STATISTICAL ESTIMATION FOR ASYMPTOTIC REGRESSION MODELS AND THE DERIVATION OF SUCH MODELS IN A NEW THEORY FOR RACIAL INTERMIXTURE.

BY

R. W. HIORNS

Thesis presented for the Degree of Doctor of Philosophy of the University of Edinburgh in the Faculty of Arts, May 1964.

ACKNOWLEDGMENTS

I wish to express my thanks to the Directors of the Computing Laboratories in the Universities of Oxford and Leeds respectively and to the staff for their help in operating my programs for the Mercury and Pegasus electronic computers.

At an early stage in the research I was fortunate enough to receive a query from Professor D.F. Roberts now of the Laboratory of Physical Anthropology, University of Washington, which led to the substantially new mathematical theory for racial intermixture described in Part I. I am grateful to Professor Roberts for his continued encouragement and interest in the application of my work.

I gratefully acknowledge a few small corrections made to Chapters One, Two and parts of Three and Four after their having been read by Professor D.G. Kendall, and to Chapter Eight after a reading by Professor M.S. Bartlett, F.R.S.

Finally, I would add my very large debt to Professor H.E. Daniels for much helpful discussion throughout and to my other supervisors Dr. D.N. Lawley and Professor A.C. Aitken, F.R.S.

NOTE CONCERNING PUBLICATIONS

It was found necessary to publish certain of the new results in this thesis before its presentation for examination. Some of the findings in Chapters One, Two, sections 3.1 and 3.3 of Chapter Three and the introduction to Chapter Four appeared in Roberts and Hiorns (1962, 1963). The material contained in Chapter Eight is at the moment in the press as the introduction to a volume of tables which I prepared for publication in the Tracts for Computers series, reference Hiorns (1964). A notification of this impending publication was given in Hiorns (1962a). The example used in Chapter Eight is a fragmentary part of the statistical analysis carried out by me in connection with a biological experiment reported in Harrison, Hiorns and Weiner (1964). Chapter Nine contains the substance of a paper which I read as a short communication to the International Congress of Mathematicians at Stockholm and for which an abstract was published as Hiorns (1962b).

In order to avoid confusion concerning the propriety of any material which has appeared in any paper with joint authorship, I would like to confirm that the content of this thesis has been restricted to work carried out exclusively by myself.

CONTENTS

PART II ASYMPTOTIC REGRESSION METHODS

Introduction

Chapters

Six — General Methods for a single non-linear parameter. The classical iterative method. Stevens' Method. Schneider's Method. Hartley-Gauss-Newton. Discussion.

Seven — Special Methods. Small sample techniques and grouping. Gomes' Method. Patterson's Method. Internal regression and general quadratic ratio estimators.

Eight — Tables for the application of Stevens' Method. Introduction. Description of the Method. Computational procedure. Use of the tables for other curves. Example. Data for which the ranges of the tabulation are exceeded. The checking of the tabulation.

Nine — Extension of Stevens' Method. Many non-linear parameter relations. Confidence limits and regions for non-linear parameters.

References

PART I

A MATHEMATICAL THEORY FOR

RACIAL INTERMIXTURE

INTRODUCTION

The application of mathematical methods of analysis has yielded results of some interest and practical importance in the field of population genetics. Following the work of Fisher (1930), Wright (1931) and Haldane (1924-32), many research workers have, with some success, tackled problems using mathematical models which describe the behaviour of gene frequencies within populations.

The single problem of the effect of intermixture between several populations on gene frequencies has not been investigated with due generality. This became apparent at the outset of the present research, following the discovery of the asymptotic regression model which emerges from the special case treated by Glass and Li (1953). In the hope that other interesting regression models might appear, a fuller and more general treatment of the problem was undertaken.

The results of this fuller investigation are presented here in Part I as a theory of intermixture and the interest in asymptotic regression models is satisfied by those types which arise; the basis for these models is given by the appropriate analysis and the whole of this theory is claimed as a new contribution to the literature. In addition to the accepted techniques for establishing differential and difference equations, lesser known results like those of Sauvage and Sylvester are employed as tools of analysis to provide new results. Attendant problems of statistical estimation concerned with the application of these models in the presence of sample data are dealt with as they arise, while a fuller account of asymptotic regression is left to Part II

CHAPTER ONE

TWO POPULATIONS WITH DISCRETE GENERATIONS

In this chapter the first of several models with many common genetic features is considered. It will be convenient to state here the assumptions and describe these features. Superimposed upon them will be the fact that generations are discrete, or identically, that migration takes place only at the end of each generation and also that the model for the present will describe intermixture between just two populations.

The genetic considerations attached to the model are that the populations are panmictic, that migrants are drawn at random from them and that intermixture is the sole process for modifying the gene frequencies of the individuals in each population.

By way of notation, let q_n^{I} be the allele frequency of a single locus in population I at the end of the n generations and let q_n^{II} represent the frequency at the corresponding locus in population II. Let $m^{I}(n)$ be the rate of gene flow into population II from

population I at the end of the n^{th} generation and $m^{I}(n)$ be the rate at that time of gene flow in the reverse direction. Here the gene flow rate is understood to be that proportion of the recipient's population genes which is received by immigration; the proportion of the donating population itself which emigrates is considered irrelevant for present purposes. These gene flow rates will be called admixture rates in the hope that confusion with other definitions may be avoided.

At the end of the first generation admixture takes place and there will, in the gametes of population I, be a component, $m^{II}(1)$, deriving from population II in which the allele frequency will be q_0^{II}; the remainder, $1 - m^{II}(1)$, of these gametes derive from population I and their gene frequencies will be unaltered. Mathematically, if q_0^{I} and q_0^{II} are the initial frequencies in populations I and II respectively, the relations between these and the frequencies after one generation are:

$$q_1^{I} = m^{II}(1)\, q_0^{II} + (1 - m^{II}(1))\, q_0^{I} \qquad (1.1)$$

and by symmetry,

$$q_1^{II} = m^{I}(1)\, q_0^{I} + (1 - m^{I}(1))\, q_0^{II} \qquad (1.2)$$

The above argument may be applied equally effectively to obtain the relations between the frequencies in the two populations

After the n^{th} generation,

$$q_n^{I} = (1 - m^{II}(n))\, q_{n-1}^{I} + m^{II}(n)\, q_{n-1}^{II} \qquad (1.3)$$

$$q_n^{II} = m^{I}(n)\, q_{n-1}^{I} + (1 - m^{I}(n))\, q_{n-1}^{II} \qquad (1.4)$$

providing two simultaneous recurrence relations. These may be readily solved by forming the sum and difference of the equations. First define

$$\Sigma_n = q_n^{I} + q_n^{II} \qquad (1.5)$$

and

$$\Delta_n = q_n^{I} - q_n^{II} \qquad (1.6)$$

Now by subtraction of (1.4) from (1.3),

$$\Delta_n = (1 - m^{I}(n) - m^{II}(n))\, \Delta_{n-1} \qquad (1.7)$$

which is a single recurrence relation, and this may be solved by successive substitution downwards, leading to the result:

$$\Delta_n = \prod (1 - m^{I}(k) - m^{II}(k))\, \Delta_0 \qquad (1.8)$$

where $\prod$ is used to denote the product of the bracketed term for $k = 1, 2, \ldots, n$.

In a similar manner, adding (1.3) and (1.4) gives again a single recurrence relation

$$\Sigma_n = \Sigma_{n-1} + (m^{I}(n) - m^{II}(n))\, \Delta_{n-1} \qquad (1.9)$$

into which the solution for Δ_{n-1} may be substituted. This gives:

$$\Sigma_n = \Sigma_{n-1} + (m^{I}(n) - m^{II}(n)) \prod (1 - m^{I}(k) - m^{II}(k)) \Delta_0 \quad (1.10)$$

for $\prod$ over $k = 1, 2, \ldots, n-1$.

Successive addition enables a general solution for Σ_n to be found and this has the form:

$$\Sigma_n = \Sigma_0 + \sum (m^{I}(\ell) - m^{II}(\ell)) \prod (1 - m^{I}(k) - m^{II}(k)) \Delta_0 \quad (1.11)$$

in which $\sum$ denotes summation over all that follows for $\ell = 1, 2, \ldots, n$ and $\prod$ now operates to form the product of the bracket which follows it for $k = 1, 2, \ldots, (\ell - 1)$.

Using (1.5) and (1.6) and defining

$$f = \begin{cases} 0 & \text{for} \quad k \neq n \\ (1 - m^{I}(k) - m^{II}(k)) & \text{for} \quad k = n \end{cases}$$

the general solution for (1.3) and 1.4) may be obtained from (1.8) and (1.11). This result is embodied in the following theorem.

THEOREM 1.1 The discrete model defined by the relations (1.3) and (1.4) has general solution:

$$q_n^{I} = \tfrac{1}{2}\left(\Sigma_0 + \Sigma\left(m^{I}(k) - m^{II}(k) + f\right)\Pi\left(1 - m^{I}(\ell) - m^{II}(\ell)\right)\Delta_0 \right. \quad (1.12)$$

and

$$q_n^{II} = \tfrac{1}{2}\left(\Sigma_0 + \Sigma\left(m^{I}(k) - m^{II}(k) - f\right)\Pi\left(1 - m^{I}(\ell) - m^{II}(\ell)\right)\Delta_0 \right. \quad (1.13)$$

in which Σ implies summation over $k = 1, 2, \ldots, n$ and Π implies product taken over $\ell = 1, 2, \ldots, (k-1)$.

THEOREM 1.2 For the special case in which the admixture rates are constant, i.e.

$$m^{I}(k) = m^{I}, \quad m^{II}(k) = m^{II} \quad (1.14)$$

for all $k = 1, 2, \ldots, n$, then the general solution to the model defined by (1.3) and (1.4) is:

$$q_n^{I} = \tfrac{1}{2}\left(\Sigma_0 + \left(m^{I} - m^{II} + 2m^{II}(1 - m^{I} - m^{II})^n\right)\Delta_0 / (m^{I} + m^{II})\right) \quad (1.15)$$

together with,

$$q_n^{II} = \tfrac{1}{2}\left(\Sigma_0 + \left(m^{I} - m^{II} - 2m^{II}(1 - m^{I} - m^{II})^n\right)\Delta_0 / (m^{I} + m^{II})\right) \quad (1.16)$$

Proof This follows from Theorem 1.1 and equation (1.14).

THEOREM 1.3 (Glass and Li) For unidirectional gene flow in the model defined above, where:

$$m^{I}(n) = m \quad \text{, a constant and}$$
$$m^{II}(n) = 0 \tag{1.17}$$

the general solution to the equations (1.3) and 1.4) is given by

$$q_n^{I} = q_0^{I} \tag{1.18}$$

and $$q_n^{II} = q_0^{II}(1-m)^n + (1-(1-m)^n)\, q_0^{I} \tag{1.19}$$

Proof This follows from Theorem 1.2 and the relations (1.17).

