INFERENCE ON A GENETIC MODEL

by

John J. Bartko, B.A., M.S.

Thesis submitted to the Graduate Faculty

of the Virginia Polytechnic Institute in

candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

May, 1962

Blacksburg, Virginia

To my parents

Pre mojich rodičov

TABLE OF CONTENTS

I. INTRODUCTION AND SUMMARY

In this thesis, methods of designing experiments and the interpretation of their results will be investigated in connection with a population genetic model introduced by Moran (1958). The deductive theory by approximate methods of such models has reached an advanced stage, but very little has been done along the line of statistical inference. Moran's model is a model of the Markov chain type. A significant amount of the work in this thesis deals with a Markov chain of the absorbing type. In particular, statistical inference for absorbing Markov chains is virtually non-existent. We quote Billingsley (1961a) "A systematic investigation of inference in such cases would be valuable." Snell (personal communication) states, "My own feeling is that the really useful things in this area have yet to be studied." Thus it is evident that more research in this field is needed and that a broad vista of investigation is available. In Moran's model the most severe assumption is that the number of individuals in the population at any time is a constant (usually denoted by M). Though this restriction may have an unappealing tone, Moran's model was selected for investigation because it is the only finite population genetic model for which the deductive theory by exact methods is well enough established to stimulate an investigation of statistical inference. The only reference known where the assumption of constant population size, for this model, is dropped is an article by Feller (1951) in which the problem is mentioned. Feller gives the form of the diffusion equation approximating the exact, discrete process but no attempt at solution is made. Thus it is hoped that this thesis will be a step towards the opening of a virtually uninvestigated field of statistical inference in population genetic models and that it will serve to illustrate the area of deductive theory needed to handle inference problems on such models.

A model in population genetics is a probability description of how genes pass from one individual (or generation) to the next, and may include such influences as mutation, selection, overlapping or non-overlapping generations and non-random mating. A brief description of these concepts follows. We shall, in this thesis, be primarily concerned with the influence of mutation.

The genetic factor with which we are concerned is of the simplest type. We assume it to be controlled by a single locus on a chromosome, at which either of two alleles "a" or A can occur. Clearly an allele is an alternative member of a pair of genes. The genotypes (genetic constitution; phenotype refers to physical description) will be haploid, that is, "a" or A as opposed to diploid individuals, aa , Aa, or AA . Haploidy is not an uncommon occurrence in nature. In the honeybee, unfertilized eggs may develop by parthenogenesis, in which case males (drones) are produced. These males are haploid. Haploidy is also found in wasps, ants, salamanders, mosses, ferns and molds.

A <u>mutation</u> is a rare instantaneous transition from one gene into its allele, say $a \rightarrow A$ or $A \rightarrow a$. Some of the designs in this thesis will involve the use of mutagents, that is, mutation producing agents. A mutation rate in one direction will be estimated while the reverse mutation rate is assumed zero. This is a realistic and practical assumption from the biological point of view. It will also be assumed that mutations occur only among the gametes produced by an individual, so that its own genotype remains unchanged throughout its lifetime. <u>Selection</u> describes non-random induced variation in the average numbers of offspring produced by different genotypes. This variation can be caused either directly by varying the number of gametes produced per genotype or indirectly by varying the life expectations. Generations may be non-overlapping, that is, no mating occurs between them (for example populations with a seasonal life cycle) or generations may be overlapping with births and deaths occurring one at a time. Random-mating or panmixia means that any individual has equal probability of mating with any other individual in the population. Non-random mating, therefore, is the possibility of gametes or zygotes uniting in non-random proportions to form new zygotes. An example of non-random mating is positive assortative mating, (likes with likes) a widely used practice in animal breeding.

The emphasis in this thesis is statistical inference on the mutation rates α_1 and α_2 of Moran's (1958) model, a population genetic model of the Markov chain type. In Chapter II an introduction to Markov chains is given along with a review of known theorems for statistical inference in Markov processes with special reference to maximum likelihood estimation procedures. Chapter III deals with the situation where both mutation rates are estimated. Methods of conducting experiments and interpretation of results are discussed. Chapter IV deals with the extremely important area of absorbing Markov chains. In this chapter one mutation rate α_1 is discussed. Several theorems are postulated for the distribution and properties of the maximum likelihood estimate of this single mutation rate α_1. Methods of conducting experiments and some illustrative examples are presented. Of special interest are results obtained by simulation methods on the IBM 650 which are extremely important in substantiating several of the theoretical discussions. Appendix I is a presentation and discussion of Hahn polynomials which were the building blocks for many of the results of the thesis. Appendix II is a listing of data obtained from the IBM 650 in the simulation study. Appendix III was also used in connection with the simulation study.

II. GENERAL DISCUSSION ON MARKOV CHAINS

2.1 A Brief Introduction to Markov Chains

Finite genetic populations, such as those discussed in this thesis, can have only a finite number of possible genetic states; the number of the various genotypes in the population at any time is limited to being a non-negative integer, and cannot exceed the total population size. A population genetic model can be described by postulating the probabilities that a given state will change to another state during a birth-death event. If the population states are ordered according to some convention, the probabilities can be tabulated as a matrix array called a "transition matrix" and the successive states form a (first order) "Markov chain" because the transition matrix is assumed to depend on the immediately preceding state only. Given the initial state, one can write down the probabilities that the population is in the various states at any subsequent time. A discussion of the above terms follows.

An r-th order Markov chain $\{x(t)\}$ satisfies the following condition:

$$\Pr[x(t) \mid x(t-1), x(t-2), \ldots, x(t-r), x(t-r-1), \ldots] = \Pr[x(t) \mid x(t-1), \ldots, x(t-r)] ,$$

that is, the distribution of $x(t)$ conditional on the whole previous history is the same as the distribution given only the r previous states. As a special case, a first order Markov chain is one for which the distribution is affected by the immediately preceding state only, and we write

$$p_{ij} = \Pr[x(t) = j \mid x(t-1) = i] . \qquad (2.1.1)$$

In this thesis, we shall be using only first order Markov chains with transition probabilities p_{ij} independent of time, and shall refer to these as "Markov chains" without further qualification. We shall assume that changes of state can only occur at integer times $t = 2, 3, \ldots,$ and the possible states are the integer values, $i, j = 0, 1, 2, \ldots, M$. Although somewhat unconventional in stochastic processes we take $t = 1$ as initial time, and the initial state $x(1)$ is assumed fixed and known.

The conditional probability p_{ij} is called the probability of transition from the state i to the state j and $P = (p_{ij})$ the matrix of transition probabilities,

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdot & \cdot & \cdot & p_{0M} \\ p_{10} & p_{11} & p_{12} & \cdot & \cdot & \cdot & p_{1M} \\ \cdot & & & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & & & & & & \\ p_{M0} & & & & & & p_{MM} \end{bmatrix} . \quad (2.1.2)$$

Clearly P is a square matrix of order $M + 1$ with non-negative elements, since $p_{ij} \geq 0$ for all i and j. Row sums are unity, i.e., $\sum_{j=0}^{M} p_{ij} = 1$ for all i. P is simply called the transition matrix.*

We state here that the convention for noting element positions in a square matrix of order $M + 1$ in this thesis is as follows:

$$\begin{bmatrix} 0,0 & 0,1 & 0,2 & \cdot & \cdot & \cdot & 0,M \\ 1,0 & 1,1 & 1,2 & \cdot & \cdot & \cdot & 1,M \\ \cdot & & & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & & & & & & \\ M,0 & M,1 & M,2 & \cdot & \cdot & \cdot & M,M \end{bmatrix} . \quad (2.1.3)$$

---

*A. A. Markov (1856-1922), Russian mathematician, arrived at the notion of Markov chains when he examined the alternation of vowels and consonants in Pushkin's poem "Onegin".

2.2 Some General Notation and Terminology

(a) Discussion. A state in a Markov chain is an absorbing state if it is impossible to leave it. A Markov chain is absorbing if (1) it has at least one absorbing state, and (2) from every state it is possible to go to an absorbing state (not necessarily in one step). For example in Chapter IV the model discussed is one in which state 0 is absorbing and the remaining states 1, 2, ..., M are non-absorbing (transient). Therefore,

$$p_{00} = 1$$

and

$$p_{0j} = 0 \quad , \qquad j = 1, 2, \ldots, M \; .$$

In this thesis Feller's (1957) definition of a transient state will be used. Feller defines a transient state as one for which the probability that the state is visited at least twice is less than one. Broadly speaking this means that it is not certain that a transient state be visited infinitely often. Note that in an absorbing Markov chain we can speak of transient states and non-absorbing states as one and the same. However, in general transient does not imply non-absorbing.

(b) Theorem 2.2.1. In an absorbing Markov chain the probability that the process will be absorbed is one. [Kemeny, Mirkil, Snell, Thompson, (1959)]

(c) A Markov chain is ergodic if the probability distributions $\{P_j(n)\}$, $[P_j(n) = \sum_i P(1) p_{ij}^{(n-1)}]$ always converge to a limiting distribution $\{P_j\}$ which is independent of the initial distribution $\{P_j(1)\}$. That is, when $\lim_{n\to\infty} P_j(n) = P_j$ $(j = 1, 2 \ldots)$.

By stationary probability for state i we mean the probability that the model is in state i irrespective of the initial state k, after many generations have elapsed.

We shall say the process is positively regular iff the transition matrix to the power t, P^t, for some finite t, has all positive (non-zero) elements. The process is called regular if P^t for some finite t has at least one row with all non-zero elements.

(d) Known Results. Extending (2.1.1), we write

$$p_{ij}^{(t)} = \Pr[x(t+\tau) = j \mid x(\tau) = i], \qquad (2.2.1)$$

$$t = 0, 1, 2 \ldots ; \quad \tau = 1, 2 \ldots$$

for the t-step transition probabilities. Then, if P (2.1.2) is the matrix of elements p_{ij}, the elements of

P^t are the t-step transition probabilities, that is $p_{ij}^{(t)}$.

Let λ_j be the j-th eigenvalue of P and $\underline{K}_j$ ($\underline{K}_j$ denotes a column vector, $\underline{K}_j'$ the corresponding row vector) the corresponding post-eigenvector. Then

$$P\underline{K}_j = \lambda_j \underline{K}_j , \qquad j = 0, 1, \ldots, M; \quad (2.2.2)$$

that is, $PK = KD_\lambda$ where K is a matrix of eigenvectors, $K = (\underline{K}_0, \underline{K}_1, \ldots, \underline{K}_M)$ and D_λ is a diagonal matrix whose elements are the eigenvalues λ_j. The columns of K are the post-eigenvectors, the rows of K^{-1} are the pre-eigenvectors and we have $P = KD_\lambda K^{-1}$ or more generally

$$P^t = KD_\lambda^t K^{-1} , \qquad t = 1, 2, \ldots \; . \quad (2.2.3)$$

These results are basic and are used widely throughout the thesis.

2.3 Some Markov Chain Theorems

The following discussion is taken from Billingsley (1961). For convenience, as much as possible of his discussion will be in the notation and wording of this thesis. Moreover, theorems and conditions will be numbered following the convention of this thesis where numbers in parentheses will be those used by Billingsley.

We now establish some notation. Given a set of transition probabilities $p_{ij}(\alpha)$, which depend on unknown parameters $\alpha \in \Omega$ where Ω is the parameter space and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ is a vector of parameters, then the likelihood function can be written

$$L(\alpha) = \prod p_{ij}^{n_{ij}}(\alpha) , \qquad (2.3.1)$$

where n_{ij} is the number of times the transition from state i to state j occurred. The log-likelihood is then

$$\log L(\alpha) = \Sigma\, n_{ij} \log p_{ij}(\alpha) . \qquad (2.3.2)$$

The maximum likelihood equations become

$$\frac{\partial}{\partial \alpha_u} \log L(\alpha) = \sum \frac{n_{ij}}{p_{ij}(\alpha)} \frac{\partial p_{ij}(\alpha)}{\partial \alpha_u} = 0 , \qquad (2.3.3)$$

$$u = 1, 2, \ldots, r .$$

For large n (n is the length of the observed chain; a realization of the chain) we can write the $r \times r$ symmetric information matrix as

$$\mathcal{I} = - \mathcal{E}\left(\frac{\partial^2 \log L(\alpha)}{\partial \alpha_u \, \partial \alpha_v}\right) . \qquad (2.3.4)$$

Condition 2.3.1 (Condition 5.1). The set D of pairs (i, j), for which the transition probabilities $p_{ij}(\alpha) > 0$, is by assumption independent of $\alpha \varepsilon \Omega$. Each $p_{ij}(\alpha)$ has continuous partial derivatives of third order throughout Ω . Moreover the $d \times r$ matrix

$$(\partial p_{ij}(\alpha)/\partial\alpha_u) \qquad (i, j)\varepsilon D , \qquad u = 1, \ldots, r ,$$

(d being the number of elements in D) has rank r throughout Ω . For each $\alpha \varepsilon \Omega$ there is only one ergodic set and there are no transient states. See section 2.2a, c for a discussion of the terms transient and ergodic.

This condition implies that $\mathcal{I}$ (2.3.4) is non-singular. It further implies the following two theorems.

Theorem 2.3.1 (Theorem 2.1). Suppose that Condition 2.3.1 holds. Then there exists a sequence $\{\hat{\alpha}\}$ of random vectors in Ω , each being a function $\hat{\alpha} = \hat{\alpha}(x(1),\ldots,x(n))$ of the observations, such that $\hat{\alpha}$ converges in probability to the true α^0 and such that $\hat{\alpha}$ is a solution of (2.3.3) with probability going to one as $n \to \infty$. Thus there is a consistent maximum likelihood estimator of α^0 . Moreover $\hat{\alpha}$ is a local maximum of (2.3.2) with probability going to one. Finally, if $\bar{\alpha}$ is a second consistent solution of (2.3.3) then the probability that $\hat{\alpha} = \bar{\alpha}$ goes to one as $n \to \infty$.

This theorem [as Billingsley notes] does not take into account certain difficulties which may arise. The conditions imposed on the transition probabilities are local in character and so hence are the results which follow from them. In summary fashion then the theorem states that if N_b is a small neighborhood of α^0 and if n is large, there is, with high probability, exactly one solution $\hat{\alpha}$ of (2.3.3) in N_b and $\log(\hat{\alpha}) \geq \log(\alpha)$ for any $\alpha \in N_b$. Now there may be other solutions of (2.3.3) far removed from α^0; the theorem provides no means of choosing the solution which is near α^0. Further, the solution $\hat{\alpha}$ need not be an absolute maximum of $\log L(\alpha)$. Even so, it is convenient to call $\hat{\alpha}$ the maximum likelihood estimator of α^0 and to write $\log L(\hat{\alpha})$ as though it were an absolute maximum. In Chapter III where the model to which this theorem applies is discussed, it is shown that the above difficulties can be avoided. For the model discussed there the solution of (2.3.3) provides the unique maximum of $\log L(\alpha)$ under some general conditions.

The next theorem provides us with the tools for statistical inference. If the vector $\hat{\alpha}$ is a consistent solution of the maximum likelihood equations (2.3.3), let $\ell(n) = (\ell_1(n), \ldots, \ell_r(n))$ be the random vector with components

$$\ell_u(n) = (\hat{\alpha}_u - \alpha_u^0) \; , \quad u = 1, 2, \ldots, r \; . \qquad (2.3.5)$$

Theorem 2.3.2 (Theorem 2.2). Suppose that Condition 2.3.1 holds. If the vector α^0 is the true value of the parameters and $\hat{\alpha}$ is a consistent solution of the maximum likelihood equations (2.3.3), then for $n \to \infty$

$$\ell(n) \mathrel{\hat{\sim}} N(0, \mathcal{I}^{-1}) \; . \qquad (2.3.6)$$

That is, for $n \to \infty$ $\ell(n)$ is asymptotically multivariate normal with mean zero and variance-covariance matrix $\mathcal{I}^{-1}$ (2.3.4).

For general interest we might mention the following. The above theorems provide us with the means of investigating the unknown parameters on which transition probabilities may depend. It is possible to make inferences about transition probabilities alone. For example, we may wish to test the hypothesis that several realizations are from the same Markov chain. Such a test uses a χ^2 goodness of fit test. Billingsley (1961) has a discussion on these goodness of fit type tests. Problems of this sort are not investigated in this thesis.

III. ESTIMATION OF BOTH MUTATION RATES IN MORAN'S MODEL

3.1 The Model under Conditions of Most Biological Interest

(a) The Model. In this chapter the estimation of both mutation rates, α_1 and α_2 in Moran's (1958) model will be discussed. We postulate $\alpha_1 , \alpha_2 > 0$ and $1 - \alpha_1 - \alpha_2 > 0$. This includes most of the cases of biological interest. The biological analogue of this situation is the estimation of spontaneous mutation rates in a natural population, that is, estimating both forward and backward mutation rates. Conditions other than $\alpha_1, \alpha_2 > 0$ and $1 - \alpha_1 - \alpha_2 > 0$ will be discussed briefly elsewhere in this chapter.

In Moran's model we assume a constant population size M of haploid individuals "a" or A. Suppose that of the M haploid individuals i are of type "a" where $i = 0, 1, 2, \ldots, M$. The number of A individuals is then $M - i$ and the proportions of "a" and A are iM^{-1} and $1 - iM^{-1}$ respectively. Also let there be a probability α_1 of a gamete "a" mutating to A and α_2 of a gamete A mutating to "a" whenever such are chosen as sex gametes for the production of offspring. We postulate that a new individual is formed by the random choice of a parent whose gamete is passed on, with possible mutation, to the offspring. Thus the probability that an offspring is of type "a" is

$$p_i = (1 - \alpha_1)iM^{-1} + (1 - iM^{-1})\alpha_2 \tag{3.1.1}$$

and of being type A is

$$q_i = iM^{-1}\alpha_1 + (1 - iM^{-1})(1 - \alpha_2) \quad . \tag{3.1.2}$$

We further assume that at each instant at which the state of the model may change, one of the gametes chosen at random dies and is replaced by a new gamete which is "a" or A with probabilities p_i , q_i as given above where i is the number of "a"'s prior to the event. Thus the birth-death model postulates that at each unit of time, one individual is chosen at random to die, and is replaced by a new individual whose genotype is determined at random from those existing before the death. Hence the number of individuals of a given genotype (the state of the population) can take any of the values 0, 1, ..., M, and can change by at most unity during one birth-death event. The model was further discussed by Moran (1958a).

Moran's model applies to a population in which breeding and mortality are occurring all the time, and in which generations overlap. Moreover it applies strictly to a haploid population. The transition matrix for the general model is defined by the elements

$$
\begin{aligned}
p_{ii+1} &= (1 - \tfrac{i}{M})[(1 - \alpha_1)\tfrac{i}{M} + \alpha_2(1 - \tfrac{i}{M})] \\
p_{ii} &= 1 - p_{ii+1} - p_{ii-1} \\
&= \tfrac{i}{M}[(1 - \alpha_1)\tfrac{i}{M} + \alpha_2(1 - \tfrac{i}{M})] \\
&\qquad + (1 - \tfrac{i}{M})[\alpha_1 \tfrac{i}{M} + (1 - \alpha_2)(1 - \tfrac{i}{M})] \\
p_{ii-1} &= \tfrac{i}{M}[\alpha_1 \tfrac{i}{M} + (1 - \alpha_2)(1 - \tfrac{i}{M})] \\
p_{ik} &= 0 \quad \text{if} \quad k > i + 1 \quad \text{or} \quad k < i - 1 ,
\end{aligned} \tag{3.1.3}
$$

taking into account the probabilities for birth and death gamete types.

The square transition matrix P (2.1.2) of order $M + 1$ with elements (3.1.3) has a tri-diagonal form

$$
P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \cdot \\ \cdot \\ i \\ \cdot \\ \cdot \\ M-1 \\ M \end{array}
\begin{array}{c}
0 \quad 1 \quad 2 \quad \cdot \quad \cdot \quad i \quad \cdot \quad \cdot \quad M-1 \quad M \\
\begin{bmatrix}
1-\alpha_2 & \alpha_2 & 0 & \cdot\ \cdot\ \cdot & & 0 & 0 \\
\ddots & \ddots & \ddots & & & & \\
 & & & & 0 & & \\
 & \ddots & \ddots & \ddots & & & \\
 & & p_{ii-1} & p_{ii} & p_{ii+1} & & \\
 & 0 & & \ddots & \ddots & \ddots & \\
 & & & & \ddots & \ddots & \ddots \\
0 & 0 & & \cdot\ \cdot\ \cdot & 0 & \alpha_1 & 1-\alpha_1
\end{bmatrix}
\end{array} . \qquad (3.1.4)
$$

Clearly there are no absorbing states provided $\alpha_1, \alpha_2 > 0$. Further $p_{00} = 1 - \alpha_2$, $p_{01} = \alpha_2$ and $p_{MM} = 1 - \alpha_1$, $p_{MM-1} = \alpha_1$ hold for all M.

We shall denote the numbers of times the transitions from state i to $i + 1$, i to i, and i to $i - 1$ are observed by a_i, b_i and c_i respectively. This notation, rather than the more general n_{ij} used in (2.3.1), will be used throughout the thesis. In general

$$\sum_{i=0}^{M} (a_i + b_i + c_i) = n - 1 \quad \text{and} \quad \sum_{i=0}^{M} n_i = n \quad \text{where } n_i \text{ is}$$

the total number of times state i is observed and n is the observed length of the chain. For example, consider the following chain of length $n = 12$ where $M = 4$ observed at equal time intervals,

$$222334332221 \qquad (3.1.5)$$

we have

i to i+1	a_i	i to i	b_i	i to i-1	c_i	n_i
						$n_1 = 1$
2 to 3	1	2 to 2	4	2 to 1	1	$n_2 = 6$
3 to 4	1	3 to 3	2	3 to 2	1	$n_3 = 4$
				4 to 3	1	$n_4 = 1$

;

furthermore, $\sum_{i=0}^{M=4} (a_i + b_i + c_i) = 11$ and $\sum_{i=0}^{4} n_i = 12$.

The probability for this outcome (the "likelihood") could be written

$$p_{23}\, p_{34}\, p_{22}^{4}\, p_{33}^{2}\, p_{21}\, p_{32}\, p_{43} \qquad (3.1.6)$$

or in general

$$\prod_i p_{ii+1}^{a_i}\, p_{ii}^{b_i}\, p_{ii-1}^{c_i} \;. \qquad (3.1.7)$$

The following relationships between the a_i's and c_i's also hold:

$a_i \equiv c_{i+1}$ if initial state of the chain is at or below i, final state is at or below i or if initial state is above i, final is above i. (3.1.8)

$a_i \equiv c_{i+1}-1$ if initial state is above i, final is at or below i.

$a_i \equiv c_{i+1}+1$ if initial state is at or below i, final is above i.

(b) Procedure for Obtaining Maximum Likelihood Estimates. In a Markov chain of the type discussed in this chapter, where there are no absorbing states, observing a single long chain (i.e., $n \to \infty$) provides us with an "infinite" amount of information. Thus, the standard procedure for conducting the experiment will be to observe a single long chain and apply the standard techniques of maximum likelihood. Clearly, replicated experiments, that is, observing many independent realizations of different chains is also a valid procedure. Replications will be discussed in Section 3.3.

Using the notation (3.1.7) we write the log-likelihood function as

$$\log L(\alpha_1\alpha_2) = \log L = \sum_{i=0}^{M-1} a_i \log p_{ii+1}$$

$$+ \sum_{i=0}^{M} b_i \log p_{ii} + \sum_{i=1}^{M} c_i \log p_{ii-1} , \qquad (3.1.9)$$

where the upper index on the first term is $M-1$ since no transition of the type M to $M+1$ is possible. Similarly the lower index on the last term is 1 since a transition 0 to -1 is not possible.

Let

$$\varphi_1 = \frac{\partial}{\partial\alpha_1}\log L , \quad \varphi_2 = \frac{\partial}{\partial\alpha_2}\log L , \qquad (3.1.10)$$

then

$$\varphi_1 = \sum_{i=1}^{M-1} \frac{-\, i a_i}{[(1-\alpha_1)i+\alpha_2(M-i)]} + \sum_{i=1}^{M} \frac{i(M-2i)b_i}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]}$$

$$+ \sum_{i=1}^{M} \frac{i c_i}{[i\alpha_1+(1-\alpha_2)(M-i)]} \qquad (3.1.11)$$

$$\varphi_2 = \sum_{i=0}^{M-1} \frac{(M-i)a_i}{[(1-\alpha_1)i+\alpha_2(M-i)]} - \sum_{i=0}^{M-1} \frac{(M-i)(M-2i)b_i}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]}$$

$$- \sum_{i=1}^{M-1} \frac{(M-i)c_i}{[i\alpha_1+(1-\alpha_2)(M-i)]} .$$

Further

$$\frac{\partial\varphi_1}{\partial\alpha_1} = - \Big[\sum_{i=1}^{M-1} \frac{i^2 a_i}{[(1-\alpha_1)i+\alpha_2(M-i)]^2}$$

$$+ \sum_{i=1}^{M} \frac{i^2(M-2i)^2 b_i}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]^2}$$

$$+ \sum_{i=1}^{M} \frac{i^2 c_i}{[i\alpha_1+(1-\alpha_2)(M-i)]^2} \Big] \qquad (3.1.12)$$

$$\frac{\partial\varphi_1}{\partial\alpha_2} = \frac{\partial\varphi_2}{\partial\alpha_1} = \sum_{i=1}^{M-1} \frac{i(M-i)a_i}{[(1-\alpha_1)i+\alpha_2(M-i)]^2}$$

$$+ \sum_{i=1}^{M-1} \frac{i(M-2i)^2(M-i)b_i}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]^2}$$

$$+ \sum_{i=1}^{M-1} \frac{i(M-i)c_i}{[i\alpha_1+(1-\alpha_2)(M-i)]^2}$$

$$\frac{\partial\varphi_2}{\partial\alpha_2} = - \Big[\sum_{i=0}^{M-1} \frac{(M-i)^2 a_i}{[(1-\alpha_1)i+\alpha_2(M-i)]^2}$$

$$+ \sum_{i=0}^{M-1} \frac{(M-2i)^2(M-i)^2 b_i}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]^2}$$

$$+ \sum_{i=1}^{M-1} \frac{(M-i)^2 c_i}{[i\alpha_1+(1-\alpha_2)(M-i)]^2} \Big] .$$

A procedure for simultaneous solution of $\varphi_1 = \varphi_2 = 0$ is to apply the Newton-Raphson iterative method for two equations in two unknowns, viz.,

$$\begin{bmatrix} \hat{\alpha}_1^{(1)} \\ \hat{\alpha}_2^{(1)} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_1^{(0)} \\ \hat{\alpha}_2^{(0)} \end{bmatrix} - \begin{bmatrix} \partial\varphi_1/\partial\alpha_1 & \partial\varphi_1/\partial\alpha_2 \\ \partial\varphi_2/\partial\alpha_1 & \partial\varphi_2/\partial\alpha_2 \end{bmatrix}^{-1} \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} \Bigg|_{\substack{\alpha_1 = \hat{\alpha}_1^{(0)} \\ \alpha_2 = \hat{\alpha}_2^{(0)}}} . \tag{3.1.13}$$

This method requires the inverse of one matrix. Using a convenient first guess (to be discussed below) for α_1 and α_2 this inverse, however, can be calculated once and iterations performed on φ_1 and φ_2. This tactic will result in somewhat slower convergence on α_1 and α_2.

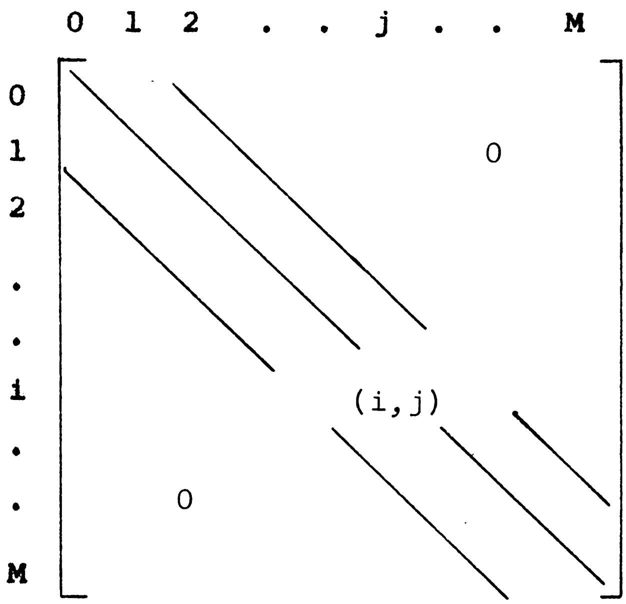
A convenient first guess for α_1 and α_2 could be obtained by selecting the most frequently occurring transitions, estimating the transition probabilities, setting these estimates equal to the right hand side (RHS) of (3.1.3) and solving for α_1 and α_2. It is known (cf., for example Bartlett 1960, p. 229) that the maximum likelihood estimate of a transition probability where no other parameters are involved implicitly is given by the ratio of the number of times the transition from state i to state j occurred to the total number of times that state i was observed. For example, in our notation the maximum likelihood estimate of p_{ii+1} is

$$\hat{p}_{ii+1} = \frac{a_i}{n_i} \quad . \tag{3.1.14}$$

Thus (3.1.14) provides the LHS of (3.1.13). Usually it is necessary to solve two equations simultaneously in order to obtain first guesses for α_1 and α_2. However, if the chain is such that the transitions $M \to M$, $M \to M-1$ or $0 \to 0$, $0 \to 1$ are observed relatively frequent, then first guesses for α_1 and α_2 respectively can be obtained straight away. (cf., 3.1.4).

