moment equations: that is, setting observations equal to their expectations. It therefore seems reasonable that such moment equations, if they will yield estimates, should give estimates that are similar to those that might be derived by maximizing the expression L (8.2.4).

The moment equations can be manipulated to give estimates for the triple catch experiment (i.e. $n = 3$). We considered the simplified case, however, where no births and no "losses on capture" occur.

It is useful to have the complete expression for L for the particular case discussed here, as this facilitates writing down the expectations and in examining the variances and covariances of the observations. For $n = 3$, $B_{jA} = 0$ $n_{jx} = s_{jx}$ $\forall$ j,x; 8.2.4 becomes:

$.B(n_{1A};\ N_{1A},\ p_{1A}).B(n_{1B},\ N_{1B},\ p_{1B})$

$.T(M_{12AA},\ M_{12AB};\ n_{1A},\ \mathbb{Q}_{1AA},\ \mathbb{Q}_{1AB})$
$.T(M_{12BA},\ M_{12BB};\ n_{1B},\ \mathbb{Q}_{1BA},\ \mathbb{Q}_{1BB})$
$.PT(U_{2A},\ U_{2B};\ N_{1A} - n_{1A},\ \mathbb{Q}_{1AA},\ \mathbb{Q}_{1AB},\ N_{1B} - n_{1B},\ \mathbb{Q}_{1BA},\ \mathbb{Q}_{1BB})$

$.B(m_{12AA};\ M_{12AA},\ p_{2A}).B(m_{12AB};\ M_{12AB},\ p_{2B})$
$.B(m_{12BA};\ M_{12BA},\ p_{2A}).B(m_{12BB};\ M_{12BB},\ p_{2B})$
$.B(u_{2A};\ U_{2A},\ p_{2A}).B(u_{2B};\ U_{2B},\ p_{2B})$

$$.T(M_{23AA}, M_{23AB}; n_{2A}, Q_{2AA}, Q_{2AB})$$

$$.T(M_{23BB}, M_{23BA}; n_{2B}, Q_{2BB}, Q_{2BA})$$

$$.PT(M_{13AA}, M_{13AB}; M_{12AA} - m_{12AA}, Q_{2AA}, Q_{2AB}, M_{12AB} - m_{12AB}, Q_{2BA}, Q_{2BB})$$

$$.PT(M_{13BB}, M_{13BA}; M_{12BB} - m_{12BB}, Q_{2BB}, Q_{2BA}, M_{12BA} - m_{12BA}, Q_{2AB}, Q_{2AA})$$

$$.PT(U_{3A}, U_{3B}; U_{2A} - u_{2A}, Q_{2AA}, Q_{2AB}, U_{2B} - u_{2B}, Q_{2BA}, Q_{2BB})$$

$$.B(m_{23AA}; M_{23AA}, p_{3A}).B(m_{23AB}; M_{23AB}, p_{3B})$$

$$.B(m_{23BA}; M_{23BA}, p_{3A}).B(m_{23BB}; M_{23BB}, p_{3B})$$

$$.B(m_{13AA}; M_{13AA}, p_{3A}).B(m_{13AB}; M_{23AB}, p_{3B})$$

$$.B(m_{13BA}; M_{13BA}, p_{3A}).B(m_{13BB}; M_{23BB}, p_{3B})$$

$$.B(u_{3A}; U_{3A}, p_{3A}).B(u_{3B}; U_3, p_{3B}) \qquad (8.3.1)$$

Equation 8.3.1 thus gives the following joint distribution:

Prob(sample at t = 1)
.Prob(structure at the time of the 2nd sample)
.Prob(sample at t = 2)
.Prob(structure at the time of the 3rd sample)
.Prob(sample at t = 3)

and the distribution of any of the unobservable random variables, or the unconditional distribution of any of the observations can be deduced from this expression. (For example, the joint distribution of N_{2A}, N_{2B} is given by the distribution of the sum of groups which make up the structure at the time of the second sample; the distribution is clearly

$$P(M_{12AA} + M_{12BA} + U_{2A}, M_{12AB} + M_{12BB} + U_{2B})$$

$$= PT(N_{2A}, N_{2B}; N_{1A}, Q_{1AA}, Q_{1AB}, N_{1B}, Q_{1BA}, Q_{1BB}))$$

The following (conditional) expectations can also be deduced:

$$n_{1A} = N_{1A}p_{1A} \tag{8.3.2}$$

$$n_{1B} = N_{1B}p_{1B} \tag{8.3.3}$$

$$N_{2A} = N_{1A}Q_{1AA} + N_{1B}Q_{1BA} \tag{8.3.4}$$

$$N_{2B} = N_{1A}Q_{1AB} + N_{1B}Q_{1BB} \tag{8.3.5}$$

$$n_{2A} = N_{2A}p_{2A} \tag{8.3.6}$$

$$n_{2B} = N_{2B}p_{2B} \tag{8.3.7}$$

$$m_{12AA} = n_{1A}Q_{1AA}p_{2A} \tag{8.3.8a}$$

$$= N_{1A}p_{1A}Q_{1AA}p_{2A} \tag{8.3.8b}$$

$$m_{12BA} = n_{1B}Q_{1BA}p_{2A} \tag{8.3.9a}$$

$$= N_{1A}p_{1B}Q_{1BA}p_{2A} \tag{8.3.9b}$$

$$m_{12AB} = n_{1A}Q_{1AB}p_{2B} \tag{8.3.10a}$$

$$= N_{1A}p_{1A}Q_{1AB}p_{2B} \tag{8.3.10b}$$

$$m_{12BB} = n_{1B}Q_{1BB}p_{2B} \tag{8.3.11a}$$

$$= N_{1B}p_{1B}Q_{1BB}p_{2B} \tag{8.3.11b}$$

$$m_{23AA} = n_{2A}Q_{2AA}p_{3A} \tag{8.3.12}$$

$$m_{23BA} = n_{2B}Q_{2BA}p_{3A} \tag{8.3.13}$$

$$m_{23AB} = n_{2A}Q_{2AB}p_{3B} \tag{8.3.14}$$

$$m_{23BB} = n_{2B}Q_{2BB}p_{3B} \tag{8.3.15}$$

$$m_{13AA} = n_{1A}Q_{1AA}(1-p_{2A})Q_{2AA}p_{3A} + n_{1A}Q_{1AB}(1-p_{2B})Q_{2BA}p_{3A}$$

$$m_{13BA} = n_{1B}Q_{1BA}(1-p_{2A})Q_{2AA}p_{3A} + n_{1B}Q_{1BB}(1-p_{2B})Q_{2BA}p_{3A}$$

$$m_{13AB} = n_{1A}Q_{1AB}(1-p_{2B})Q_{2BB}p_{3B} + n_{1A}Q_{1AA}(1-p_{2A})Q_{2AB}p_{3B}$$

$$m_{13BB} = n_{1B}Q_{1BB}(1-p_{2B})Q_{2BB}p_{3B} + n_{1B}Q_{1BA}(1-p_{2A})Q_{2AB}p_{3B}$$

These equations can then be manipulated to derive moment estimates for the parameters, as follows:

Use 8.3.4 in 8.3.6 to give

$$n_{2A} = N_{1A}Q_{1AA}p_{2A} + N_{1B}Q_{1BA}p_{2A}$$

$$= m_{12AA}/p_{1A} + m_{12BA}/p_{1B}$$

(using 8.3.8b and 8.3.9b)

similarly, using 8.3.5 in 8.3.7, and then 8.3.10a and 8.3.10b to give

$$n_{2B} = m_{12AB}/p_{1A} + m_{12BB}/p_{1B}$$

This pair of equations can then be solved to give estimates $\hat{p}_{1A}$ and $\hat{p}_{1B}$, and hence, $\hat{N}_{1A}$ and $\hat{N}_{1B}$.

Next, the last four equations (for $m_{13AA} \dots m_{13BB}$) can be expanded so that each is the sum of four terms. Then using 8.3.12 - 8.3.15, replace $Q_{2AA}p_{3A}$, $Q_{2BA}p_{3A}$, $Q_{2AB}p_{3B}$, $Q_{2BB}p_{3B}$ by m_{23AA}/n_{2A}, m_{23BA}/n_{2B}, m_{23AB}/n_{2A}, m_{23BB}/n_{2B} respectively. Similarly, using 8.3.8a - 8.3.11a, replace $Q_{1AA}p_{2A}$, $Q_{1BA}p_{2A}$ etc., by m_{12AA}/n_{1A}, m_{12BA}/n_{1B} etc. After simplification this leaves the following two equations.

$$n_{2A}n_{2B}m_{13AA} + n_{2B}m_{12AA}m_{23AA} + n_{2A}m_{12AB}m_{23BA}$$

$$= n_{1A}n_{2B}m_{23AA}Q_{1AA} + n_{1A}n_{2A}m_{23BA}Q_{1AB}$$

$$n_{2A}n_{2B}m_{13AB} + n_{2A}m_{12AB}m_{23BB} + n_{2B}m_{12AA}m_{23AB}$$

$$= n_{1A}n_{2B}m_{23AB}Q_{1AA} + n_{1A}n_{2A}m_{23BB}Q_{1AB}$$

and a similar pair of equations (in fact, the same equations with A replaced by B and vice versa) for Q_{1BB} and Q_{1BA}.

The entire set of estimates can now be given.