CHAPTER TWO

TWO POPULATIONS WITH CONTINUOUS MIGRATION

The model of Chapter One is now changed only in one important respect. Instead of the assumption that migration takes place at the end of each generation, or at discrete times, continuous migration throughout each generation is now considered. Over a large number of generations the difference in effect of the discrete and continuous assumptions is small. The general model and analysis which follows, yields as a special case the solution for unidirectional gene flow given by Glass and Li (1953).

To define the new notation required for this continuous model, let $m^{I}(t)$ and $m^{II}(t)$ be the rates of admixture, in the sense of the last chapter, at time t ; more strictly $m^{I}(t)$ and $m^{II}(t)$ are constant over the small interval $(t, t+\delta t)$ so that the amounts of migration which take place over that interval are $m^{I}(t)\delta t$ and $m^{II}(t)\delta t$ respectively. The gene frequencies in the two populations will be defined at time t as $q^{I}(t)$ and $q^{II}(t)$.

This model implies a pair of relations describing the frequencies in terms of the admixture rates; these are:

$$q^{I}(t+\delta t) = (1 - m^{II}(t)\delta t)\, q^{I}(t) + m^{II}(t)\,\delta t\, q^{II}(t) \qquad (2.1)$$

$$q^{II}(t+\delta t) = m^{I}(t)\,\delta t\, q^{I}(t) + (1 - m^{I}(t)\,\delta t)\, q^{II}(t) \qquad (2.2)$$

Again, defining $\Sigma(t)$ and $\Delta(t)$ as in the previous section as the sum and difference of $q^{I}(t)$ and $q^{II}(t)$, forming the difference of the pair of relations gives:

$$\Delta(t+\delta t) = (1 - \delta t\, m^{I}(t) - \delta t\, m^{II}(t))\,\Delta(t) \qquad (2.3)$$

or $$\Delta(t+\delta t) - \Delta(t) = -(m^{I}(t) + m^{II}(t))\,\Delta(t)\,\delta t$$

Dividing by δt and letting $\delta t \to 0$

$$\frac{d\Delta(t)}{dt} = -(m^{I}(t) + m^{II}(t))\,\Delta(t) \qquad (2.4)$$

which is a first order differential equation in $\Delta(t)$ with solution:

$$\Delta(t) = \Delta_0\, e^{-\int_0^t (m^{I}(t') + m^{II}(t'))\,dt'} \qquad (2.5)$$

Adding equations (2.1) and (2.2) gives:

$$\Sigma(t+\delta t) = \Sigma(t) + \delta t\,(m^{I}(t) - m^{II}(t))\,\Delta(t)$$

substituting for $\Delta(t)$, rearranging and taking the limit as

$$\frac{d\Sigma(t)}{dt} = (m^{I}(t) - m^{II}(t))\,\Delta_0\, e^{-\int_0^t (m^{I}(t') + m^{II}(t'))\,dt'} \qquad (2.6)$$

and this differential equation in $\Sigma(t)$ has solution:

$$\Sigma(t) = \Sigma_0 + \Delta_0 \int_0^t (m^{I}(t'') - m^{II}(t''))\, e^{-\int_0^{t''} m^{I}(t') + m^{II}(t')\,dt'}\, dt'' \qquad (2.7)$$

~~Where~~ In the above Σ_0 and Δ_0 are as defined in the previous section, the sum and difference respectively of the initial frequencies, ~~i.e. those~~ at time $t=0$.

THEOREM 2.1 The continuous model defined by the relations (2.1) and (2.2) has the general solution:

$$q^{I}(t) = \tfrac{1}{2}\Big(\Sigma_0 + \Delta_0 \int_0^t (m^{I}(t'') - m^{II}(t'')) \, e^{-\int_0^{t''} (m^{I}(t') + m^{II}(t'))\,dt'} dt'' + \Delta_0 \, e^{-\int_0^t (m^{I}(t') + m^{II}(t'))\,dt'}\Big) \tag{2.8}$$

and

$$q^{II}(t) = \tfrac{1}{2}\Big(\Sigma_0 + \Delta_0 \int_0^t (m^{I}(t'') - m^{II}(t'')) \, e^{-\int_0^{t''} (m^{I}(t') + m^{II}(t'))\,dt'} dt'' - \Delta_0 \, e^{-\int_0^t (m^{I}(t') + m^{II}(t'))\,dt'}\Big) \tag{2.9}$$

Proof follows by addition and subtraction of (2.5) and (2.7).

THEOREM 2.2 When one of the admixture rates is zero and the other a general function, the solutions are:

for $m^{I}(t) = m(t), \quad m^{II}(t) = 0,$ (2.10)

$$q^{I}(t) = 0 \tag{2.11}$$

$$q^{II}(t) = \tfrac{1}{2}\left(\Sigma_0 + \Delta_0 - 2\Delta_0 \, e^{-\int_0^t m(t')\,dt'}\right) \tag{2.12}$$

Proof follows by substitution of (2.10) into (2.8) and (2.9).

THEOREM 2.3 When the admixture rates are equal for,

$$m^{I}(t) = m^{II}(t) = m(t) \tag{2.13}$$

the solutions are:

$$q^{I}(t) = \tfrac{1}{2}\left(\Sigma_0 + \Delta_0 e^{-2\int_0^t m(t')\,dt'}\right) \tag{2.14}$$

$$q^{II}(t) = \tfrac{1}{2}\left(\Sigma_0 - \Delta_0 e^{-2\int_0^t m(t')\,dt'}\right) \tag{2.15}$$

Proof follows by substituting (2.13) into (2.8) and (2.9)

THEOREM 2.4 When the admixture rates are different but constant

$$\text{for} \quad m^{I}(t) = m^{I} \qquad \text{and} \quad m^{II}(t) = m^{II} \tag{2.16}$$

the solutions are:

$$q^{I}(t) = \tfrac{1}{2}\left(\Sigma_0 + \frac{\Delta_0(m^{I}-m^{II})}{m^{I}+m^{II}} + \frac{2m^{II}\Delta_0 e^{-(m^{I}+m^{II})t}}{m^{I}+m^{II}}\right) \tag{2.17}$$

$$q^{II}(t) = \tfrac{1}{2}\left(\Sigma_0 + \frac{\Delta_0(m^{I}-m^{II})}{m^{I}+m^{II}} - \frac{2m^{I}\Delta_0 e^{-(m^{I}+m^{II})t}}{m^{I}+m^{II}}\right) \tag{2.18}$$

Proof follows by substitution of (2.16) into (2.8) and (2.9).

THEOREM 2.5 (Glass and Li) For unidirectional gene flow, where one admixture rate is constant and the other zero, i.e.

$$\text{for} \qquad m^{I}(t) = m, \qquad m^{II}(t) = 0 \tag{2.19}$$

the solutions are:

$$q^{I}(t) = q_0^{I} \tag{2.20}$$

$$q^{II}(t) = q_0^{I} - (q_0^{I} - q_0^{II})e^{-mt} \tag{2.21}$$

Proof This follows by using (2.19) and Theorem 2.4.

THEOREM 2.6 For unidirectional hyperbolically increasing gene flow the solutions, where,

$$m^{I}(t) = 0, \qquad m^{I}(t) = \frac{c}{a-t} \tag{2.22}$$

in which a and c are suitable constants, are given by:

$$q^{I}(t) = q_0^{I} \tag{2.23}$$

$$q^{II}(t) = q_0^{I} - (q_0^{I} - q_0^{II})\frac{(a-t)^c}{a^c} \tag{2.24}$$

Proof follows by using (2.22) and Theorem 2.2.

THEOREM 2.7 **For equal hyperbolically increasing admixture rates, i.e. for**

$$m^{I}(t) = m^{II}(t) = \frac{c}{a-t} \tag{2.25}$$

the solution is:

$$q^{I}(t) = \tfrac{1}{2}\left(\Sigma_0 + \Delta_0 \frac{(a-t)^{2c}}{a^{2c}}\right) \tag{2.26}$$

$$q^{II}(t) = \tfrac{1}{2}\left(\Sigma_0 - \Delta_0 \frac{(a-t)^{2c}}{a^{2c}}\right) \tag{2.27}$$

Proof follows by using (2.25) and Theorem 2.3.

THEOREM 2.8 **For unidirectional parabolically increasing gene flow, i.e. for**

$$\begin{aligned} m^{I}(t) &= at^2 + bt + c \\ m^{II}(t) &= 0 \end{aligned} \tag{2.28}$$

where a, b and c are constants, the solution is:

$$q^{I}(t) = q_0^{I} \tag{2.29}$$

$$q^{II}(t) = q_0^{I} - 2\Delta_0 e^{-(\frac{1}{3}at^3 + \frac{1}{2}bt^2 + ct)} \tag{2.30}$$

Proof follows by using (2.28) and Theorem 2.2

THEOREM 2.9 **For equal parabolically increasing admixture rates, i.e. for,**

$$m^{I}(t) = m^{II}(t) = at^2 + bt + c \qquad (2.31)$$

the solution is:

$$q^{I}(t) = \tfrac{1}{2}\left(\Sigma_0 + \Delta_0 e^{-2\left(\frac{1}{3}at^3 + \frac{1}{2}bt^2 + ct\right)}\right) \qquad (2.32)$$

$$q^{II}(t) = \tfrac{1}{2}\left(\Sigma_0 - \Delta_0 e^{-2\left(\frac{1}{3}at^3 + \frac{1}{2}bt^2 + ct\right)}\right) \qquad (2.33)$$

Proof **follows by using (2.31) and Theorem 2.3.**

It is clear from the above results, particularly where constant admixture rates are present, as in Theorems 2.4 and 2.5, that the limits of frequencies as $t \to \infty$ **are the same as the corresponding ones for the discrete model considered in Chapter One.**

CHAPTER THREE

MORE THAN TWO POPULATIONS WITH DISCRETE GENERATIONS AND CONSTANT MIGRATION: DETERMINATION OF GENE FREQUENCIES

Returning to the model with discrete generations of Chapter One, it is possible to extend the arguments of that section from two populations to any general number of populations. Corresponding to equations (1.3) and (1.4) for two populations are the N equations with constant admixture rates,

$$\begin{aligned} q_{1n} &= m_{11} q_{1(n-1)} + m_{12} q_{2(n-1)} + \dots + m_{1N} q_{N(n-1)} \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ q_{Nn} &= m_{N1} q_{1(n-1)} + m_{N2} q_{2(n-1)} + \dots + m_{NN} q_{N(n-1)} \end{aligned} \qquad (3.1)$$

where the notation has been generalised so that m_{ij} is the component of admixture in population i deriving from population j at the end of the n^{th} generation for

$$i = 1, 2, \dots, N \qquad \text{and} \qquad j = 1, 2, \dots, N$$

and q_{in} is the frequency in population i at the end of the n^{th} generation.

When the condition that,

$$m_{ii} = 1 - \sum_{j \neq i} m_{ij} \qquad (3.2)$$

for all i is added to the N equations above, they completely describe the model. In view of this it will be convenient to represent these equations in matrix form so that further reference to the model may be simplified. The equivalent matrix equation to the set (3.1) is:

$$q_n = M q_{n-1} \qquad (3.3)$$

where q_n is the vector of frequencies $\{q_{in}\}$
q_{n-1} " " " " $\{q_{i(n-1)}\}$

and M is the matrix of admixture rates $\{m_{ij}\}$.