(c) <u>Uniqueness Theorems</u>. Nothing has been said so far about the existence of maximum likelihood solutions of the Newton-Raphson system (3.1.13). For this discussion we turn to Billingsley's (1961) results presented in Section 2.3.

In light of Condition 2.3.1, for the Markov chain discussed in this chapter, the set D exists. The set D of integer pairs (i, j) are those of the tri-diagonal matrix



where the number of elements d of D is $3M+1$. Each transition probability (3.1.3) has continuous partial derivatives of third order. The parameter space Ω is the open unit square $(0 < \alpha_1 < 1,\ 0 < \alpha_2 < 1)$. This square contains the useful values of the mutation rates (probabilities) α_1, α_2. The $d \times 2$ $(r = 2)$ matrix

$$\left(\frac{\partial p_{ij}}{\partial \alpha_u}\right) \qquad u = 1, 2 \qquad \text{has rank } 2 ,$$

for consider $p_{00} = 1 - \alpha_2$ and $p_{MM} = 1 - \alpha_1$ then

$$\begin{vmatrix} \dfrac{\partial p_{00}}{\partial \alpha_1} & \dfrac{\partial p_{00}}{\partial \alpha_2} \\[2ex] \dfrac{\partial p_{MM}}{\partial \alpha_1} & \dfrac{\partial p_{MM}}{\partial \alpha_2} \end{vmatrix} = \begin{vmatrix} 0 & -1 \\ -1 & 0 \end{vmatrix} = -1 \neq 0 ,$$

hence at least one of the 2×2 determinants of the $d \times 2$ matrix does not vanish, thus the rank is $2(= r)$. Further there is only one ergodic set $\{0, 1, \ldots, M\}$ and there are no transient states. Recall Feller's (1957) definition of a transient state as one for which the probability that the state is visited at least twice is less than one. This would hold for an absorbing Markov chain, but here we have no absorbing states and for an infinitely long chain each state can and will be visited infinitely often. Thus there are no transient states and Condition 2.3.1 is satisfied.

Having satisfied Condition 2.3.1 we can make use of Theorems 2.3.1 and 2.3.2. If (α_1^0, α_2^0) are true values of the parameters (α_1, α_2) and $\hat{\alpha}_1, \hat{\alpha}_2$ are the maximum likelihood estimates, then from the above theorems we can state that

$$(\hat{\alpha}_1 - \alpha_1^0, \hat{\alpha}_2 - \alpha_2^0) \mathrel{\dot{\sim}} N(0, \underset{\sim}{I}^{-1}) \quad , \qquad (3.1.15)$$

that is $(\hat{\alpha}_1 - \alpha_1^0, \hat{\alpha}_2 - \alpha_2^0)$ is asymptotically distributed (as $n \to \infty$) as the multivariate normal with mean zero and variance-covariance matrix $\underset{\sim}{I}^{-1}$ where

$$\mathcal{I} = \begin{bmatrix} -\varepsilon \dfrac{\partial\varphi_1}{\partial\alpha_1} & -\varepsilon \dfrac{\partial\varphi_1}{\partial\alpha_2} \\ \\ -\varepsilon \dfrac{\partial\varphi_2}{\partial\alpha_1} & -\varepsilon \dfrac{\partial\varphi_2}{\partial\alpha_2} \end{bmatrix} . \tag{3.1.16}$$

Recall that Billingsley's results were general and did not guarantee that the solution of $\varphi_1 = 0$, $\varphi_2 = 0$ (3.1.10) would be unique nor that the consistent solution would correspond to the absolute maximum of log L (3.1.9). The following discussion shows that under some general conditions log L has a unique maximum, at the solution of the maximum likelihood equations $\varphi_1, \varphi_2 = 0$, and therefore an absolute maximum of log L in Ω. The solution must of necessity be the consistent solution.

We state the following theorem.

Theorem 3.1.1. Assume that there is at least one solution $(\hat{\alpha}_1, \hat{\alpha}_2)$ of

$$\begin{aligned} \frac{\partial}{\partial\alpha_1} \log L(\alpha_1, \alpha_2) &= 0 \\ \frac{\partial}{\partial\alpha_2} \log L(\alpha_1, \alpha_2) &= 0 \end{aligned} \tag{3.1.17}$$

within the domain Ω; further for any two different i , other than $i = M/2$, at least one a_i, b_i or $c_i > 0$ then (i) the solution is unique in Ω and also in the unit square, (ii) it maximizes $\log L(\alpha_1, \alpha_2)$, and (iii) it provides the consistent estimate for which the asymptotic normality expressed in (2.3.6) applies.

Proof: Following Hobson (1926, p. 213) we can write the function $\log L(\alpha_1, \alpha_2)$, defined for all values of α_1, α_2 lying within the domain Ω as

$$\begin{aligned}\log L(\alpha_1, \alpha_2) &= \log L(\hat{\alpha}_1, \hat{\alpha}_2) + (\alpha_1 - \hat{\alpha}_1)\frac{\partial \log L}{\partial \alpha_1}\bigg|_{\hat{\alpha}_1, \hat{\alpha}_2} \\ &+ (\alpha_2 - \hat{\alpha}_2)\frac{\partial \log L}{\partial \alpha_2}\bigg|_{\hat{\alpha}_1, \hat{\alpha}_2} + \tfrac{1}{2}[(\alpha_1 - \hat{\alpha}_1)^2\frac{\partial^2 \log L}{\partial \alpha_1^2}\bigg|_{\bar{\alpha}_1, \bar{\alpha}_2} \\ &+ 2(\alpha_1 - \hat{\alpha}_1)(\alpha_2 - \hat{\alpha}_2)\frac{\partial^2 \log L}{\partial \alpha_1\, \partial \alpha_2}\bigg|_{\bar{\alpha}_1, \bar{\alpha}_2} \\ &+ (\alpha_2 - \hat{\alpha}_2)^2\frac{\partial^2 \log L}{\partial \alpha_2^2}\bigg|_{\bar{\alpha}_1, \bar{\alpha}_2}] \ , \end{aligned} \tag{3.1.18}$$

where $\bar{\alpha}_1 = \hat{\alpha}_1 + \theta(\alpha_1 - \hat{\alpha}_1)$, $\bar{\alpha}_2 = \hat{\alpha}_2 + \theta(\alpha_2 - \hat{\alpha}_2)$, $0 < \theta < 1$. If $\hat{\alpha}_1$, $\hat{\alpha}_2$ is any solution of $\frac{\partial \log L}{\partial \alpha_1} = 0$, $\frac{\partial \log L}{\partial \alpha_2} = 0$, then (3.1.18) becomes

$$\log L(\alpha_1, \alpha_2) = \log L(\hat{\alpha}_1, \hat{\alpha}_2) + \tfrac{1}{2}[(\alpha_1-\hat{\alpha}_1)^2 \frac{\partial^2 \log L}{\partial \alpha_1^2}\Big|_{\bar{\alpha}_1,\bar{\alpha}_2}$$

$$+ 2(\alpha_1-\hat{\alpha}_1)(\alpha_2-\hat{\alpha}_2)\frac{\partial^2 \log L}{\partial \alpha_1 \, \partial \alpha_2}\Big|_{\bar{\alpha}_1,\bar{\alpha}_2}$$

$$+ (\alpha_2-\hat{\alpha}_2)^2 \frac{\partial^2 \log L}{\partial \alpha_2^2}\Big|_{\bar{\alpha}_1,\bar{\alpha}_2}] . \tag{3.1.19}$$

Let us write x for $\alpha_1 - \hat{\alpha}_1$, y for $\alpha_2 - \hat{\alpha}_2$ and φ_{11} for $\frac{\partial^2 \log L}{\partial \alpha_1^2}\Big|_{\bar{\alpha}_1,\bar{\alpha}_2}$, φ_{12} for $\frac{\partial^2 \log L}{\partial \alpha_1 \, \partial \alpha_2}\Big|_{\bar{\alpha}_1,\bar{\alpha}_2}$ and φ_{22} for $\frac{\partial^2 \log L}{\partial \alpha_2^2}\Big|_{\bar{\alpha}_1,\bar{\alpha}_2}$. Then we would like to be able to say that for all $(\alpha_1, \alpha_2) \neq (\hat{\alpha}_1, \hat{\alpha}_2)$, i.e., $(x, y) \neq (0, 0)$,

$$x^2\varphi_{11} + 2xy\varphi_{12} + y^2\varphi_{22} < 0 \tag{3.1.20}$$

so that from

$$\log L(\alpha_1,\alpha_2) = \log L(\hat{\alpha}_1,\hat{\alpha}_2) + \tfrac{1}{2}(x^2\varphi_{11} + 2xy\varphi_{12} + y^2\varphi_{22}) \tag{3.1.21}$$

we can write

$$\log L(\alpha_1,\alpha_2) < \log L(\hat{\alpha}_1, \hat{\alpha}_2) \quad . \tag{3.1.22}$$

From (3.1.10) and (3.1.12) we can write

$$\varphi_{11} = -[\sum_{i=0}^{M} [i^2 k_i + i^2(M - 2i)^2 \ell_i + i^2 m_i]]$$

$$\varphi_{12} = \sum_{i=0}^{M} [i(M-i)k_i + i(M-2_i)^2(M-i)\ell_i + i(M-i)m_i]$$

$$\varphi_{22} = -[\sum_{i=0}^{M} [(M-i)^2 k_i + (M-2i)^2(M-i)^2\ell_i + (M-i)^2 m_i]]$$

where $k_i = a_i/[(1-\bar{\alpha}_1)i + \bar{\alpha}_2(M-i)]^2 \geq 0$

$\ell_i = b_i/[i^2(1-\bar{\alpha}_1) + (M-i)[i(\bar{\alpha}_1+\bar{\alpha}_2)+(1-\bar{\alpha}_2)(M-i)]]^2 \geq 0$

$m_i = c_i/[i\bar{\alpha}_1 + (1-\bar{\alpha}_2)(M-i)]^2 \geq 0$. (3.1.23)

Thus $x^2\varphi_{11} + 2xy\varphi_{12} + y^2\varphi_{22}$ becomes

$$-[\sum_{i=0}^{M} k_i[ix-(M-i)y]^2 + \sum_{i=0}^{M} \ell_i[ix-(M-i)y]^2(M-2i)^2 + \sum_{i=0}^{M} m_i[ix-(M-i)y]^2]$$

$$= -\sum_{i=0}^{M} [k_i + (M-2i)^2\ell_i + m_i][ix-(M-i)y]^2 \quad (3.1.24)$$

where for convenience the summation is now taken from 0 to M .

Now for all $(x, y) \neq (0, 0)$, provided that for two different i, $k_i + (M-2i)^2\ell_i + m_i > 0$ holds, note that k_i, ℓ_i, m_i are functions of a_i, b_i and c_i respectively and by the assumptions of the theorem at least one of these is > 0 , then at least one term in (3.1.24) will be strictly negative (even allowing for $ix - (M - i)y = 0$ to hold for one value of i).

Thus (3.1.22) holds for all α_1, α_2 in Ω and any solution $\hat{\alpha}_1, \hat{\alpha}_2$ in Ω . Suppose now that another solution $\breve{\bar{\alpha}}_1, \breve{\bar{\alpha}}_2$ say existed in Ω , then from (3.1.22) we would have

$$\log L(\breve{\bar{\alpha}}_1, \breve{\bar{\alpha}}_2) < \log L(\hat{\alpha}_1, \hat{\alpha}_2)$$

and similarly, by interchanging the roles of $\hat{\alpha}_1, \hat{\alpha}_2$ with $\breve{\bar{\alpha}}_1, \breve{\bar{\alpha}}_2$, $\log L(\hat{\alpha}_1, \hat{\alpha}_2) < \log L(\breve{\bar{\alpha}}_1, \breve{\bar{\alpha}}_2)$ which yields a contradiction. Thus there can be at most one solution of $\varphi_1 = \varphi_2 = 0$ in Ω . This proves (i). (3.1.22) proves (ii). The fact that at least one solution of $\varphi_1 = \varphi_2 = 0$ (in Ω) must be consistent, by Theorems 2.3.1 and 2.3.2, ensures that (iii) holds.

We state the following theorem taken from Kaplan (1956, p. 126) in our notation.

Theorem 3.1.2. Let Ω be a bounded domain of the α_1, α_2 plane. Let $\log L(\alpha_1, \alpha_2)$ be defined and continuous in the closed region E formed of Ω plus its boundary. Then $\log L(\alpha_1, \alpha_2)$ has an absolute maximum and an absolute minimum in E.

Corollary 3.1.3. If $a_0 > 0$, $b_0 > 0$, $b_M > 0$, $c_M > 0$ all hold for a particular realization, then the conclusions of Theorem 3.1.1 hold.

Proof: Note from (3.1.9) that

$$\log L = \log L(\alpha_1,\alpha_2) = \sum_{i=0}^{M-1} a_i \log p_{ii+1}$$

$$+ \sum_{i=0}^{M} b_i \log p_{ii} + \sum_{i=1}^{M} c_i \log p_{ii-1} . \qquad (3.1.25)$$

From (3.1.3) when $\alpha_1 = 0$, $p_{MM-1} = 0$,

when $\alpha_2 = 0$, $p_{01} = 0$,

when $\alpha_1 = 1$, $p_{MM} = 0$,

and when $\alpha_2 = 1$, $p_{00} = 0$.

Thus if a_0, b_0, b_M, c_M are all positive (3.1.25) becomes $\log L(\alpha_1,\alpha_2) = -\infty$ for all points on the boundary of Ω , namely of the forms $(0, \alpha_2)$, $(\alpha_1, 0)$, $(1, \alpha_2)$ or $(\alpha_1, 1)$. Therefore $\log L$ does not have an absolute maximum on the boundary of Ω , and Theorem 3.1.2 ensures that the absolute maximum occurs within Ω . Of necessity, therefore, at least one solution of $\varphi_1 = \varphi_2 = 0$ exists in Ω . Thus a_0, b_0, b_M, $c_M > 0$ implies both requirements of Theorem 3.1.1, and the proof is complete.

Corollary 3.1.4. As n (the length of chain; the number of observations) increases the conditions of Corollary 3.1.3 will hold with probability increasing to one. Hence the conclusions of Theorem 3.1.1 hold asymptotically with probability one.

We have thus given a theorem and two corollaries whose applicability can be verified after an experiment is completed, and which also can be used to design an experiment having desirable asymptotic properties.

(d) <u>Application of the Theorems</u>. In order to determine the information matrix $\underset{\sim}{I}$ (3.1.16) we need to find the expectations of the transition numbers a_i, b_i and c_i which are the random variables contained in the elements (cf., 3.1.12) of the information matrix. A discussion of these expectations follows.

Suppose we have a chain of length n with initial state k, where n_i, a random variable, is the total number of times state i is observed and a_i, b_i, c_i are as previously defined (cf., 3.1.5). Let

$$y_{it} = \begin{cases} 1 & \text{if } x(t) = i \\ 0 & \text{if } x(t) \neq i \end{cases} ,$$

then

$$n_i = \sum_{t=1}^{n} y_{it}$$

and

$$\mathcal{E}(n_i) = \sum_{t=1}^{n} \mathcal{E}(y_{it}) = \sum_{t=1}^{n} \Pr(x(t) = i) \quad .$$

Since the initial state is $x(1) = k$ then $\Pr(x(t) = i) = p_{k,i}^{(t-1)}$ (the $t-1$ step transition probability, cf., Section 2.2.d), where $\Pr(x(1) = i) = \delta_{k,i}$ (the Kronecker delta) $= p_{k,i}^{(0)}$ (say). Then

$$\mathcal{E}(n_i) = \sum_{t=0}^{n-1} p_{k,i}^{(t)} = (k,i) \text{ element in } \sum_{t=0}^{n-1} P^t \ ,$$

where $P^0 = I$, the identity matrix.

We can find the $\mathcal{E}(a_i)$ in a similar manner. Let

$$y_1 = \begin{cases} 1 & \text{if } (x(1), x(2)) = (i,i+1), \text{ this has probability } \delta_{k,i}\, p_{ii+1} \\ 0 & \text{otherwise} \end{cases}$$

$$y_2 = \begin{cases} 1 & \text{if } (x(2), x(3)) = (i,i+1), \text{ this has probability } p_{ki}\, p_{ii+1} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$y_j = \begin{cases} 1 & \text{if } (x(j), x(j+1) = (i, i+1), \text{ this has probability } p_{ki}^{(j-1)}\, p_{ii+1} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$y_{n-1} = \begin{cases} 1 & \text{if } (x(n-1), x(n)) = (i,i+1), \text{ this has probability } p_{ki}^{(n-2)}\, p_{ii+1} \\ 0 & \text{otherwise} \end{cases} .$$

Then $a_i = y_1 + y_2 + \ldots + y_{n-1}$ and

$$\mathcal{E}(a_i) = p_{ii+1} \sum_{j=1}^{n-1} p_{ki}^{(j-1)}$$

$$= p_{ii+1} \sum_{t=0}^{n-2} p_{ki}^{(t)}$$

$$= p_{ii+1}[(k,i)\text{element of } \sum_{t=0}^{n-1} P^t - p_{k,i}^{(n-1)}]$$

$$= p_{ii+1}[\mathcal{E}(n_i) - p_{k,i}^{(n-1)}] .$$

Similar results hold for $\mathcal{E}(b_i)$ and $\mathcal{E}(c_i)$. In summary fashion then we have

$$
\begin{aligned}
\mathcal{E}(n_i) &= (k,i) \text{ element of } \sum_{t=0}^{n-1} P^t \\
\mathcal{E}(a_i) &= p_{ii+1}[(k,i) \text{ element of } \sum_{t=0}^{n-2} P^t] \\
&= p_{ii+1}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \\
\mathcal{E}(b_i) &= p_{ii}[(k,i) \text{ element of } \sum_{t=0}^{n-2} P^t] \\
&= p_{ii}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \\
\mathcal{E}(c_i) &= p_{ii-1}[(k,i) \text{ element of } \sum_{t=0}^{n-2} P^t] \\
&= p_{ii-1}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \quad .
\end{aligned}
\tag{3.1.27}
$$

Since P is the transition matrix, the elements of P^t are the t-step transition probabilities discussed in Section 2.2d.

Before proceeding we introduce the following theorem.

Theorem 3.1.5. Transforming the transition matrix P (3.1.4) with elements (3.1.3) by the matrix R , where R has the typical element $R_{ij} = \binom{i}{j}$ and R^{-1} has the typical element $(-1)^{i+j} \binom{i}{j}$, $i, j = 0, 1, \ldots, M$, then $R^{-1}PR$ has non-zero terms only in the leading and first super diagonals. The i-th row is

$$(0,0,\ldots,0,1-i[\frac{\alpha_1+\alpha_2}{M}+\frac{i-1}{M^2}(1-\alpha_1-\alpha_2)],$$

$$(1-\frac{i}{M})[(1-\alpha_1)\frac{i}{M}+\alpha_2(1-\frac{i}{M})],0,\ldots,0) \ , \qquad (3.1.28)$$

the quantity

$$\lambda_i = 1 - i[\frac{\alpha_1+\alpha_2}{M} + \frac{i-1}{M^2}(1-\alpha_1-\alpha_2)] \ , \qquad (3.1.29)$$

in the diagonal position is the i-th eigenvalue of P . The quantity in the super diagonal is the transition probability p_{ii+1} . Hannan in an appendix to Moran (1958) has proved the theorem for the case where $\alpha_1 = \alpha_2 = 0$. The above result is a generalization to the case where both mutation rates are present. Karlin and McGregor (1960) and Gani (1961) have found the eigenvalues (3.1.29) by another method. While this theorem gives an elementary way of finding the eigenvalues, the proof is not given as the eigenvalues are derived incidentally in Theorem 3.1.6.

If P can be written $P = KD_\lambda K^{-1}$ (see Section 2.2d) where K is a matrix of eigenvectors of P and D_λ is the diagonal matrix of eigenvalues λ_i, where λ_i is given by (3.1.29), that is if $P^t = KD_\lambda{}^t K^{-1}$ then

$$\mathcal{E}(n_i) = (k,i) \text{ element in } KD_{\frac{1-\lambda^n}{1-\lambda}} K^{-1} , \qquad (3.1.30)$$

where k is the initial (starting) state of the Markov chain and where the element in the first position (0,0 cf., 2.1.3) of $D_{\frac{1-\lambda^n}{1-\lambda}}$ is $1 + 1 + \ldots + 1 = n$, and the other terms are sums of geometric series.

In order to discuss the result (3.1.30) more fully we present the following very important theorem. Fundamental to the theorem and its proof is the use of Hahn polynomials whose properties are discussed in Appendix I.

Theorem 3.1.6. For the matrix P (3.1.4) defined by elements (3.1.3)

(i) The eigenvalues are

$$\lambda_j = 1 - j\left[\frac{\alpha_1+\alpha_2}{M} + \frac{j-1}{M^2}(1-\alpha_1-\alpha_2)\right]; \; j=0,1,\ldots,M . \qquad (3.1.31)$$

(ii) The post-eigenvectors are the columns of the matrix

$$K = (\underline{K}_0, \underline{K}_1, \ldots, \underline{K}_M) = Q , \qquad (3.1.32)$$

where Q has the Hahn polynomial (9.1.2) $Q_j(i, a, b, M+1)$

in the (i,j) position, $i,j = 0, 1, \ldots, M$.

$$Q = \begin{bmatrix} Q_0(0,a,b,M+1) & Q_1(0,a,b,M+1) & \cdots & Q_M(0,a,b,M+1) \\ Q_0(1,a,b,M+1) & Q_1(1,a,b,M+1) & \cdots & Q_M(1,a,b,M+1) \\ \vdots & & & \vdots \\ Q_0(M,a,b,M+1) & & & Q_M(M,a,b,M+1) \end{bmatrix} .$$

From (9.1.9) we have

$$Q = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & Q_1(1,a,b,M+1) & & Q_M(1,a,b,M+1) \\ \vdots & & & \vdots \\ 1 & Q_1(M,a,b,M+1) & & Q_M(M,a,b,M+1) \end{bmatrix} . \tag{3.1.33}$$

(iii) $$a = \frac{M\alpha_2}{1-\alpha_1-\alpha_2} - 1 , \quad b = \frac{M\alpha_1}{1-\alpha_1-\alpha_2} - 1 . \tag{3.1.34}$$

(iv) The pre-eigenvectors are the rows of the matrix

$$K^{-1} = Q^{-1} . \tag{3.1.35}$$

Proof: Parts (ii) and (iv) of the theorem are either true or false together; the inverse of the post-eigenvector matrix gives the pre-vectors. It will therefore be sufficient to prove that (i) and (ii) are correct, [(iii) will enter with this proof] and this is done by proving

$$PK = KD_\lambda$$

for the particular definitions used here. Recall that D_λ is the diagonal matrix of eigenvalues. Write g_{ij} and h_{ij} for the typical elements of the left- and right-hand sides respectively; then we have to show that $g_{ij} = h_{ij}$ for $i,j = 0, 1, \ldots, M$.

Multiplying out $PK = PQ$ we find

$$g_{ij} = \sum_{k=0}^{M} p_{ik}\, Q_j(k)$$

$$= p_{ii-1}\, Q_j(i-1) + p_{ii}\, Q_j(i) + p_{ii+1}\, Q_j(i+1) ,$$

since $p_{ik} = 0$ if $|i-k| > 1$.

Again, multiplying out KD_λ we get

$$h_{ij} = \sum_{k=0}^{M} Q_k(i)\lambda_{kj}$$

$$= Q_j(i)\lambda_j \quad .$$

Equating g_{ij} and h_{ij} where λ_j is given by (i), we have

$$p_{ii-1}Q_j(i-1) + p_{ii}Q_j(i) + p_{ii+1}Q_j(i+1)$$

$$= Q_j(i)[1 - j(\frac{\alpha_1+\alpha_2}{M} + \frac{j-1}{M^2}(1-\alpha_1-\alpha_2))]$$

or

$$p_{ii-1}Q_j(i-1) - Q_j(i)(1-p_{ii}) + p_{ii+1}Q_j(i+1)$$

$$=-Q_j(i)[j(\frac{\alpha_1+\alpha_2}{M} + \frac{j-1}{M^2}(1-\alpha_1-\alpha_2))] .$$

Recall that $1 - p_{ii} = p_{ii+1} + p_{ii-1}$, thus

$$p_{ii-1}Q_j(i-1) - Q_j(i)(p_{ii-1}+ p_{ii+1}) + p_{ii+1}Q_j(i+1)$$

$$= - Q_j(i)[j(\frac{\alpha_1+\alpha_2}{M} + \frac{j-1}{M^2}(1-\alpha_1-\alpha_2))] . \quad (3.1.36)$$

The equality of g_{ij} and h_{ij} follows by noting that (3.1.36) is the difference equation (9.1.4). From (9.1.4) $\omega_j = j(j + a + b + 1)$, and by (iii) of the theorem we obtain for ω_j

$$\omega_j = j[\frac{\alpha_1+\alpha_2}{M} + \frac{j-1}{M^2}(1-\alpha_1-\alpha_2)]\frac{M^2}{1-\alpha_1-\alpha_2} .$$

Thus the RHS of (3.1.36) becomes

$$- \omega_jQ_j(i)\frac{M^2}{1-\alpha_1-\alpha_2} . \quad (3.1.37)$$

Also from (9.1.4) $B(i) = (M - i)(a + 1 + i)$ (recall that the Hahn polynomial for this case uses $M+1$ rather than M)

using the definition of "a" from (iii) then

$$B(i) = p_{ii+1} \frac{M^2}{1-\alpha_1-\alpha_2} .$$

$D(i) = i(M + 1 + b - i)$ and in a similar fashion it follows

that $$D(i) = p_{ii-1} \frac{M^2}{1-\alpha_1-\alpha_2} .$$

Thus (3.1.36) becomes

$$D(i)Q_j(i-1) - [B(i) + D(i)]Q_j(i) + B(i)Q_j(i+1) = -\omega_j Q_j(i)$$

which is the difference equation (9.14) and thus g_{ij} and h_{ij} are equal for all relevant i,j. This completes the proof of the theorem.

The theorem is not completely new. It restates the eigenvalues found by Karlin and McGregor (1960) and Gani (1961).

We can now write (3.1.30), using the above results, as

$$\mathcal{E}(n_i) = (k,i) \text{ element in } QD_{\frac{1-\lambda^n}{1-\lambda}} Q^{-1} , \qquad (3.1.38)$$

where k is the initial state and λ_j is given by (3.1.31). The inverse of Q can be found by use of the orthogonality relation (9.1.5), that is

$$Q'D_V Q = D_\delta$$

or

$$Q^{-1} = D_{\delta^{-1}} Q' D_V \quad . \tag{3.1.39}$$

$D_{\delta^{-1}}$ is a diagonal matrix of order $M+1$ with elements

$$\begin{cases} 1 \quad \text{in the } (0,0) \text{ position} \\ \dfrac{\binom{M}{u}\Gamma(b+1)\Gamma(u+a+1)\Gamma(u+a+b+1)(2u+a+b+1)}{\binom{M+1+a+b+u}{u}\Gamma(a+1)\Gamma(a+b+1)\Gamma(u+b+1)\Gamma(u+1)(a+b+1)} \\ \qquad\qquad \text{in the } (u,u) \text{ position, } u = 1, 2, \ldots, M \;. \end{cases} \tag{3.1.40}$$

D_V is a diagonal matrix of order $M+1$ with elements

$$\frac{\binom{a+v}{v}\binom{M+b-v}{M-v}}{\binom{M+a+b+1}{M}} \;, \quad v = 0, 1, 2, \ldots, M \;, \tag{3.1.41}$$

and Q' is the transpose of Q .

Recall from (3.1.34) that $a = \dfrac{M\alpha_2}{1-\alpha_1-\alpha_2} - 1$ and $b = \dfrac{M\alpha_1}{1-\alpha_1-\alpha_2} - 1$. Thus (3.1.38) becomes

$$\mathcal{E}(n_i) = (k,i) \text{ element in } QD_{\frac{1-\lambda^n}{1-\lambda}} D_{\delta^{-1}} Q'D_V$$

$$= \sum_{s=0}^{M} Q_s(k)\, d_{ss}\, Q_s(i)\, d_{ii} \tag{3.1.42}$$

where

$Q_j(i)$ is the Hahn polynomial defined in (3.1.32)

d_{ii} is the (i,i) element of D_V (3.1.41)

$\bar{d}_{ss}$ is the (s,s) element of $D_{\frac{1-\lambda^n}{1-\lambda}} D_{\delta^{-1}}$,

that is

$$\bar{d}_{ss} = \begin{cases} n \quad \text{for} \quad s = 0 , \\ \dfrac{1-\lambda_s^{\,n}}{1-\lambda_s} \dfrac{\binom{M}{s}\Gamma(b+1)\Gamma(s+a+1)\Gamma(s+a+b+1)(2s+a+b+1)}{\binom{M+1+a+b+s}{s}\Gamma(a+1)\Gamma(a+b+1)\Gamma(s+b+1)\Gamma(s+1)(a+b+1)} \\ \qquad\qquad\qquad s = 1, 2, \ldots, M . \end{cases} \tag{3.1.43}$$

As $n \to \infty$, $\bar{d}_{ss}$ converges to a finite limit for $s = 1, 2, \ldots, M$ $[(1 - \lambda_s^{\,n})/(1 - \lambda_s)$ becomes $1/(1 - \lambda_s)$, cf., 3.1.30], but $\bar{d}_{00}$ diverges. Thus from (3.1.42)

$$\mathcal{E}(n_i) \sim Q_0(k)\bar{d}_{00}Q_0(i)d_{ii}$$

but from (9.1.9) $Q_0(k) = Q_0(i) = 1$, and hence

$$\mathcal{E}(n_i) \sim n\, d_{ii} \tag{3.1.44}$$

asymptotically as $n \to \infty$.