Let $|M_{ij}|$ be the determinant

$$\begin{vmatrix} m_{ijAA} & m_{ijAB} \\ m_{ijBA} & m_{ijBB} \end{vmatrix}$$

then

$$\hat{p}_{1A} = |M_{12}| / \begin{vmatrix} n_{2A} & m_{12BA} \\ n_{2B} & m_{12BB} \end{vmatrix}$$

$$\hat{p}_{1B} = |M_{12}| / \begin{vmatrix} m_{12AA} & n_{2A} \\ m_{12AB} & n_{2B} \end{vmatrix}$$

$$\hat{N}_{1A} = n_{1A}/\hat{p}_{1A}$$

$$\hat{N}_{1B} = n_{1B}/\hat{p}_{1B}$$

$$\hat{Q}_{1AA} = (n_{2A}\begin{vmatrix} m_{13AA} & m_{23BA} \\ m_{13AB} & m_{23BB} \end{vmatrix} + m_{12AA}|M_{23}|)/(n_{1A}|M_{23}|)$$

$$\hat{Q}_{1AB} = (n_{2B}\begin{vmatrix} m_{23AA} & m_{13AA} \\ m_{23AB} & m_{13AB} \end{vmatrix} + m_{12AB}|M_{23}|)/(n_{1A}|M_{23}|)$$

$\hat{Q}_{1BA}$ = expression for $\hat{Q}_{1AB}$ with A and B interchanged

$\hat{Q}_{1BB}$ = expression for $\hat{Q}_{1AA}$ with A and B interchanged

$$\hat{N}_{2A} = \hat{N}_{1A}\hat{Q}_{1AA} + \hat{N}_{1B}\hat{Q}_{1BA}$$

$$\hat{N}_{2B} = \hat{N}_{1A}\hat{Q}_{1AB} + \hat{N}_{1B}\hat{Q}_{1BB}$$

Estimates of the Q_{2xy} are not available from this experiment, nor are estimates for N_{3A}, N_{3B} or p_{3A}, p_{3B}; however

$$\widehat{Q_{2AA}p}_{3A} = m_{23AA}/n_{2A}$$
$$\widehat{Q_{2BA}p}_{3A} = m_{23BA}/n_{2B}$$
$$\widehat{Q_{2AB}p}_{3B} = m_{23AB}/n_{2A}$$
$$\widehat{Q_{2BB}p}_{3B} = m_{23BB}/n_{2B}$$

8.3.1 Accuracy and Precision of the estimates.

A proof that these estimates maximize the expression (8.3.1) was not attempted. Nor were expressions for the asymptotic variances and covariances of the estimates derived. This latter problem is tedious but not difficult: the variances and covariances of the $\{m_{ijxy}, n_{jx}\}$ must be determined. The standard delta technique (cf. Darroch (1959) or Seber (1962) Appendix 1) can then be used to derive the asymptotic variances and covariances of the estimates. Derivation of the variances and covariances of the $\{m_{ijxy}, n_{jx}\}$ is done by repeated application of rules for conditional variance and covariance of the variables, all of which have marginal distributions which are either of the binomial form, or the sum of two or more binomials. The parameters of the appropriate conditional distribution are easily determined by examining the expression 8.3.1. For example, m_{23AA} has a binomial (marginal) distribution conditional on n_{2A} with parameters $B(n_{2A}, Q_{2AA}p_{2A})$. n_{2A} in turn has a marginal distribution which is the sum of two independent binomial distributions with parameters $B(N_{1A}, Q_{1AA}p_{2A})$

and $B(N_{1B}, Q_{1BA}p_{2A})$. Thus, the unconditional variance of m_{23AA} is

$$\begin{aligned}&\mathrm{var}(m_{23AA}|N_{1A}, N_{1B})\\ &= N_{1A}Q_{2AA}p_{3A}Q_{1AA}p_{2A}(1-Q_{2AA}p_{3A}Q_{1AA}p_{2A})\\ &+ N_{1B}Q_{2AA}p_{3A}Q_{1BA}p_{2A}(1-Q_{2AA}p_{3A}Q_{1BA}p_{2A})\end{aligned}$$

Similarly m_{23AA} and m_{23AB} have a joint trinomial distribution conditional on n_{2A}, so

$$\begin{aligned}&\mathrm{cov}(m_{23AA}m_{23AB}|N_{1A}N_{1B})\\ &= -N_{1A}Q_{2AA}p_{3A}Q_{1AA}p_{2A}Q_{2AB}p_{3B}Q_{1AA}p_{2A}\\ &-N_{1B}Q_{2AA}p_{3A}Q_{1BA}p_{2A}Q_{2AB}p_{3B}Q_{1BA}p_{2A}\end{aligned}$$

However there are 171 such expressions to be determined and none of the covariances appear to be zero!

To determine if these estimates were reasonably unbiased, and to discover how intensively the experimenter must sample to get adequate precision, the experiment was simulated on the computer and estimates were formed from the simulated data. The results of this study are given in tables 8.1 - 8.3. A flow chart for the method of simulation is given in figure 8.1.

8.3.2 Explanation of Tables and Simulation.

The basis of the simulation is made clear by the flow chart, figure 8.1. In the main part of the simulation a stochastic realization of a single capture history is generated. This capture history is recorded as M(a, b, c) where

M(1, 1, 1) => never captured

M(1, 2, 1) => captured in A at t = 2

M(1, 3, 2) => captured in B at t = 2 and in A at t = 3.

etc.