For this model, in which the matrix M has constant elements, it is a straightforward matter to obtain an expression for q_n in terms of the vector of initial frequencies, q_0. From (3.3),

$$q_n = M q_{n-1} = M(M q_{n-2}) = \ldots = M^n q_0 \qquad (3.4)$$

and this has the same form as (3.3). M^n is the matrix of accumulated admixture over n generations, having an important property. It contains as elements, the admixture rates which would need to be exercised over a single generation to achieve the change in the vector of initial frequencies equivalent to n generations under the operation of the constant rates which are the elements of M.

3.1 Determination of future gene frequencies from exact data.

The relation (3.4) may be used to obtain the vector from known n, q_0 and M by repeated matrix multiplication n times. In general, this is not the quickest or the most accurate method for determining q_n from the relation. Other properties possessed by the matrix M are of use here. The rows of the matrix all sum to unity so that at least one latent root is unity and the matrix elements are all in the range (0,1) so that the latent roots λ_i $(i = 1,2,\ldots,N)$ are such that $|\lambda_i| \leq 1$ with the largest root unity. These results were proved by Fréchet (1937-8, p.105) for stochastic matrices, which have similar properties to M here; reference is made for this, and for some of what follows, to Bartlett (1956, pp.24-30).

In general the λ_i may be considered to be distinct and it will appear later what difficulties arise when this is not so. Under this single assumption of distinctness, there exists a spectral resolution of M into N terms so that:

$$M = \sum \lambda_i M_i \qquad (3.5)$$

where the M_i are component matrices of the spectral set, with the properties

$$\sum M_i = I \qquad (3.6)$$

where I is the unit matrix, and,

$$M_i M_j = 0 \quad (i \neq j) \tag{3.7}$$

$$= M_i \quad (i = j) \tag{3.8}$$

It follows at once from these properties that

$$M^n = \Sigma \lambda_i^n M_i \tag{3.9}$$

For small N but large n it is in general more accurate and convenient to evaluate the latent roots and the spectral set of M, rather than carry out a large number of multiplications of M. This is particularly true for $N = 3$ where the latent roots are directly available, the largest being unity and the other two being the roots of a quadratic equation.

To determine the matrices of the spectral set, M_i, it is first necessary to find vectors s_i and t_i associated with each latent root λ_i such that,

$$M s_i = \lambda_i s_i \tag{3.10}$$

and

$$t_i' M = \lambda_i t_i' \tag{3.11}$$

the spectral matrices are then:

$$M_i = (s_i t_i')/(s_i' t_i) \tag{3.12}$$

where a dash $(')$ denotes the transpose of a vector throughout.

To illustrate this method of decomposition and its application to the determination of future gene frequencies, consider the following data from three Nilotic populations in southern Sudan.

The matrix of admixture rates is:

$$M = \begin{pmatrix} .9850 & .0125 & .0025 \\ .0138 & .9775 & .0087 \\ 0 & .0098 & .9902 \end{pmatrix}$$

in which the populations referred to in order are Nuer, Dinka and Shilluk.

The latent roots of M are,

$\lambda_1 = 1$, $\lambda_2 = 0.986937$ and $\lambda_3 = 0.965763$

and the spectral matrices are:

$$M_1 = \begin{pmatrix} .302\,388 & .328\,683 & .768\,929 \\ .302\,388 & .328\,683 & .768\,929 \\ .302\,388 & .328\,683 & .768\,929 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} .419\,578 & .058\,890 & -.478\,469 \\ .162\,804 & .022\,850 & -.185\,654 \\ -.488\,945 & -.068\,626 & .557\,571 \end{pmatrix}$$

$$M_3 = \begin{pmatrix} .278\,033 & -.387\,572 & .109\,539 \\ -.465\,191 & .648\,467 & -.183\,276 \\ .186\,557 & -.260\,057 & .073\,500 \end{pmatrix}$$

and using (3.4 and (3.5), together with $q_0 = \begin{pmatrix} .5750 \\ .5870 \\ .5047 \end{pmatrix}$,

$$q_n = \sum \lambda_i^n M_i q_0 \qquad (3.13)$$

$$= \begin{pmatrix} .546\ 435 \\ .546\ 435 \\ .546\ 435 \end{pmatrix} + (0.986\ 937)^n \begin{pmatrix} .033\ 165 \\ .012\ 869 \\ -.038\ 648 \end{pmatrix} + (0.965\ 763)^n \begin{pmatrix} -.004\ 600 \\ .007\ 697 \\ .003\ 087 \end{pmatrix} \qquad (3.14)$$

In this form it is a simple matter to determine q_n for any n and Table 3.1 shows the results of seven evaluations.

TABLE 3.1

	5	10	15	20	30	40	60
Nuer	.5736	.5723	.5709	.5696	.5672	.5649	.5610
Dinka	.5650	.5632	.5616	.5602	.5578	.5559	.5532
Shilluk	.5076	.5104	.5129	.5152	.5193	.5228	.5285

It is clear that the second and third terms of the expression for q_n (3.14) will tend to zero as $n \to \infty$. The first term then represents the vector of asymptotic frequencies; the quality of these frequencies is to be expected and is a necessary result of some migration taking place between all three populations. This property, that the first term expresses the asymptotic frequencies depends upon the assumption that unity is the only latent root of modulus one. By Fréchet's result all the remaining roots are then in modulus less than unity; hence their n th powers converge to zero.

3.2 Estimation of future gene frequencies from sample data.

The determination of the last section takes no account of any statistical properties which the known values may possess; it provides exact answers to exact known frequencies and admixture rates. In practice it is likely that both frequencies and rates are obtained as estimates from sampling, since the entire population is too large for a complete census to be carried out.

However, the admixture rates can often be obtained to a much greater degree of accuracy than the gene frequencies and for the moment they are considered to be known exactly. Each gene frequency is an independent binomially distributed estimate of a proportion obtained from a sample. The samples used for this purpose are large and usually contain several hundred individuals; the binomial distribution in these circumstances, may be replaced by the normal distribution, to a good approximation.

By analogy with the relation (3.4) it is worthwhile to describe the statistical model. This is done by letting q_o represent a vector of random variables, each of which is distributed independently of any other and whose expected values may be exhibited in a vector called $E(q_o)$ and whose variances are contained in the vector $V(q_o)$. In view of the presumed exactness of M, and therefore M^n, the expected value of the vector q_n, $E(q_n)$ is given by:

$$E(q_n) = M^n E(q_o) \qquad (3.15)$$

and in view of the known result for the variance of a sum of independently distributed random variables (e.g. Feller Theorem 2 p.216),

$$V(q_n) = m^{(n)} V(q_0) \qquad (3.16)$$

where $m^{(n)}$ is a matrix consisting of the squares of the elements of the matrix M^n.

It is straightforward to take the square roots of the elements of $V(q_n)$ to obtain standard errors of the estimates of $E(q_n)$ and, using the approximate normality of the distributions of the elements of q_0, to form confidence limits for these estimates.

In the example considered earlier, the admixture rates were, in fact, estimates by sampling and are not therefore exact. The fact that each rate is nearly zero or unity makes the assumption of exactness valid since the binomial sampling variances, being proportional to $m(1-m)$ are all quite small. For the gene frequencies, all of which are in the range 0.5 to 0.6, the sampling variances are much higher and the statistical model considered above is appropriate, at least where approximate variances for predicted values are required.

When it is not admissible to assume the exactness of M, the analysis is more complex. In order to investigate the sampling properties of estimates q_m when the elements of M are statistically distributed, the nature of the distribution of these elements is now studied.

The usual method for estimating the gene flow rates by sampling, requires an enumeration for a sample from each population of those individuals who have taken part in each possible type of intermixture to form the present generation. From these enumerations proportions of individuals are obtained directly to represent the corresponding rates of gene flow between the populations.

For two populations, under the assumption of a constant independent probability for each individual to migrate, the distribution of each rate is binomial, or for large samples, approximately normal. For many populations however, the distribution of each rate is multinomial, or again for large samples, normal.

Let M represent the matrix of observed rates obtained from sampling so that if the expectation of this matrix is written $\mathcal{U}$ then

$$M = \mathcal{U} + \epsilon \qquad (3.17)$$

where ϵ represents the matrix of errors. From this,

$$M^n = (U + \epsilon)^n \tag{3.18}$$

or considering only products in the expansion of the right-hand side with not more than one factor ϵ, using an assumption of smallness of the errors,

$$M^n \doteq U^n + \sum_{r=0}^{n-1} U^r \epsilon U^{n-r-1} \tag{3.19}$$

The difficulty of the second term requires that special forms for the matrix U be considered separately. In many situations, for example, U is nearly scalar because of the retention of most of the gene pool within each population. When this is so, U commutes approximately with any other matrix and, therefore, with ϵ so that (3.19) becomes

$$M^n \doteq U^n + n\epsilon U^{n-1} \tag{3.20}$$

Alternatively, the matrix ϵ may contain elements which are of similar size and in this case write,

$$\epsilon \doteq \varepsilon u \tag{3.21}$$

where ε is a scalar representing the error in each gene flow rate, and u is the matrix of unit elements. When, in addition, the matrix of gene flow rates is symmetrical, indicating that the gene flow from one population to another equals that in the reverse direction, further

simplification of (3.19) is possible. This is due to the commutativity between any symmetric M and U and as a result,

$$M^n = U^n + n\varepsilon U U^{n-1} \qquad (3.22)$$

An extension of this approach now leads to a relaxation of the condition of similarity on the elements of ϵ. Let ε_{min} and ε_{max} be the minimum and maximum elements respectively of ϵ. Using the fact that the elements of U are non-negative, bounds for the size of the second term in the expression for M^n in (3.19) follow since

$$n\varepsilon_{min} U U^{n-1} \leq \sum_{r=0}^{n-1} U^r \epsilon U^{n-r-1} \leq n\varepsilon_{max} U U^{n-1} \qquad (3.23)$$

The application of these results for common special types of admixture matrix, is now possible in order to study the sampling properties of the estimates of gene frequencies. First the above results are summarized. Simple forms for the error in M^n seem to be available

(i) when M is near scalar, as direct errors, and

(ii) when M is near symmetric, as error bounds; direct errors can be found for (ii) if the error matrix for M has equal elements.

The estimation of gene frequencies after n generations from given sample values, at time zero, and from sample estimates of the

gene flow rates, is achieved by using the relation,

$$q_n = M^n q_0 \tag{3.24}$$

If χ_n represents the vector of true frequencies after n generations for a constant admixture matrix U of gene flow rates, and χ_0 a vector of true initial frequencies, then write

$$\chi_n = U^n \chi_0 \tag{3.25}$$

and suppose,

$$\begin{aligned} E(q_n) &= \chi_n \\ E(q_0) &= \chi_0 \\ E(M) &= U \end{aligned} \tag{3.26}$$

with errors ϵ_n, ϵ_0 and ϵ defined so that,

$$\begin{aligned} q_n &= \chi_n + \epsilon_n \\ q_0 &= \chi_0 + \epsilon_0 \\ M &= U + \epsilon \end{aligned} \tag{3.27}$$

From (3.24),

$$q_n = (U+\epsilon)^n (\chi_0 + \epsilon_0)$$

$$= (U^n + \sum_{r=0}^{n-1} U^r \epsilon U^{n-r-1} + \ldots)(\chi_0 + \epsilon_0)$$

and ignoring terms containing more than one ϵ term,

$$= U^n \chi_0 + \left(\sum_{r=0}^{n-1} U^r \epsilon U^{n-r-1}\right)\chi_0 + U^n \epsilon_0$$

When the errors ϵ and ϵ_0 are independent,

$$\mathrm{var}(q_n) = \mathrm{var}\{(U+\epsilon)^n \chi_0\} + \mathrm{var}(U^n \epsilon_0) \qquad (3.28)$$

where the terms on the right represent the components of variance of q_n due to ϵ and ϵ_0 respectively.