From (2.2.1), Theorem 3.1.6 and (3.1.39) we can write

$$\lim_{n\to\infty} p_{ki}^{(n)} = \lim_{n\to\infty} [(k,i) \text{ element of } QD_\lambda^{\,n} D_{\delta^{-1}} Q'D_V] . \tag{3.1.45}$$

D_λ is the $M+1$ diagonal matrix of eigenvalues (3.1.31). $\lambda_0 = 1$ and $\lambda_1 = 1 - (\alpha_1+\alpha_2)M^{-1}$. λ_1 is the largest non-unit eigenvalue of P . Thus for $D_\lambda^{\,n}$ as $n \to \infty$ we need $n \gg M$ otherwise $\lambda_1^{\,n}$ will not be negligible. This assumes that α_1 and α_2 are not themselves very small. If $\alpha_1 = O(\frac{1}{M})$, $\alpha_2 = O(\frac{1}{M})$ then n would need to be much larger than M^2 for the theory to work. With these conditions in mind, $D_\lambda^{\,n}$ is the diagonal matrix with elements $\lambda_s^{\,n}$ where

$$\lim_{n\to\infty} \lambda_s^{\,n} = \begin{cases} 1 & s = 0 \\ 0 & s = 1, 2, \ldots, M . \end{cases} \tag{3.1.46}$$

Hence (3.1.45) becomes

$$\begin{aligned} \lim_{n\to\infty} p_{ki}^{(n)} &= \sum_{s=0}^{M} Q_s(k)(\lambda^n d_{\delta^{-1}})_s Q_s(i) d_{ii} \\ &= Q_0(k)(\lambda^n d_{\delta^{-1}})_0 Q_0(i) d_{ii} + \sum_{s=1}^{M} Q_s(k)(\lambda^n d_{\delta^{-1}}) Q_s(i) d_{ii} \\ &= (\lambda^n d_{\delta^{-1}})_0 d_{ii} + 0 \\ &= d_{ii} \quad , \end{aligned} \tag{3.1.47}$$

by (9.1.9), (3.1.40) and (3.1.46). Thus d_{ii} is the stationary probability for state i , that is, the probability that the model is in state i , irrespective of the initial state k , after many generations have elapsed. Thus d_{ii} describes the behavior of the model (population) after the stationary distribution has been attained and provides in general a measure of the effect on evolution of the environmental influences included in the model. As a further general remark on this model it may be of some interest to note that the largest non-unit eigenvalue of the transition matrix P , is the value which governs the rate that the population approaches its stationary distribution. From (3.1.31) this is $\lambda_1 = 1 - (\alpha_1+\alpha_2)M^{-1}$.

Interpreting (3.1.44) we see that the asymptotic values of the expectations $\mathcal{E}(n_i)$ are (number of observations) times (the stationary probabilities). This is known from general theory concerning positively regular Markov chains, (cf., Bartlett, 1960). Moreover, the stationary probabilities are, from (3.1.34) and (3.1.41)

$$d_{ii} = \frac{\dbinom{\frac{M\alpha_2}{1-\alpha_1-\alpha_2} - 1 + i}{i}\dbinom{M + \frac{M\alpha_1}{1-\alpha_1-\alpha_2} - 1 - i}{M-i}}{\dbinom{\frac{M}{1-\alpha_1-\alpha_2} - 1}{M}} ,$$

$$i=0,1,\ldots,M , \qquad (3.1.48)$$

which was found previously by Moran (1958) using a different approach. Moran writes his stationary probability elements as P_i using Gamma rather than combinatorial notation. We can write Moran's P_i as

$$\frac{\left(\frac{M(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2} - 1\right)!\, M!\, \left(\frac{M\alpha_2}{1-\alpha_1-\alpha_2} + i-1\right)!\, \left(\frac{M(1-\alpha_2)}{1-\alpha_1-\alpha_2} - i-1\right)!}{\left(\frac{M}{1-\alpha_1-\alpha_2} - 1\right)!\, \left(\frac{M\alpha_1}{1-\alpha_1-\alpha_2}\right)!\, i!\,(M-i)!\, \left(\frac{M\alpha_2}{1-\alpha_1-\alpha_2} - 1\right)!} \; ; \tag{3.1.49}$$

with suitable regrouping (3.1.48) and (3.1.49) are seen to be equal.

It is obvious from general reasoning that $\sum_{i=0}^{M} \mathcal{E}(n_i) = n$, the length of the chain, but this may also be verified from (3.1.42),

$$\sum_{i=0}^{M} \mathcal{E}(n_i) = \sum_{i=0}^{M} \sum_{s=0}^{M} Q_s(k)\bar{d}_{ss} Q_s(i) d_{ii}$$

$$= \sum_{s=0}^{M} Q_s(k)\bar{d}_{ss} \sum_{i=0}^{M} Q_s(i) d_{ii} \quad .$$

From (9.1.5) note that $d_{ii} = \rho(i)$, and from (9.1.9) that $Q_0(i) = 1$, thus $\sum_{i=0}^{M} Q_s(i) d_{ii}$ can be written

$$\sum_{i=0}^{M} Q_s(i) Q_0(i) \rho(i) \quad . \tag{3.1.50}$$

Further, from (9.1.5) for $s \neq 0$ (3.1.50) is zero; for $s = 0$ it is one, thus

$$\sum_{i=0}^{M} \mathcal{E}(n_i) = Q_0(k)\bar{d}_{00} = \bar{d}_{00} \quad ,$$

where from (3.1.43) $\bar{d}_{00} = n$.

Having found $\mathcal{E}(n_i)$ we can now find $\mathcal{E}(a_i)$, $\mathcal{E}(b_i)$ and $\mathcal{E}(c_i)$ which we need for the information matrix $\mathcal{I}$ (3.1.16). From (3.1.27) and (3.1.47)

$$\mathcal{E}(a_i) = p_{ii+1}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}]$$

for $n \to \infty$

$$\mathcal{E}(a_i) \sim np_{ii+1}d_{ii} \quad .$$

Similarly (3.1.51)

$$\mathcal{E}(b_i) = p_{ii}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \sim np_{ii}d_{ii}$$

and

$$\mathcal{E}(c_i) = p_{ii-1}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \sim np_{ii-1}d_{ii} \quad ,$$

where d_{ii} is the stationary probability given by (3.1.48).

Hence, from (3.1.3), (3.1.12) and (3.1.51) we can write for $n \to \infty$ $[n \gg M$ or $n \gg M^2$ if $\alpha_1, \alpha_2 = O(\frac{1}{M})]$

$$-\mathcal{E}\left(\frac{\partial\varphi_1}{\partial\alpha_1}\right) \sim \frac{n}{M^2}\Big[\sum_{i=1}^{M-1}\frac{i^2(M-i)d_{ii}}{[(1-\alpha_1)i+\alpha_2(M-i)]}$$

$$+\sum_{i=1}^{M}\frac{i^2(M-2i)^2 d_{ii}}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]}$$

$$+\sum_{i=1}^{M}\frac{i^3 d_{ii}}{[i\alpha_1+(1-\alpha_2)(M-i)]}\Big]$$

$$-\mathcal{E}\left(\frac{\partial\varphi_1}{\partial\alpha_2}\right) = -\mathcal{E}\left(\frac{\partial\varphi_2}{\partial\alpha_1}\right) \sim -\frac{n}{M^2}\Big[\sum_{i=1}^{M-1}\frac{i(M-i)^2 d_{ii}}{[(1-\alpha_1)i+\alpha_2(M-i)]}$$

$$+\sum_{i=1}^{M-1}\frac{i(M-2i)^2(M-i)d_{ii}}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]} \qquad (3.1.52)$$

$$+\sum_{i=1}^{M-1}\frac{i^2(M-i)d_{ii}}{[i\alpha_1+(1-\alpha_2)(M-i)]}\Big]$$

$$-\mathcal{E}\left(\frac{\partial\varphi_2}{\partial\alpha_2}\right) \sim \frac{n}{M^2}\Big[\sum_{i=0}^{M-1}\frac{(M-i)^3 d_{ii}}{[(1-\alpha_1)i+\alpha_2(M-i)]}$$

$$+\sum_{i=0}^{M-1}\frac{(M-2i)^2(M-i)^2 d_{ii}}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]}$$

$$+\sum_{i=1}^{M-1}\frac{i(M-i)^2 d_{ii}}{[i\alpha_1+(1-\alpha_2)(M-i)]}\Big] \quad .$$

Finally for $n \to \infty$ [$n \gg M$ or $n \gg M^2$ if $\alpha_1,\alpha_2 = O(\frac{1}{M})$] from (3.1.15) and (3.1.52) we can write

$(\hat{\alpha}_1-\alpha_1^{\,0},\ \hat{\alpha}_2-\alpha_2^{\,0})$ is asymptotically multivariate normal with mean zero and variance-covariance matrix

$$\mathcal{I}^{-1} = \begin{bmatrix} -\mathcal{E}(\dfrac{\partial\varphi_1}{\partial\alpha_1}) & -\mathcal{E}(\dfrac{\partial\varphi_1}{\partial\alpha_2}) \\ -\mathcal{E}(\dfrac{\partial\varphi_2}{\partial\alpha_1}) & -\mathcal{E}(\dfrac{\partial\varphi_2}{\partial\alpha_2}) \end{bmatrix}^{-1} . \tag{3.1.53}$$

Suppose we wish to test

$$H_0: \quad \alpha_1 = \alpha_1^{\,0}$$
$$\alpha_2 = \alpha_2^{\,0}$$

against

$$H_1: \quad \alpha_1 \neq \alpha_1^{\,0}$$
$$\alpha_2 \neq \alpha_2^{\,0} \quad .$$

The following test statistic can be used

$$\chi^2_{(2)} = (\hat{\alpha}_1-\alpha_1^{\,0},\ \hat{\alpha}_2-\alpha_2^{\,0})\ \mathcal{I}^{-1}\begin{pmatrix}\hat{\alpha}_1-\hat{\alpha}_1^{\,0} \\ \hat{\alpha}_2-\hat{\alpha}_2^{\,0}\end{pmatrix} , \tag{3.1.54}$$

where $\mathcal{I}$ is evaluated using $\hat{\alpha}_1$, $\hat{\alpha}_2$. Reject H_0 if the calculated χ^2 (3.1.54) is greater than the tabulated $\chi^2_{(2)}$ at the appropriate significance level.

3.2 Some Diffusion Theory Results

Suppose that M becomes large and α_1, α_2 tend to zero in such a way that $\alpha_1 = \beta_1 M^{-1}$, $\alpha_2 = \beta_2 M^{-1}$ where β_1 and β_2 are fixed. Then (Moran, 1958) d_{ii} (3.1.48) is asymptotically (as $M \to \infty$) equal to

$$\frac{1}{B(\beta_1,\beta_2)} x^{\beta_2 - 1} (1-x)^{\beta_1 - 1} , \qquad (3.2.1)$$

where $x = iM^{-1}$ and $B(\beta_1,\beta_2) = \Gamma(\beta_1)\Gamma(\beta_2)/\Gamma(\beta_1+\beta_2)$. This is a density approximating the discrete distribution d_{ii} . This distribution will be a good approximation as long as $\beta_1 > 0$, $\beta_2 > 0$ are not too small. If β_1 and β_2 are much smaller than unity, the distribution (3.2.1) will be U shaped. If they are equal to unity the distribution is uniform. When both are greater than unity, however, there will be a mode in the distribution. This is the interesting case. In the integrals to follow we shall require for convergence that β_1 and β_2 both be greater than unity.

Now from (3.1.52) let us write

$$- \mathcal{E}\left(\frac{\partial\varphi_1}{\partial\alpha_1}\right) = \frac{n}{M^2} \sum_{i=0}^{M} \left[\frac{i^2(M-i)}{[(1-\alpha_1)i+\alpha_2(M-i)]} \right.$$

$$+ \frac{i^2(M-2i)^2}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]}$$

$$\left. + \frac{i^3}{[i\alpha_1+(1-\alpha_2)(M-i)]} \right] d_{ii} \qquad (3.2.2)$$

and put $x = iM^{-1}$, then (3.2.2) becomes

$$n \sum_{i=0}^{M} \left[\frac{x}{1-x} \, \frac{1-3x(1-x)}{1-2x(1-x)} d_{ii} \right] \qquad (3.2.3)$$

ignoring the α_1, α_2 in the coefficients of d_{ii} and further, replacing $\sum_{i=Mx=0}^{i=Mx=M}$ by $\int_{x=0}^{x=1}$ and d_{ii} by (3.2.1), (3.2.3) becomes

$$- \mathcal{E}\left(\frac{\partial\varphi_1}{\partial\alpha_1}\right) \sim \frac{n}{B(\beta_1,\beta_2)} \int_0^1 \frac{1-3x(1-x)}{1-2x(1-x)} x^{\beta_2} (1-x)^{\beta_1-2} \, dx \equiv A ,$$

and similarly, (3.2.4)

$$- \mathcal{E}\left(\frac{\partial\varphi_1}{\partial\alpha_2}\right) = - \mathcal{E}\left(\frac{\partial\varphi_2}{\partial\alpha_1}\right) \sim -\frac{n}{B(\beta_1,\beta_2)} \int_0^1 \frac{1-3x(1-x)}{1-2x(1-x)} x^{\beta_2-1} (1-x)^{\beta_1-1} \, dx \equiv B ,$$

$$- \mathcal{E}\left(\frac{\partial\varphi_2}{\partial\alpha_2}\right) \sim \frac{n}{B(\beta_1,\beta_2)} \int_0^1 \frac{1-3x(1-x)}{1-2x(1-x)} x^{\beta_2-2} (1-x)^{\beta_1} \, dx \equiv C ,$$

where we require $\beta_1, \beta_2 > 1$ for convergence. If these conditions do not hold then in (3.2.3) a more careful approximation would be needed, probably obtained by not ignoring the α_1, α_2 in the coefficients of d_{ii}. We further write (3.2.4) as

$$A \sim \frac{n}{B(\beta_1,\beta_2)}[B(\beta_1-1,\beta_2+1) - \int_0^1 \frac{1}{1-2x(1-x)} x^{\beta_2+1}(1-x)^{\beta_1-1}\,dx]$$

(3.2.5)

$$B \sim \frac{-n}{B(\beta_1,\beta_2)}[B(\beta_1,\beta_2) - \int_0^1 \frac{1}{1-2x(1-x)} x^{\beta_2}(1-x)^{\beta_1}\,dx]$$

$$C \sim \frac{n}{B(\beta_1,\beta_2)}[B(\beta_1+1,\beta_2-1) - \int_0^1 \frac{1}{1-2x(1-x)} x^{\beta_2-1}(1-x)^{\beta_1+1}\,dx] .$$

If β_1 and β_2 are small and integers, the integrals can be evaluated without too much difficulty directly; otherwise, they can be evaluated numerically using, for example, Simpson's Rule. Consider the following integral where $\beta_1 = 2$, $\beta_2 = 1$,

$$\int_0^1 \frac{x(1-x)^2 dx}{1-2x(1-x)} = \tfrac{1}{2}\int_0^1 \frac{x\,dx}{(x-\frac{1}{2})^2+\frac{1}{4}}$$

$$+ \int_0^1 \frac{-2x^2 dx}{1-2x+2x^2} + \int_0^1 \frac{x^3 dx}{1-2x+2x^2} = \frac{\pi}{8} - \frac{1}{4} = 0.1427 .$$

(3.2.6)

For comparison we evaluate this integral using Simpson's Rule where the interval (0,1) is partitioned into divisions of length $\frac{1}{6}$, thus

$$\int_0^1 \frac{x^{\beta_2}(1-x)^{\beta_1}dx}{1-2x(1-x)} = \frac{1}{18}\{\frac{36}{10}[(\frac{2}{6})^{\beta_2}(\frac{4}{6})^{\beta_1} + (\frac{4}{6})^{\beta_2}(\frac{2}{6})^{\beta_1}]$$

$$+ 4[\frac{36}{26}(\frac{1}{6})^{\beta_2}(\frac{5}{6})^{\beta_1} + 2(\frac{1}{2})^{\beta_1+\beta_2} + \frac{36}{26}(\frac{5}{6})^{\beta_2}(\frac{1}{6})^{\beta_1}]\} ,$$

where for our case $\beta_1 = 2$, $\beta_2 = 1$, thus by Simpson's Rule

$$\int_0^1 \frac{x(1-x)^2dx}{1-2x(1-x)} = 0.1427 \quad . \tag{3.2.7}$$

Hence the method of Simpson's Rule gives excellent results.*

Recall that $\alpha_1 = \beta_1 M^{-1}$, $\alpha_2 = \beta_2 M^{-1}$; maximum likelihood estimates for α_1 and α_2 can be obtained in the manner discussed in Section 3.1 (cf., 3.1.13). Using

---

* Since $[1 - 2x(1- x)]^{-1}$ can be written $\sum_{j=0}^{\infty}[2x(1-x)]^j$, then $\int_0^1[1-2x(1-x)]^{-1}x^{\beta_2}(1-x)^{\beta_1}dx = \sum_{j=0}^{\infty}2^j\int_0^1 x^{\beta_2+j}(1-x)^{\beta_1+j}dx$ $= \sum_{j=0}^{\infty} 2^j B(\beta_2+j+1,\beta_1+j+1)$. For $\beta_1 = \beta_2 = 0$, $\sum_{j=0}^{\infty} 2^j B(j+1,j+1) = \pi/2$ which is believed to be a new identity.

these estimates in terms of β_1 and β_2 we can evaluate the elements of the information matrix I_β (3.2.5), and finally we can say that

$(\hat{\beta}_1-\beta_1^{\,0},\ \hat{\beta}_2-\beta_2^{\,0})$ is asymptotically (as $n \to \infty$) multivariate normal with mean zero and variance-covariance matrix

$$I_\beta^{-1} = M^2 \begin{bmatrix} A & B \\ B & C \end{bmatrix}^{-1} .$$

With the factor M^2, I_β^{-1} is of the order M^2/n which strengthens the requirement that $n \gg M^2$. This requirement was discussed in detail in the last section.

In the following table we consider $\alpha_1=\alpha_2=\alpha$. For $M=1$ and $M=2$ the variance elements were obtained by using (3.1.52). For the last three entries the variance elements were obtained by using (3.2.5) where for the first of these three $\beta_1=\beta_2=2$ and population size was M so that $\alpha_1=\alpha_2=\alpha=2/M$; for the next entry $\beta_1=\beta_2=4$ with population size $2M$ so that $\alpha=2/M$ and for the last entry $\beta_1=\beta_2=8$ with population size $4M$ so that $\alpha=2/M$. n was considered the same for all population sizes. Column three then is the ratio of var$(\hat{\alpha})$ for $M=1$ to the other variances in Column two.

Column four is a result of the following. For $M=1$ suppose we do an experiment long enough to get Var $\hat{\alpha} = 2\alpha/n = \sigma^2$ say. This means, we need $n = 2\alpha/\sigma^2$ observations, which will take a time T, say. For the population of size $M = 1$, one birth-death event corresponds to one generation. However, if the generation time is not affected by population size, the same number of generations can be observed in time T for the larger populations yielding an increase in the number of individual birth-death events by a factor M. For $M = 2$ then we would get on the average $n_2 = 2n$ observations. Hence $\mathrm{Var}(\hat{\alpha}) = 2\alpha/n_2 = \frac{1}{2}\sigma^2$ for $M = 2$. For $M = 2/\alpha$ we get on the average $n_{2/\alpha} = 2n/\alpha$ observations with corresponding variance $0.204\sigma^2$. In like manner we obtain for $M = 4/\alpha$, $8/\alpha$ the values $0.30\,\sigma^2$ and $0.37\,\sigma^2$ respectively. The values in Column four are the ratios of σ^2 for $M = 1$ to the other above variances where all observations were considered over the same time T.

For the two mutation rate case then it appears from Column three of the table that many observations on a small population M is more efficient than the same number of observations on a large population. This implies that it is

Table 3.2.1

M	Var($\hat{\alpha}$)	Relative Efficiencies (n equal)	Relative Efficiencies (generations fixed)
1	$\frac{2\alpha(1-\alpha)}{n} \approx \frac{2\alpha}{n}$	1	1
2	$\frac{2\alpha(1+\alpha-2\alpha^2)}{n}\left[\frac{2\alpha^2(1-\alpha)+1}{4\alpha^2(1-\alpha)+1}\right]$ $\approx \frac{2\alpha(1+\alpha)}{n} \approx \frac{2\alpha}{n}$	$\frac{1-\alpha}{1+\alpha} \approx 1-2\alpha \approx 1$	2
$2/\alpha$	$\frac{0.817}{n}$	$2.45\ \alpha$	4.90
$4/\alpha$	$\frac{2.397}{n}$	$0.83\ \alpha$	3.32
$8/\alpha$	$\frac{5.874}{n}$	$0.34\ \alpha$	2.72

more important to pass through a few states many times than to pass through many states a few times.

Column four and the above discussion, however, indicates that for a fixed time T the larger the population the more observations we get, for in a population of size M one generation consists of M birth-death events. Column four also indicates that between population size $M = 2$ and $M = 2/\alpha$ an optimum size exists. There is, however, a great deal of difference between $M = 2$ and $2/\alpha$, markedly so if α is small.

3.3 Replicated Experiments

In Section 3.1 the mutation rates α_1, α_2 were estimated from data obtained by observing a single long realization $(n \to \infty)$ of the Markov chain. Note from (3.1.44) that for $n \to \infty$ $\mathcal{E}(n_i)$ is independent of the initial state k. The discussion in this section is on replicated independent experiments with finite n. From (3.1.42) we note that $\mathcal{E}(n_i)$ obtained from this Hahn polynomial expression depends on the initial state k.

Suppose that we have R replicated independent realizations of a Markov chain, that is, we have observed R realizations of the same type Markov chain. The length of each realization and its initial state can be the same for each replication. In order to obtain estimates for α_1 and α_2 it is not necessary that they be the same. However, in discussing $\mathcal{E}(n_i)$ and I_R, the information matrix for these replicated experiments, it will be more convenient if all realizations have the same initial state k and same length n.

In any case estimates for α_1 and α_2 can be obtained by using the Newton-Raphson scheme (3.1.13). In (3.1.11) and (3.1.12) we replace a_i, b_i and c_i with $\sum_{r=1}^{R} a_i(r)$, $\sum_{r=1}^{R} b_i(r)$ and $\sum_{r=1}^{R} c_i(r)$ respectively, where for example $a_i(r)$ is the value of a_i in the r-th replicate and $\sum_{r=1}^{R} a_i(r)$ is the total number of times the transition from state i to $i + 1$ occurred over the R replicates.

Suppose for the elements of the information matrix $\underset{\sim}{I}$ (3.1.16) we consider each realization of the same length n (n finite) and with the same starting state k. From (3.1.42) with finite n,

$$\mathcal{E}(n_i) = \sum_{s=0}^{M} Q_s(k)\bar{\alpha}_{ss} Q_s(i) d_{ii} \qquad (3.3.1)$$

which depends on the initial state k.

In taking the expectations of (3.1.12) to obtain the elements of the information matrix $\underset{\sim}{I}_R$ we proceed as follows:

$$- \mathcal{E}\left(\frac{\partial\varphi_1}{\partial\alpha_1}\right) = \mathcal{E}\Bigg[\sum_{i=1}^{M-1} \frac{i^2 \sum_{r=1}^{R} a_i(r)}{[(1-\alpha_1)i+\alpha_2(M-i)]^2}$$

$$+ \sum_{i=1}^{M} \frac{i^2(M-2i)^2 \sum_{r=1}^{R} b_i(r)}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]^2}$$

$$+ \sum_{i=1}^{M} \frac{i^2 \sum_{r=1}^{R} c_i(r)}{[i\alpha_1+(1-\alpha_2)(M-i)]^2} \Bigg] , \qquad (3.3.2)$$

where the a_i, b_i, c_i terms have been replaced with

$$\sum_{r=1}^{R} a_i(r) , \quad \sum_{r=1}^{R} b_i(r) , \quad \sum_{r=1}^{R} c_i(r) , \qquad (3.3.3)$$

respectively. Since the initial states are the same for each realization $\mathcal{E} \sum_{r=1}^{R} a_i(r) = R\mathcal{E}(a_i)$. Similar expressions hold for the b_i and c_i . Further note that $- \mathcal{E}(\frac{\partial\varphi_1}{\partial\alpha_2})$, $- \mathcal{E}(\frac{\partial\varphi_2}{\partial\alpha_1})$ and $- \mathcal{E}(\frac{\partial\varphi_2}{\partial\alpha_2})$ follow in like fashion as (3.3.2). Thus for finite n the same for each realization and also the same initial state k for all independent realizations (3.3.2) is simply

$$R\left[\sum_{i=1}^{M-1} \frac{i^2 \mathcal{E}(a_i)}{[(1-\alpha_1)i+\alpha_2(M-i)]^2} \right.$$

$$+ \sum_{i=1}^{M} \frac{i^2(M-2i)^2 \mathcal{E}(b_i)}{[i^2(1-\alpha_1)+(M-i)[i(\alpha_1+\alpha_2)+(1-\alpha_2)(M-i)]]^2}$$

$$\left. + \sum_{i=1}^{M} \frac{i^2 \mathcal{E}(c_i)}{[i\alpha_1+(1-\alpha_2)(M-i)]^2} \right] ; \qquad (3.3.4)$$

similar expressions hold for the other elements of $\mathcal{I}_R$. Therefore under these conditions, the replicated experiment has information matrix $\mathcal{I}_R = R\,\mathcal{I}$, where here $\mathcal{I}$ is the information matrix (3.1.16) for a single replicate. We can further write that $(\hat{\alpha}_1-\alpha_1^0, \hat{\alpha}_2-\alpha_2^0)$ is asymptotically as $R \to \infty$ multivariate normal with mean zero and variance-covariance matrix $(\mathcal{I}_R)^{-1}$.

Now as to the method of conducting an experiment the following scheme is proposed. If the estimate of α_1 say is of more interest than α_2 then the initial state $k = M$ should be selected (cf., 3.1.4). If α_2 is of more interest then the initial state should be $k = 0$. If the two parameters are of the same order then in the first case the variance of $\hat{\alpha}_1$ will be less than the variance of $\hat{\alpha}_2$ and similarly in the second case the variance of $\hat{\alpha}_2$ will be less than the variance of $\hat{\alpha}_1$. If both parameters are of equal interest then it may be best to select the neighborhood $k = M/2$ as the initial state in which case the variances of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ will be approximately equal and lying between the two extremes mentioned above provided the two parameters are of about the same order. For example, in the following table we have a comparison of the variances of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ under different initial states for $M = 2$. The figures are entered apart from the replication factor R^{-1} . It was assumed that $\hat{\alpha}_1 = \hat{\alpha}_2 = 0.1$ and $n = 10$, $\mathcal{E}(n_i)$ for $k = 0, 1, 2$ was obtained from (3.3.1); $\mathcal{E}(a_i)$, $\mathcal{E}(b_i)$ and $\mathcal{E}(c_i)$ were obtained by the method of (3.1.27). The variance elements were obtained by inverting the matrix with elements of the form (3.3.4).

Table 3.3.1

$M = 2$; $\hat{\alpha}_1 = \hat{\alpha}_2 = 0.1$; $n = 10$	
initial state k	Variance (apart from factor R^{-1})
0	$\mathrm{Var}(\hat{\alpha}_1) = 0.1317$ $\mathrm{Var}(\hat{\alpha}_2) = 0.0125$
1	$\mathrm{Var}\ \hat{\alpha}_1 = 0.0283$ $\mathrm{Var}\ \hat{\alpha}_2 = 0.0283$
2	$\mathrm{Var}\ \hat{\alpha}_1 = 0.0125$ $\mathrm{Var}\ \hat{\alpha}_2 = 0.1317$

3.4 Conditions on the Mutation Rates of Moran's Model

Recall in Section 3.1 that we postulated $\alpha_1, \alpha_2 > 0$ and $1 - \alpha_1 - \alpha_2 > 0$. This includes most of the cases of biological interest. Associated with these conditions and the transition matrix P was the Hahn polynomial $Q_j(i,a,b,M+1)$ (3.1.33) where

$$a = \frac{M\alpha_2}{1-\alpha_1-\alpha_2} - 1 \quad \text{and} \quad b = \frac{M\alpha_1}{1-\alpha_1-\alpha_2} - 1 \quad .$$

The stochastic process discussed in this chapter is recognized as the discrete time analogue of an example of a classical birth and death process [Karlin and McGregor, 1957] with birth rates

$$\upsilon_i = \upsilon(1 - \frac{i}{M})[\frac{i}{M}(1-\alpha_1) + (1 - \frac{i}{M})\alpha_2] \tag{3.4.1}$$

and death rates

$$\mu_i = \upsilon\frac{i}{M}[\frac{i}{M}\alpha_1 + (1 - \frac{i}{M})(1-\alpha_2)] \ , \tag{3.4.2}$$

corresponding to a population size i of "a" gametes, $0 \leq i \leq M$.