After $N_{1A} + N_{1B}$ capture histories have been generated, the appropriate sums are calculated to give the statistics $\{m_{ijxy}, n_{jx}\}$, from which estimates are calculated. This "experiment" was then simulated several hundred times to give an idea of the variances of the estimates. Several such sets of simulations were performed, using a different set of parameters for each set of simulations. The results of four such sets of simulations are reported in tables 8.1 - 8.3.

Table 8.1 simply lists the parameter values used in each set of simulations. To maintain comparability, migration and survival rates were kept constant in all sets. In the first set a small population was considered, but with high sampling intensities. In set two, the same population sizes were used but combined with a more moderate sampling intensity. In sets 3 and 4 the population sizes were increased by a factor of 2 and 4, and the same fairly low sampling intensities were used. The survival rates were chosen more or less haphazardly for these four sets of simulations. It happens that the probability that an individual, alive at time i, survives to time i+1 is the same regardless of the location of the individual at time i. That is, $\varphi_{1A} = \varphi_{1B} = .925$ and $\varphi_{2A} = \varphi_{2B} = .850$. Also, the conditional probability that an individual is in the same population at $t = i$ and $t = i+1$, given that it survives, is approximately the same for i and $i+1$. That is, $\phi_{1BB}/\phi_{1B} \doteqdot \phi_{2BB}/\phi_{2B}$, $\phi_{1AA}/\phi_{1A} \doteqdot \phi_{2AA}/\phi_{2A}$. These restrictions are not necessary for the development of the estimates, but were imposed to reflect what might be a realistic situation (namely, a migration system near equilibrium. The decreased survival rate between $t = 1, 2$ and $t = 2, 3$ would result by allowing more time to elapse between sample 2 and 3 than between 1 and 2). Several further sets of simulations were carried out where these restrictions did not prevail, and results consistent with those for these four sets were noted. In particular, all such simulations supported the 'rule of thumb' proposed at the end of this chapter.

It was expected that the precision of the estimates would be related to the expected values of the $\{m_{ijxy}, n_{jx}, N_{jx}\}$. Therefore, in table 8.2 we list these expected values. It was found that when the size of the populations was small, precision was poor even at very high sampling intensities. Thus the expected values of the N_{jx} must be considered when relating precision to the choice of parameters, since a very low survival rate may reduce the population to such a low level that very poor estimates result. If approximately constant effort is expended by the experimenter at each sampling occasion (so that the p_{jx} stay roughly constant for all j) then sharp decreases in the n_{jx} for increasing j will indicate this low survival rate. Such information could be used to improve the estimates, since the estimates of $\phi_{12AA}p_{2A}$ etc., are much more precise than the separate estimates for ϕ_{12AA} and p_{2A} etc. However, in this case the model and estimates should be re-derived to incorporate effort explicitly.

The results of the simulations are summarized by giving means and standard deviations for the estimates over 500 simulations for each set. The estimate of N_{1A} and N_{1B} was not formed, because at low sampling intensities $\hat{p}_{1A}$ and $\hat{p}_{1B}$ were frequently zero. By estimating $1/N_{1A}$ one avoids this problem of infinite bias. Very frequently (particularly in set 2) inadmissable estimates of probabilities (negative or greater than 1) resulted. Normally, the experimenter would reject such estimates, or take 0 or 1 as the estimate. For the purpose of the simulation however, this was not done as this introduces bias and reduces the variance.

The estimates for the second set are thoroughly bad. More often than not, parameter estimates gave inadmissable values, and the variance of the estimates is so large as to render them thoroughly meaningless. Note however, that the expected capture histories are very low. In most of the simulations, one or more of the m_{ijxy} was zero. Nevertheless, the estimates of $\mathbb{Q}_{2AA}p_{3A}$ etc., are quite precise and unbiased. The same was true of the $\mathbb{Q}_{1AA}p_{2A}$ etc., although these figures have not been given in table 8.3.

The best estimates were obtained in set 4 and set 1. The estimates are not percetpibly biased, and the precision is good enough that estimates in these cases can be considered to be trustworthy. Note that the expectations of the m_{ijxy} are all reasonably large; all greater than 5 and most are greater than 10. In none of the simulations did a zero observation occur.

It seems therefore that in general, the precision is not merely dependent upon sampling intensity, but improves with increasing population sizes. It seems that the m_{ijxy} should <u>all</u> have expectations of 5 or more if good estimates are to result. An ad hoc rule of thumb (based on these, and several further sets of simulations, not reported here) seems to be that if all the observed m_{ijxy} are greater than 5, reasonably good estimates are obtained.

Table 8.1

Parameter Values Used in Simulations of the Triple-Catch Experiment.