Consider now the special types of admixture matrix mentioned above. First suppose that U is scalar: from (3.20),

$$q_n = U^n \chi_0 + n\epsilon U^{n-1} \chi_0 + U^n \epsilon_0$$

or

$$q_n = U^n \chi_0 + n\epsilon \chi_{n-1} + U^n \epsilon_0 \qquad (3.29)$$

where χ_{n-1} is defined in an obvious way as the expected vector of frequencies after $n-1$ generations. From (3.29) the variance of q_n is the expectation of ϵ_n^2 or, using the independence of ϵ and ϵ_0, (3.28) becomes

$$\mathrm{var}(q_n) = \mathrm{var}(n\epsilon \chi_{n-1}) + \mathrm{var}(U^n \epsilon_0)$$

or in suffix notation,

$$\mathrm{var}(q_{ni}) = n^2 \sum_{j=1}^{N} \chi_{n-1,j}^2 \, \mathrm{var}(\epsilon_{ij}) + n^2 \sum_{j \neq k} \chi_{n-1,j} \chi_{n-1,k} \, \mathrm{cov}(\epsilon_{ij}, \epsilon_{ik}) + \sum_{j=1}^{N} U_{ij}^{(n)} \, \mathrm{var}(\epsilon_{0j}) \qquad (3.30)$$

in which the $\mathrm{var}(\epsilon_{ij})$ and $\mathrm{cov}(\epsilon_{ij}, \epsilon_{ik})$ are the multinomial sampling

variances and covariances of the elements of M, the $var(\epsilon_{oj})$ are the binomial variances of the elements of q_o and $\{ U^{(n)}_{ij} \}$ is the matrix whose elements are the squares of those of U^n, using the independence of errors ϵ_{oj} at different loci j.

Equation (3.30) may be written more concisely for matrix computation, by defining N matrices $V_i = \{ V_{ijk} \}$ such that

$$V_{ijk} = cov (\epsilon_{ij} , \epsilon_{ik}) \qquad (3.31)$$

The well known results for multinomial sampling variances (see e.g. Rao (p.35)) now allow the expression of the elements of each V_i in terms of the admixture rates m_{ij}; briefly the relevant variances and covariances are

$$\begin{aligned} var (m_{ij}) &= m_{ij} (1 - m_{ij}) / w_i \\ cov (m_{ij} , m_{ik}) &= - m_{ij} m_{ik} / w_i \end{aligned} \qquad (3.32)$$

where w_i is the number of individuals in the sample from which the admixture rates for the i^{th} population were obtained. V_i may now be written

$$V_i = (diag (m_i) - m_i m_i') / w_i \qquad (3.33)$$

where m_i is the vector $\{ m_{ij} \}$, the i^{th} row of M, and $diag(m_i)$ is the matrix whose diagonal elements are those of this vector.

From (3.30)

$$\text{var}(q_{ni}) = n^2 \chi'_{n-1} V_i \chi_{n-1} + \mathcal{U}_i^{(n)} \text{var}(\epsilon_o) \qquad (3.34)$$

in which $\mathcal{U}_i^{(n)}$ is the i^{th} row of the matrix $\{ \mathcal{U}_{ij}^{(n)} \}$.

For a near symmetric matrix M, possibly an upper bound to $\text{var}(q_n)$ would be of greatest value and several bounds may be obtained readily. It is convenient to make use of the fact that the multinomial covariances which appear in considering the variance of the product of the matrix ϵ with any other matrix, make only a negative contribution to this variance and in considering an upper bound they may be disregarded. Using the relation (3.23) the required bound can at once be obtained for

$$\text{var}(q_n) \leq \text{var}(\mathcal{U}^n \epsilon_o) + \text{var}(n \epsilon_{max} u \mathcal{U}^{n-1} \chi_o) \qquad (3.35)$$

and this relation also presumes the non-negative property for the elements of χ_o and $\mathcal{U}$ which is always satisfied. The second term on the right now appears as a vector of constant elements, each equal to

$$\text{tr} \{ n \, \text{diag} (\mathcal{U}^{n-1} \chi_o) \}^2 \, \text{var} (\epsilon_{max})$$

or

$$n^2 (\chi_{n-1} u)^2 \, \text{var} (\epsilon_{max})$$

where u is the vector of unit elements and $\text{var}(\epsilon_{max})$ is the largest variance of an element in M; if w_{min} is the smallest sample size,

so that

$$w_{min} = \min_i (w_i)$$

then $var(\varepsilon_{max})$ may be replaced by $(1/4\, w_{min})$ because of the well known result that the multinomial variate m has maximum sampling variance at $m = \frac{1}{2}$. The simple upper bound is then

$$var(q_{ni}) \leq \mathcal{U}_i^{(n)}\, var(\epsilon_o) + n^2 (\chi_{n-1}\, u)^2 / 4 w_{min} \qquad (3.36)$$

The second term can be taken one stage further to give an upper bound for all frequencies, since each element of χ_{n-1} is less than one, the second term is less than $(9n^2/4w_{min})$. As this bound involves only n, the number of generations for which frequencies are being predicted, and w_{min} the smallest sample size for a determined admixture rate, one of these quantities can be found in terms of the other. For practical purposes, a frequency estimated by the method of this chapter which has a variance of 0.25 (the binomial maximum value) or more, will have no value so it seems reasonable to require

$$\frac{9n^2}{4 w_{min}} \leq 0.25$$

or

$$w_{min} \geq 9n^2$$

which implies for all i that

$$w_i \geq 9n^2 \qquad (3.37)$$

This simple condition requires that for a contribution of less than 0.25 from the errors in the admixture rates to the upper bound of the variance of q_{ni}, the samples for the determination of these rates should be at least nine times the square of the number of generations, e.g. to predict q_5 from q_0 would require sample sizes of at least 225 for the determination of the m_{ij}, and to obtain q_{10} would require samples sizes to be at least 900.

It should be stressed that this condition still does not ensure reasonable estimates q_n because the first term on the right of (3.36), the variance due to the error in the initial frequencies q_0, is not negligible. On the other hand, the condition (3.37) imposes an upper bound which will be too high for many real situations as the generality has been carried too far: the bound provided by (3.36) is stricter and would be worth the small amount of computation required.

An application of an upper bound of the type given above is perhaps worth some discussion. When it is required to test an estimated frequency q_{ni} against some known constant c such a test may be carried out by observing that the estimate is approximately normally distributed, being no more than a linear combination of independent approximately normal observations. Common practice here is to compare the ratio of the difference between the estimate and the constant to the variance of the estimate with a standardised normal

variable u (i.e. with mean zero and unit variance) by means of a table of the normal distribution. Suppose that δ is the difference between the constant and the estimate which would be important for discrimination, and that u_α is that value of u which is exceeded only with probability α. Then the requirement for the variance of the estimate may be written as

$$\text{var}(q_{ni}) \leq \delta / u_\alpha \qquad (3.38)$$

and this can only be true for <u>all</u> cases if the upper bound of (3.36) is also less than δ / u_α. From above it follows that this is so only if

$$n_i \geq q n^2 u_\alpha / 4\delta \qquad (3.39)$$

for all i. For example, if $n = 5$, $\alpha = 0.05$, $\delta = 0.1$, $n_i \geq 1100$ approximately.

Finally, for the application of any of the relations concerning the variance of q_n given above, it will be necessary in practice to use observed values q_0 instead of γ_0 and M instead of U.

Example. Using the matrix M and vector q_0 given in the last section together with sample sizes of 200, 288 and 255 for the estimates in the rows of M and q_0, variances and bounds for future gene frequencies are given in the table below.

TABLE 3.2

generations n	frequencies q_n	variance due to error in q_0	variance due to error in M	bounds
5	0.573 625	0.001 057	0.000 000 06	0.005
	0.564 951	0.000 689	0.000 000 11	
	0.507 653	0.000 892	0.000 000 13	
10	0.572 267	0.000 928	0.000 000 22	0.021
	0.563 150	0.000 577	0.000 000 36	
	0.510 370	0.000 817	0.000 000 44	
20	0.569 639	0.000 743	0.000 000 69	0.083
	0.560 162	0.000 447	0.000 001 07	
	0.515 186	0.000 701	0.000 001 27	
30	0.567 172	0.000 623	0.000 001 25	0.186
	0.557 815	0.000 386	0.000 001 81	
	0.519 299	0.000 615	0.000 002 10	
60	0.560 934	0.000 447	0.000 002 46	0.745
	0.553 233	0.000 341	0.000 003 15	
	0.528 494	0.000 464	0.000 003 45	

The variance components are calculated from the two terms on the right-hand side of equation (3.34) and the bounds for the second term given in the end column, are obtained from (3.35); in this case

the upper bounds are rather rough and other possible bounds derived from simpler expressions are not tabulated here as they are even poorer as to an indication of the likely variances. Perhaps some comment is necessary on the reduction of variance due to q_0 with n: this is due to the fact that smaller fractions of the initial frequencies are taken for higher generations. The increasing variance due to M expresses the increasing uncertainty in values of q_n for larger n.

3.3 Determination of past frequencies from exact data.

From the relation (3.4), by multiplying through by $(M^n)^{-1}$, the inverse of the accumulated admixture matrix,

$$q_0 = (M^n)^{-1} q_n \qquad (3.40)$$

and this relation may be used to determine past frequencies from present ones. The method of section 3.1 still holds because in general the latent roots of M^{-1} will be distinct and a resolution will, therefore, be available. Of course, the latent roots are no longer confined to the range (0,1) ~~and~~ although unity is ~~not~~ necessarily one root; the only effect of these conditions in the last section, was to exhibit the asymptotic frequencies and these are not relevant here.

The fact that one of the latent roots is ~~not~~ known to be unity makes it only a little ~~more difficult~~ easier to find these roots, but there are several iterative methods (see e.g. Aitken (1937)) which will usually prove adequate. The arguments and example of section 3.1 serve to show how the frequencies in the vector q_0 may be determined given the exact values of M and q_n.

This method is applied four times to the three Nilotic populations to determine the frequencies q_0 at four loci from given frequencies q_n at the same loci, and from the matrix M of the last

section, taking the number of generations, n, as twenty; the results are shown below in Table 3.3.