For the case discussed in this chapter, that is $\alpha_1, \alpha_2 > 0$ and $1 - \alpha_1 - \alpha_2 > 0$ the birth and death rates (3.4.1) and (3.4.2) oppose each other, one exhibiting attraction, the other repulsion toward the same end state.

For $\alpha_1, \alpha_2 > 0$ and $1 - \alpha_1 - \alpha_2 < 0$ the birth and death rates extend their force in the same direction. For this case the Hahn polynomial is $Q_j(i,a,b,M+1)$ where

$$a = -\left(\frac{M\alpha_2}{\alpha_1+\alpha_2-1} + 1\right) \quad \text{and} \quad b = -\left(\frac{M\alpha_1}{\alpha_1+\alpha_2-1} + 1\right) .$$

For $\alpha_1, \alpha_2 > 0$ and $\alpha_1 + \alpha_2 = 1$ the birth and death rates become linear in i rather than quadratic. The polynomial in this case is the Krawtchouk polynomial, another member of the family of orthogonal polynomials, [cf., Erdélyi, 1953]. For further discussion on these conditions see Karlin and McGregor, 1960.

Another case which we might mention, although there obviously is no inference involved, is when $\alpha_1 = \alpha_2 = 0$ that is Moran's model without mutation. Here state M and state 0 are absorbing states. The two absorbing states correspond to fixation in homozygous populations of "a" or A gametes. Karlin and McGregor (1960) discuss this case using Hahn polynomials, while Watterson (1961) uses Tchebichef polynomials. See Appendix I for further discussion of their results.

IV. INFERENCE ON AN ABSORBING MARKOV CHAIN

4.1 Estimating Mutation Rate from a Single Chain (Theory)

(a) The Model. This chapter will be a discussion of one of the mutation rates α_1 of Moran's model. At first sight this may appear to be a simpler problem than that of the two mutation rate case discussed in Chapter III. This, however, is not true, in that inferences will be obtained from realizations of an absorbing Markov chain whose peculiarities provide some unique difficulties.

In this section we shall discuss inference on the mutation rate α_1 , where $\alpha_2 = 0$, using results found by observing a single long Markov chain. Replicated experiments will be discussed in the next section. We postulate $\alpha_1 > 0$ and $1 - \alpha_1 > 0$. The case where the chain length n is predetermined, that is fixed, and also the case where n is a random variable determined by some sequential stopping rule will be discussed. Although the results will be general, emphasis will be placed on an experiment where the initial state is $k = M$, and stopping state will be the absorbing state 0 , so that n , the chain length, is a random variable. By Theorem 2.2.1, for a long chain the absorbing state will be reached with probability one.

In most biological experiments (see for example Falconer 1949) dealing with mutation, a mutagenic agent is introduced into the population under study and the effect of the agent is measured in terms of mutation rate. Compared to the mutagenic rate, the reverse mutation rate is negligible and we shall assume it to be zero. Now if in Moran's model we put $\alpha_2 = 0$ and obtain estimates of α_1 we are in the same type of situation but have a precisely defined model. With $\alpha_2 = 0$ the transition probabilities (3.1.3) become

$$\begin{aligned}
p_{ii+1} &= (1 - \tfrac{i}{M})(\tfrac{i}{M})(1-\alpha_1) \\
p_{ii} &= (1 - \tfrac{i}{M}) - \tfrac{i}{M}(1-\alpha_1)(1 - \tfrac{2i}{M}) \\
p_{ii-1} &= \tfrac{i}{M}[1 - (1-\alpha_1)\tfrac{i}{M}] \\
p_{ik} &= 0 \quad \text{if} \quad |i - k| > 1 \ .
\end{aligned} \tag{4.1.1}$$

The square transition matrix P of order $M+1$ with elements (4.1.1) has a tri-diagonal form,

$$P = \begin{array}{cc} & \begin{array}{cccccccc} 0 & 1 & 2 & \cdot & \cdot & \cdot & M-1 & M \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \cdot \\ \cdot \\ i \\ \cdot \\ \cdot \\ M-1 \\ M \end{array} & \left[\begin{array}{cccccccc} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ & \ddots & \ddots & \ddots & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & & & & & \\ & & & & & & & \\ & & p_{ii-1} & p_{ii} & p_{ii+1} & & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & \ddots & \ddots & \ddots \\ & & & & & & & \\ 0 & 0 & 0 & \cdot & \cdot & \cdot\ 0 & \alpha_1 & 1-\alpha_1 \end{array} \right] \end{array} . \qquad (4.1.2)$$

Clearly state $i = 0$ is absorbing, the other states are transient. Further, $p_{MM} = 1-\alpha_1$, $p_{MM-1} = \alpha_1$ hold for all M .

(d) <u>Procedure for Maximum Likelihood Solution</u>. Using the notation of (3.1.7) we write the log-likelihood function as

$$\log L(\alpha_1) = \log L = \sum_{i=0}^{M-1} a_i \log p_{ii+1}$$

$$+ \sum_{i=0}^{M} b_i \log p_{ii} + \sum_{i=1}^{M} c_i \log p_{ii-1} . \qquad (4.1.3)$$

See (3.1.5) for a discussion of the a_i, b_i and c_i notation.

Let

$$\varphi = \frac{d \log L}{d\alpha_1} , \qquad (4.1.4)$$

then

$$\varphi = \sum_{i=1}^{M-1} \frac{a_i}{p_{ii+1}} \frac{(M-i)(-i)}{M^2} + \sum_{i=1}^{M} \frac{b_i}{p_{ii}} \frac{i(M-2i)}{M^2} + \sum_{i=1}^{M} \frac{c_i}{p_{ii-1}} \frac{i^2}{M^2}$$

$$= \sum_{i=1}^{M-1} \frac{-a_i}{(1-\alpha_1)} + \sum_{i=1}^{M} \frac{ib_i(1-2iM^{-1})}{[M-i-(1-\alpha_1)(i)(1-2iM^{-1})]}$$

$$+ \sum_{i=1}^{M} \frac{ic_i}{[M-i(1-\alpha_1)]} \quad ,$$

and (4.1.5)

$$\varphi' = -[\sum_{i=1}^{M-1} \frac{a_i}{(1-\alpha_1)^2} + \sum_{i=1}^{M} \frac{i^2 b_i(1-2iM^{-1})^2}{[(M-i)-(1-\alpha_1)(i)(1-2iM^{-1})]^2}$$

$$+ \sum_{i=1}^{M} \frac{i^2 c_i}{[M-i(1-\alpha_1)]^2}] \quad ,$$

all lower indices are one. For example, for this absorbing Markov chain where 0 is the absorbing state no transition from state 0 to state 1 is possible, and hence $a_0 = 0$. The maximum likelihood estimate $\hat{\alpha}_1$ of α_1, can be found iteratively by using the Newton-Raphson scheme in the following way,

$$\hat{\alpha}_1^{(1)} = \hat{\alpha}_1^{(0)} - \frac{\varphi(\hat{\alpha}_1^{(0)})}{\varphi'(\hat{\alpha}_1^{(0)})} \quad . \qquad (4.1.6)$$

A convenient first guess for α_1 could be obtained by estimating the transition probability p_{MM-1} $(=\alpha_1)$. As discussed in Chapter III, the maximum likelihood estimate of p_{MM-1} where no other parameters are involved implicitly is

$$\hat{p}_{MM-1} = \frac{c_M}{n_M} \quad , \tag{4.1.7}$$

so that a first guess for α_1 is readily available.

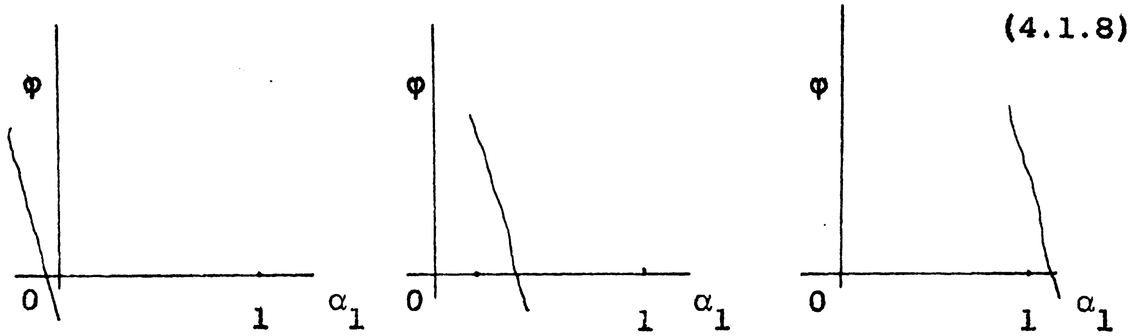
Convergence occurs when $\varphi(\hat{\alpha}_1) = 0$. A discussion of this system follows.

(c) <u>Uniqueness</u> <u>Theorems</u>. Theorem 4.1.1. There is only one solution (at most) of $\varphi(\hat{\alpha}_1) = 0$ (4.1.5) in $(0 < \hat{\alpha}_1 < 1)$.

Proof: From (4.1.5) $-\varphi' \geq 0$ in (0,1), and hence φ is monotonic decreasing and the theorem is immediate.

φ must appear as one of the following

(4.1.8)



no root — one root $\hat{\alpha}_1$ in (0,1) — no root

The first and third possibilities correspond to the likelihood being a maximum in (0,1) at $\alpha_1 = 0,1$ respectively. Note that the theorem does not say that a root exists. It does say that if there is a solution in (0,1) then there is only one solution in this interval.

Theorem 4.1.2. For a given realization of length n, if $c_M > 0$, and either $b_M > 0$ or one $a_i > 0$, then there is exactly one root of $\varphi(\hat{\alpha}_1) = 0$ in (0,1).

Proof: From (4.1.5)

$$\lim_{\alpha_1 \to 0} \varphi(\alpha_1) = \sum_{i=1}^{M-1} -a_i + \sum_{i=1}^{M} \frac{ib_i(1-2iM^{-1})}{[M-i-i(1-2iM^{-1})]} + \sum_{i=1}^{M} \frac{ic_i}{M-i} . \tag{4.1.9}$$

The first two terms of (4.1.9) are finite while the last term is $+\infty$ since by the assumptions of the theorem $c_M > 0$.

Similarly,

$$\lim_{\alpha_1 \to 1} \varphi(\alpha_1) = - \frac{\sum_{i=1}^{M-1} a_i}{0} + \sum_{i=1}^{M} \frac{ib_i(1-2iM^{-1})}{M-i} + \sum_{i=1}^{M} \frac{ic_i}{M} . \tag{4.1.10}$$

The first term is $-\infty$ if at least one $a_i > 0$; the last term is finite. The middle term is finite or $-\infty$ if $b_M > 0$.

Thus under the conditions of the theorem $\varphi(0) = +\infty$, $\varphi(1) = -\infty$ and there is exactly one root of $\varphi(\hat{\alpha}_1) = 0$ in $(0,1)$.

Corollary 4.1.3. If the initial state is $k = M$ and a sequential stopping rule is employed such that 0 (the absorbing state) is the stopping state and $b_M > 0$ or at least one $a_i > 0$, then there is exactly one root of $\varphi(\hat{\alpha}_1) = 0$ in $(0,1)$ with probability one.

Proof: The sequential rule implies $c_M > 0$. The result follows from the proof of Theorem 4.1.2, noting that with probability 1 (cf., Theorem 2.2.1) none of the transition numbers a_i, b_i, or c_i in (4.1.9) or (4.1.10) become $+\infty$.

The following table shows the number of experiments out of 500 realizations obtained by simulation methods on the IBM 650 for each of the populations $M = 2, 4, 6, 10, 20$, $\alpha_1 = 0.1$, which did not satisfy the conditions of Theorem 4.1.2. That is, b_M and all a_i were zero.

Table 4.1.1

M	no. of experiments out of 500 not satisfying Theorem 4.1.2
2	16
4	2
6	2
10	0
20	0

Clearly the theorem applies in an overwhelming proportion of realizations. For example the 16 experiments for $M = 2$ were of the type 2.1 ... 1.0 that is, the initial state was $k = M = 2$, then the next transition was to state 1 followed by a finite number of transitions 1 to 1, and then to the absorbing state 0. For this situation the φ function is $1/(1+\alpha_1) + 1/\alpha_1$ which never crosses the α_1 axis. For a further extensive discussion of this simulation study see the latter part of this chapter.

Before discussing expectations of the transition numbers a_i, b_i and c_i which we shall need in taking the expectation of φ' (4.1.5), we present the following theorem.

(d) Application of the Theorems. Theorem 4.1.4. Transforming the transition matrix P with elements (4.1.1) by the matrix R, where R has the typical element $R_{ij} = \binom{i}{j}$ and R^{-1} has the typical element $(-1)^{i+j}\binom{i}{j}$, $i,j = 0, 1, \ldots, M$, then $R^{-1}PR$ has non-zero terms only in the leading and first super diagonals. The i-th row is

$$(0,\ldots,0\ 1-i[\frac{\alpha_1}{M}+\frac{i-1}{M^2}(1-\alpha_1)],(1-\frac{i}{M})(1-\alpha_1)(\frac{i}{M}),0,\ldots,0) , \tag{4.1.11}$$

the quantity

$$\lambda_i = 1 - i[\frac{\alpha_1}{M} + \frac{i-1}{M^2}(1-\alpha_1)] \tag{4.1.12}$$

in the diagonal position is the i-th eigenvalue of P. The quantity in the super diagonal is the transition probability p_{ii+1}. For further discussion on this theorem see Theorem 3.1.6.

Since P is the transition matrix, the elements of P^t are the t-step transition probabilities discussed in Section 2.2. If P can be written $P = KD_\lambda K^{-1}$, where K is the matrix of eigenvectors and D_λ is the matrix of eigenvalues λ_i (4.1.12), that is, if $P^t = KD_\lambda{}^t K^{-1}$, then

$$\mathcal{E}(n_i) = (k,i) \text{ element in } K(\sum_{t=0}^{n-1} D_\lambda{}^t)K^{-1}$$

$$= (k,i) \text{ element in } K\, D_{\frac{1-\lambda^n}{1-\lambda}} K^{-1} , \tag{4.1.13}$$

(cf., 3.1.27, 3.1.30), where the (0,0) element of $D_{\frac{1-\lambda^n}{1-\lambda}}$ is $1 + 1 + \ldots + 1 = n$, and the other terms are sums of geometric series.

In order to discuss the result (4.1.13) more fully we present the following theorem. Fundamental to the theorem and its proof is the use of Hahn polynomials which are discussed in Appendix I.

Theorem 4.1.5. For the matrix P defined by elements (4.1.1)

(i) The eigenvalues are

$$\lambda_j = 1 - j\left[\frac{\alpha_1}{M} + \frac{j-1}{M^2}(1-\alpha_1)\right], \quad j=0,1,\ldots,M . \tag{4.1.14}$$

(ii) The post-eigenvectors are the columns of the matrix

$$K = (\underline{K}_0, \underline{K}_1, \ldots, \underline{K}_M) = CQ \tag{4.1.15}$$

where

$$C = \begin{bmatrix} 1 & & & & & & \\ 1 & 1 & & & 0 & & \\ 1 & 1 & 1 & & & & \\ \cdot & & & \cdot & & & \\ \cdot & & & & \cdot & & \\ \cdot & & & & & \cdot & \\ 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix}, \tag{4.1.16}$$

and Q has the Hahn polynomial (cf., Appendix I) $Q_{j-1}(i-1,0,b,M)$ in the (i,j) position, $i,j = 0,1,\ldots,M$.

$$Q = \begin{bmatrix} Q_{-1}(-1,0,b,M) & Q_0(-1,0,b,M) & Q_1(-1,0,b,M) & \cdots & Q_{M-1}(-1,0,b,M) \\ Q_{-1}(0,0,b,M) & Q_0(0,0,b,M) & Q_1(0,0,b,M) & \cdots & Q_{M-1}(0,0,b,M) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ Q_{-1}(M-1,0,b,M) & Q_0(M-1,0,b,M) & Q_1(M-1,0,b,M) & \cdots & Q_{M-1}(M-1,0,b,M) \end{bmatrix}$$

From (9.1.8) and (9.1.9) we have

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 1 & & 1 \\ 0 & 1 & Q_1(1,0,b,M) & Q_2(1,0,b,M) & \cdot & Q_{M-1}(1,0,b,M) \\ 0 & 1 & Q_1(2,0,b,M) & Q_2(2,0,b,M) & \cdot & Q_{M-1}(2,0,b,M) \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & Q_1(M-1,0,b,M) & Q_2(M-1,0,b,M) & \cdots & Q_{M-1}(M-1,0,b,M) \end{bmatrix} . \tag{4.1.17}$$

(iii) $b = \dfrac{M\alpha_1}{1-\alpha_1}$, note that for this case $a = 0$. (4.1.18)

(iv) The pre-eigenvectors are the rows of the matrix

$$K^{-1} = Q^{-1}C^{-1} , \tag{4.1.19}$$

where Q^{-1} is the inverse of Q and

$$C^{-1} = \begin{bmatrix} 1 & & & & & & 0 \\ -1 & 1 & & & & & \\ & -1 & 1 & & & & \\ & & -1 & 1 & & & \\ & & & \cdot & \cdot & & \\ & 0 & & & \cdot & \cdot & \\ & & & & & \cdot & \cdot \\ & & & & & -1 & 1 \end{bmatrix} . \qquad (4.1.20)$$

Proof: Parts (ii) and (iv) of the theorem are either true or false together; the inverse of the post-eigenvector matrix gives the pre-vectors. It will therefore be sufficient to prove that (i) and (ii) are correct [(iii) will also enter with this proof] and this is done by proving

$$PK = KD_\lambda$$

for the particular definitions used here. Recall that D_λ is the matrix of eigenvalues. Write g_{ij} and h_{ij} for the typical elements of the left- and right-hand sides respectively; then we have to show that $g_{ij} = h_{ij}$ for $i,j = 0, 1, \ldots, M$.

Multiplying out $PK = PCQ$ we find

$$g_{ij} = \sum_{\ell=0}^{M} [Q_{j-1}(\ell-1) \sum_{u=\ell}^{M} p_{iu}]$$

and with the substitution for the transition probabilities p_{iu} from (4.1.1) we obtain

$$g_{ij} = \begin{cases} \sum_{\ell=0}^{i-1} Q_{j-1}(\ell-1) + (1-\frac{i}{M}[1-\frac{i}{M}(1-\alpha_1)])Q_{j-1}(i-1) \\ \qquad\qquad + (1-\frac{i}{M})(\frac{i}{M})(1-\alpha_1)Q_{j-1}(i), \quad j > 0 \\ 1, \quad j = 0, \quad [\text{cf., } (9.1.8)] . \end{cases}$$

Again, multiplying out $KD_\lambda = CQD_\lambda$ we get

$$h_{ij} = \lambda_j \sum_{\ell=0}^{i} Q_{j-1}(\ell-1)$$

$$= \{1 - j[\frac{\alpha_1}{M} + \frac{j-1}{M^2}(1-\alpha_1)]\} \sum_{\ell=0}^{i} Q_{j-1}(\ell-1)$$

$$= \begin{cases} \sum_{\ell=0}^{i} Q_{j-1}(\ell-1) - j[\frac{\alpha_1}{M} + \frac{j-1}{M^2}(1-\alpha_1)] \sum_{\ell=0}^{i} Q_{j-1}(\ell-1), \quad j > 0 \\ 1 \quad j = 0 . \end{cases} \tag{4.1.21}$$

The equality of g_{ij} and h_{ij} follows from Corollary 9.1.4.

By the corollary,

$$\sum_{\ell=0}^{i} Q_{j-1}(\ell-1) = \begin{cases} \dfrac{(1-\alpha_1)(M-i)(i)[Q_{j-1}(i-1) - Q_{j-1}(i)] + Mi\alpha_1 Q_{j-1}(i-1)}{j[(j-1)(1-\alpha_1) + M\alpha_1]}, \quad j>0 \\ 1, \quad j = 0 . \end{cases}$$

Hence

$$j\left[\frac{\alpha_1}{M} + \frac{j-1}{M^2}(1-\alpha_1)\right] \sum_{\ell=0}^{i} Q_{j-1}(\ell-1) =$$

$$\begin{cases} \frac{i}{M}\left[1 - \frac{i}{M}(1-\alpha_1)\right] Q_{j-1}(i-1) - (1-\alpha_1)\frac{i}{M}\left(1 - \frac{i}{M}\right) Q_{j-1}(i) \ , & j > 0 \\ 0 \ , \quad j = 0 \ , & \end{cases} \tag{4.1.22}$$

then from (4.1.21) and (4.1.22)

$$h_{ij} = \begin{cases} \sum_{\ell=0}^{i} Q_{j-1}(\ell-1) - \frac{i}{M}\left[1 - \frac{i}{M}(1-\alpha_1)\right] Q_{j-1}(i-1) \\ \qquad\qquad + (1-\alpha_1)\frac{i}{M}\left(1 - \frac{i}{M}\right) Q_{j-1}(i) \ , \quad j > 0 \\ 1 \ , \quad j = 0 \ , \end{cases}$$

and finally

$$h_{ij} = \begin{cases} \sum_{\ell=0}^{i-1} Q_{j-1}(\ell-1) + \left(1 - \frac{i}{M}\left[1 - \frac{i}{M}(1-\alpha_1)\right]\right) Q_{j-1}(i-1) \\ \qquad\qquad + \left(1 - \frac{i}{M}\right)\left(\frac{i}{M}\right)(1-\alpha_1) Q_{j-1}(i) \ , \quad j > 0 \\ 1 \ , \quad j = 0 \ . \end{cases}$$

Thus g_{ij} and h_{ij} are equal for all relevant i,j . This completes the proof of the theorem.

We can now write (4.1.13), using the above results, as

$$\mathcal{E}(n_i) = (k,i) \text{ element in } CQD_{\frac{1-\lambda^n}{1-\lambda}} Q^{-1}C^{-1} \tag{4.1.23}$$

where k is the initial state and λ_j is given by (4.1.14). The inverse of Q can be found by use of the orthogonality relation (9.1.5), that is

$$Q'D_V Q = D_\delta$$

or

$$Q^{-1} = D_{\delta^{-1}} Q' D_V \quad . \tag{4.1.24}$$

$D_{\delta^{-1}}$ is a diagonal matrix of order $M+1$ with elements

$$\begin{cases} 1 & \text{in the } (0,0) \text{ position} \\ \dfrac{\binom{M-1}{u-1}(2u+b-1)}{\binom{M+b+u-1}{u-1}(b+1)}, & \text{in the } (u,u) \text{ position, } u=1,2,\ldots,M. \end{cases} \tag{4.1.25}$$

D_V is a diagonal matrix of order $M+1$ with elements

$$d_{vv} = \begin{cases} 1 & \text{in the } (0,0) \text{ position} \\ \dfrac{\binom{M+b-v}{M-v}}{\binom{M+b}{M-1}}, & \text{in the } (v,v) \text{ position, } v=1,2,\ldots,M \; . \end{cases} \tag{4.1.26}$$

Q' is simply the transpose of Q and recall from (4.1.18) that $b = M\alpha_1/(1-\alpha_1)$.

Thus from (4.1.24) we can write (4.1.23) as

$$\mathcal{E}(n_i) = (k,i) \text{ element in } CQD_{\frac{1-\lambda^n}{1-\lambda}} D_{\delta^{-1}} Q'D_V C^{-1} \quad . \tag{4.1.27}$$

The (k,i) element of (4.1.27) is

$$\sum_{w=0}^{k} c_{kw}\left[\sum_{s=0}^{M} Q_{s-1}(w-1) \sum_{v=0}^{M} \bar{d}_{ss} Q_{s-1}(v-1) d_{vv} c_{vi}^{-1}\right] , \tag{4.1.28}$$

where

c_{kw} is an element of (4.1.16) the C matrix, recall that $c_{kw} = 0$ for $w > k$,

$Q_{j-1}(i-1)$ is an element of (4.1.17), the Hahn polynomial matrix Q ,

c_{vi}^{-1} is an element of (4.1.20), the C^{-1} matrix; also note that

$$c_{vi}^{-1} = \begin{cases} -1 & v = i + 1 \\ 1 & v = i \\ 0 & \text{otherwise} , \end{cases} \tag{4.1.29}$$

$\bar{d}_{ss}$ is an element of $D_{\frac{1-\lambda^n}{1-\lambda}} D_{\delta^{-1}}$, that is

$$\bar{d}_{ss} = \begin{cases} n \quad \text{for} \quad s = 0 \\ \dfrac{1-\lambda_s^{\,n}}{1-\lambda_s} \binom{M-1}{s-1}(2s+b+1)/\binom{M+b+s-1}{s-1}(b+1) \quad , \quad s=1,2,\ldots,M \ , \end{cases} \tag{4.1.30}$$

d_{vv} is an element of D_V defined in (4.1.26).

From (4.1.29)

$$\begin{aligned} \sum_{v=0}^{M} Q_{s-1}(v-1)d_{vv}c_{vi} &= -\ Q_{s-1}(i)d_{i+1i+1}+Q_{s-1}(i-1)d_{ii} \\ &= -\ \Delta_{(i)}[d_{ii}Q_{s-1}(i-1)] \ , \quad i \neq M \\ &= Q_{s-1}(M-1)d_{MM} \ , \quad i = M \ . \end{aligned} \tag{4.1.31}$$

Hence (4.1.28) becomes

$$\mathcal{E}(n_i) = \begin{cases} -\ \Delta_{(i)} \displaystyle\sum_{w=0}^{k} \{ \sum_{s=0}^{M} Q_{s-1}(w-1)\bar{d}_{ss}[d_{ii}Q_{s-1}(i-1)]\} \ , \quad i \neq M \\ d_{MM} \displaystyle\sum_{w=0}^{k} [\sum_{s=0}^{M} Q_{s-1}(w-1)\bar{d}_{ss}Q_{s-1}(M-1)] \ , \quad i = M \end{cases} \tag{4.1.32}$$

recalling that the c_{kw} are 1 for $w \leq k$. Thus given the initial state k we have the expression for $\mathcal{E}(n_i)$, where again n_i is the total number of times state i is observed in a realization of the Markov chain. Note that in the final summation step of $\mathcal{E}(n_i)$ Corollary 9.1.4 can be used, i.e.,

$$\sum_{w=0}^{k} Q_{s-1}(w-1) = \frac{[(M-k)k[Q_{s-1}(k-1)-Q_{s-1}(k)]+bkQ_{s-1}(k-1)]}{s(s-1+b)} .$$

It is obvious from general reasoning that $\sum_{i=0}^{M} \mathcal{E}(n_i) = n$, the length of the chain, but this may also be verified from (4.1.32),

$$\begin{aligned}\sum_{i=0}^{M} \mathcal{E}(n_i) = &- \sum_{w=0}^{k} [\sum_{s=0}^{M} Q_{s-1}(w-1)\bar{d}_{ss} d_{MM} Q_{s-1}(M-1)] \\ &+ \sum_{w=0}^{k} [\sum_{s=0}^{M} Q_{s-1}(w-1)\bar{d}_{ss} d_{00} Q_{s-1}(-1)] \\ &+ \sum_{w=0}^{k} [\sum_{s=0}^{M} Q_{s-1}(w-1)\bar{d}_{ss} d_{MM} Q_{s-1}(M-1)] ,\end{aligned} \quad (4.1.33)$$

where the first two terms are the result of $\sum_{i=0}^{M-1} \mathcal{E}(n_i)$ [that is, $\sum_{a}^{b} \Delta f(x) = f(b+1) - f(a)$] and the last term is the expression for $\mathcal{E}(n_M)$. Thus

$$\sum_{i=0}^{M} \mathcal{E}(n_i) = \sum_{w=0}^{k} [\sum_{s=0}^{M} Q_{s-1}(w-1)\bar{d}_{ss} d_{00} Q_{s-1}(-1)] . \quad (4.1.34)$$

Note from (9.1.8) that $Q_{s-1}(-1) = 1$ for $s = 0$ and zero otherwise, thus

$$\sum_{i=0}^{M} \mathcal{E}(n_i) = \sum_{w=0}^{k} Q_{-1}(w-1)\bar{d}_{00}d_{00}$$

$$= \bar{d}_{00}d_{00}$$

$$= n \ , \quad [\text{cf., } (4.1.30), (4.1.26)] \ . \quad (4.1.35)$$

Having found an expression for $\mathcal{E}(n_i)$ we now discuss the expectations of φ and φ' (4.1.5). Recall from (3.1.27) that

$$\begin{aligned}\mathcal{E}(a_i) &= p_{ii+1}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \\ \mathcal{E}(b_i) &= p_{ii} \ [\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \\ \mathcal{E}(c_i) &= p_{ii-1}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \ .\end{aligned} \qquad (4.1.36)$$

We write

$$\mathcal{E}(\varphi) = \mathcal{E}\left(\frac{d \log L}{d\alpha_1}\right) = 0$$

and

$$\mathcal{E}(-\varphi') = \mathcal{E}\left(\frac{-d^2 \log L}{d\alpha_1{}^2}\right)$$

$$= \sum_{i=1}^{M-1} \frac{[\mathcal{E}(n_i)-p_{ki}^{(n-1)}]p_{ii+1}}{(1-\alpha_1)^2}$$

$$+ \sum_{i=1}^{M} \frac{i^2[\mathcal{E}(n_i)-p_{ki}^{(n-1)}](1-2iM^{-1})^2 p_{ii}}{[(M-i)-(1-\alpha_1)(i)(1-2iM^{-1})]^2} + \sum_{i=1}^{M} \frac{i^2[\mathcal{E}(n_i)-p_{ki}^{(n-1)}]p_{ii-1}}{[M-i(1-\alpha_1)]^2} ,$$

$$= \sum_{i=1}^{M-1} \frac{[\mathcal{E}(n_i)-p_{ki}^{(n-1)}]i(M-i)}{M^2(1-\alpha_1)} + \sum_{i=1}^{M} \frac{[\mathcal{E}(n_i)-p_{ki}^{(n-1)}]i^2(M-2i)^2}{M^3[(M-i)-(1-\alpha_1)i(1-2iM^{-1})]}$$

$$+ \sum_{i=1}^{M} \frac{[\mathcal{E}(n_i)-p_{ki}^{(n-1)}]i^3}{M^2[M-i(1-\alpha_1)]}$$

$$= \mathcal{I} \ . \qquad (4.1.37)$$

Exactly what this $\mathcal{I}$ means is a question. By the Crámer-Rao inequality (Kendall and Stuart, Vol. II, p. 8 et.seq.)

$$\mathrm{Var}(\hat{\alpha}_1) \geq [\frac{d}{d\alpha_1} \mathcal{E}(\hat{\alpha}_1)]^2/\mathcal{I} \ ,$$

but there seems little hope of finding either the bias term $\mathcal{E}(\hat{\alpha}_1) - \alpha_1$ or an exact expression for the variance of $\hat{\alpha}_1$ by theoretical methods. Some encouraging results, however, were obtained from a simulation study on the IBM 650 which is discussed later in this chapter.