Parameter	Value in Simulation Set			
	1	2	3	4
N_{1A}	250	250	250	1000
$(1/N_{1A})$	.004	.004	.002	.001
p_{1A}	.60	.20	.20	.20
N_{1B}	200	200	400	800
$(1/N_{1B})$	.005	.005	.0025	.00125
p_{1B}	.55	.15	.15	.15
Q_{1AA}		.773	for all sets	
Q_{1AB}		.152		
Q_{1BA}		.303		
Q_{1BB}		.622		
p_{2A}	.65	.25	.25	.25
p_{2B}	.60	.20	.20	.20
Q_{2AA}		.738	for all sets	
Q_{2AB}		.112		
Q_{2BA}		.280		
Q_{2BB}		.570		
p_{3A}	.70	.30	.30	.30
p_{3B}	.65	.25	.25	.25

Table 8.2

Expected Population Sizes and Capture Histories (for explanation and discussion, see text; for parameter values, see table 8.1)

	Simulation Set			
	1	2	3	4
N_{1A}	250.0	250.0	500.0	1000.0
n_{1A}	150.0	50.0	100.0	200.0
N_{1B}	200.0	200.0	400.0	800.0
n_{1B}	110.0	30.0	60.0	120.0
N_{2A}	253.85	253.85	507.70	1015.40
m_{12AA}	75.37	9.66	19.32	38.65
m_{12BA}	21.66	2.27	4.54	9.09
n_{2A}	165.0	63.46	126.92	253.85
N_{2B}	162.40	162.40	324.80	649.60
m_{12BB}	41.05	3.73	7.46	14.93
m_{12AB}	13.68	1.52	3.04	6.08
n_{2B}	97.44	32.48	64.96	129.92
N_{3A}	232.81	232.81	465.62	931.24
m_{13AA}	22.75	6.93	13.86	27.71
m_{13BA}	11.39	2.73	5.53	11.05
m_{23AA}	85.24	14.05	28.10	56.20
m_{23BA}	19.10	2.73	5.46	10.91
n_{3A}	162.97	69.84	139.69	279.38

Table 8.2 continued ...

Expected Population Sizes and Capture Histories
(for explanation and discussion, see text; for
parameter values, see table 8.1)

	Simulation Set			
	1	2	3	4
N_{3B}	121.0	121.0	242.0	484.0
m_{13BB}	10.99	2.31	4.64	9.27
m_{13AB}	6.33	1.68	3.36	6.71
m_{23BB}	36.10	4.63	9.26	18.51
m_{23AB}	12.01	1.78	3.55	7.11
n_{3B}	78.65	30.25	60.50	121.0

Table 8.3

Means and S.D. of Parameter Estimates

Over 500 Simulations

(for explanation, see text).

True Value of Parameter	SIMULATION SET							
	1		2		3		4	
	r=1 mean	s=1 s.d.	r=1 mean	s=0 s.d.	r=2 mean	s=0 s.d.	r=4 mean	s=0 s.d.
$1/N_{1A} = 1/250r$	.00404	.00038	.00556	.0144	.00238	.00317	.00112	.00089
$p_{1A} = .4s + .20$	.605	.067	.276	.769	.237	.307	.223	.170
$1/N_{1B} = 1/200r$	.00502	.00060	.00657	.0106	.00446	.0370	.00131	.00042
$p_{1B} = .4s + .15$	.555	.074	.198	.336	.255	.196	.156	.052
$Q_{1AA} = .773$	.773	.778	.072	.728	1.48	.252	.773	.158
$Q_{1AB} = .152$	.153	.063	.285	2.19	.136	.260	.151	.135
$Q_{1A} = .925$	.931	.050	1.013	.924	.927	.204	.924	.127
$Q_{1BB} = .738$	.627	.100	.938	4.35	.708	.471	.657	.260
$Q_{1BA} = .112$	.300	.090	.284	2.32	.261	.325	.282	.193
$Q_{1B} = .925$	.927	.062	.966	2.56	.970	.295	.939	.176
$N_{2A}^* = 253.85r$ (9.27r)	254.5	27.25	187.7	757.9	499.5	184.4	998.4	230.1
$p_{2A} = .4s + .25$	.653	.073	.305	.912	.302	.469	.272	.084

Table 8.3 continued ...