TABLE 3.3

	q_0 vectors at loci				q_n vectors at loci			
	M	S	d	B	M	S	d	B
Nuer	.5803	.2559	.1766	.1324	.5750	.2254	.1782	.1270
Dinka	.5786	.1197	.1566	.1158	.5670	.1505	.1759	.1155
Shilluk	.4900	.1501	.2709	.0894	.5047	.1477	.2523	.0941

3.4 Determination of past frequencies from sample data.

The difficulties of estimating past frequencies from sample data are somewhat greater than those of the problem of prediction of future frequencies from sample data which were discussed in section 3.2. It is possible, however, to use some of the results obtained in that section once a relation is established between the errors in an observed admixture matrix and in its inverse in the present context. This relation is now derived.

As previously, let ϵ represent the error matrix for an observed admixture matrix M, whose expected value is given by U. Let ϵ_{-1} denote the error matrix for M^{-1} computed from M by inversion and it follows that U^{-1}, the inverse of U, is the expected value of M^{-1}. These definitions imply,

$$M = U + \epsilon \tag{3.41}$$

$$M^{-1} = U^{-1} + \epsilon_{-1} \tag{3.42}$$

and,
$$MM^{-1} = I \tag{3.43}$$

Substitution in (3.43) and ignoring the product $\epsilon\epsilon_{-1}$, as a small quantity, gives:

$$UU^{-1} + U\epsilon_{-1} + \epsilon U^{-1} = I$$

or,
$$U\epsilon_{-1} + \epsilon U^{-1} = 0$$

and,
$$\epsilon_{-1} = -U^{-1}\epsilon U^{-1} \tag{3.44}$$

Consider next the variance of q_o, the frequency predicted from a sample vector q_n containing error ϵ_n, and M, by the relation (3.40). Substitution from above leads to:

$$\begin{aligned} q_o = (M^{-1})^n q_n &= (U^{-1} + \epsilon_{-1})^n (\chi_n + \epsilon_n) \\ &= (U^{-1} - U^{-1}\epsilon U^{-1})^n (\chi_n + \epsilon_n) \\ &= (U^{-1})^n (U - \epsilon)^n (U^{-1})^n (\chi_n + \epsilon_n) \\ &= (U^{-1})^n (U-\epsilon)^n \chi_o + (U^{-1})^n (U-\epsilon)^n (U^{-1})^n \epsilon_n \end{aligned}$$

and, ignoring product terms with more than one of the factors ϵ and ϵ_n,

$$= (U^{-1})^n (U-\epsilon)^n \chi_o + (U^{-1})^n \epsilon_n \qquad (3.45)$$

Following (3.28) write,

$$\text{var}(q_o) = (U^{-1})^{(n)} \text{var}\{(U-\epsilon)^n \chi_o\} + \text{var}\{(U^{-1})^n \epsilon_n\} \qquad (3.46)$$

in view of the independence between ϵ and ϵ_n where $(U^{-1})^{(n)}$ contains the squares of the elements of $(U^{-1})^n$. It is now clear that the second term on the right may be evaluated directly. The first term $\text{var}(q_o|\epsilon)$ contains a factor $\text{var}\{(U-\epsilon)^n \chi_o\}$ which is identical with the term $\text{var}\{(U+\epsilon)^n \chi_o\}$ discussed in section 3.2 when computational formulae were derived for special cases of the admixture matrix. These formulae may be applied once more in order to evaluate variances or limits for equation (3.46).

3.5 Admixture matrices with repeated latent roots.

In this section a more general case is considered; when the admixture matrix need not have distinct roots, it is not possible to spectrally decompose the matrix in the manner described in section 3.1 in order to simplify the estimation of certain frequencies. This is not to say that a decomposition does not exist, for it does, but the method is rather less straightforward. Such special cases of the admixture matrix are sufficiently common in practice to require the method to be given here. A simple example is that with three populations into one of which there is gene flow, but out of which there is none; it will become clear that this model uses an admixture matrix two of whose latent roots are unity, and so cannot be treated by the method given. It will be dealt with later by application of the more general method given below.

Define the matrix $\Phi(\lambda)$ by the relation,

$$\Phi(\lambda) = \lambda I - M \qquad (3.47)$$

where I is the unit matrix, so that the roots of the determinantal polynomial equation are the latent roots of the matrix M:

$$|\Phi(\lambda)| = 0 \qquad (3.48)$$

This equation is usually referred to as the characteristic equation. The adjoint of the matrix $\Phi(\lambda)$ will be required, and this is defined as $A(\lambda)$ which is the transpose of the matrix whose elements are the

cofactors of the elements of $\Phi(\lambda)$.

The general case of repeated latent roots is now considered. Suppose that the first s out of the N roots are distinct and that of these the i^{th} root λ_i appears r_i times ($i = 1,2,\ldots,s$). It follows that

$$\sum_{i=1}^{s} r_i = N \tag{3.49}$$

For the case in which each r_i is unity, the N roots will be distinct with no repetitions, and $s = N$; the general result given in the following theorem will lead to the spectral resolution of previous sections as a special case.

THEOREM 3.1 (Confluent form of Sylvester's Theorem) Any polynomial function $P(M)$ of the general matrix M may be expressed as a sum of polynomial terms in the distinct latent roots of M as follows:

$$P(M) = \sum_{i=1}^{s} \frac{1}{(r_i-1)!} \left[\frac{d^{r_i-1}}{d\lambda^{r_i-1}} \frac{P(\lambda)A(\lambda)}{\psi_i(\lambda)} \right]_{\lambda=\lambda_i} \tag{3.50}$$

where, $$\psi_i(\lambda) = \prod_{\substack{j=1 \\ j\neq i}}^{s} (\lambda - \lambda_j)^{r_j} \tag{3.51}$$

Proof is given by Fraser, Duncan and Collar (1946 par.3.10) or by Sylvester (1883)

THEOREM 3.2 Any power of the general matrix M may be expressed as a sum of polynomial terms in the distinct latent roots of M :

$$M^n = \sum_{i=1}^{s} \frac{1}{(r_i-1)!} \left[\frac{d_i^{r_i-1}}{d\lambda^{r_i-1}} \frac{\lambda^n A(\lambda)}{\psi_i(\lambda)} \right]_{\lambda=\lambda_i} \qquad (3.52)$$

Proof follows from substitution of,

$$P(\lambda) = \lambda^n \qquad (3.53)$$

in the result of Theorem 3.1

Where the latent roots are distinct, the above theorems lead to the general result which follows:

THEOREM 3.3 Any power of the general matrix M, which has all latent roots distinct, may be expressed as a sum of matrices, each containing that power of one distinct root as a factor,

$$M^n = \sum_{i=1}^{s} \lambda_i^n \frac{A(\lambda_i)}{\psi_i(\lambda_i)} \qquad (3.54)$$

Proof follows directly from Theorem 3.2.

It is a straightforward matter to show that the result of this last theorem 3.3 corresponds with that given in the earlier sections 3.1 and 3.2.

Example The general resolution of Theorem 3.2 is now applied to the common admixture matrix referred to earlier which has two latent roots equal to unity, and results from migration out of two populations into a third population but no migration in the reverse direction, or between the two populations. Such a matrix may be denoted by,

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{3.55}$$

in which m_1 and m_2 are the admixture rates out of populations 1 and 2, into population 3, and,

$$m_3 = 1 - m_1 - m_2 \tag{3.56}$$

Directly, it follows that,

$$\Phi(\lambda) = \begin{pmatrix} \lambda-1 & 0 & 0 \\ 0 & \lambda-1 & 0 \\ -m_1 & -m_2 & \lambda-m_3 \end{pmatrix} \tag{3.57}$$

and,

$$|\Phi(\lambda)| = (\lambda-1)(\lambda-1)(\lambda-m_3) = 0 \tag{3.58}$$

so that the latent roots are,

$$1, \quad 1 \quad \text{and} \quad m_3$$

In accordance with the notation defined earlier, let,

$$\lambda_1 = 1, \quad \lambda_2 = m_3 \quad \text{and} \quad \lambda_3 = 1 \tag{3.59}$$

so that $s=2$, $r_1=2$ and $r_2=1$ clearly (3.49) holds since $N=3$; for this matrix

$$A(\lambda) = \begin{pmatrix} (\lambda-1)(\lambda-m_3) & 0 & 0 \\ 0 & (\lambda-1)(\lambda-m_3) & 0 \\ m_1(\lambda-1) & m_2(\lambda-1) & (\lambda-1)^2 \end{pmatrix} \tag{3.60}$$

$$A'(\lambda) = \begin{pmatrix} 2\lambda-1-m_3 & 0 & 0 \\ 0 & 2\lambda-1-m_3 & 0 \\ m_1 & m_2 & 2(\lambda-1) \end{pmatrix} \tag{3.61}$$

so that,

$$\left[A(\lambda)\right]_{\lambda=1} = 0, \quad \text{the null matrix} \tag{3.62}$$

$$\left[A(\lambda)\right]_{\lambda=m_3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ m_1(m_3-1) & m_2(m_3-1) & (m_3-1)^2 \end{pmatrix} \tag{3.63}$$

$$\left[A'(\lambda)\right]_{\lambda=1} = \begin{pmatrix} 1-m_3 & 0 & 0 \\ 0 & 1-m_3 & 0 \\ m_1 & m_2 & 0 \end{pmatrix} \tag{3.64}$$

and,

$$\psi_1(\lambda) = (\lambda-m_3) \tag{3.65}$$

$$\psi_2(\lambda) = (\lambda-1)^2 \tag{3.66}$$

Applying the result of Theorem 3.2,

$$M^n = \left[\frac{d}{d\lambda} \frac{\lambda^n A(\lambda)}{\psi_i(\lambda)} \right]_{\lambda=1} + \left[\frac{\lambda^n A(\lambda)}{\psi_i(\lambda)} \right]_{\lambda=m_3} \tag{3.67}$$

$$= \left[\frac{n\lambda^{n-1} A(\lambda)}{\psi_i(\lambda)} \right]_{\lambda=1} + \left[\frac{\lambda^n A'(\lambda)}{\psi_i(\lambda)} \right]_{\lambda=1} + \left[\frac{-\lambda^n A(\lambda)}{\{\psi_i(\lambda)\}^2} \right]_{\lambda=1} + \left[\frac{\lambda^n A(\lambda)}{\psi_i(\lambda)} \right]_{\lambda=m_3}$$

and, using results (3.60) to (3.66) the first and third terms vanish so that,

$$M^n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{m_1}{1-m_3} & \frac{m_2}{1-m_3} & 0 \end{pmatrix} + m_3^n \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{m_1}{1-m_3} & -\frac{m_2}{1-m_3} & 1 \end{pmatrix} \tag{3.68}$$

and this may be written

$$= M_1 + \lambda_2^n M_2 \tag{3.69}$$

to correspond with the notation of (3.9), and it is clear that

since

$$M_1 + M_2 = I \tag{3.70}$$

$$M_1 M_2 = M_2 M_1 = 0 \tag{3.71}$$

$$M_1^2 = M_1 \tag{3.72}$$

$$M_2^2 = M_2 \tag{3.73}$$

and verifying that the relations (3.6) to (3.9) hold for this matrix, this will be of some importance to the discussion of other properties of these resolutions of admixture matrices in the following section.