Up to this point we have dealt with the case where n is fixed, [cf., for example (4.1.13), (4.1.30)]. We now discuss the situation where $n \to \infty$. In this case several of the above quantities have slightly different values.

We note that from (4.1.32), for $i \neq M$

$$\mathcal{E}(n_i) = -\Delta_{(i)} \sum_{w=0}^{k} \{ \sum_{s=0}^{M} Q_{s-1}(w-1)\bar{d}_{ss}[d_{ii}Q_{s-1}(i-1)]\}$$

$$= -\Delta_{(i)} \sum_{w=0}^{k} Q_{-1}(w-1)\bar{d}_{00}[d_{ii}Q_{-1}(i-1)]$$

$$-\Delta_{(i)} \sum_{w=0}^{k} \sum_{s=1}^{M} Q_{s-1}(w-1)\bar{d}_{ss}[d_{ii}Q_{s-1}(i-1)]$$

and provided $i \neq 0$

$$= 0 -\Delta_{(i)} \sum_{w=0}^{k} \sum_{s=1}^{M} Q_{s-1}(w-1)\bar{d}_{ss}[d_{ii}Q_{s-1}(i-1)] , \tag{4.1.38}$$

by the conventions (9.1.8).

Similarly, when $i = M$ we have

$$\mathcal{E}(n_M) = d_{MM} \sum_{w=0}^{k} \sum_{s=1}^{M} Q_{s-1}(w-1)\bar{d}_{ss}Q_{s-1}(M-1) . \tag{4.1.39}$$

From (4.1.30),

$$\lim_{n\to\infty} \bar{d}_{ss} = \frac{1}{1-\lambda_s}\binom{M-1}{s-1}(2s+b-1)/\binom{M+b+s-1}{s-1}(b+1) , \quad s \neq 0$$

and so

$$\lim_{n\to\infty} \mathcal{E}(n_i) = -\Delta_i \sum_{w=0}^{k} \sum_{s=1}^{M} Q_{s-1}(w-1)$$

$$[\frac{1}{1-\lambda_s}\binom{M-1}{s-1}(2s+b-1)/\binom{M+b+s-1}{s-1}(b+1)][d_{ii}Q_{s-1}(i-1)] \;,$$

$$0 < i < M$$

$$= d_{MM} \sum_{w=0}^{k} \sum_{s=1}^{M} Q_{s-1}(w-1)[\frac{1}{1-\lambda_s}\binom{M-1}{s-1}(2s+b-1)/\binom{M+b+s-1}{s-1}(b+1)]Q_{s-1}(M-1) \;,$$

$$i = M \;.$$

(4.1.40)

Note that Corollary 9.1.4 can be used in summing $\sum_{w=0}^{k} Q_{s-1}(w-1)$.

Of course, as $n \to \infty$, $\mathcal{E}(n_0) \to \infty$, but this does not enter into the formula for $\underset{\sim}{I}$ (4.1.37), nor are a_0, b_0 involved in the function φ and φ' (4.1.5) used in obtaining $\hat{\alpha}_1$. For $n \to \infty$ we can now write (4.1.37) as

$$\lim_{n\to\infty} \underset{\sim}{I} = \lim_{n\to\infty} [\sum_{i=1}^{M-1} \frac{\mathcal{E}(n_i)\,i(M-i)}{M^2(1-\alpha_1)} + \sum_{i=1}^{M} \frac{\mathcal{E}(n_i)\,i^2(M-2i)^2}{M^3[(M-i)-(1-\alpha_1)i(1-2iM^{-1})]}$$

$$+ \sum_{i=1}^{M} \frac{\mathcal{E}(n_i)\,i^3}{M^2[M-i(1-\alpha_1)]} \;] \;, \qquad (4.1.41)$$

noting that $p_{ki}^{(n-1)} \to 0$, $i \neq 0$, $p_{ki}^{(n-1)} \to 1$, $i = 0$. We now discuss the asymptotic behavior of the estimate and manner of conducting the experiment.

Keeping M, k fixed but with $n \to \infty$ it is clear from general considerations that once absorption has occurred, nothing of value is obtained by prolonging the experiment. With probability one, only a finite number of useful observations will be obtained, and no asymptotic theory of consistency or normality of the estimate $\hat{\alpha}_1$, will hold. For this situation we postulate the following theorems whose validity are very strongly felt but proofs of which have not been found. An outline of the anticipated proofs is presented. It will be noted that the gaps are a result of inadequacies which exist in the inference theory of positively regular Markov chains. Before presenting the theorems it could be mentioned that the current difficulties in making inferences from observations on a single absorbing Markov chain can be handled by performing independent replicated experiments. Such experiments are discussed in the next section. The following theorems are stated for the particular situation studied in this chapter. It is felt, however, that more general theorems hold.

(e) Some Postulated Theorems. Postulated Theorem 4.1.6. For the transition matrix P , (4.1.2), with one absorbing state, 0 , and no other closed sets, suppose that a realization is commenced at state $k = M$ and that $n = M$ observations are taken. Then, provided $\hat{\alpha}_1$ is the maximum likelihood estimate of α_1, the distribution of $\hat{\alpha}_1$ is asymptotically normal with mean α_1, and variance $-[\mathcal{E}(\frac{d^2 \log L}{d\alpha_1^2})]^{-1} = I^{-1}$ as $M \to \infty$, where

$$L = \text{likelihood} = \prod_i p_{ii+1}^{a_i}\, p_{ii}^{b_i}\, p_{ii-1}^{c_i} .$$

Proof outline: By the assumption that $n = k = M$, and since only unit transitions are possible in P, the absorbing state, 0, cannot be reached in the $M - 1$ steps after the initial observation $x_1 = M$. Thus it is immaterial how the elements p_{0j} in P are defined. Consider, therefore, a transition matrix

$$P^* = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ i \\ \vdots \\ M-1 \\ M \end{array}
\begin{array}{c} \begin{array}{ccccccc} 0 & 1 & \cdot & \cdot & \cdot & M-1 & M \end{array} \\
\begin{bmatrix}
0 & 0 & 0 & \cdots & & 0 & 1 \\
\ddots & \ddots & \ddots & & & & \\
 & & & & & & \\
 & p_{ii-1} & p_{ii} & p_{ii+1} & & & \\
 & & \ddots & \ddots & \ddots & & \\
 & & & & & & \\
0 & 0 & 0 & \cdots & 0 & \alpha_1 & 1-\alpha_1
\end{bmatrix} \end{array}$$

where from state 0 an instantaneous return to state M occurs. While P was <u>regular</u> but absorbing, P^* is <u>positively regular</u> and satisfies Billingsley's conditions (Section 2.3).

Consider now two experiments, one of which is performed with P as model, the other with P^* as model, and for each the initial state is $k = M$. Denote the maximum likelihood estimates by $\hat{\alpha}_1$ and $\hat{\alpha}_1^*$ respectively. Then for the positively regular chain, we have asymptotic normality according to Theorem 2.3.2, that is,

$$\lim_{\substack{n\to\infty \\ M \text{ fixed}}} \Pr\{\sqrt{n}\,(\hat{\alpha}_1^* - \alpha_1) \leq y\} = \Phi(y/\sigma_M) \tag{4.1.42}$$

where Φ is the standard normal distribution function,

$$\sigma^2 = \lim_{n\to\infty} n\, I^{*-1} \quad ,$$

and

$$I^* = -\,\mathcal{E}\left(\frac{d^2 \log L^*}{d\alpha_1^{\,2}}\right) \quad ,$$

with L^* being the likelihood of n observations drawn from P^* .

Suppose now that the convergence (4.1.42) is uniform over M , and that $\sigma_M^2 \to \sigma$, with $0 < \sigma < \infty$, as $M \to \infty$.

Then from (4.1.42),

$$\lim_{M=n\to\infty} \Pr\{\sqrt{n}(\hat{\alpha}_1{*}-\alpha_1)\leq y\} = \lim_{M\to\infty}\ \lim_{n\to\infty} \Pr\{\sqrt{n}(\hat{\alpha}_1^*-\alpha_1)\leq y\}$$

$$= \Phi(y/\sigma) \quad . \tag{4.1.43}$$

But on the left of (4.1.43), we have $\hat{\alpha}_1{*}$ drawn from a chain with $k = M = n$, which has identically the distribution of $\hat{\alpha}_1$ made under the same conditions on the P matrix. Thus, if (4.1.43) is valid, we have

$$\lim_{M=n\to\infty} \Pr\{\sqrt{n}(\hat{\alpha}_1-\alpha_1) \leq y\} = \Phi(y/\sigma) \ . \tag{4.1.44}$$

Further, if the similar interchange of double limits

$$\lim_{M=n\to\infty} n(\mathcal{I}\,{*})^{-1} = \lim_{M\to\infty}\ \lim_{n\to\infty} n(\mathcal{I}\,{*})^{-1} = \lim_{M\to\infty} \sigma_M^2 = \sigma^2 \tag{4.1.45}$$

holds, we have (because for $k = M = n$, $\mathcal{I}\,{*} = \mathcal{I}$)

$$\lim_{M=n\to\infty} n\,\mathcal{I}^{-1} = \lim_{M=n\to\infty} n(\mathcal{I}\,{*})^{-1} = \sigma^2 \ . \tag{4.1.46}$$

From (4.1.44) we have that $\hat{\alpha}_1$ is asymptotically normal with mean α_1 , and variance (from (4.1.46)) equal

$$\mathcal{I}^{-1} = -\left[\mathcal{E}\left(\frac{d^2\log L}{d\alpha_1^{\,2}}\right)\right]^{-1} \ .$$

Note that the difficulties in the theorem are due to unresolved questions about limiting operations (4.1.43) and (4.1.46) in the positively regular theory.

Postulated Theorem 4.1.7. For the transition matrix P, (4.1.2), with one absorbing state, 0, and no other closed sets, suppose that a realization is commenced at $k = M$ and continued until the absorbing state is first reached. Then, provided $\hat{\alpha}_1$ is the maximum likelihood estimate of α_1, the distribution of $\hat{\alpha}_1$ is asymptotically normal with mean α_1 and variance $\lim_{n\to\infty} \mathcal{I}^{-1}$ as $M \to \infty$, where $\mathcal{I}$ (4.1.37) is the fixed sample size (n) information.

Proof outline. As in Theorem 4.1.6, it makes no difference how the elements p_{0j} are defined since by definition of the experimental procedure, the realization is terminated as soon as state 0 is reached. Thus the estimate $\hat{\alpha}_1$ has the same distribution properties as $\hat{\alpha}_1{}^*$ made on the P^* process under the same conditions.

We postulate the following result for the positively regular chain estimate $\hat{\alpha}_1{}^*$:

$$\lim_{M\to\infty} \Pr\{(\hat{\alpha}_1{}^* - \alpha_1)a(M) \leq y\} = \Phi(y/\sigma) \quad , \qquad (4.1.47)$$

where $a(M)$ is some standardizing factor, perhaps, but not necessarily, $\sqrt{M}$, and

$$\sigma^2 = \lim_{M\to\infty} \{[a(M)]^2(\mathcal{I}^*)^{-1}\} \text{ , with } 0 < \sigma^2 < \infty \text{ .}$$

Here $\mathcal{I}^*$ is the information for the sequential stopping rule on chain P^*, and is identical to the same quantity for the P matrix, namely

$$\mathcal{I}^* = \lim_{n\to\infty} \mathcal{I} \quad . \tag{4.1.48}$$

It follows from (4.1.47), if true, that the sequential stopping rule applied to the P chain estimate yields

$$\lim_{M\to\infty} \Pr\{(\hat{\alpha}_1 - \alpha_1)a(M) \leq y\} = \Phi(y/\sigma) \quad , \tag{4.1.49}$$

where $\quad \sigma^2 = \lim_{M\to\infty} \{[a(M)]^2 \lim_{n\to\infty} \mathcal{I}^{-1}\}$.

This proves the theorem.

Again, we see that the difficulties inherent in (4.1.47) relate to the positively regular case.

Corollary 4.1.8. If, in Theorem 4.1.7, $a(M) \to \infty$ as $M \to \infty$, then $\hat{\alpha}_1$ is a consistent estimate of α_1.

Proof. The proof is immediate from the fact that $\hat{\alpha}_1$ is asymptotically unbiased and has variance of order $O(\frac{1}{a(M)^2})$ as $M \to \infty$ by Theorem 4.1.7 itself. It is felt, however, that the corollary holds regardless of the validity of the postulated Theorem 4.1.7. That is, it may be possible to find a proof of the corollary without relying on Theorem 4.1.7 specifically.

4.2 Estimating Mutation Rate from a Single Chain (Simulation Study)

4.2.1 Background

In connection with the postulated Theorem 4.1.7 and its Corollary 4.1.8, a simulation study on the IBM 650 was performed, the summary results of which appear in the following figures and tables, the more extensive results being in Appendices II, III. Five values of M were studied, $M = 2, 4, 6, 10$, and 20. The program was written with $\alpha_1 = 0.1$. For each value of M, 500 independent realizations were generated in the following manner. In all cases, the initial state was $k = M$ and the chain was continued until the absorbing state zero occurred. The 500 maximum likelihood estimates of $\alpha_1 = 0.1$ solved from the data of these realizations appear in Appendix II listed in increasing order. Note that $M = 4$ and 6 have 502 values, the last two values being $\hat{\alpha}_1 = 1.0$. The corresponding experiments did not satisfy the conditions of Theorem 4.1.2 (that is, b_M and all a_i were zero for these experiments) resulting in a maximum of the likelihood at $\alpha_1 = 1$, not, however, at a turning point. Two further realizations were made to give in all 500 replicates of a form yielding $\hat{\alpha}_1$ values satisfying the likelihood equations. However, for $M = 2$ there were 16 such extreme realizations, and these were not replaced. For $M = 10$ and 20, all realizations provided admissible likelihood equation estimates.

In Table 4.1.2 $\bar{\hat{\alpha}}_1$ is the mean of the 500 estimates of $\hat{\alpha}_1$ and $s^2_{\hat{\alpha}_1}$ is their sample variance. The numbers which appear in parentheses for $M = 4, 6$ are based on the 502 estimates. These values are included for general interest.

The figures show the observed distribution of all estimates obtained. Note that there are gaps in what one might expect to be virtually continuous distributions. These are especially pronounced in Fig. 1 where $M = 2$. The extreme right bar is for the 16 values of one coming from realizations of the form 2 1 .... 1 0 with state 2 occurring once only. The second bar from the right is for 26 values of $\hat{\alpha}_1 = 0.577350$ which arose from experiments of the type 2 2 1 .... 1 0 . The third tall bar from the right is for 29 values of $\hat{\alpha}_1 = 0.390388$ which arose from experiments of the type 2 2 2 1 .... 1 0 . Thus $M = 2$ has many peculiar characteristics due to the comparatively limited number of possible realizations. These peculiarities become less pronounced as M increases.

Table 4.1.2

M	$\bar{\hat{\alpha}}_1$ ($\alpha_1 = 0.1$)	Estimated Bias $\bar{\hat{\alpha}}_1 - \alpha_1$	Estimated Variance $s^2_{\hat{\alpha}_1}$	Postulated Asymptotic Variance $\lim\limits_{n\to\infty} I^{-1}$	% Error $\left(\frac{s^2_{\hat{\alpha}_1} - \lim\limits_{n\to\infty} I^{-1}}{s^2_{\hat{\alpha}_1}}\right) \cdot 100$
2	0.212050	0.112050	0.038427	0.004905	87.2355
4	0.168281	0.068281	0.014846	0.003036	79.5500
	(0.171595)		(0.017537)		
6	0.153653	0.053657	0.010560	0.002479	76.5246
	(0.157025)		(0.013366)		
10	0.131192	0.031192	0.003421	0.002025	40.8068
20	0.114452	0.014452	0.001481	0.001496	(-1.0128)

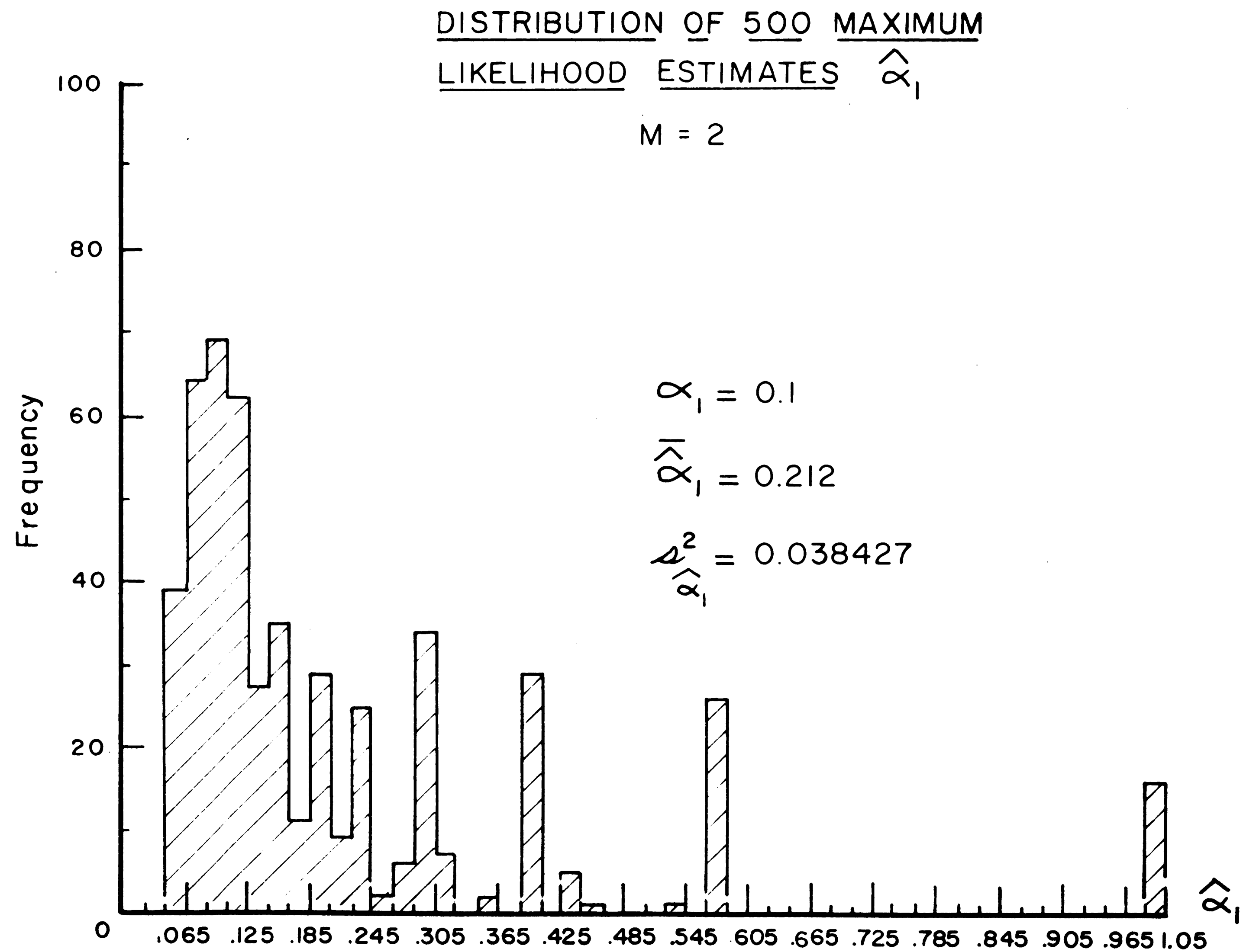
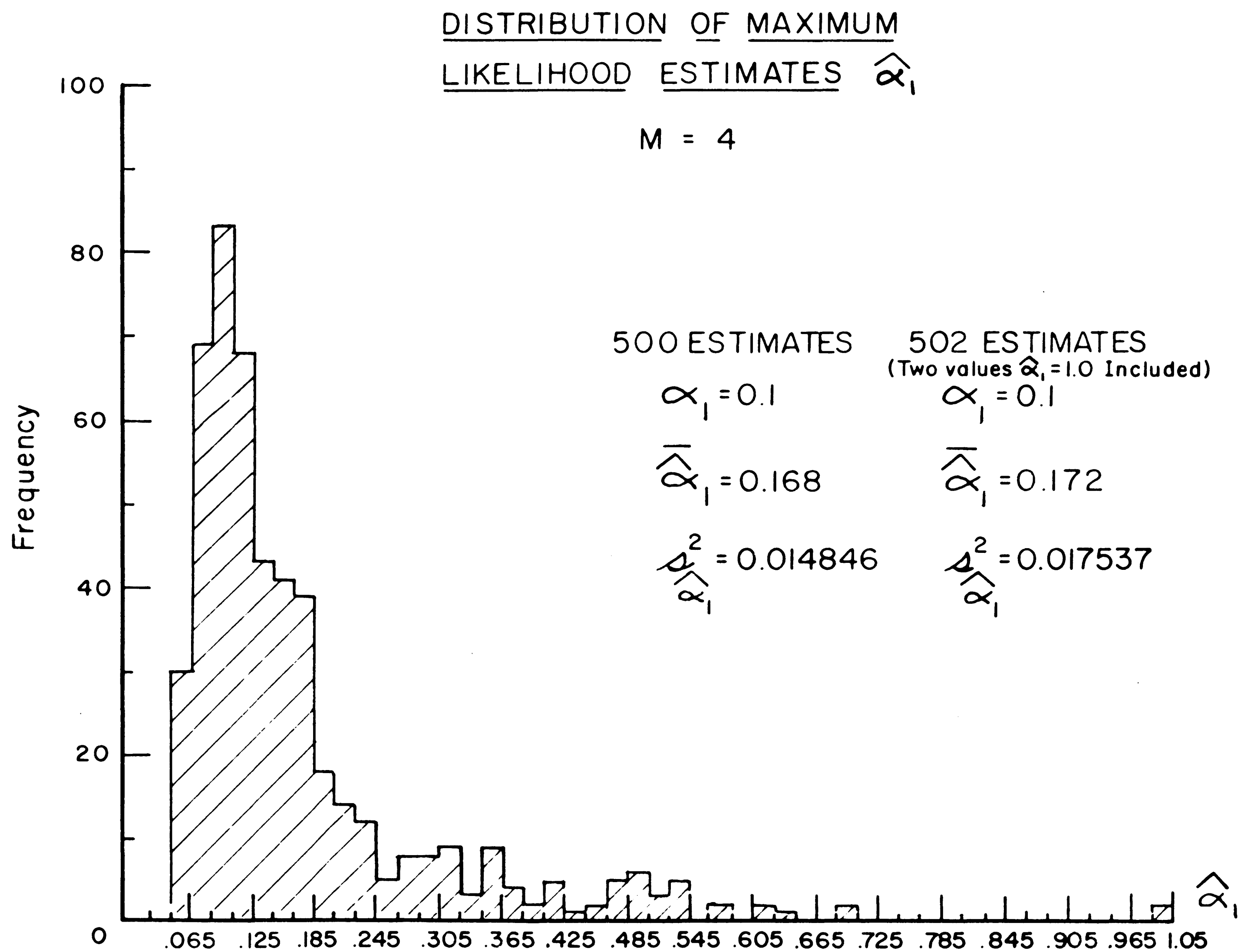

DISTRIBUTION OF 500 MAXIMUM
LIKELIHOOD ESTIMATES $\hat{\alpha}_1$
M = 2
$\alpha_1 = 0.1$
$\bar{\hat{\alpha}}_1 = 0.212$
$s^2_{\hat{\alpha}_1} = 0.038427$
Frequency
100
80
60
40
20
0
.065 .125 .185 .245 .305 .365 .425 .485 .545 .605 .665 .725 .785 .845 .905 .965 1.05
$\hat{\alpha}_1$

Fig. 1



DISTRIBUTION OF MAXIMUM
LIKELIHOOD ESTIMATES $\hat{\alpha}_1$
M = 4
500 ESTIMATES
$\alpha_1 = 0.1$
$\bar{\hat{\alpha}}_1 = 0.168$
$s^2_{\hat{\alpha}_1} = 0.014846$
502 ESTIMATES
(Two values $\hat{\alpha}_1 = 1.0$ Included)
$\alpha_1 = 0.1$
$\bar{\hat{\alpha}}_1 = 0.172$
$s^2_{\hat{\alpha}_1} = 0.017537$
Frequency
100
80
60
40
20
0
.065 .125 .185 .245 .305 .365 .425 .485 .545 .605 .665 .725 .785 .845 .905 .965 1.05
$\hat{\alpha}_1$