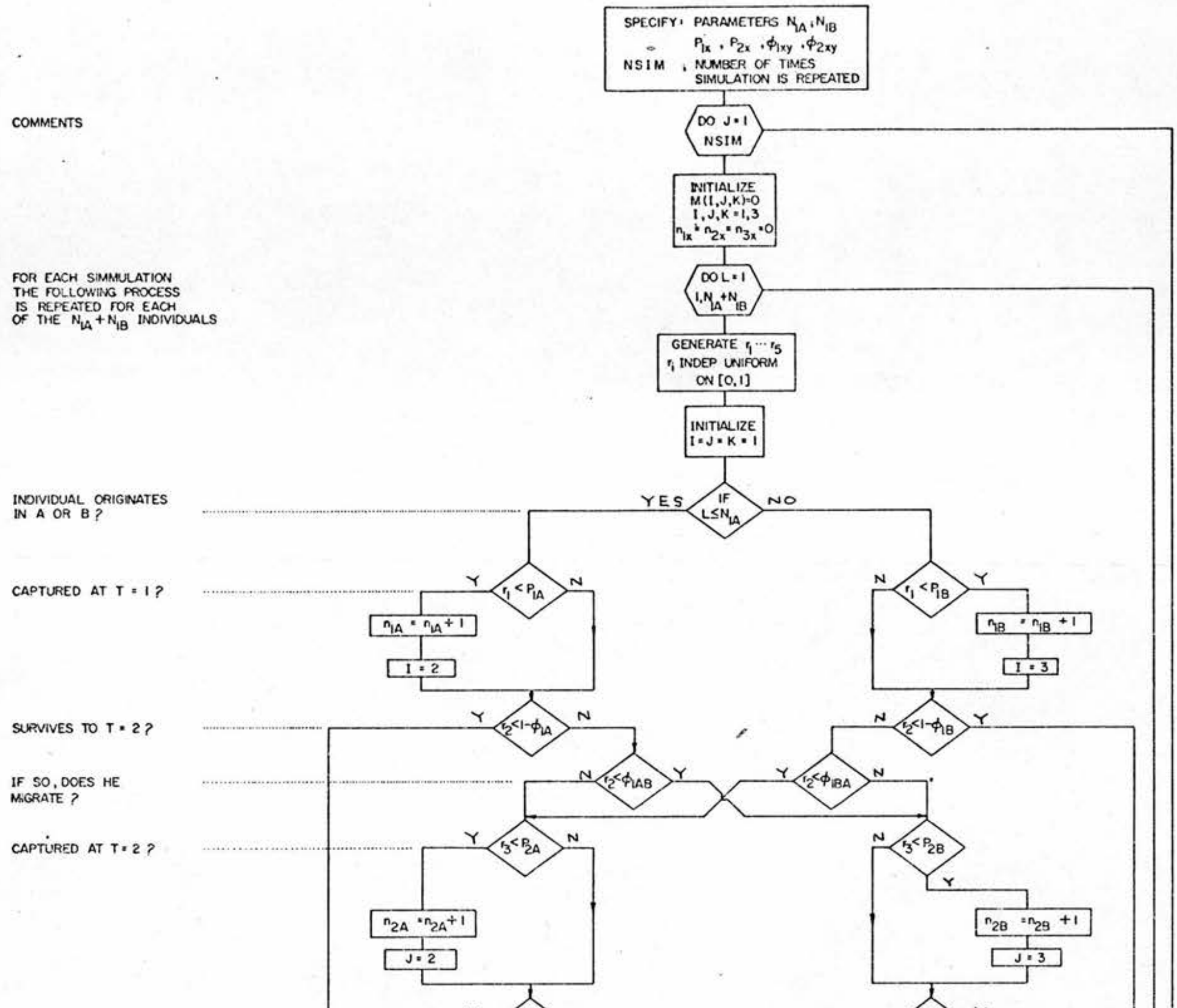
Means and S.D. of Parameter Estimates
Over 500 Simulations
(for explanation, see text).

True Value of Parameter			SIMULATION SET							
			1		2		3		4	
			r=1	s=1	r=1	s=0	r=2	s=0	r=4	s=0
			mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
N_{2B}* = 162.40r (8.82r)			163.6	26.36	258.8	1291.0	351.6	235.7	676.8	264.4
P_{2B} = .4s + .20			.611	.099	.286	1.837	.219	.569	.247	.356
	s=1	s=0								
$Q_{2AA}P_{3A}$ =	.5166	.2214	.515	.039	.221	.048	.223	.036	.223	.026
$Q_{2AB}P_{3B}$ =	.0728	.0280	.071	.021	.028	.022	.029	.017	.028	.012
$Q_{2BB}P_{3B}$ =	.3205	.1425	.371	.050	.140	.061	.141	.046	.143	.030
$Q_{2BA}P_{3A}$ =	.1960	.0840	.195	.041	.084	.049	.084	.035	.084	.025

* Expected Population Sizes.

The standard deviation of this (unobservable) random variable is given in brackets.

FIGURE 8-1
SIMULATION OF TRIPLE CATCH EXPERIMENT
FLOW CHART



Bibliography

BIBLIOGRAPHY

1. Adke, S.R., (1969): A Birth-Death and Migration Process. J. Appl. Prob. 6 p.687.

2. Anscombe, F.J., (1948): The Transformation of Poisson, Binomial and Negative-Binomial Data. Biometrika 35 p.246.

3. Armitage, P., (1952): The Statistical Theory of Bacterial Populations Subject to Mutation. J. R. Statist. Soc. B 14 p.1.

4. Bailey, N.T.J., (1951): On Estimating the Size of Mobile Populations from Recapture Data. Biometrika 38 p.293.

5. Bailey, N.T.J., (1964): The Elements of Stochastic Processes with Application to the Natural Sciences. 1st edition. John Wiley and Sons.