CHAPTER FOUR

MORE THAN TWO POPULATIONS WITH DISCRETE GENERATIONS AND CONSTANT MIGRATION: DETERMINATION OF ADMIXTURE RATES.

The natural consequence of the previous chapter, which dealt with the estimation of gene frequencies future or past, would seem to be the consideration of the problem of determination of the admixture rates from known frequencies at different points in time. Where there are several such points in time, the problem can be treated as asymptotic regression in view of certain results of Chapters One and Two; this problem is taken up again in the next chapter. However, an interesting problem is met with by using only two points in time but several loci at which frequencies are known. A further problem which occurs in dealing with a matrix of accumulated admixture over several generations, is that of determining the amount of admixture per generation, assuming constancy of the intermixture.

In this section the determination of admixture rates from given exact frequencies at two points in time is dealt with, leaving

to the following sections other considerations, e.g. many points in time, statistically distributed sample estimates of gene frequencies.

It will be convenient to let the two points in time be zero, after no generations, and n, after n generations so that q_0 and q_n may refer to vectors of frequencies, in an obvious manner, at these two points in time. If these frequencies are known at k loci, there will be k vectors q_n and these may be represented by the k columns of a newly defined matrix Q_n; similarly, the k vectors q_0 will be the columns of a matrix Q_0. In this case, where the intermixture is in progress between N populations, Q_n and Q_0 will each have N rows, and k columns whilst M, the admixture matrix will be square of order N.

Referring to the previously established relation (3.4) viz:

$$q_n = M^n q_0 \qquad (4.1)$$

it is clear that since this holds for every locus, the k similar relations may be expressed by the single matrix equation:

$$Q_n = M^n Q_0 \qquad (4.2)$$

in which both Q_n and Q_0 are known, and M^n may be determined by carrying out matrix multiplications so that,

$$M^n = Q_n Q_0' (Q_0 Q_0')^{-1} \qquad (4.3)$$

where matrices meant to be transposed are primed. (Of course, it would be possible to determine M^n from the relation

$$M^n = Q_n Q_n' (Q_0 Q_n')^{-1} \tag{4.4}$$

which is also available in view of the exact knowledge of Q_0 and Q_n. Where this exactness is not present, it will be shown later that (4.3) is a more suitable expression, and for this reason only it is preferred at this stage. The results are identical.)

Having obtained a simple estimate of M^n the problem of determining from it a value for M, the admixture per generation, is now considered. The matrix M^n possesses several properties which are of interest, and these are identically those of M considered in section 3.1: the latent roots, $\lambda^{(n)}$ will, in general, be distinct but always such that $|\lambda_i^{(n)}| \leq 1$. The distinctness of the latent roots allows the spectral decomposition of M^n, in the simple form

$$M^n = \sum \lambda_i^{(n)} M_i^{(n)} \tag{4.5}$$

but the uniqueness of this decomposition, together with the known form for the n^{th} power of M in terms of its roots,

$$M^n = \sum \lambda_i^n M_i \tag{4.6}$$

shows that $\lambda_i^{(n)} = \lambda_i^n$ and $M_i^{(n)} = M_i$ (4.7)

or, $\lambda_i = \sqrt[n]{\lambda_i^{(n)}}$ (4.8)

The matrix M may be determined from M^n by applying spectral resolution to the latter matrix taking n^{th} roots of its latent roots and using:

$$M = \sum \lambda_i M_i \qquad (4.9)$$

since the spectral matrices M_i, are the same for M^n as for M.

Example An example of the procedure described above is now given using the data of Table 3.2 as known exact values of Q_0 and Q_n, where $n = 20$. Ten decimal places were retained throughout the calculation but only four are shown for economy of space.

TABLE 4.1

$$M^n = \begin{matrix} M \\ s \\ d \\ B \end{matrix}\begin{pmatrix} .5750 & .2254 & .1782 & .1270 \\ .5670 & .1505 & .1759 & .1155 \\ .5047 & .1477 & .2523 & .0941 \end{pmatrix} \begin{pmatrix} .5803 & .5786 & .4900 \\ .2559 & .1197 & .1501 \\ .1766 & .1566 & .2709 \\ .1324 & .1158 & .0894 \end{pmatrix}$$

$$\times \left[\begin{pmatrix} .5803 & .2559 & .1766 & .1324 \\ .5786 & .1197 & .1566 & .1158 \\ .4900 & .1501 & .2709 & .0894 \end{pmatrix} \begin{pmatrix} .5803 & .5786 & .4900 \\ .2559 & .1197 & .1501 \\ .1766 & .1566 & .2709 \\ .1324 & .1158 & .0894 \end{pmatrix} \right]^{-1}$$

$$= \begin{pmatrix} .4396 & .4023 & .3752 \\ .4139 & .3870 & .3584 \\ .3877 & .3601 & .3462 \end{pmatrix} \begin{pmatrix} 67.05 & -46.90 & -26.25 \\ -46.90 & 77.80 & -27.98 \\ -26.25 & -27.98 & 60.90 \end{pmatrix}$$

$$= \begin{pmatrix} .7651 & .1810 & .0538 \\ .1963 & .6695 & .1341 \\ .0177 & .1461 & .8363 \end{pmatrix}$$

$$= (1.000)\begin{pmatrix} .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \end{pmatrix} + (.7727)\begin{pmatrix} .4196 & .0589 & -.4785 \\ .1628 & .0229 & -.1857 \\ -.4889 & -.0686 & .5576 \end{pmatrix}$$

$$+ (.4984)\begin{pmatrix} .2780 & -.3876 & .1095 \\ -.4652 & .6485 & -.1833 \\ .1866 & -.2601 & .0735 \end{pmatrix}$$

The 20th roots of the latent roots, i.e., $\sqrt[20]{1} = 1$

$\sqrt[20]{(0.7727)} = .9869$ and $\sqrt[20]{(0.4984)} = .9658$ are most conveniently obtained by using logarithms, so that the expression for M is:

$$M = (1.000)\begin{pmatrix} .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \\ .3024 & .3287 & .3689 \end{pmatrix} + (9869)\begin{pmatrix} .4196 & .0589 & -.4785 \\ .1628 & .0229 & -.1857 \\ -.4889 & -.0686 & .5576 \end{pmatrix}$$

$$+ (.9658)\begin{pmatrix} .2780 & -.3876 & .1095 \\ -.4652 & .6485 & -.1833 \\ .1866 & -.2601 & .0735 \end{pmatrix}$$

and, summing the expressions on the right-hand side,

$$M = \begin{pmatrix} .9850 & .0125 & .0025 \\ .0138 & .9775 & .0087 \\ 0 & .0098 & .9902 \end{pmatrix}$$

4.1 Linear programming approach.

In this section, the notation and assumptions coincide with those of the previous section except that now Q_n and Q_0 are not known exactly but are reckoned to have been obtained from sampling; the result is that the set of equations (4.2), viz:

$$Q_n = M^n Q_0 \qquad (4.10)$$

are not satisfied exactly. The determination of the accumulated admixture matrix M^n, must be such that (4.10) is most closely satisfied according to some criterion. Noticing that the simultaneous relations represented in (4.10) are, in fact, all linear relations, suggests that the linear programming technique might be appropriate. This is confirmed by the fact that the required matrix M^n must contain positive elements and further that the rows must sum to unity by the definition of gene flow adopted in Chapter One. These constraints are linear and, therefore, of the type which are well catered for by linear programming.

Of possible criteria for ascertaining the closeness of degree to which the equations (4.10) are satisfied, one that seems appropriate is the minimisation of the sum of absolute deviations between the left and right-hand sides of the equations. As this is a linear criterion, expressed in the manner to be described, it may form the basis of a linear programming approach. The reasonable nature of this criterion

suggests that this method would provide an interesting solution, if not the most satisfactory one. It will be necessary to discuss this last point further later as well as to consider disadvantages of the method.

Using suffix notation, the equation (4.10) is equivalent to:

$$q_{nij} = \sum_{r=1}^{N} m_{nir}\, q_{orj} \tag{4.11}$$

for $i = 1,2,\ldots,N$ and $j = 1,2,\ldots,k$ and in which q_{orj} and q_{nij} refer to exact values: let q^*_{orj} and q^*_{nij} be defined as the corresponding sampled values. The substitution of the latter values into the equation (4.11) will result in a deviation ϵ_{ij}, so that:

$$q^*_{nij} = \sum_{r=1}^{N} m_{nir}\, q^*_{orj} + \epsilon_{ij}$$

or,

$$\epsilon_{ij} = q^*_{nij} - \sum_{r=1}^{N} m_{nir}\, q^*_{orj} \tag{4.12}$$

These deviations are not all positive and it is necessary to define two non-negative error variables which may be used in combination to represent a negative deviation. To do this, let:

$$\epsilon_{1ij} = \max(\epsilon_{ij}, 0) \tag{4.13}$$

and

$$\epsilon_{2ij} = -\min(\epsilon_{ij}, 0)$$

so that these definitions imply that:

$$\epsilon_{1ij} \geq 0 \qquad (4.14)$$

and $$\epsilon_{2ij} \geq 0$$

and further that:

$$\epsilon_{ij} = \epsilon_{1ij} - \epsilon_{2ij} \qquad (4.15)$$

The sum of absolute deviations, S_{ij}, may now be expressed in terms of the newly defined error variables,

$$S_{ij} = \epsilon_{1ij} + \epsilon_{2ij} \qquad (4.16)$$

With the aid of this quantity, the Linear Programming problem is now formulated taking each population i separately as any row of M^n may be determined independently of any other.

Problem To minimise $\sum_{j=1}^{k} S_{ij}$ with respect to choice of non-negative m_{nir} subject to the conditions:

$$\begin{aligned} q_{nij} - \Sigma\, m_{nir}\, q_{orj} - \epsilon_{1ij} + \epsilon_{2ij} &= 0 \\ \epsilon_{1ij} &\geqslant 0 \\ \epsilon_{2ij} &\geqslant 0 \end{aligned} \qquad (4.17)$$

for $r = 1,2,\ldots,N$ and $j = 1,2,\ldots,k$, together with the additional constraint:

$$\sum_{r=1}^{N} m_{nir} = 1 \qquad (4.18)$$

which is necessary for the conservation of proportionate gene flow, or equivalently for each row of M^n to sum to unity.

Solution (Description of Method) The Standard Simplex Computational Procedure due to G.B. Dantzig (1949) may be used. This method begins with a trial solution, taking the null point in the parameter space, and minimises the criterion sum by moving in a step-by-step manner. Each move is to an adjacent vertex of the Simplex polyhedron whose faces are the hyperplanes defined by the equations in (4.17) and (4.18). This iterative process ceases when a move to an adjacent vertex cannot reduce the criterion sum further. The solution values, being the coordinates of the final vertex reached, satisfy N out of the $(k+1)$ equalities, this vertex being the point of intersection of N faces of the polyhedron.