Fig. 2

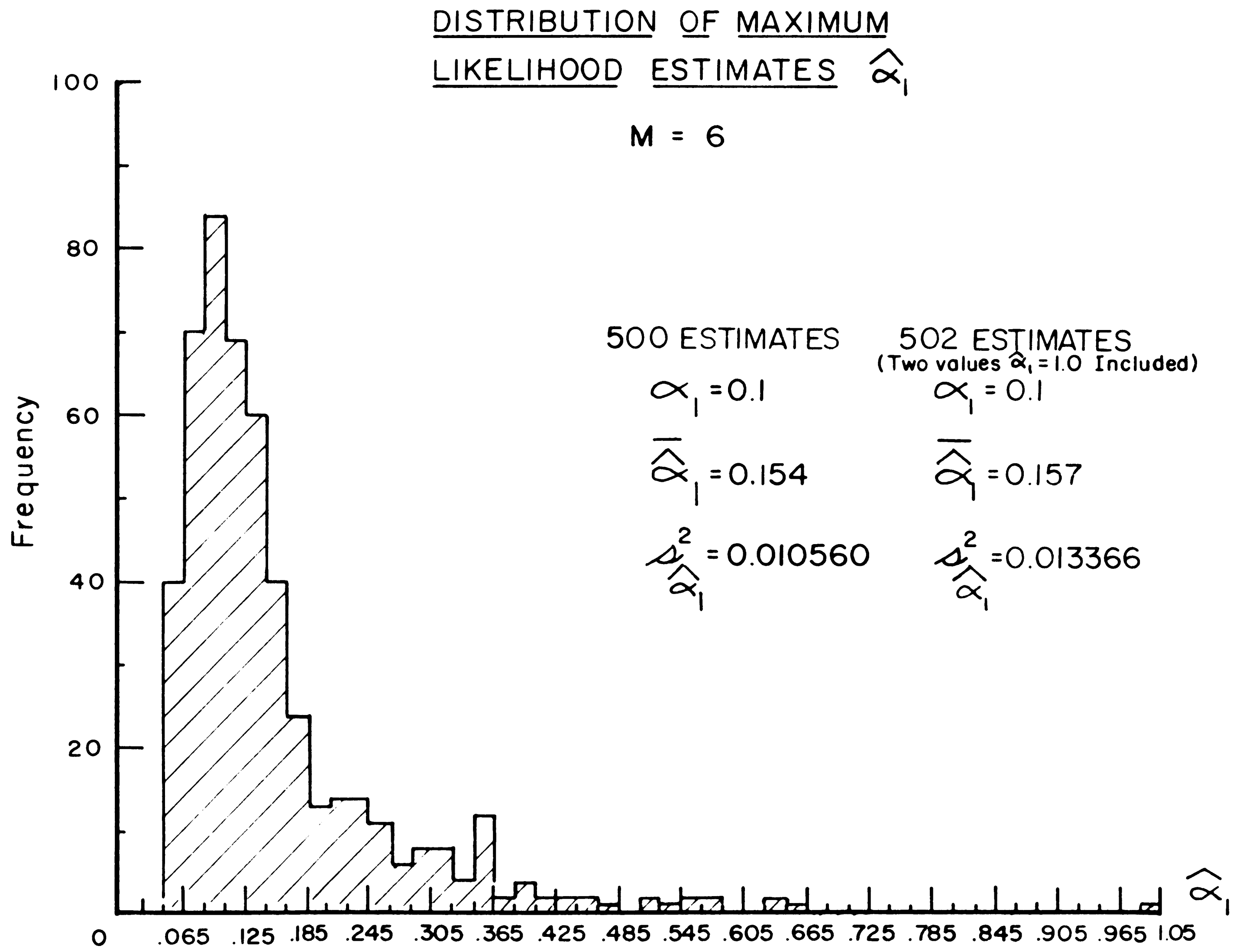

DISTRIBUTION OF MAXIMUM
LIKELIHOOD ESTIMATES $\hat{\alpha}_1$
M = 6
500 ESTIMATES
$\alpha_1 = 0.1$
$\bar{\hat{\alpha}}_1 = 0.154$
$s^2_{\hat{\alpha}_1} = 0.010560$
502 ESTIMATES
(Two values $\hat{\alpha}_1 = 1.0$ Included)
$\alpha_1 = 0.1$
$\bar{\hat{\alpha}}_1 = 0.157$
$s^2_{\hat{\alpha}_1} = 0.013366$
Frequency
100
80
60
40
20
0
.065 .125 .185 .245 .305 .365 .425 .485 .545 .605 .665 .725 .785 .845 .905 .965 1.05
$\hat{\alpha}_1$

Fig. 3

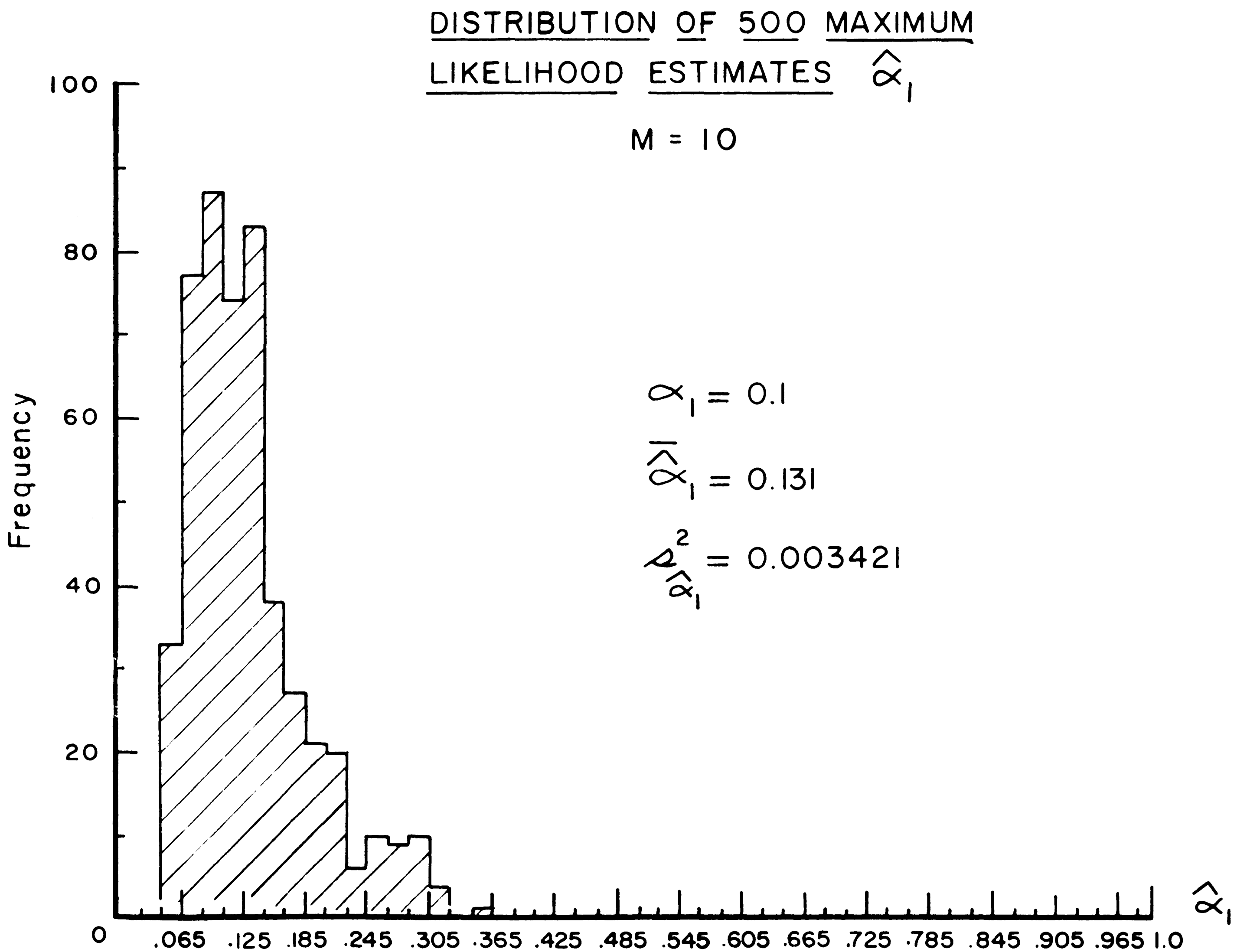
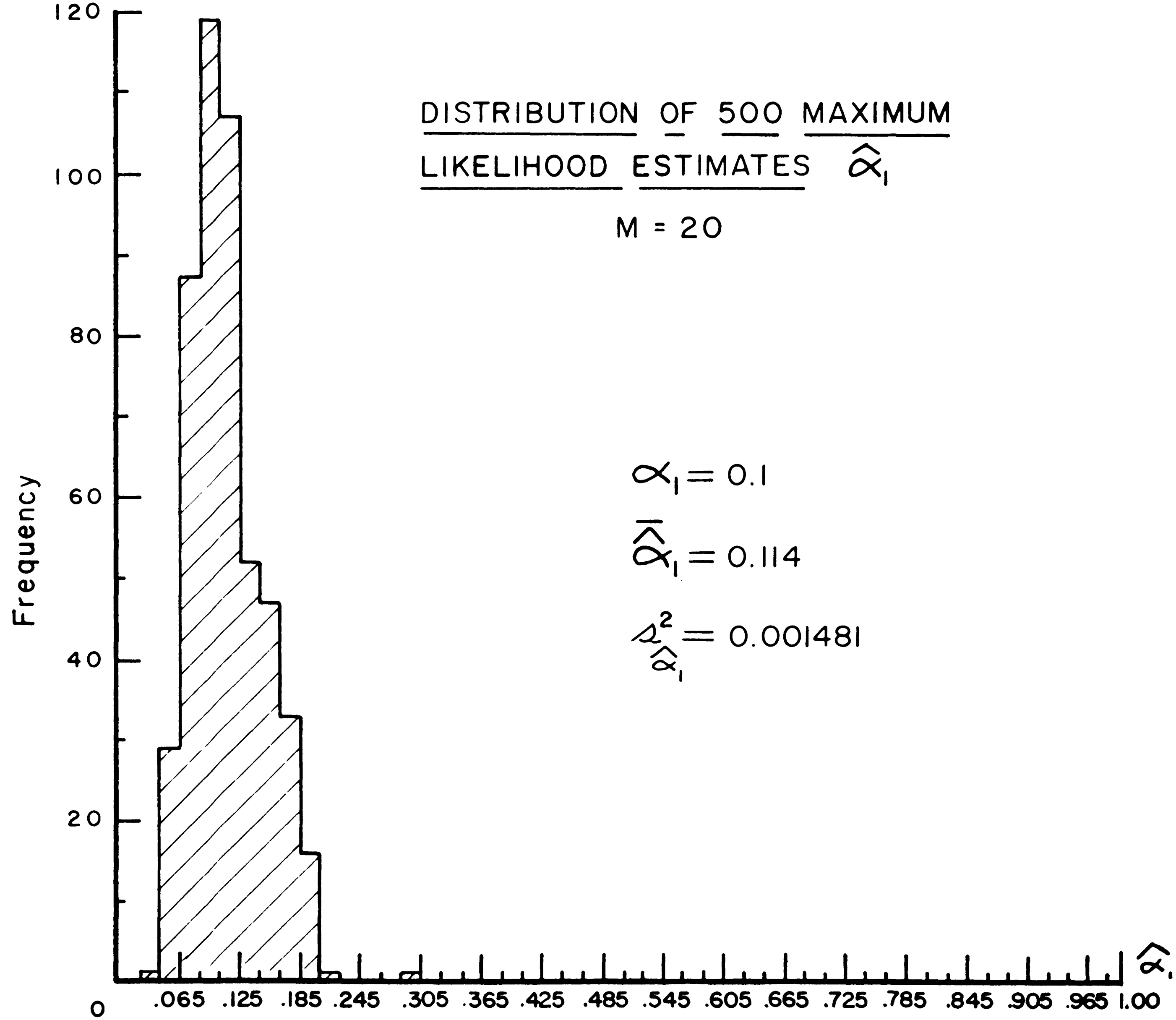

DISTRIBUTION OF 500 MAXIMUM
LIKELIHOOD ESTIMATES $\hat{\alpha}_1$
M = 10
$\alpha_1 = 0.1$
$\bar{\hat{\alpha}}_1 = 0.131$
$\sigma^2_{\hat{\alpha}_1} = 0.003421$
Frequency
100
80
60
40
20
0
.065 .125 .185 .245 .305 .365 .425 .485 .545 .605 .665 .725 .785 .845 .905 .965 1.0
$\hat{\alpha}_1$

Fig. 4



DISTRIBUTION OF 500 MAXIMUM LIKELIHOOD ESTIMATES $\hat{\alpha}_1$
M = 20
$\alpha_1 = 0.1$
$\bar{\hat{\alpha}}_1 = 0.114$
$s^2_{\hat{\alpha}_1} = 0.001481$
Frequency
120
100
80
60
40
20
0
.065 .125 .185 .245 .305 .365 .425 .485 .545 .605 .665 .725 .785 .845 .905 .965 1.00
$\hat{\alpha}_1$

Fig. 5

4.2.2 Conclusions from the simulation study

From the figures we note that skewness decreases as M increases and the empirical distributions of the $\hat{\alpha}_1$ appear to have a mode close to the true value $\alpha_1 = 0.1$. Moreover, from the figures and Table 4.1.2 it is apparent that the bias and variance of the estimates decreases rapidly as M increases. This gives very strong evidence that Corollary 4.1.8 is correct, and that the estimate is consistent as M increases.

For the considerably sharper Theorem 4.1.7, we find that the observed variances $s^2_{\hat{\alpha}_1}$ and the postulated asymptotic variance $\lim_{n\to\infty} I^{-1}$ (4.1.41) are in extraordinarily good agreement for $M = 20$, but less so for smaller population sizes. This is to be expected if the Theorem 4.1.7 is correct, but we do not claim that the realizations generated to date are sufficiently convincing that in fact $\mathrm{var}(\hat{\alpha}_1)$ and $\lim_{n\to\infty} I^{-1}$ are asymptotically equal. Higher values of M need be investigated before such a conclusion could be established beyond doubt. Similarly, the asymptotic normality property is not yet established, although the decrease in skewness is suggestive. A χ^2 goodness of fit test on

$M = 20$ data still indicates a significant departure from normality. Finally, the postulated standardizing factor $a(M)$ of Theorem 4.1.7 could be established by plotting $\lim_{n\to\infty} I^{-1}$ or $s^2_{\hat{\alpha}_1}$ as a function of M and observing the rate of approach to zero. However, here again values larger than $M = 20$ would be needed. It was not possible to examine larger populations on the IBM 650 because of the prohibitive amount of time required.

4.2.3 Comments on the design of the experiment

In discussing the non-absorbing, two mutation case (Section 3.2) it was found that, per observation, the most efficient experiment was one carried out on the smallest possible population ($M = 1$), but per generation, the same conclusion did not hold. In the present situation with ultimate absorption being a certainty, the expected number of birth-death events required for the transition from state M to state 0 can be calculated either by finding the limit

$$\lim_{n\to\infty} \sum_{i=1}^{M} \mathcal{E}(n_i) \tag{4.1.50}$$

using the Hahn polynomial expression (4.1.32), or by a method proposed by Kemeny and Snell (1960). The latter consists of evaluating the inverse $(I - P_\Delta)^{-1}$ where I is the $M \times M$ identity matrix and P_Δ is the $M \times M$ matrix obtained by deleting the 0 column and row of P (4.1.2), thus leaving the transient state probabilities only. The elements in the (M,i) position of $(I - P_\Delta)^{-1}$ are the required $\lim\limits_{n\to\infty} \mathcal{E}(n_i)$, and summing along the last row of $(I - P_\Delta)^{-1}$ yields the expected number of transitions. The expected number of observations $\mathcal{E}(N)$ is, of course, one more than this quantity allowing for the transition from state 1 to state 0. For our triple diagonal matrix P, in fact all elements below the diagonals in $(I - P_\Delta)^{-1}$ are equal to the diagonal elements themselves, and so we present in Appendix III, only the upper triangular portions of $(I - P_\Delta)^{-1}$ for $M = 2, 4, 6, 10$ and 20. The individual values have been used to calculate $\lim\limits_{n\to\infty} I^{-1}$ (4.1.41) as tabulated in Table 4.1.2.

The expected length of chain to absorption, $\mathcal{E}(N)$, is given in Table 4.1.3, together with the bias and variance estimates calculated on a per generation basis. It is clear that not only are the larger populations more efficient, in total, if carried to absorption (see Table 4.1.2) but also

Table 4.1.3

M	$\mathcal{E}(N)$ Expected Length of Chain to Absorption	$\frac{\text{Bias}}{\mathcal{E}(N)/M}$	$\frac{s^2_{\hat{\alpha}_1}}{\mathcal{E}(N)/M}$
2	21.818181 + 1 (23)	0.009821	0.003368
4	49.868406 + 1 (51)	0.005369	0.001674
6	82.618032 + 1 (84)	0.003850	0.000758
10	158.753792 + 1 (160)	0.001953	0.000214
20	391.368270 + 1 (392)	0.000737	0.000075

on a per observation and a per generation basis. It is conceded that there is a slight increase in the ratio $\mathcal{E}(N)/M = \mathcal{E}$(generations to absorption) as M increases and if the number of generations were kept fixed at a level where small populations reach absorption on average but larger ones do not, the estimated bias and variance for the latter would be somewhat underestimating the true values. However, the effect is thought to be sufficiently small not to nullify the conclusion. Again further calculations would be interesting.

4.3 Replicated Experiments

One method of overcoming the difficulties encountered in attempting to make inferences from a single long realization of an absorbing Markov chain is to perform independent replicated experiments. Such experiments are discussed in this section.

4.3.1 Geometric Stopping Rule

Suppose we consider the following. Let the initial state be $k = M$ (i.e., consisting of all "a"'s corresponding to a genetic pure line) and observe the population at equal time intervals until the $M - 1$ state occurs for the first time, then stop. The number of times that the chain remains in state M, b_M, has the geometric distribution

$$f(b_M) = p_{MM}^{b_M}\, p_{MM-1} = (1 - \alpha_1)^{b_M}\, \alpha_1 \quad , \tag{4.3.1.1}$$

where $b_M = 0, 1, 2, \ldots$. For R independent replicate populations we can write the joint probability function (likelihood function), where for convenience we shall write $b_M(r)$ for the value of b_M in the r-th replicate,

$$L = (1 - \alpha_1)^{\sum_{r=1}^{R} b_M(r)}\, \alpha_1^{R} \quad . \tag{4.3.1.2}$$

The log-likelihood is

$$\log L = \sum_{r=1}^{R} b_M(r) \log(1-\alpha_1) + R \log \alpha_1 \tag{4.3.1.3}$$

and

$$\frac{d \log L}{d\alpha_1} = - \sum_{r=1}^{R} \frac{b_M(r)}{1-\alpha_1} + \frac{R}{\alpha_1} \quad , \tag{4.3.1.4}$$

thus the maximum likelihood estimate of α_1 is

$$\hat{\alpha}_1 = \frac{R}{\sum_{r=1}^{R} b_M(r) + R} \,. \tag{4.3.1.5}$$

This estimate, however, is a biased one. Let us then look at an estimate which is unbiased and whose distribution theory is known. $\sum_{r=1}^{R} b_M(r)$ is distributed as the negative binomial

$$f\left(\sum_{r=1}^{R} b_M(r)\right) = \binom{\sum_{r=1}^{R} b_M(r)+R-1}{R-1}(1-\alpha_1)^{\sum_{r=1}^{R} b_M(r)}\alpha_1^{R} \tag{4.3.1.6}$$

because the convolution of independent negative binomial (and in particular, geometric) variables is again a negative binomial. The unbiased estimate for α_1 in this case (Haldane, 1943) is in our notation

$$\bar{\alpha}_1 = \frac{R-1}{\sum_{r=1}^{R} b_M(r)+R-1} \,. \tag{4.3.1.7}$$

Finney (1949) found an unbiased estimate of the variance of $\bar{\alpha}_1$. In our notation it is

$$\mathrm{var}(\overline{\alpha}_1) \doteq s^2 = \frac{\overline{\alpha}_1(1 - \overline{\alpha}_1)}{\sum_{r=1}^{R} b_M(r) + R - 2} \quad . \qquad (4.3.1.8)$$

A normal approximation is a satisfactory indicator of the error of estimation of α_1 only when R is large. For small R a method of Finney's (1949) for reading exact confidence limits on α_1 directly from <u>Biometrika Table 41</u> is shown below.

It may not have been generally realized that methods and tables for determining exact confidence limits for binomial sampling may be adapted very easily to inverse binomial sampling, i.e., the negative binomial distribution. The proof of the following rule may be found in Finney (1949). A more explicit proof may be found in Bartko (1960). The rule may be stated:

(i) The upper limit on α_1 is found by entering <u>Biometrika Table 41</u> with $C = R - 1$, $n = \sum_{r=1}^{R} b_M(r) + R - 1$.

(ii) The lower confidence limit on α_1 is found by entering the table with $C = R$, $n = \sum_{r=1}^{R} b_M(r) + R$, where n is the notation used in <u>Table 41</u>.

Consider the following example which illustrates the principles of the above discussion. A chain with 11 states, i.e., $M = 10$ (actually for this geometric stopping rule the size of M, aside from its being constant, is immaterial) was considered. α_1 was arbitrarily set at 0.1. The initial state was $k = M = 10$ as stated above and when state $M - 1 = 9$ was first observed the chain was terminated. By use of random number tables where $p_{10,9} = 0.1$ and $p_{10,10} = 0.9$, $R = 30$ such independent chains were constructed. The results were

30 times in state 9

294 times in state 10

264 times the transition from state 10 to 10 occurred .

Thus $\sum_{r=1}^{R=30} b_M(r) = 264$ and from (4.3.1.7)

$$\bar{\alpha}_1 = 0.0989 \qquad (4.3.1.9)$$

and from (4.3.1.8)

$$s = 0.01747 \quad . \qquad (4.3.1.10)$$

Using a normal approximation, 95% confidence limits on α_1 are

$$\bar{\alpha}_1 - (1.96)s < \alpha_1 < \bar{\alpha}_1 + (1.96)s \quad ,$$

which for our case is

$$0.065 < \alpha_1 < 0.133 \quad . \tag{4.3.1.11}$$

Although it would appear in this case that R is sufficiently large for the approximation to hold, the rules stated above for Biometrika Table 41 will be used in this example for purposes of illustration. For the upper limit enter the table with $C = 29$, $n = 293$ and find approximately the value 0.14. For the lower bound enter the table with $C = 30$, $n = 294$, to which corresponds the value of about 0.07. Consequently, an exact two sided 95% confidence interval for α_1 is

$$0.07 < \alpha_1 < 0.14 \quad . \tag{4.3.1.12}$$

4.3.2 Fixed chain length

Let us here consider making inferences on α_1 from data obtained by observing R independent replicate chains each of the same finite length n . For an absorbing chain a finite, useful, n can be accomplished by setting $n \leq k$ so that absorption does not occur.

We estimate α_1 by using the Newton-Raphson scheme (4.1.6). The procedure is to replace a_i, b_i and c_i in φ and φ' (4.1.5) with

$$\sum_{r=1}^{R} a_i(r) \ , \quad \sum_{r=1}^{R} b_i(r) \quad \text{and} \quad \sum_{r=1}^{R} c_i(r) \quad \text{respectively,}$$

where for example $a_i(r)$ is the value of a_i in the r-th replicate. In obtaining the estimate of α_1 it is immaterial whether all replications have the same initial state k or not. However, in computing $\mathcal{I}$ (4.1.37) we need to find $\mathcal{E}(n_i)$ which from (4.1.32) we see depends on k. Suppose that the initial state is the same for all replications. Recall from (4.1.36) for example that

$$\mathcal{E}(a_i) = p_{ii+1}[\mathcal{E}(n_i) - p_{ki}^{(n-1)}] \ . \tag{4.3.2.1}$$

Thus for our case where all initial states are the same,

$$\mathcal{E} \sum_{r=1}^{R} a_i(r) = R\mathcal{E}(a_i) \ . \tag{4.3.2.2}$$

Similar expressions hold for b_i and c_i. Thus from $\mathcal{I}$ (4.1.37) and (4.3.2.2) the variance element for the replicated experiments $(\mathcal{I}_R)^{-1}$ is

$$(\mathcal{I}_R)^{-1} = (R\,\mathcal{I})^{-1} \ , \tag{4.3.2.3}$$

where $\mathcal{I}$ now refers to the information (4.1.37) obtained from a single replicate, and we write for the replicated experiments with finite n and same initial state k that $(\hat{\alpha}_1 - \alpha_1)$ ($\hat{\alpha}_1$ the maximum likelihood estimate of α_1) will be asymptotically normal with mean zero and variance $(\mathcal{I}_R)^{-1}$ (4.3.2.3). These results hold as $R \to \infty$ by the usual theory for maximum likelihood estimates from independent experiments.

4.3.3 Chain length a random variable

Suppose that the chain length n for each replication is a random variable determined by some sequential stopping rule. Let us consider the absorbing state 0 as the stopping state. That is, we observe the chain until it is absorbed. And further, let us choose k the same for each replication and large enough (implies M large) so that we can regard n as very large and use $\mathcal{E}(n_i)$ as given by (4.1.40).

We replace $\mathcal{E}(n_i)$ in (4.1.41) with $\mathcal{E} \sum_{r=1}^{R} n_i(r)$ and since k is the same for all replications

$\mathcal{E} \sum_{r=1}^{R} n_i(r) = R\mathcal{E}(n_i)$. Thus (4.1.41) becomes for the replicated experiments

$$\mathcal{I}_R = R\,\mathcal{I} \quad . \tag{4.3.3.1}$$

Thus we can write $(\hat{\alpha}_1 - \alpha_1)$ will be asymptotically normal with mean zero and variance $(\mathcal{I}_R)^{-1}$ as R the number of independent replicates $\to \infty$.

4.4 Sample Calculations on the Absorbing Chain

Using (4.1.1) P for $M = 2$ is

$$P = \begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4}(1+\alpha_1) & \frac{1}{2} & \frac{1}{4}(1-\alpha_1) \\ 0 & \alpha_1 & 1-\alpha_1 \end{bmatrix} \end{array} \quad . \tag{4.4.1}$$

From (4.1.14) i.e.,

$$\begin{aligned} \lambda_j &= 1 - j[\tfrac{1}{2}\alpha_1 + \tfrac{1}{4}(j-1)(1-\alpha_1)] \\ \lambda_0 &= 1 \\ \lambda_1 &= (2-\alpha_1)/2 \\ \lambda_2 &= (1-\alpha_1)/2 \quad . \end{aligned} \tag{4.4.2}$$

From (4.1.17)

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1-b \end{bmatrix} \qquad (4.4.3)$$

$$\begin{matrix} (1) & (2) & (-b) & \text{column sums ,} \end{matrix}$$

where $(-1-b)$ in the (2,2) position $= Q_1(1;\ 0,\ b,\ 2)$

$$= \sum_{\ell=0}^{1} \frac{(-1)_\ell\,(-1)_\ell\,(2+b)_\ell}{(1)_\ell\,(-1)_\ell\,\ell!} \ . \qquad (4.4.4)$$

The figures at the bottom of the columns of Q are the column sums, that is from Corollary 9.1.5

$$\sum_{x=0}^{M} Q_{m-1}(x-1) = \frac{M(-1)^{m-1}\,(b)_{m-1}}{m!}\ , \quad m \neq 0\ .$$

For example, $-\,b = \sum_{x=0}^{2} Q_1(x-1) = \dfrac{2(-1)\,b}{2!}$. (4.4.5)

Note: Recall from (2.1.3) that the convention in this thesis is to call the first element position of a matrix the (0,0) position. Also recall from (4.1.17) that the Hahn polynomial $Q_{j-1}(i-1)$ occupies the (i,j) position, $i,j = 0, 1, \ldots, M$ of the matrix Q . Thus in finding (4.4.5) which by our convention is the sum of column 2 in

(4.4.3) we used $\sum_{x=0}^{2} Q_1(x-1)$ since $Q_1(x-1)$ denotes the second column of Q .

From (4.1.24)

$$Q^{-1} = D_{\delta^{-1}} Q' D_V . \qquad \text{For } M = 2$$

$$Q^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (b+1)^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1-b \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1+b}{2+b} & 0 \\ 0 & 0 & \frac{1}{2+b} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1+b}{2+b} & \frac{1}{2+b} \\ 0 & \frac{1}{2+b} & \frac{-1}{2+b} \end{bmatrix} . \tag{4.4.6}$$

As a sample calculation on $\mathcal{E}(n_i)$ for $M = 2$, n fixed and initial state $k = M = 2$, we have from (4.1.27)

$\mathcal{E}(n_i) =$ (2,i) element in $CQD_{\frac{1-\lambda^n}{1-\lambda}} D_{\delta^{-1}} Q' D_V C^{-1}$ and

from (4.1.32) for $i = 2 = M$

$$\mathcal{E}(n_2) = d_{22} \sum_{w=0}^{2} [\sum_{s=0}^{2} Q_{s-1}(w-1)\bar{d}_{ss}Q_{s-1}(1)]$$

$$= d_{22} \sum_{w=0}^{2} [Q_{-1}(w-1)\bar{d}_{00}Q_{-1}(1) + Q_0(w-1)\bar{d}_{11}Q_0(1) + Q_1(w-1)\bar{d}_{22}Q_1(1)] , \quad (4.4.7)$$

by (4.4.3) and (9.1.9) we write (4.4.7)

$$d_{22} \sum_{w=0}^{2} [Q_0(w-1)\bar{d}_{11} - Q_1(w-1)\bar{d}_{22}(1+b)] , \quad (4.4.8)$$

where from (4.4.3) we note that $Q_1(1) = -(1+b)$,

$$= d_{22} \{\bar{d}_{11} \sum_{w=0}^{2} Q_0(w-1) - \bar{d}_{22}(1+b) \sum_{w=0}^{2} Q_1(w-1)\}$$

$$= d_{22} [2\bar{d}_{11} - \bar{d}_{22}(1+b)(-b)] , \quad (4.4.9)$$

where from (4.4.5) $\sum_{w=0}^{2} Q_1(w-1) = -b$.

From (4.1.26)

$$d_{00} = 1$$

$$d_{11} = (1+b)/(2+b)$$

$$d_{22} = 1/(2+b) \quad ;$$

from (4.1.30)

$$\bar{d}_{00} = n$$

$$\bar{d}_{11} = 2/\alpha_1[1 - (\tfrac{1}{2})^n(2 - \alpha_1)^n]$$

$$\bar{d}_{22} = 2/(1+\alpha_1)^2[1 - (\tfrac{1}{2})^n(1- \alpha_1)^n](1-\alpha_1)$$

and (4.1.18) $\quad b = 2\alpha_1/(1-\alpha_1)$.