6. Bailey, N.T.J., (1968): Stochastic Birth Death and Migration Processes for Spatially Distributed Populations. Biometrika 55 p.189.

7. Barnard, G.A., G.M. Jenkins, and C.B. Winsten, (1962): Likelihood Inference and Time Series. J.R.Stat. Soc. A 125 p.321.

8. Bartlett, M.S., (1949): Some Evolutional Stochastic Processes. J.R. Statist. Soc. B 11 p.211.

9. Bartlett, M.S., (1955): An Introduction to Stochastic Processes. Cambridge University Press.

10. Bartlett, M.S., (1956): Deterministic and Stochastic Models for Recurrent Epidemics. Proc. Third Berkeley Symposium. 4 p.81.

11. Bateman, H., (1910): The Solution of a System of Differential Equations Occurring in the Theory of Radio-active Transformations. Proc. Cambridge Phil. Soc. 15 p.423.

12. Bather, J.A., (1963): Two Non-Linear Birth and Death Processes. J. Australian Math. Soc. 3 p.104.

13. Beverton, R.J.H., and S.J. Holt, (1957): On the Dynamics of Exploited Fish Populations. Fisheries Investigations, Series II, volume XIX, H.M. Stationery Office, London.

14. Bharucha-Reid, A.T., (1960): Elements of the Theory of Markov Processes and their Applications. 1st edition. McGraw-Hill.

15. Billingsley, P., (1961): Statistical Inference for Markov Processes. Statistical Research Monographs. Volume 2, University of Chicago Press.

16. Box, M.J., D. Davies and W.H. Swann, (1969): Non-Linear Optimization Techniques. I.C.I. Monographs 5. Oliver and Boyd.

17. Chapman, D.G., and C.O. Junge, (1956): The Estimation of the Size of a Stratified Animal Population. Ann. Math. Stat. 27 p.375.

18. Cormack, R.M., (1968): The Statistics of Capture-Recapture Methods. Oceanogr. Mar. Biol. Ann. Rev. 6 p.455.

19. Cox, D.R. and H.D. Miller, (1965): The Theory of Stochastic Processes. 1st edition. Methuen.

20. Darroch, J.N.,(1958): The Multiple Recapture Census I. Biometrika 46 p.336.

21. Darroch, J.N., (1959): The Multiple Recapture Census II. Biometrika 45 p.343.

22. Darroch, J.N., (1961): The Two Sample Capture-Recapture when Tagging and Sampling are Stratified. Biometrika 48 p.241.

23. David, F.N., and J. Newman, (1938): Extension of the Markoff Theorem on Least Squares. Statistical Research Memoirs 2 p.105.

24. Dempster, J.P., (1957): The population Dynamics of the Moroccan Locust in Cyprus. Anti-Locust Bulletin no. 27.

25. Dietz, K. and F. Downton, (1968): Carrier-borne Epidemics with Immigration. I - Immigration of Both Susceptibles and Carriers. J. Appl. Prob. 5 p.31.

26. Doob, J.L., (1953): Stochastic Processes. John Wiley and Sons. New York.

27. Draper, N.R., and H. Smith, (1966): Applied Regression Analysis. John Wiley and Sons. New York.

28. Ehrenfest, P. and T., (1907): Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. Physikalische Zeitschrift, 8 p.311.

29. Feller, W., (1957): An Introduction to Probability Theory and Its Applications. Vol. I, 2nd edition. Wiley International.

30. Finkbeiner, D.T., (1960): Introduction to Matrices and Linear Transformations. 1st edition. W.H. Freeman and Co.

31. Handbook of Mathematical Functions. M. Abramovitz and I.A. Stegun, eds. National Bureau of Standards, Applied Mathematics Series, No. 55, 1964.

32. Ishida, K., (1960): The Stochastic Model for Unimolecular Gas Reaction. Bull. Chem. Soc. Japan 33, p.1030.

33. Iwao, S., (1963): On A Method for Estimating the Rate of Population Interchange between Two Areas. Res. Popul. Ecol., 5 p.44.

34. Iwao, S., K. Kiritani, and N. Hokyo, (1966): Application of a Marking and Recapture Method for the Analysis of Larval Adult Populations of an Insect. Res. Popul. Ecol. 8 p.147.

35. Jackson, C.H.N., (1940): The Analysis of a Tsetse-Fly Population. Ann. Eugenics 10 p.337.

36. Jackson, C.H.N., (1948): An Analysis of a Tsetse-Fly Population III. Ann. Eugenics 14 p.91.

37. Jackson, J.R., (1957): Networks of Waiting Lines. Operat. Res. 5 p.518.

38. Jolly, G.M., (1965): Explicit Estimates from Capture-Recapture Data with Both Death and Immigration. Biometrika 52 p.225.