Example. Consider one evaluation of a row of M for which the relations with frequencies substituted are:

$$\begin{aligned}
.329 &= .026\, m_1 + .551\, m_2 + \epsilon_1 \\
.789 &= .683\, m_1 + .710\, m_2 + .973\, m_3 + \epsilon_2 \\
.477 &= .420\, m_1 + .100\, m_2 + .615\, m_3 + \epsilon_3 \\
.154 &= .389\, m_1 + .253\, m_2 + .038\, m_3 + \epsilon_4 \\
.374 &= .532\, m_1 + .509\, m_2 + .801\, m_3 + \epsilon_5
\end{aligned} \tag{4.19}$$

with solutions after nine iterations,

$$\begin{aligned}
m_1 &= 0.097\ 002 \\
m_2 &= 0.596\ 876 \\
m_3 &= 0.310\ 506
\end{aligned} \tag{4.20}$$

and,

$$\Sigma\, |\epsilon_i| = 0.459\ 384$$

$$\begin{aligned}
\epsilon_{13} &= 0.186\ 086 = \epsilon_3 \\
\epsilon_{24} &= 0.045\ 438 = -\epsilon_4 \\
\epsilon_{25} &= 0.227\ 861 = -\epsilon_5
\end{aligned} \tag{4.21}$$

where the remaining error values are zero. The first two equations are satisfied exactly together with the conservation of proportionate gene flow. This calculation was carried out using the library program Simpfix for Ferranti Pegasus.

The general method of formulating the linear programming problem for this absolute deviations criterion seems to be due to Charnes, Cooper and Ferguson (1955), who first used the device of a difference between two non-negative variables. The method is further

discussed by W.D. Fisher (1961). H. Wagner, (1959, 1962), applied the technique in conventional regression situations, particularly to non-linear models where the knowledge of the functions involved is minimal, using only assumptions of curvature.

An approach which caters for non-linearity was devised by Fourier (c.1820), who suggested the use of "least lines" and he arrived at an iterative procedure not unlike the Simplex Method described above.

The outcome of this approach is, however, limited by the lack of standard errors which are optimally available by the alternative least squares method. Considerations of reliability or significance of the estimates reached, are frequently important and when this is so, the linear programming technique is weak. However, the illustration above shows the flexibility of this technique and how constraints may be conveniently included; this feature is not present, in the same form at least, in the least squares method. The special use made of this feature in the present context is, of course, the determination of admixture rates which are necessarily non-negative and which sum to unity. Of these two requirements, as will be seen later, only the second may be included in least squares theory.

4.2 Weighted regression: all variables subject to error.

Suppose that in the equation the gene frequencies which result from sampling are substituted so that the statistical model may be written:

$$Q_n + \epsilon_n = M^n(Q_0 + \epsilon_0) \qquad (4.22)$$

where ϵ_0 and ϵ_n are matrices whose elements contain the respective errors in the frequencies in the matrices Q_0 and Q_n. The values of these frequencies, to be used in the estimation of M^n, are determined by sampling so that the elements of ϵ_0 and ϵ_n may be considered to be independently and binomially distributed with zero expected values. However, the sampling variances of these errors are dependent upon the values of the frequencies themselves. Using the usual formula for the estimate of the variance of a binomially distributed sample value, these errors may be determined explicitly. It is sometimes a feature of the type of data used, that the gene frequencies vary little at a given locus between populations; this implies that the errors in the columns are uniform in magnitude. In view of differences in the orders of magnitude of the frequencies, and therefore their errors, between loci the sampled values at any locus should be weighted. Doing this, in the standard manner, by dividing each value by the square root

of its variance, allows the model (4.22) to be written:

$$Q_n^* + \epsilon_n^* = M^n(Q_o^* + \epsilon_o^*) \qquad (4.23)$$

where the asterisk refers to the standardisation. The error terms are now independent of their associated frequencies and the above model may be re-arranged:

$$Q_n^* = M^n Q_o^* + (M^n \epsilon_o^* - \epsilon_n^*) \qquad (4.24)$$

The usual Least Squares Method may now be applied as the bracketed error term is independent of the sampled frequencies. In view of the independence and standardisation of the errors, they may be assumed to have constant variances, and the bracketed error term may now be represented by E. The elements of M^n are to be determined by the standard method to minimise the estimates of variance of the errors in E.

Computationally, if W_n is the diagonal matrix of weights for the columns of Q_n and W_o, that for Q_o,

$$\text{if} \quad W = W_o = W_n$$
$$Q_n W = Q_n^*$$
$$Q_o W = Q_o^* \qquad (4.25)$$

so that, $$Q_n W = M^n Q_o W + E$$

and the least squares solution for M^n is given by:

$$M^n = Q_n W (Q_0 W)' \{(Q_0 W)(Q_0 W)'\}^{-1}$$
$$= Q_n WW' Q_0 \{Q_0 WW' Q_0'\}^{-1}$$

or writing $WW' = W^2$ for the square of the diagonal matrix W,

$$M^n = Q_n W^2 Q_0' (Q_0 W^2 Q_0')^{-1} \qquad (4.26)$$

Before proceeding to a discussion of the errors in the least squares estimates, it is required to extend the above method to ensure that they fulfil the biological condition which requires the conservation of gene flow within the closed group of populations under study.

4.3 Use of Lagrange multipliers to ensure conservation of gene flow.

If, as in the previous section, the sampled values of the gene frequencies are represented in the matrix equation and further that they have been weighted as described:

$$Q_n^* = M^n Q_o^* + E \qquad (4.27)$$

The sums of squares to be minimised ~~is~~ are the diagonal elements of

$$EE' = (Q_n^* - M^n Q_o^*)(Q_n^* - M^n Q_o^*)' \qquad (4.28)$$

and the conservation of gene flow requires that,

$$M^n u = u \qquad (4.29)$$

where u is the unit vector, so that the minimisation is conditional upon this equation.

The application of the Least Squares Method to this situation requires the introduction of a vector of Lagrange Multipliers, λ which are to be estimated in addition to M^n.

Vector differentiation, with respect to λ, of

$$EE' - \lambda (M^n u - u)'$$

and equating the derivative to zero gives the vector equation represented above (4.29). Matrix differentiation of this expression provides the matrix equation:

$$-2 (Q_n^* - M^n Q_o^*) Q_o^{*\prime} - \lambda u' = 0$$

Absorbing the factor $^1/_2$ into the multipliers, the simultaneous equations which are satisfied by the least squares estimates of M^n (and λ) may be summarised:

$$(Q_n^* - M^n Q_o^*) Q_o^{*\prime} + \lambda u' = 0 \qquad (4.30)$$

$$M^n u - u = 0 \qquad (4.31)$$

where $Q_o^* Q_o^{*\prime}$ is non-singular, equation (4.30) may be multiplied through by its inverse from the right,

$$Q_n^* Q_o^{*\prime} (Q_o^* Q_o^{*\prime})^{-1} - M^n + \lambda u' (Q_o^* Q_o^{*\prime})^{-1} = 0 \qquad (4.32)$$

and if this operation is followed by a multiplication by u, again from the right,

$$Q_n^* Q_o^{*\prime} (Q_o^* Q_o^{*\prime})^{-1} u - M^n u + \lambda u' (Q_o^* Q_o^{*\prime})^{-1} u = 0$$

so that the substitution of $M^n u$ from equation (4.31) eliminates M^n and providing the scalar $u' (Q_o^* Q_o^{*\prime})^{-1} u$ is now non-zero, the inverse of this product may be multiplied from the right, and re-arranging:

$$\lambda = \frac{u - Q_n^* Q_o^{*\prime} (Q_o^* Q_o^{*\prime})^{-1} u}{u' (Q_o^* Q_o^{*\prime})^{-1} u} \qquad (4.33)$$

the quantity appearing in the denominator being scalar. This may now be substituted into (4.32) to provide a direct computational formula

for the estimate of $M^{\wedge}$,

$$M^{\wedge} = Q_{\wedge}^{*} Q_{0}^{*\prime} (Q_{0}^{*} Q_{0}^{*\prime})^{-1} + \frac{(u - Q_{\wedge}^{*} Q_{0}^{*\prime} (Q_{0}^{*} Q_{0}^{*\prime})^{-1} u) u' (Q_{0}^{*} Q_{0}^{*\prime})^{-1}}{u' (Q_{0}^{*} Q_{0}^{*\prime})^{-1} u} \quad (4.34)$$

From the form of this expression it is clear that if the product which forms the first term has rows which sum to unity, the second term will be zero. In this case the estimate is that used with weighting earlier in section 4.2. Further, if there are slight departures from unity in these row sums, the contribution of the second term to $M^{\wedge}$ will be small.

In discussing the error in the estimate $M^{\wedge}$ it will be practically convenient to refer to the "simple estimate" as that given by (4.34) above, for which the second term is negligible, and to the "full estimate" when this is not the case.

4.4 Errors in the least squares estimation of the gene flow rates.

The vector of diagonal elements of the ~~trace~~ of the matrix EE', $\mathrm{d}(EE')$, contains the sums of squares which were minimised in the estimation of M^n. Using the Gauss-Markov Theorem it is possible to produce a matrix, V, whose elements are the variances of the corresponding elements of M^n. This requires the introduction of $\mathrm{d}((Q_o^* Q_o^{*\prime})^{-1})$ to represent the vector of diagonal elements of ~~trace~~ of the inverse matrix in the outer brackets:

$$V = \mathrm{d}((Q_o^* Q_o^{*\prime})^{-1})\,\mathrm{d}(EE')' / (k - N + 1) \qquad (4.35)$$

The divisor $(k - N + 1)$ reflects that each of the N estimates in any row of M^n was based on k observed sets of frequencies and that the minimisation process used was subject to one extraneous condition.

The matrix V may be obtained directly from the data by substituting for EE' in the above expression. From (4.28),

$$EE' = (Q_n^* - M^n Q_o^*)(Q_n^* - M^n Q_o^*)' \qquad (4.36)$$

but as a result of minimisation,

$$(Q_n^* - M^n Q_o^*)\, Q_o^{*\prime} + \lambda u' = 0$$

so that,

$$\begin{aligned} EE' &= (Q_n^* - M^n Q_o^*)(Q_n^{*\prime} - Q_o^{*\prime} M^{n\prime}) \\ &= (Q_n^* - M^n Q_o^*) Q_n^{*\prime} + \lambda u' M^{n\prime} \end{aligned} \qquad (4.37)$$

Substitution of M^n from (4.34) and then of $\delta(EE')$ into (4.35) provides a computational formula for V. For the "simple estimate" of the previous section, the second term of (4.37) is zero.

The technique of calculation is illustrated in the examples which appear in the following section.