Thus from (4.4.9) and the above

$$\mathcal{E}(n_2) = \frac{2(1-\alpha_1)}{\alpha_1}[1- (\frac{2-\alpha_1}{2})^n]+ \frac{2\alpha_1}{1+\alpha_1}[1- (\frac{1-\alpha_1}{2})^n] \quad . \quad (4.4.10)$$

From (4.1.32) we have for $\mathcal{E}(n_1)$

$$\mathcal{E}(n_1) = \sum_{w=0}^{2}[\sum_{s=0}^{2} Q_{s-1}(w-1)\bar{d}_{ss}(-\Delta[d_{11}Q_{s-1}(0)])]$$

$$= \sum_{w=0}^{2}[\sum_{s=0}^{2} Q_{s-1}(w-1)\bar{d}_{ss}(-d_{22}Q_{s-1}(1)+Q_{s-1}(0)d_{11})]$$

$$= 2\bar{d}_{11}(d_{11}-d_{22}) - b\,\bar{d}_{22}[d_{11}+ (1+b)d_{22}]$$

$$= 4[1 - (\frac{2-\alpha_1}{2})^n] - \frac{4\alpha_1}{1+\alpha_1}[1 - (\frac{1-\alpha_1}{2})^n] \quad , \quad (4.4.11)$$

and finally

$$\mathcal{E}(n_0) = \bar{d}_{00} - 2\bar{d}_{11}d_{11} + \bar{d}_{22}d_{11}b$$

$$= n - \frac{2(1+\alpha_1)}{\alpha_1}[1-(\frac{2-\alpha_1}{2})^n]+ \frac{2\alpha_1}{1+\alpha_1}[1-(\frac{1-\alpha_1}{2})^n] \quad . (4.4.12)$$

Note that $\mathcal{E}(n_0) + \mathcal{E}(n_1) + \mathcal{E}(n_2) = n$.

Further, if we consider $n \to \infty$, then we use

$$\mathcal{E}(n_i) = (2,i) \text{ element in } CQD_{\frac{1}{1-\lambda}}(CQ)^{-1} .$$

Then from (4.1.40)

$$\mathcal{E}(n_2) = \bar{a}_{22}[2\bar{a}_{11} - \bar{a}_{22}(1+b)(-b)] ,$$ note that

this is the same as (4.4.9) except that now

$$\bar{a}_{00} = n$$
$$\bar{a}_{11} = 2/\alpha_1$$
$$\bar{a}_{22} = 2(1-\alpha_1)/(1+\alpha_1)^2 .$$

Thus

$$\mathcal{E}(n_2) = 2/\alpha_1(1+\alpha_1) . \tag{4.4.13}$$

In like fashion

$$\mathcal{E}(n_1) = 4/(1+\alpha_1) \tag{4.4.14}$$

and

$$\mathcal{E}(n_0) = n - \frac{2(1+2\alpha_1)}{\alpha_1(1+\alpha_1)} . \tag{4.4.15}$$

Note that $\mathcal{E}(n_1)$, $\mathcal{E}(n_2)$ remain finite as $n \to \infty$; recall also that $\mathcal{E}(n_0)$ does not enter into any of our major computing formulas. $\mathcal{E}(n_0) + \mathcal{E}(n_1) + \mathcal{E}(n_2) = n$, thus these expectations are also appropriate for the sequential stopping rule.

V. SUGGESTIONS FOR FUTURE RESEARCH

It is anticipated that an extensive simulation program on an electronic computor will be undertaken at a later date as a continuation of the study already begun in this thesis. This is a pressing and exciting area of research. Such a study would be an invaluable factor in further ascertaining the validity of the postulated Theorems 4.1.6 and 4.1.7 and the properties of the maximum likelihood estimates $\hat{\alpha}_1$. For the present study $(M = 2, 4, 6, 10, 20)$ we have shown consistency and that skewness is less pronounced as M increases. From the figures (1, 2, 3, 4, 5) it appears that the empirical distributions of the $\hat{\alpha}_1$ have a mode close to the true value $\alpha_1 = 0.1$. However, normality has not been demonstrated. With the larger study (it is anticipated to investigate up to $M = 50$) it is hoped that a great many of these questions such as normality and the postulated standardizing factor $a(M)$ of Theorem 4.1.7 will be answered and clarified.

In connection with this study or apart from it, it would be valuable to investigate the unresolved questions about the limiting operations (4.1.43), (4.1.46), and (4.1.47) in the positively regular theory of Markov processes.

With the successful proof of these postulated theorems then more general theorems relating to absorbing Markov chains could be investigated, for example, transition probabilities which depend on several unknown parameters. It may also be valuable to investigate the situation where transitions occur in steps greater than unity, and chains with more than one absorbing state.

It might be valuable to investigate other integral approximations for the elements (3.1.52) of the matrix I for the two mutation rate case and for the quantity I (4.1.37) for the absorbing chain.

Finally for research not following directly from the problems of this thesis, the question of investigating population genetic models where the population is not assumed constant remains open for future research.

VI. ACKNOWLEDGMENTS

The author wishes to express his appreciation to the following:

Dr. G. A. Watterson for suggesting the problem, for his invaluable guidance and for his contributions.

Dr. Boyd Harshbarger for his advice and encouragement to the graduate student.

Members of the Department of Statistics, particularly Dr. R. J. Freund for his assistance with the computer programs.

Mrs. Lela Barnhisel for her careful typing of the manuscript.

This study was supported in part by a research grant from the National Institutes of Health, under contract numbers 2G, 3G, and 4G.

VII. REFERENCES

Anderson, T.W. and Goodman, L.A., "Statistical Inference about Markov Chains," Ann. Math. Stat., 28 (1957), pp. 89-110.

Bartko, J.J., "The Negative Binomial Distribution," Master's Thesis, Virginia Polytechnic Institute, 1960.

Bartko, J.J., "The Negative Binomial Distribution: A Review of Properties and Applications," Va. Jour. Science, 12 (1961), pp. 18-37.

Bartlett, M.S., An Introduction to Stochastic Processes. Cambridge University Press, New York, (1960).

Bhat, B.R., "Maximum Likelihood Estimation for Positive Regular Markov Chains," Sankhyá, 22 (1960), pp. 394-344.

Bhat, B.R., "Some Properties of Regular Markov Chains," Ann. Math. Stat., 32 (1961), pp. 59-71.

Bhat, B.R. and Gani, J., "A Note on Sufficiency in Regular Markov Chains," Biometrika, 47 (1960), pp. 452-457.

Billingsley, P., Statistical Inference for Markov Processes. Statistical Research Monographs Vol. II, University of Chicago Press, (1961).

Billingsley, P., "Statistical Methods in Markov Chains," Ann. Math. Stat., 32 (1961a), pp. 12-40.

Chung, K.L., "Contributions to the Theory of Markov Chains. II," Trans. Amer. Math. Soc., 76 (1954), pp. 397-419.

Crámer, H., Mathematical Methods of Statistics. Princeton University Press, (1946).

Crow, J.F., "The Estimation of Spontaneous and Radiation-Induced Mutation Rates in Man," Eugen. Quart. 3 (1956), pp. 201-208.

Erdélyi, A., (Ed) Higher Trancendental Functions, Vol. II. McGraw-Hill, New York, (1953).

Falconer, D.S., "The Estimation of Mutation Rates from Incompletely Tested Gametes, and the Detection of Mutations in Mammals," Jour. Genetics, 49 (1948-49), pp. 226-236.

Falconer, D.S. Introduction to Quantitative Genetics. Ronald Press, New York, (1960).

Feller, W., "Diffusion Processes in Genetics," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, (1951), pp. 227-246.

Feller, W., An Introduction to Probability Theory and Its Applications. Wiley, New York, (1957).

Finney, D. J., "On a Method of Estimating Frequencies," Biometrika, 36 (1949), pp. 233-235.

Gani, J., "Some Theorems and Sufficiency Conditions for the Maximum-Likelihood Estimator of an Unknown Parameter in a Simple Markov Chain," Biometrika, 42 (1955), pp. 342-359.

Gani, J., (corrigenda) "Some Theorems and Sufficiency Conditions for the Maximum-Likelihood Estimator of an Unknown Parameter in a Simple Markov Chain," Biometrika, 43 (1956), pp. 497-498.

Gani, J., "On the Stochastic Matrix in a Genetic Model of Moran," Biometrika, 48 (1961), pp. 203-206.

Goodman, L.A., "On Some Statistical Tests for m th Order Markov Chains," Ann. Math. Stat., 30 (1959), pp. 154-164.

Grenander, U., "Stochastic Processes and Statistical Inference," Arkiv. för Mathematic, Band 1 (1950), pp. 195-277.

Hahn, W., "Über Orthogonalpolynome, die q-Differenzengleichungen genügen," Math. Nach. 2 (1949), pp. 4-34.

Haldane, J.B.S., "On a Method of Estimating Frequencies," Biometrika, 33 (1943), pp. 222-225.

Hobson, E.W., The Theory of Functions of a Real Variable, Vol. II. Dover, New York, (1926).

Huzurbazar, V.S., "The Likelihood Equation, Consistency and The Maximum-Likelihood Function," Ann. Eug., 14 (1948), pp. 185-200.

Kaplan, W., Advanced Calculus. Addison-Wesley Publishing Company, Reading, Mass., (1956).

Karlin, S., and McGregor, J., "The Classification of Birth-and-Death Processes," Trans. Amer. Math. Soc., 86 (1957), pp. 366-400.

Karlin, S. and McGregor, J., On a Genetics Model of Moran, Technical Report No. 3, November 1960. Applied Mathematics and Statistics Laboratories, Stanford University, California.

Karlin, S. and McGregor, J., "The Hahn Polynomials, Formulas and an Application," Scripta Math, 26 (1961), pp. 33-46.

Kemeny, J.G., Snell, J.L., Mirkil, and Thompson, Finite Mathematical Structures. D. Van Nostrand, Princeton, (1959).

Kemeny, J.G. and Snell, J.L., Finite Markov Chains. D. Van Nostrand, Princeton, (1960).

Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics, Vol. I. Hafner, New York, (1958).

Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics, Vol. II. Hafner, New York, (1961).

Kimura, M., "Some Problems of Stochastic Processes in Genetics," Ann. Math. Stat., 28 (1957), pp. 882-901.

Li, C.C. Population Genetics. University of Chicago Press, (1955).

Lush, J.L., The Genetics of Populations. Mimeographed notes, Iowa State University, (1948).

Moran, P.A.P., "Random Processes in Genetics," Proc. Camb. Phil. Soc., 54 (1958), pp. 60-71.

Moran, P.A.P., "The Rate of Approach to Homozygosity," Ann. Human Genetics, 23 (1958a), pp. 1-5.

Pearson, E.S. and Hartley, H.O., (Ed). Biometrika Tables for Statisticians, Vol. I. Cambridge University Press, (1956).

Rao, C.R., Advanced Statistical Methods in Biometric Research. Wiley, New York, (1952).

Rao, C.R., "Maximum Likelihood Estimation for the Multinomial Distribution," Sankhyá, 18 (1957), pp. 139-148.

Richardson, C.H., An Introduction to the Calculus of Finite Differences. D. Van Nostrand, Princeton, (1954).

Silvey, L.D., "A Note on Maximum-Likelihood in the Case of Dependent Random Variables," Jour. Roy. Stat. Soc., 23 (1961), pp. 444-452.

Sinnott, E.W., Dunn, L.C., and Dobzhansky, TH., Principles of Genetics, Fourth Edition. McGraw-Hill, New York, (1950).

Snell, J.L., Personal Communication

Takács, L., Stochastic Processes. Wiley, New York, (1960).

Watterson, G.A., Probability Theory Applied to Genetic Populations. Ph.D. Thesis, Australian National University, (1960).

Watterson, G.A., "Markov Chains with Absorbing States: A Genetic Example," Ann. Math. Stat., 32 (1961), pp. 716-729.

Weber, M. and Erdélyi, A., "On the Finite Difference Analogue of Rodrigues' Formula," Amer. Math. Monthly, 59 (1952), pp. 163-168.

Whiting, P.W., "The Evolution of Male Haploidy," Quart. Rev. Biology, 20 (1945), pp. 231-260.

Whittle, P., "Some Distribution and Moment Formulae for the Markov Chain," Jour. Roy. Stat. Soc., 17 (1955), pp. 235-242.

The vita has been removed
from the scanned document

IX. APPENDICES

APPENDIX I

9.1 Properties of Hahn Polynomials

The Hahn polynomials form a family of orthogonal polynomials. They were introduced by Hahn (1949), discussed by Weber and Erdélyi (1952) and further discussed by Karlin and McGregor (1961). In the following presentation results which are believed to be new are so labelled. Results taken from other works will be given in the notation of this thesis.

The Hahn polynomials may be defined in terms of the generalized hypergeometric series

$$ {}_3F_2(a_1,a_2,a_3;\ b_1,b_2;\ z) = \sum_{\ell=0}^{\infty} \frac{(a_1)_\ell\,(a_2)_\ell\,(a_3)_\ell\, z^\ell}{(b_1)_\ell\,(b_2)_\ell\,\ell!} $$

where $(a)_0 = 1$, $(a)_\ell = a(a+1)\ldots(a+\ell-1) = \Gamma(a+\ell)/\Gamma(a)$ for $\ell \geq 1$. The series terminates if one of the a_i is zero or a negative integer. For real $a > -1$, $b > -1$ and for positive integral M, the Hahn polynomials

$Q_m(x) = Q_m(x;\ a,b,M)$, $m = 0, 1, 2, \ldots, M-1$ are defined by

$$ Q_m(x) = {}_3F_2(-m,\ -x,\ m+a+b+1;\ a+1,\ -M+1;\ 1) \quad . \qquad (9.1.1) $$

Explicit formula (Erdélyi and Weber, 1952)

$$Q_m(x) = Q_m(x;\ a,b,M)$$

$$= \sum_{\ell=0}^{m} \frac{(-m)_\ell\ (-x)_\ell\ (m+a+b+1)_\ell}{(a+1)_\ell\ (-M+1)_\ell\ \ell!} \ . \tag{9.1.2}$$

Recurrence relation (Erdélyi and Weber, 1952)

$$-xQ_m(x) = d_m Q_{m-1}(x) - (b_m+d_m)Q_m(x) + b_m Q_{m+1}(x) \ , \tag{9.1.3}$$

where

$$b_m = \frac{(m+a+b+1)(m+a+1)(M-1-m)}{(2m+a+b+1)(2m+a+b+2)}$$

$$d_m = \frac{m(m+b)(m+a+b+M)}{(2m+a+b)(2m+a+b+1)} \ ,$$

and (9.1.3) is valid for all complex values of x if $m = 0, 1, 2, \ldots, M-2$ but is valid only for $x = 0, 1, 2, \ldots, M-1$ when $m = M-1$.

Difference equation (Karlin and McGregor, 1961)

$$-\omega_m Q_m(x) = D(x)Q_m(x-1) - [B(x)+D(x)]Q_m(x) + B(x)Q_m(x+1) \ , \tag{9.1.4}$$

where

$$B(x) = (M-1-x)(a+1+x)$$

$$D(x) = x(M+b-x)$$

$$\omega_m = m(m+a+b+1) \ ,$$

and (9.1.4) is valid for $m = 0, 1, \ldots, M-1$ and all complex values of x .

<u>Orthogonality relation</u> (Karlin and McGregor, 1961)

$$\sum_{x=0}^{M-1} Q_m(x)Q_n(x)\rho(x;\ a,b,M) = \delta_{m,n}\frac{1}{\psi_{m,M}(a,b,M)} \quad (9.1.5)$$

where

$$\delta_{m,n} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

$$\rho(x;\ a,b,M) = \rho(x) = \frac{\binom{a+x}{x}\binom{M+b-1-x}{M-1-x}}{\binom{M+a+b}{M-1}}$$

$$\psi_{m,M}(a,b,M) = \psi_{m,M} = \frac{\binom{M-1}{m}\Gamma(b+1)\Gamma(m+a+1)\Gamma(m+a+b+1)(2m+a+b+1)}{\binom{M+a+b+m}{m}\Gamma(a+1)\Gamma(a+b+1)\Gamma(m+b+1)\Gamma(m+1)(a+b+1)} .$$

In particular $\psi_{0,M}(a,b,M) = 1$. It is also true that

$$\sum_{x=0}^{M-1} Q_m(x)Q_n(x)\frac{(a+1)_x(1-M)_x}{x!\ (1-M-b)_x} = \delta_{m,n}\frac{1}{\rho(0,a,b,M)\psi_{m,M}(a,b,M)} , \quad (9.1.6)$$

where $\rho(x)$ and $\psi_{m,M}$ are defined above. The equivalence of (9.1.5) and (9.1.6) is established by noting that

$$\frac{(a+1)_x(1-M)_x}{x!\ (1-M-b)_x}\rho(0,a,b,M) = \rho(x;\ a,b,M) \quad . \quad (9.1.7)$$

Conventions (New)

$$Q_{-1}(x;\ a,b,M) = \begin{cases} 0 & \text{if } x = 0,\ 1,\ \ldots,\ M-1 \\ 1 & \text{if } x = -1 \end{cases}$$

$$Q_m(-1;\ a,b,M) = 0\ , \quad m = 0,\ 1,\ \ldots,\ M-1\ . \tag{9.1.8}$$

Special Values

$$Q_0(x;\ a,b,M) = 1 \qquad x \geq 0$$

$$Q_m(0;\ a,b,M) = 1 \qquad m = 0,\ 1,\ \ldots,\ M-1$$

$$Q_m(M-1;\ a,b,M) = (-1)^m \binom{m+b}{m} \Big/ \binom{m+a}{m} \tag{9.1.9}$$

$$Q_m(M-1-x;\ a,b,M) = \frac{Q_m(x;\ a,b,M)}{Q_m(M-1;\ a,b,M)}$$

Theorem 9.1.1 (New). For the Hahn polynomials where $a = 0$, that is $Q_m(x;\ 0,b,M)$ then for $m \neq -1$,

$$\sum_{x=x_1}^{x_2} Q_m(x)$$

$$= \frac{B(x_2)[Q_m(x_2)-Q_m(x_2+1)]+b(x_2+1)Q_m(x_2)-\{B(x_1-1)[Q_m(x_1-1)-Q_m(x_1)]+bx_1Q_m(x_1-1)\}}{(m+1)(m+b)}\ . \tag{9.1.10}$$

For $m = -1$

$$\sum_{x=x_1}^{x_2} Q_m(x) = \begin{cases} 0 & x_1 > -1 \\ 1 & x_1 = -1 \end{cases}\ .$$

Proof: From (9.1.4) with $a = 0$

$$B(x) = (M-1-x)(1+x) \quad ; \quad B(-1) = 0$$

$$D(x) = x(M+b-x)$$

$$= bx + B(x-1)$$

$$D(x+1) = (x+1)(M+b-x-1)$$

$$= b(x+1) + B(x)$$

$$\omega_m = m(m+b+1) \quad .$$

The difference equation (9.1.4),

$$-\omega_m Q_m(x) = D(x)Q_m(x-1)-[B(x)+D(x)]Q_m(x)+B(x)Q_m(x+1) \quad ,$$

after substitution with the above identities becomes

$$-\omega_m Q_m(x) = [bx+B(x-1)]Q_m(x-1)-[B(x)+bx+B(x-1)]Q_m(x)+B(x)Q_m(x+1)$$

$$= B(x-1)[Q_m(x-1)-Q_m(x)]-B(x)[Q_m(x)-Q_m(x+1)]$$

$$+ bx[Q_m(x-1)-Q_m(x)] \quad .$$

Note that

$$bx[Q_m(x-1)-Q_m(x)] = bxQ_m(x-1)-b(x+1)Q_m(x)+bQ_m(x) \quad ,$$

and therefore

$$-\omega_m Q_m(x) = -\Delta(B(x-1)[Q_m(x-1)-Q_m(x)])-\Delta[bxQ_m(x-1)]+bQ_m(x) \quad .$$

Hence

$$(\omega_m+b)Q_m(x) = \Delta(B(x-1)[Q_m(x-1)-Q_m(x)]+bxQ_m(x-1)) \quad , \tag{9.1.11}$$

and

$$(\omega_m+b)\sum_{x=x_1}^{x_2} Q_m(x) = B(x_2)[Q_m(x_2)-Q_m(x_2+1)]+b(x_2+1)Q_m(x_2)$$
$$- (B(x_1-1)[Q_m(x_1-1)-Q_m(x_1)]+bx_1Q_m(x_1-1)),$$

noting that for finite differences

$$\sum_{x=a}^{b} \Delta f(x) = f(b+1) - f(a) \quad .$$

Finally

$$\sum_{x=x_1}^{x_2} Q_m(x)$$

$$= \frac{B(x_2)[Q_m(x_2)-Q_m(x_2+1)]+b(x_2+1)Q_m(x_2)-(B(x_1-1)[Q_m(x_1-1)-Q_m(x_1)]+bx_1Q_m(x_1-1))}{(m+1)(m+b)}$$

since $\omega_m+b = (m+1)(m+b)$.

For $m = -1$ the proof is immediate from (9.1.8).

Corollary 9.1.2 (New)

$$\sum_{x=0}^{i} Q_m(x) = \begin{cases} \dfrac{B(i)[Q_m(i)-Q_m(i+1)]+b(i+1)Q_m(i)}{(m+1)(m+b)}, & m \neq 1 \\ 0 \ , \quad m = -1 \ . \end{cases} \qquad (9.1.12)$$

Proof: The proof for both identities follows immediately from Theorem 9.1.1 where for the first identity we note that $B(-1) = 0$. (cf., 9.1.4)

Corollary 9.1.3 (New)

$$\sum_{x=0}^{M-1} Q_m(x) = \begin{cases} \dfrac{M(-1)^m (b)_m}{(m+1)!} , & m \neq -1 \\ 0 , & m = -1 , \end{cases} \tag{9.1.13}$$

where $(b)_m = b(b+1) \ldots (b+m-1)$.

Proof: From Corollary 9.1.2 we have

$$\sum_{x=0}^{M-1} Q_m(x) = \frac{b\,M\;Q_m(M-1)}{(m+1)(m+b)} ,$$

noting that $B(M-1) = 0$. Further, from (9.1.9)

$$Q_m(M-1;\ 0,b,M) = (-1)^m \binom{m+b}{m} ,$$

thus
$$\sum_{x=0}^{M-1} Q_m(x) = \frac{b\,M\,(-1)^m \binom{m+b}{m}}{(m+1)(m+b)} , \tag{9.1.14}$$

and finally
$$\sum_{x=0}^{M-1} Q_m(x) = \frac{M(-1)^m (b)_m}{(m+1)!} .$$

For $m = -1$ the proof is immediate from Corollary 9.1.2.

Corollary 9.1.4 (New)

$$\sum_{x=0}^{i} Q_{m-1}(x-1) = \begin{cases} \dfrac{[(M-i)i[Q_{m-1}(i-1)-Q_{m-1}(i)]+biQ_{m-1}(i-1)]}{m(m+b-1)} , & m\neq 0 \\ 1 , \quad m = 0 . \end{cases} \tag{9.1.15}$$

Proof: The proof follows by noting that from (9.1.11)

$$(\omega_{m-1}+b)Q_{m-1}(x-1) = \Delta[B(x-2)[Q_{m-1}(x-2)-Q_{m-1}(x-1)] + b(x-1)Q_{m-1}(x-2)] \quad ,$$

$\sum_{a}^{b} \Delta f(x) = f(b+1) - f(a)$ and that $(\omega_{m-1} + b) = m(m+b-1)$.

For $m = 0$ the proof is immediate from (9.1.8).

Corollary 9.1.5 (New)

$$\sum_{x=0}^{M} Q_{m-1}(x-1) = \begin{cases} \dfrac{M(-1)^{m-1}(b)_{m-1}}{m!} , & m \neq 0 \\ 1 \quad , & m = 0 \quad . \end{cases} \tag{9.1.16}$$

Proof: From Corollary 9.1.4

$$\sum_{x=0}^{M} Q_{m-1}(x-1) = \frac{b\,M\,Q_{m-1}(M-1)}{m(m+b-1)}$$

where $Q_{m-1}(M-1) = (-1)^{m-1} \binom{m+b-1}{m-1}$ from (9.1.9). Therefore

$$\sum_{x=0}^{M} Q_{m-1}(x-1) = \frac{M(-1)^{m-1}(b)_{m-1}}{m!} \quad .$$

For $m = 0$ the proof is immediate from (9.1.15).

Corollary 9.1.6 (New)

Let $b = M\alpha_1/1-\alpha_1$, then

$$\sum_{x=0}^{i} Q_{m-1}(x-1) = \begin{cases} \dfrac{(1-\alpha_1)(M-i)i[Q_{m-1}(i-1)-Q_{m-1}(i)]+M\alpha_1 iQ_{m-1}(i-1)}{m[(m-1)(1-\alpha_1)+M\alpha_1]}, & m\neq 0 \\ 1, & m = 0 . \end{cases}$$

Proof: The proof is immediate from Corollary 9.1.4.

9.2 A Relation Between Hahn and Tchebichef Polynomials (New)

Erdélyi (1953, Vol. II, p. 224), defines the Hahn polynomials in the following way:

$$p_m(x;\ \beta,\gamma,\delta) = \frac{(\beta)_m(\gamma)_m}{m!}\ {}_3F_2(-m,-x,\beta+\gamma-\delta+m;\ \beta,\gamma;\ 1)\ . \tag{9.2.1}$$

To put (9.2.1) into the $Q_m(x;\ a,b,M)$ form used in this thesis, we make the following substitutions [cf., (9.1.2)]

$$\beta+\gamma-\delta+m = m+a+b+1$$

$$\beta = a+1$$

$$\gamma = -M+1$$

$$\delta = 1-M-b\ .$$

Then

$$Q_m(x;\ a,b,M) = \frac{p_m(x;\ \beta,\gamma,\delta)}{\dfrac{(\beta)_m\ (\gamma)_m}{m!}} \tag{9.2.2}$$

$$= \frac{p_m(x;\ a+1,\ -M+1,\ 1-M-b)}{\dfrac{(a+1)_m\ (1-M)_m}{m!}}\ .$$

For $a = b = 0$

$$Q_m(x;\ 0,0,M) = \frac{p_m(x;\ 1,\ 1-M,\ 1-M)}{(1-M)_m} \quad . \tag{9.2.3}$$

However, Erdélyi also gives

$$p_m(x;\ 1,\ 1-M,\ 1-M) = t_m(x) \quad , \tag{9.2.4}$$

where $t_m(x)$ is a Tchebichef polynomial defined by

$$t_m(x) = m!\ \Delta^m[\binom{x}{m}\binom{x-M}{m}] \ ,\quad m = 0,\ 1,\ \ldots,\ M-1 \ .$$

The orthogonal property for the Tchebichef polynomials is

$$\sum_{x=0}^{M-1} t_m(x)t_n(x) = (2m+1)^{-1}M(M^2-1^2)(M^2-2^2)\ldots(M^2-m^2)\delta_{mn}$$

$$m,n = 0,\ 1,\ \ldots,\ M-1 \quad . \tag{9.2.5}$$

Hence from (9.2.4) and (9.2.3)

$$Q_m(x;\ 0,0,M) = \frac{t_m(x)}{(1-M)_m} \quad , \tag{9.2.6}$$

thus from (9.2.6)

$$\sum_{x=0}^{M-1} Q_m(x)Q_n(x) = \sum_{x=0}^{M-1} \frac{t_m(x)\ t_n(x)}{(1-M)_m\ (1-M)_n}$$

$$= \frac{(2m+1)^{-1}M(M^2-1^2)\ldots(M^2-m^2)\delta_{mn}}{(1-M)_m\ (1-M)_n}$$

$$= \begin{cases} \dfrac{M}{(2m+1)} \dfrac{\binom{M+m}{m}}{\binom{M-1}{m}} \ , & m = n \\[2ex] 0 \ , & m \neq n \ . \end{cases} \tag{9.2.7}$$

From (9.1.5) where $\rho(x;\ a,b,M)$ and $\psi_{m,M}(a,b,M)$ are defined, we can write (9.2.7) as

$$\sum_{x=0}^{M-1} Q_m(x)Q_n(x) = \delta_{mn} \frac{1}{\psi_{m,M}(0,0,M)\rho(0,0,0,M)}$$

$$= \delta_{mn} \frac{M}{\psi_{m,M}(0,0,M)} , \qquad (9.2.8)$$

where in $Q_m(x;\ a,b,M)$ $a = b = 0$. This verifies again the equivalence of (9.1.5) and (9.1.6) in this special case.

The connection (9.2.6) between Hahn polynomials $Q_m(x,0,0,M)$ and Tchebichef polynomials explains the apparently different results obtained by Karlin and McGregor (1960) using the former and Watterson (1961) using the latter, for Moran's model without mutation.

APPENDIX II

Table of $\hat{\alpha}_1$ from the Absorbing Chain

The following table is a listing of 500 maximum likelihood estimates listed in increasing order (502 for $M = 4,6$) of the mutation rate α_1 ($=0.1$). These estimates were obtained from data of experiments generated by simulation methods on the IBM 650. Five population sizes were studied, $M = 2$, 4, 6, 10 and 20. The experiments were generated by setting $k = M$ as the initial state where the number of transitions until absorption (state 0) occurred were punched along with other information by the machine.

For $M = 2$ 16 of the estimates were 1.0. These estimates were obtained from experiments of the type 2 1 .... 1 0 , that is, from state 2 we immediately went to state 1 , remained in state 1 for a finite number of times, it does not matter how many where the estimate of α_1 is concerned, and then passed directly on to absorption. For this case the conditions of Theorem 4.1.2 were not met, i.e., b_M and a_i were zero. The φ function (4.1.5) is $1/(1+\alpha_1) + 1/\alpha_1$ which never crosses the α_1 axis.