39. Kac, M., (1947): Random Walk and the Theory of Brownian Motion. Am. Math. Monthly, 54 p.369.

40. Karlin, S., and J.L. McGregor (1957a): The Differential Equations of Birth and Death Processes and the Stieltjes Moment Problem. Trans. Amer. Math. Soc. 85 p.489.

41. Karlin, S., and J.L. McGregor, (1957b): The Classification of Birth and Death Processes. Trans. Amer. Math. Soc. 86 p.366.

42. Karlin, S., and J.L. McGregor, (1961): Occupation Time Laws for Birth and Death Processes. Proc. of the Fourth Berkeley Symposium 2 p.249.

43. Kermack, W.O., and A.G. McKendrick, (1933): Contributions to the Mathematical Theory of Epidemics. Proc. Roy. Soc. A 141 p.94.

44. Kingman, J.F.C., (1969): Markov Population Processes. J. Appl. Prob. 6 p.1.

45. Lederman, W., and G.E.H. Reuter, (1954): Spectral Theory for the Differential Equations of Simple Birth and Death Processes. Phil. Trans. Roy. Soc. London A 246 p.321.

46. Lefkovitch, L.P., (1965): The Study of Population Growth in Organisms Grouped by Stages. Biometrics 21 p.1.

47. Leslie, P.H., (1945): On the Use of Matrices in Certain Population Mathematics. Biometrika 33 p.183.

48. Leslie, P.H., (1948): Some Further Notes on the Use of Matrices in Population Mathematics. Biometrika 35 p.213.

49. Leslie, P.H., (1952): The Estimation of Population Parameters II. The Estimation of Total Numbers. Biometrika 39 p.363.

50. Matis, J.H., and H.O. Hartley, (1970): Stochastic Compartmental Analysis: Model and Least Squares Estimation from Time Series Data. Mimeographed

publication (pre-print of a paper read at the 7th International Biometric Conference, Hanover, August 1970.)

51. McQuarrie, D.A., (1967): Stochastic Approach to Chemical Kinetics. J. Appl. Prob. 4 p.413.

52. Moran, P.A.P., (1958): Random Processes in Genetics. Proc. Camb. Philos. Soc. 54 p.60.

53. Nissen-Meyer, S., (1966): Analysis of Antibiotics on Bacteria by Means of Stochastic Models. Biometrics 22 p.761.

54. Noble, B., (1964): Numerical Methods Vol. I. Oliver and Boyd.

55. Pollard, J.H., (1966): On the Use of the Direct Matrix Product in Analysing Certain Stochastic Population Models. Biometrika 53 p.397.

56. Puri, P.S., (1968): Interconnected Birth and Death Processes. J. Appl. Prob. 5 p.334.

57. Pyke, R., (1955): The Statistical Theory of Some Population Migration Models. M.Sc. Thesis, University of Washington.

58. Reuter, G.E.H., (1961): Competition Processes. Proc. of the Fourth Berkeley Symposium 2 p.421.

59. Schaeffer, M.B., (1951): Estimation of the Size of Animal Populations by Marking Experiments. U.S. Fish and Wildlife Service Fisheries Bull. 69 p.191.

60. Seber, G.A.F., (1962): The Multi-Sample Single Recapture Census. Biometrika 49 p.339.

61. Seber, G.A.F., (1965): A Note on the Multiple-Recapture Census. Biometrika 52 p.249.

62. Shiyomi, M., (1967): A Model of the Plant-to-Plant Movement of Aphids I. Description of the Model. Res. Popul. Ecology 9 p.53.

63. Shiyomi, M., (1968): A Model of the Plant-to-Plant Movement of Aphids II. Estimation of Parameters and the Application. Res. Popul. Ecology 10 p.105.

64. Skellam, J.G., (1951): Random Dispersal in Theoretical Populations. Biometrika 38 p.196.

65. Snedecor, G.W., and W.G. Cochran, (1967): Statistical Methods. 6th edition. Iowa State University Press.

66. Soper, H.E., (1929): Interpretation of Periodicity in Disease-Prevalence. J.R. Statist. Soc. 92 p.34.

67. Sprott, D.A., and J.D. Kalbfleisch, (1969): Examples of Likelihoods and Comparison with Point Estimates and Large Sample Approximations. J. Amer. Stat. Ass'n. 64.

68. Sprott, D.A., and J.G. Kalbfleisch, (1965): Use of the Likelihood Function in Inference. Psychol. Bull. 64 p.15.

69. Turner, M.E., R.J. Munroe and L.D. Homer, (1963): Generalized Kinetic Regression Analysis: Hypergeometric Kinetics. Biometrics 19 p.406.

70. Usher, M.B., and M.H. Williamson, (1970): A Deterministic Matrix Model for Handling the Birth, Death and Migration Processes of Spatially Distributed Populations. Biometrics 26 p.1.

71. Whittle, P., (1967): Non-linear Migration Processes. Proc. 36th Session of the I.S.I. 1 p.

72. Whittle, P., (1968): Equilibrium Distributions for an Open Migration Process. J. Appl. Prob. 5 p.567.