However, the value one was included in these estimates since 1 is the maximum value of the likelihood of these experiments. They were also included because of their frequency and also because they are an integral part of the peculiarities of the case $M = 2$.

The second group consisting of 26 estimates was $\hat{\alpha}_1 = 0.577350$ which were the result of experiments of the type 2 2 1 .... 1 0 .

For $M = 4$ and 6 the last two estimates 1.0 are a result of experiments which did not satisfy the conditions of Theorem 4.1.2. They were included for general interest for a total of 502 estimates. The solutions $\hat{\alpha}_1$ of the likelihood equations in these cases were greater than one but in the parameter space the maximum of the likelihood is at $\alpha_1 = 1$.

The estimates are listed as six decimal place numbers, with five place accuracy.

MAXIMUM LIKELIHOOD ESTIMATES OF THE MUTATION RATE α_1 (=0.1)

FROM REPLICATES OF POPULATIONS OF SIZE M . (6 DECIMAL PLACES)

M=2	M=4	M=6	M=10	M=20
0.052357	0.050207	0.046912	0.045372	0.044672
52357	50261	47870	48369	54405
52357	51440	50944	49136	54462
52357	53617	51276	50507	56186
52357	54037	51850	50763	57033
52357	54318	52700	50863	57267
52494	54479	52737	52207	57659
52494	54667	53152	52945	57931
52540	54784	54960	53914	58400
53262	55050	55360	54123	58991
53262	55737	55947	54147	58999
53474	56305	56076	54655	59432
54443	56324	56352	55334	59458
55234	57682	56467	57421	59535
55448	57911	56630	57957	59963
56967	57940	56993	58086	60500
58444	57986	57083	58293	60709
58444	58609	57102	59183	60761
58444	58932	57179	60225	61340
58444	59067	57343	60912	61872
58444	59854	57828	61571	63529
58444	60635	57869	61727	63532
59444	61436	58178	61902	63632
59601	61464	58211	61973	63752
60398	62211	58359	62206	64301
60398	62431	59299	62310	64392
60398	62512	59327	62408	64400
60398	64429	59492	62653	64484

Table 1

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.062047	0.064658	0.059519	0.063345	0.064596
62047	64788	59853	64024	64939
62047	65569	60007	64228	65102
62272	65798	60639	64555	66450
62272	65842	61101	64981	66717
62348	66609	61751	65128	66829
63668	66888	61807	65518	66838
63668	67405	62821	65709	67020
64266	67902	63479	65819	67207
64390	68273	63891	66093	67542
64390	68286	63923	66158	67565
66120	68588	64112	66175	67567
66120	68609	65236	66229	67592
66120	68861	65332	66368	67737
66120	69463	65342	66421	67878
66120	69754	67075	66698	68067
66120	69754	67253	66968	68078
66391	69884	67322	67132	68979
66482	70856	67924	67294	69062
68083	71646	67954	67923	69091
68661	72401	68236	68917	69320
68661	72902	68256	69281	69948
69557	73033	68948	70400	70078
69557	73096	69026	70661	70290
70761	73292	69323	70843	70381
70761	73452	69758	71019	70459
70761	73614	70370	71164	70669
70761	73688	70372	71459	70920
70761	73904	71298	71641	70999
70761	73962	71315	72279	71175
70761	74282	71341	72655	72194
71092	74335	71851	72737	72299
72929	74525	72009	72978	72357

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.072929	0.074694	0.072161	0.073162	0.072492
73699	75192	72311	73311	72583
73699	76038	72635	73941	72932
74740	76660	72703	74655	73220
74740	76946	73122	74922	73343
76095	76970	73129	74962	73459
76095	77081	73383	75115	73522
76095	77264	73690	75177	74155
76095	77894	74653	75603	74816
76095	77998	75101	75667	74910
76095	78376	75161	75688	74960
76095	78523	75638	75939	74978
76095	78648	75664	76438	75051
76095	78652	75709	76533	75411
76505	78785	75824	76735	75423
76505	78913	76294	76761	75701
78646	79196	76347	77244	75774
79531	79263	76576	77307	75826
79531	79315	76779	77492	75876
79531	79590	77432	77722	76031
79531	79652	77742	78129	76141
79764	79753	77862	78132	76446
79764	79829	78018	78221	76712
80755	79986	78196	78239	76905
81383	80369	78604	78309	76933
81383	80679	78952	78328	77327
82291	80904	79743	78364	77373
82291	81826	79771	78428	77756
82291	81890	79810	78520	77943
82291	82123	79937	78757	78030
82291	82330	80327	78991	78772
82291	82534	80349	79464	79147
82291	82604	80514	79856	79339

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.082291	0.082744	0.080579	0.080289	0.079885
82806	82924	81017	80336	80098
82806	83171	81334	80644	80469
82806	83187	81488	82320	80470
82806	84693	81670	82605	80796
82806	85081	81924	82724	80994
82980	85519	81975	82790	81094
83068	85544	82070	82849	81164
83156	85624	82166	83305	81511
85331	85624	82261	83483	81848
85331	85894	82589	83915	81862
85331	86211	83009	84198	81968
86360	86379	83962	84446	82648
86360	86404	83975	84743	82862
86360	86953	84072	84800	83216
86656	87428	84777	84999	83404
87818	87647	85088	85104	83665
89570	87692	85131	85247	84191
89570	88053	85296	85283	84211
89570	88087	85424	85449	84550
89570	88226	85650	86081	84567
89570	88261	86288	86265	84714
89570	88319	86497	86358	84867
89570	88884	86668	86500	85459
89570	89055	86716	86512	85582
89570	89237	87017	86522	85663
89570	89283	87116	87485	85701
89570	89370	87304	87569	86173
89570	89450	87429	88132	86200
89570	89774	88080	88199	86481
89570	90005	88486	88410	86559
89570	90247	88738	88649	86589
89570	90910	88881	88794	87103

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.0 89570	0.0 90980	0.0 89046	0.0 88927	0.0 87170
90454	91320	89529	89095	87325
90713	91327	89578	89542	87445
91777	91670	89633	89556	87599
93253	92906	89779	89821	87728
94462	93026	90460	90188	88240
94462	93508	90587	90333	88373
94462	93713	90772	90490	88522
94462	94128	90847	90653	88644
94847	94254	90971	91192	88655
97141	95566	91297	91267	89060
97141	95730	91690	91525	89084
97141	95853	91805	91525	89146
98073	95967	92782	91705	89147
98073	96247	92926	92008	89428
98242	97411	92997	92219	89897
98242	97806	93094	92292	89945
98242	98257	94511	92542	90167
98242	98339	94674	93142	90309
98242	98453	94849	93152	90312
98242	98484	96008	93459	90327
98242	98528	96093	93780	90365
98242	98543	96542	93942	90379
98242	98649	97086	94103	90394
98242	98845	97308	94195	90419
98242	98853	97411	94352	90432
98242	99261	97533	94482	90438
98242	99671	98297	94907	90593
98242	100481	98337	95223	91042
99106	100514	98527	95255	91066
99106	100567	98572	95788	91333
99106	100815	98826	96007	91367
99106	101112	98846	96132	92040

M=2	M=4	M=6	M=10	M=20
0.099106	0.101714	0.099229	0.096603	0.092446
99106	101719	99483	96656	92454
99106	101745	99905	97239	92555
99549	101820	100302	97282	92699
104228	101876	100557	97336	92769
104228	101993	100795	97363	93183
104228	102303	100856	97744	93991
104228	102390	100878	97752	94214
104228	102700	100880	98669	94872
104228	102772	100893	98731	95047
104741	103283	100988	98959	95281
104741	103333	101283	99136	95423
106412	103413	101563	99283	95680
106412	103413	101755	99413	95797
106412	103474	101928	99512	96004
106412	103537	102124	99573	96201
107544	103685	102377	99598	96321
107544	104197	102383	100277	96351
108741	104299	102542	100288	97048
108741	104331	102687	100290	97110
108741	104442	103188	100325	97165
108741	104514	103412	100331	97200
108741	105204	103415	101154	97286
108741	105319	103435	101402	97339
108741	105723	103470	101483	97415
108741	105793	103513	101638	97725
108741	106077	103574	101973	97814
108741	106152	104017	102283	98132
108741	106832	104184	102613	98576
108741	106894	104192	102793	98636
108741	106931	104360	103490	98738
108741	107023	104365	103563	98991
108741	107067	104715	103805	99026

Table 1 (continued)

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.108741	0.107225	0.104774	0.103827	0.099087
108741	107274	105672	103988	99374
108741	107734	105836	104409	99816
108741	107804	106115	104594	99818
108741	107847	106380	105647	99861
108741	108164	106484	105745	99862
109294	108249	107022	106454	99911
109902	109031	107170	106630	99918
109902	109071	107185	107232	99966
109902	109290	107232	107991	100031
109902	109469	107591	108204	100413
109902	109481	107727	108642	100419
110299	109594	108126	108800	100513
110299	109598	108431	108972	100531
113623	109635	108873	109014	100793
114480	109849	109311	109038	100878
114480	110310	109508	109201	100888
116223	110707	109630	109394	100990
116223	110737	109833	109548	101092
116223	111126	110546	109914	101174
116223	111331	110861	109967	101178
116223	111406	110906	110346	101437
116927	111415	110915	110497	101594
116927	111452	110920	110561	101706
118988	112337	111322	110685	101749
118988	113701	111430	111230	102001
121699	113958	111554	111485	102452
121699	114198	111832	111493	102712
121699	114488	111961	111835	102813
121699	114666	112102	112247	103019
121699	115723	112220	112292	103488
121699	116688	112289	112316	103560
121699	116828	112578	112541	103561

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.121699	0.117316	0.112620	0.112604	0.103604
123308	117490	112668	112863	103894
123308	117522	113010	112914	103983
123308	118220	113240	113189	104040
123308	118262	113562	113332	104070
123308	118663	113691	113410	104073
123308	118791	114014	113598	104076
123862	119067	114189	113637	104344
128094	119083	114195	114157	104493
131301	119427	114284	114362	104710
131301	120234	114737	115528	105141
131301	120615	115184	115897	105325
131301	121134	115650	115959	105479
134906	121399	115957	116543	105679
134906	121947	116367	116674	105711
138071	122066	116507	116808	105881
138071	123048	116678	116940	105988
138071	123174	116783	117684	106037
138071	123191	117010	117727	106301
138071	123696	117178	118161	106307
138071	123868	117535	118292	106478
138071	124243	119505	118728	106664
138071	124543	119590	119102	106717
138071	124645	120212	119258	106771
138071	125270	120860	119523	106861
138071	126426	120872	119535	106874
138071	126923	122553	119840	106875
138071	127070	122564	119995	107227
140388	127690	122751	120314	107423
140388	127800	123397	120404	107626
140388	127912	123637	121004	107815
140388	128196	124023	121064	107933
140388	128285	124565	121545	107988

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.141193	0.129072	0.124567	0.121578	0.108004
141193	129369	124732	121681	108119
148090	129378	124966	122105	108395
150325	129543	124985	122169	108579
150806	129916	125107	122190	108701
150806	130182	125286	122998	108877
150806	132214	126048	123195	109013
150806	132576	126087	124163	109283
150806	132691	126361	124651	109486
150806	132810	126756	124688	109788
150806	133934	126799	124693	110034
150806	134082	127275	124820	110098
150806	134239	127525	125357	110103
150806	135143	127869	125703	110289
150806	137043	128024	126465	110312
152298	137323	128323	126577	110458
155687	137401	128332	127228	110621
159365	137626	128613	127291	110673
159365	137660	128771	129112	110708
159365	139943	129376	129183	110749
159365	140097	129563	129611	110921
159365	140199	129846	129631	111458
159365	140413	129893	130189	111578
159365	140527	130850	130192	111689
159365	140610	131063	130193	111935
159365	140918	131124	130258	112410
159365	141463	131163	130788	112620
159365	141520	131215	130843	112687
159365	141549	131309	131498	112763
159869	142522	131317	131778	113381
162865	142671	131899	132121	113405
162865	142671	132064	132483	113482
162865	143878	132132	132533	113498

M=2	M=4	M=6	M=10	M=20
0.162865	0.144341	0.132986	0.132900	0.113875
164096	145950	133010	133021	114193
164096	146408	133037	133060	114562
164725	147370	133261	133255	114938
165107	147424	133263	133717	115003
171725	147672	133319	133752	115217
176980	148019	133601	133903	115488
176980	148888	134199	134003	115493
176980	148947	134772	134254	115661
176980	149238	134965	134384	115703
176980	149919	135291	134525	115709
176980	150126	135405	134643	116010
176980	150593	135586	134685	116300
176980	150766	135848	134973	116482
176980	150964	136556	135049	116860
187652	152052	136788	135111	117184
187652	152188	138475	135119	117629
188089	152419	138536	135127	117939
188089	153055	138708	135144	117974
188089	153778	141197	135174	118019
188089	154269	141232	135185	118109
188089	154537	141693	135205	118372
188089	154598	141977	135221	118409
188089	154712	142005	135335	118552
188089	154802	142234	135442	118825
188089	155190	143092	135449	119322
188089	155989	143282	135581	119456
188089	156388	143859	135777	119469
188089	156996	144079	135801	119474
188089	158773	144873	135882	119778
188089	159076	144993	135980	120122
188089	159792	145063	136019	120396
188089	159811	145151	136043	120616

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.188089	0.159887	0.145558	0.136101	0.120660
188089	160508	145722	136130	120674
188089	160906	146202	136218	120986
188089	160906	146669	136230	121506
188089	161689	146712	136554	121681
193713	161932	146732	137330	121812
193713	162289	146816	137448	121839
193713	163822	148662	137486	121912
193713	164670	149071	138245	122190
195728	166754	149853	138328	122196
197397	167057	150049	139205	122505
209119	167364	150806	139283	122545
209119	167577	151830	139405	122847
213850	168273	151993	139418	123246
213850	168753	152858	139660	123267
213850	169062	153875	139807	123499
213850	169512	153915	140643	123584
213850	169810	155123	140865	124393
224440	170150	155244	141563	125309
224440	170166	155457	142093	125640
228714	170173	156995	142234	125797
228714	170830	158052	142552	126534
228714	171112	158571	142767	126826
228714	171322	158768	143045	127059
228714	172532	158783	144017	127984
228714	172542	159207	144433	128316
228714	172642	159850	144845	129108
228714	172642	160085	144910	129311
228714	173421	160130	144931	130131
228714	174031	160546	146010	130140
228714	174928	160686	146297	130371
228714	175715	161854	147033	130500
228714	176020	162626	147335	130560

Table 1 (continued)

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.238516	0.176220	0.162723	0.147672	0.131056
238516	177081	163584	148031	131499
238516	177499	163895	149068	131538
238516	178795	163984	149099	131671
238516	178941	164963	149807	132086
238516	179334	166850	150229	132173
238516	179564	167328	151186	132576
242133	179581	168908	152252	132952
242133	180497	169199	152614	133147
244017	182011	169459	154295	133243
244017	182292	170154	154473	133981
244017	182641	170277	154625	134217
259536	182839	170510	154853	134606
262785	182839	170739	155698	134781
269381	182839	171309	155885	135389
269381	185544	173319	156117	135765
269381	187086	173410	156147	136149
269381	187091	174722	156326	136489
269381	187307	174855	156554	137599
269381	187864	178677	157572	137620
287185	188402	179176	157981	137995
287185	188562	179749	158887	138105
289898	189240	180387	159156	138343
289898	190702	181304	160363	138377
289898	195839	182348	160757	139100
289898	196664	183441	161594	139188
289898	197731	183932	161966	139532
289898	199182	184340	162837	139745
289898	199443	184356	162983	140903
289898	200193	186988	163011	141556
289898	201165	187698	163404	143023
289898	201390	188438	163446	143351
289898	204028	189234	164298	143449

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.289898	0.206825	0.192535	0.164800	0.143554
289898	210223	192636	165297	144174
289898	211274	192681	165668	146280
289898	212718	193483	166308	146335
289898	213349	194247	166799	146364
289898	213872	198733	167578	146598
289898	216102	201919	169445	146792
289898	217673	203026	169581	147176
289898	218852	203898	169911	147415
289898	219485	207219	170693	148144
289898	219485	207777	171044	148969
289898	220521	208330	172725	148981
289898	222172	209385	173130	150432
289898	222967	212039	173341	150519
289898	225081	214471	173731	150558
289898	225627	217023	174457	150816
289898	228198	217368	175186	151057
289898	230467	218403	175595	151359
289898	231773	220275	175964	152508
289898	232288	221799	176009	152712
289898	233222	222368	177834	152999
309017	234771	223824	177835	153822
309017	234779	224538	178436	154008
309017	241250	227900	181159	154290
309017	244245	229237	182709	154524
309017	244605	231973	184208	154811
316431	246090	234139	184398	154924
322829	246665	236718	184718	154953
361590	255497	237347	186781	155907
361590	261491	237539	187285	156000
390388	263252	238657	188781	156734
390388	266799	238973	188821	157396
390388	267008	239336	188994	157526

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.390388	0.270962	0.241433	0.189230	0.158269
390388	272875	241617	190780	158501
390388	277043	242356	191213	158618
390388	280464	243834	192635	158678
390388	281824	249161	192696	158696
390388	283545	249396	193935	159584
390388	286617	249860	194061	160054
390388	287148	250275	197396	160976
390388	288336	251620	199937	161210
390388	289625	253163	199959	161320
390388	295692	253242	200351	161979
390388	297826	255044	200402	162177
390388	297935	259027	201231	162779
390388	301183	261091	201547	163664
390388	305368	261936	201611	165028
390388	305474	266753	204519	165521
390388	312079	267961	205585	166277
390388	312224	270172	205702	167074
390388	313393	271314	205830	168507
390388	316307	271693	206163	168866
390388	321325	275356	207159	169299
390388	322140	287360	207190	169503
390388	322730	288203	207429	169774
390388	329768	290273	207611	170578
390388	331479	290846	208598	171109
390388	336605	292880	209279	171938
434259	345704	294010	209790	172130
434259	346114	295496	212822	172962
434259	348483	300451	212888	173273
434259	348523	305597	213220	173680
434259	356672	305940	216609	174954
463325	359280	309315	217723	175474
540312	359280	311609	219404	176403

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
0.540312	0.361863	0.315188	0.220163	0.176535
577350	363936	316770	220893	177678
577350	365763	317115	222045	177862
577350	367441	317822	226432	178201
577350	379724	331537	237083	178289
577350	383265	333354	239271	179159
577350	397409	344464	239354	179444
577350	401341	344756	239414	179824
577350	414213	347234	239936	180459
577350	415222	349628	245224	180577
577350	418280	350156	247437	181075
577350	419453	351304	248117	181363
577350	421402	353485	248246	183695
577350	425056	356255	250351	184530
577350	451391	356792	250389	185683
577350	461981	357510	254460	185867
577350	466858	359231	257506	185972
577350	468782	359258	258241	187172
577350	479258	360919	262737	189615
577350	479403	362860	266273	191558
577350	482251	368644	269881	191813
577350	485415	375233	270863	192897
577350	491995	390298	270923	193121
577350	493000	397664	273538	193150
577350	497482	401062	277186	195877
577350	498389	403583	277792	196411
577350	499618	412111	278662	198796
1000000	519960	416312	280343	198981
1000000	523588	430728	286891	200404
1000000	523588	436075	289054	203976
1000000	533884	457572	289106	206453
1000000	536051	459809	292683	207105
1000000	536051	474712	293201	207836

Table 1 (continued)

M=2	M=4	M=6	M=10	M=20
1.000000	0.536051	0.507698	0.295183	0.208450
1000000	538237	522755	296851	208913
1000000	572276	536729	299478	210172
1000000	582749	553373	301580	212858
1000000	608981	557106	302950	213675
1000000	622739	567056	305047	213751
1000000	631759	575397	307691	214758
1000000	694521	626273	318542	215323
1000000	701818	627231	323784	225486
1000000	720920	647186	355405	296158
	1000000	1000000		
	1000000	1000000		

APPENDIX III

$(I - P_\Delta)^{-1}$ For $M = 2, 4, 6, 10, 20$

The following matrices of order $M \times M$ are the inverses of $(I - P_\Delta)$ where I is the identity matrix and P_Δ is the matrix P (4.1.2) where the first row and first column of P are deleted. All values below the main diagonal are the same as the diagonal element, hence the triangular presentation. Given an initial state k these matrices provide us with the expected value of the total number of times state i is entered $\mathcal{E}(n_i)$ before absorption occurs. For example, for the case where the initial state is $k = M$ and $M = 2$ then

$$\mathcal{E}(n_2) = 18.181818$$

and

$$\mathcal{E}(n_1) = 3.636364 \quad ,$$

which are the diagonal elements of the matrix array. The total length of chain to absorption is about $18 + 3.6 + 1 \approx 23$. The value one allows for the transition from state 1 to state 0 .

The matrix $(I - P_\Delta)^{-1}$ can be used to find $\mathcal{E}(n_i)$ only when the chain is considered until absorption. For n finite, not a random variable, then the method of Hahn polynomials (4.1.32) must be utilized. For a further discussion on $(I - P_\Delta)^{-1}$ see Kemeny and Snell (1960).

MATRICES $(I - P_\Delta)^{-1}$

M = 2

$$
\begin{array}{c}
\begin{array}{cc} \text{column } 1 & 2 \end{array} \\
\begin{bmatrix} 3.636364 & 8.181818 \\ & 18.181818 \end{bmatrix}
\end{array}
$$

M = 4

$$
\begin{array}{c}
\begin{array}{cccc} \text{column } 1 & 2 & 3 & 4 \end{array} \\
\begin{bmatrix} 5.161290 & 3.167155 & 2.923528 & 4.933453 \\ & 6.803518 & 6.280171 & 10.597788 \\ & & 10.382735 & 17.520864 \\ & & & 27.520863 \end{bmatrix}
\end{array}
$$

$$M = 6$$

$$\begin{array}{c}
\begin{array}{cccccc} \text{column } 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{bmatrix}
7.058823 & 3.781513 & 2.750191 & 2.320471 & 2.227656 & 2.784569 \\
 & 8.067227 & 5.867073 & 4.950338 & 4.752332 & 5.940414 \\
 & & 9.503436 & 8.018516 & 7.697787 & 9.622233 \\
 & & & 11.768511 & 11.297789 & 14.122235 \\
 & & & & 16.097796 & 20.122244 \\
 & & & & & 30.122242
\end{bmatrix}
\end{array}$$

M = 10

column 1	2	3	4	5
10.989009	5.427498	3.568765	2.634753	2.069479
	11.525060	7.578121	5.594785	4.394449
		12.144331	8.965931	7.042332
			12.872181	10.110514
				13.746877

6	7	8	9	10
1.687075	1.406981	1.187140	0.999697	0.809755
3.582431	2.987665	2.520842	2.122814	1.719479
5.741031	4.787887	4.039780	3.401920	2.755555
8.242266	6.873859	5.799818	4.884058	3.956086
11.206693	9.346122	7.885790	6.640666	5.378939
14.829881	12.367777	10.435311	8.787630	7.117980
	16.228781	13.693033	11.530975	9.340089
		18.157318	15.290373	12.385202
			21.138326	17.122043
				27.122042

M = 20

column 1	2	3	4	5
20.942404	9.838325	6.141844	4.297418	3.193952
	20.827343	13.002043	9.097465	6.761471
		20.709171	14.490105	10.769422
			20.587669	15.301287
				20.462578

6	7	8	9	10
2.461093	1.940132	1.551726	1.251813	1.013969
5.210038	4.107183	3.284945	2.650039	2.146532
8.298356	6.541769	5.232138	4.220885	3.418917
11.790375	9.294603	7.433867	5.997069	4.857627
15.767397	12.429772	9.941391	8.019946	6.496157
20.333607	16.029413	12.820040	10.342508	8.377433
	20.200423	16.156395	13.033731	10.557322
		20.062644	16.184991	13.109844
			19.919819	16.135055
				19.771419

M = 20

column 11	12	13	14	15
0.821397	0.662921	0.530828	0.419641	0.325383
1.738865	1.403377	1.123742	0.888364	0.688824
2.769599	2.235248	1.789855	1.414953	1.097133
3.935071	3.175859	2.543042	2.010378	1.558816
5.262413	4.247110	3.400838	2.688499	2.084621
6.786399	5.477067	4.385714	3.467085	2.688324
8.552286	6.902253	5.526920	4.369254	3.387853
10.620035	8.571061	6.863204	5.425641	4.206959
13.070701	10.548908	8.446949	6.677655	5.177751
16.016451	12.926321	10.350641	8.182601	6.344664
19.616810	15.832047	12.677377	10.021979	7.770889
	19.455235	15.578614	12.315525	9.549269
		19.285752	15.246168	11.821645
			19.107170	14.815407
				18.917972

M = 20

column 16	17	18	19	20
0.245127	0.176712	0.118583	0.006973	0.002981
0.518924	0.374093	0.251036	0.147615	0.006311
0.826523	0.595841	0.399841	0.235115	0.100512
1.174331	0.846577	0.568098	0.334054	0.142808
1.570446	1.132136	0.759723	0.446734	0.190979
2.025244	1.460001	0.979737	0.576107	0.246286
2.552233	1.839907	1.234675	0.726016	0.310372
3.169305	2.284755	1.533191	0.901548	0.385413
3.900649	2.811982	1.886988	1.109590	0.474349
4.779741	3.445720	2.312259	1.359659	0.581254
5.854185	4.220289	2.832036	1.665299	0.711915
7.193924	5.186108	3.480152	2.046405	0.874838
8.905814	6.420212	4.308299	2.533375	1.083018
11.161160	8.046093	5.399351	3.174937	1.357285
14.251819	10.274152	6.894497	4.054115	1.733134
18.716105	13.492462	9.054151	5.324038	2.276026
	18.498718	12.413614	7.299476	3.120526
		18.261566	10.738200	4.590580
			17.997728	7.694029
				17.694029

ABSTRACT

This Dissertation deals with statistical inference on the mutation rates α_1 and α_2 of a population genetic model introduced by Moran [Proc. Camb. Phil. Soc. 54 (1958), pp. 60-71]. The deductive theory by approximate methods of such models has reached an advanced stage but little has been done along the line of statistical inference. Moran's model is a model of the Markov chain type. It was selected for investigation because it is the only finite population genetic model for which the deductive theory by exact methods is well enough established to stimulate an investigation of statistical inference.

The first broad area of discussion of this dissertation deals with the simultaneous consideration of the mutation rates α_1 and α_2. Maximum likelihood estimates for α_1 and α_2 are obtained iteratively from the Newton-Raphson scheme for simultaneous solution of two equations in two unknowns. Several theorems are given which ensure that the log likelihood function involving α_1 and α_2 has a unique maximum in the parameter space of useful values.

The transition matrix consists of conditional probability elements involving the unknown parameters α_1 and α_2 . These elements are the probability of a transition from one state to another in at most unit steps. The eigenvalue expression along with the corresponding pre- and post-eigenvector matrices are given. The post-eigenvector matrix has elements consisting of Hahn polynomials. The pre-eigenvector matrix is obtained by inverting the post-eigenvector matrix for which an expression is given. The Hahn polynomials form a family of orthogonal polynomials. They were introduced by Hahn [_Math_. _Nach_. _2_ (1949), pp. 4-34], and further discussed by Karlin and McGregor [_Scripta Math_, _26_ (1961), pp. 33-46]. These polynomials form the foundation and are basic to many of the results of the dissertation. The expression for the expected value of the number of transitions from one state to another is given and this expression is also in terms of Hahn polynomials.

Finally for this positively regular transition matrix involving both of the mutation rates α_1 and α_2 , asymptotic multivariate normality of the maximum likelihood estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$ is discussed along with hypothesis testing. Also discussed are large sample approximations, methods of designing and conducting experiments and replicated experiments.

The second broad area of this dissertation deals with an absorbing Markov chain. That is, α_2 is set equal to zero and investigation on α_1 only is carried out. For this case the above transition matrix becomes an absorbing one (regular) and inferences are obtained from realizations on this absorbing chain whose peculiarities provide some unique difficulties. The eigenvalue expression with the corresponding post-eigenvector matrix whose elements are also Hahn polynomials and the expression (in terms of Hahn polynomials) for the expected number of transitions from one state to another are all given.

Of particular interest are several postulated theorems on the maximum likelihood estimate $\hat{\alpha}_1$ of the mutation rate α_1 of the absorbing Markov chain in which an attempt is made at establishing the properties and normality of $\hat{\alpha}_1$. The estimate is again obtained iteratively. An outline of the proofs of the postulated theorems is presented. Gaps in the proof are a result of unresolved questions in positive regular Markov chain theory.

In connection with the above theory and postulated theorems a simulation study on the IBM 650 was undertaken. This study substantiated many of the assumptions of the postulated theorems. The study, however, was not extensive enough to be conclusive. A further study is proposed.

Replicated experiments are also discussed. Of particular interest here is a geometric type stopping rule in which the negative binomial is employed. Methods of conducting and designing experiments are discussed.

An appendix discusses the Hahn polynomial system along with many of its important properties.