4.4.1 Deduction of error in the estimate of M.

The method described in section 4.3 may be applied directly to obtain M from M^n and it is now necessary to consider the possibility of obtaining the error in the overall determination of M. Suppose that this error is represented by the matrix δM and $(M+\delta M)^n$ may be expanded as follows, ignoring matrix products containing more than one δM factor,

$$(M + \delta M)^n = M^n + \sum_{r=0}^{n-1} M^r(\delta M) M^{n-r-1}$$

or

$$\delta M^{(n)} = \sum_{r=0}^{n-1} M^r(\delta M) M^{n-r-1} \qquad (4.38)$$

where $\delta M^{(n)}$ is the error in M^n. In view of the lack of commutability, in general, of the matrices M and δM, it is difficult to extract an estimate of δM from this expression. However, in the very special and, it must be admitted, unlikely event of the matrices M and δM being commutative, it is possible to solve the matrix equation (4.38) for δM. In this case, by interchanging pairs of

matrices in the product,

$$\begin{aligned} \delta M^{(n)} &= \delta M \sum_{r=0}^{n-1} M^r M^{n-r-1} \\ &= n(\delta M) M^{n-1} \end{aligned} \tag{4.39}$$

however, if the interchanging is carried out differently,

$$\delta M^{(n)} = n M^{n-1} (\delta M) \tag{4.40}$$

showing that $\delta M^{(n)}$ and M are also commutative. It is easy to show that the converse is not true: commutativity of $\delta M^{(n)}$ and M implies a similar property between δM and M^n not between δM and M.

From (4.39) or (4.40) it is clear that:

$$\begin{aligned} \delta M = (\delta M^{(n)})(M^{n-1})^{-1} &= (\delta M^{(n)})(M^{-1})^{n-1} \\ \text{or} \quad &= (M^{-1})^{n-1}(\delta M^{(n)}) \end{aligned} \tag{4.41}$$

which provides a computational formula only under the assumption that δM and M are commutative. It is extremely difficult to provide solutions for more general or more likely assumptions than this. The best that would seem possible would be to find the estimate δM from the known matrix M and $\delta M^{(n)}$ (whose elements are the square roots of those of V), to test the commutativity condition and to adopt or discard the estimate as a result of this test.

4.5 Maximum Likelihood Estimation of admixture rates.

We assume that each q_{0i} has a known sampling variance σ_{0i}^2 and being an estimate from a large sample, is independently normally distributed. The quantity

$$x = q_n - \sum_{i=1}^{N} m_i q_{0i} \qquad (4.42)$$

is, therefore, normally distributed for any constants m_i. If the $E(q_{0i})$ for $i = 1,2,\ldots, N$ refer to the expected initial frequencies in the N populations at a locus and $E(q_n)$ to the expected frequency at that locus after n generations, the expected value of x is zero because of the equation

$$E(q_n) = \sum m_i E(q_{0i}) \qquad (4.43)$$

derived in the last chapter. The variance of x is given by,

$$\text{var}(x) = \sigma_n^2 + \sum m_i^2 \sigma_i^2 \qquad (4.44)$$

and it is clear that if a new variate, x' is defined as

$$x' = x / \sqrt{\{\text{var}(x)\}} \qquad (4.45)$$

x' will be normally distributed with mean zero and variance unity.

It is straightforward to write down the expression for the likelihood of the observed frequencies in place of the expected values

given above. This is, at the j^{th} locus:

$$e^{\ell_j} = \frac{1}{\sqrt{2\pi}} e^{-x'^2/2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{(q_{nj} - \Sigma m_i q_{oij})^2}{2(\sigma^2_{nj} + \Sigma m_i^2 g_{ij}^2)}} \qquad (4.46)$$

and over all loci,

$$e^{L} = e^{\Sigma \ell_i} = \frac{1}{k\sqrt{2\pi}} e^{-\sum_j \frac{(q_{nj} - \Sigma m_i q_{oij})^2}{2(\sigma^2_{nj} + \Sigma m_i^2 g_{ij}^2)}} \qquad (4.47)$$

The principle of Maximum Likelihood may now be applied to obtain estimates of the values m_i from the frequencies observed at k loci. We seek those values of m_i which maximise the logarithm of the likelihood function, by means of differentiation and equating derivatives to zero in the usual manner. First the log likelihood,

$$L = \text{constant} - \sum_j' \frac{(q_{nj} - \Sigma m_i q_{oij})^2}{2(\sigma_{nj}^2 + \Sigma m_i^2 g_{oij}^2)}$$

Differentiating gives N equations for $r = 1, 2, \ldots, N$ like,

$$\frac{\partial L}{\partial m_r} = \Sigma \frac{q_{ori}(q_{nj} - \Sigma m_i q_{oij})}{\sigma_{nj}^2 + \Sigma m_i^2 g_{oij}^2} + \Sigma' \frac{m_r g_{orj}^2 (q_{nj} - \Sigma m_i q_{oij})^2}{(\sigma_{nj}^2 + \Sigma m_i^2 g_{oij}^2)^2} = 0 \qquad (4.48)$$

This set of equations may be expressed more concisely if,

$$w_j^2 = 1 / (\sigma_{nj}^2 + \Sigma m_i^2 g_{ij}^2)$$

and $q^*_{oij} = w_j q_{oij}$ for all i, j

so that (4.48) is equivalent to:

$$\Sigma\, q^{*}_{or_j}\,(q^{*}_{nj} - \Sigma m_i q^{*}_{oij}) + \Sigma m_r \sigma^2_{rj} w^2 (q^{*}_{nij} - \Sigma m_i q^{*}_{oij})^2 = 0 \qquad (4.49)$$

The second term in this equation may, in common cases, be very small. To see this it is enough to realise that the magnitude of the squared bracketed quantity in this term will be small for reasonable data, this being in fact, a squared deviation between the sides of equation (4.43).

This second term is clearly a sum of non-negative quantities and it is possible to derive limits for its value. Let σ^2_{min} and σ^2_{max} be the minimum and maximum values taken by the variances σ^2_{rj}, then it is straightforward to show that:

$$\frac{\sigma^2_{min}}{4\sigma^2_{max}} \leq \sigma^2_{rj} w_j^4 \leq \frac{\sigma^2_{max}}{\sigma^2_{min}} \left(\frac{N}{N+1}\right)^2 \qquad (4.50)$$

Applying these inequalities to the second term of equation (4.49) and writing,

$$q_{nj} - \Sigma m_i q_{oij} = \epsilon_j \qquad (4.51)$$

leads to the inequalities:

$$\frac{1}{4} m_r \frac{\sigma^2_{min}}{\sigma^2_{max}} \Sigma \epsilon_j^2 \leq \Sigma m_r \sigma^2_{rj} w^2 (q^{*}_{nj} - \Sigma m_i q^{*}_{oij})^2 \leq m_r \frac{\sigma^2_{max}}{\sigma^2_{min}} \left(\frac{N}{N+1}\right)^2 \qquad (4.52)$$

However, only in extreme conditions will these weak inequalities be of

use in establishing the order of magnitude of the second term. The best course in practice is to evaluate the term directly.

The estimates of m which satisfy the set of equations (4.49) may be arrived at by a simple iterative method for reasonable data. This method is sufficiently powerful for the data used in the examples that $m_1 = m_2 = m_3 = 0$ was a sufficiently good first approximation to lead to convergence in from three to five iterations.

The first step is to evaluate weights, w_j, using the first approximate values of the admixture rates, m_i. From these the modified frequencies q^*_{nj} and q^*_{oij} (for all i, j) may be calculated and the N linear equations,

$$\Sigma\, q^*_{orj} (q^*_{nj} - \Sigma m_i q^*_{oij}) = 0 \qquad (4.53)$$

solved directly by elimination, or matrix inversion, to arrive at a new set of values for the rates m_i. These new values are used to obtain new weights and the above cycle of operations may be repeated until stability is reached in the successive sets of admixture rates.

Where the smallness of the second term is in question, the values,

$$c_j = \Sigma m_j \sigma^2_{rj} w^2 (q^*_{nj} - \Sigma m_i q^*_{oij})^2$$

may be evaluated at each stage, and then instead of (4.53) the following

equations may be solved:

$$\Sigma\ q^{*}_{or_j}\ ((q^{*}_{n_j} + c_j) - \Sigma m_i q^{*}_{oi_j}) \quad = \quad 0 \qquad (4.54)$$

Standard errors for the maximum likelihood estimates arrived at by using the negative inverse of the second derivatives $\partial^2 L / \partial m_j^2$ or in the case when the values c_j are small, directly from weighted least squares theory. In the latter case, the estimation corresponds with weighted least squares and approximate standard errors may be obtained directly from the result of the Gauss-Markov Theorem. The square of the standard error will be given by the variance of an estimate,

$$\text{var}(m_j) \quad = \quad -1 \ / \ (\partial^2 L / \partial m_j^2)$$

$$\frac{\partial^2 L}{\partial m_j^2} = - \Sigma \frac{q^{2}_{or_j}}{(\sigma^2_{nj} + \Sigma m_i^2 \sigma^2_{oi_j})} - 2 \Sigma \frac{m_r \sigma^2_{or_j} q_{or_j} (q_{nj} - \Sigma m_i q_{oi_j})}{(\sigma^2_{nj} + \Sigma m_i^2 \sigma^2_{oi_j})}$$

$$+ \Sigma \frac{\sigma^2_{or_j} (q_{nj} - \Sigma m_i q_{oi_j})^2}{(\sigma^2_{nj} + \Sigma m_i^2 \sigma^2_{oi_j})} + 2 \Sigma \frac{m_r^2 \sigma^4_{or_j} (q_{nj} - \Sigma m_i q_{oi_j})^2}{(\sigma^2_{nj} + \Sigma m_i^2 \sigma^2_{oi_j})^2}$$

$$= - \Sigma q^{*2}_{r_j} - 2 \Sigma m_r \sigma^2_{or_j} q^{*}_{r_j} \epsilon^{*}_j$$

$$+ 2 \Sigma \sigma^2_{or_j} \epsilon^{*2}_j + 2 \Sigma m_r^2 \sigma^4_{r_j} w^2 \epsilon^{*2}_j$$

(where $\epsilon^{*}_j = q^{*}_{n_j} - \Sigma m_i q^{*}_{i_j}$) (4.55)

Example 1.

Nordestinos 1.

Data q_{01}	q_{02}	q_{03}	q_{11}	
.635(900)	0(238)	.046(3000)	.233(288)	Numbers
.049(900)	.483(238)	.101(3000)	.116(288)	in samples
.804(280)	1(86)	.510(618)	.517(296)	given in
.179(900)	0(238)	.398(3000)	.257(288)	brackets.

Using equation (4.53),

iteration	m_1	m_2	m_3
0	0	0	0
1	0.302 751	0.059 878	0.473 775
2	0.303 648	0.061 276	0.474 282
3	0.303 655	0.061 268	0.474 289
4	0.303 655	0.061 268	0.474 289

Using equation (4.54),

iteration	m_1	m_2	m_3
0	0	0	0
1	0.302 751	0.059 878	0.473 775
2	0.303 648	0.061 276	0.474 283
3	0.303 655	0.061 268	0.474 290
4	0.303 655	0.061 268	0.474 290

The results are identical except for one digit in the sixth decimal place in the estimate of m_3. This shows that the contribution from the second term in equation (4.54) is small and has a negligible effect upon the estimation for the above data. Approximate standard

errors are found, therefore, by using the weighted least squares approach, which requires the quantity $\Sigma \epsilon^{*2}_j$ together with the elements on the diagonal of the matrix $\{Q^*_o Q^{*'}_o\}^{-1}$. These latter elements are,

$d_1 =$ 0.001 510 18, $d_2 =$ 0.001 252 82 and $d_3 =$ 0.004 883 99

and $\Sigma \epsilon^{*2}_j =$ 0.322 981 01 $= 2s^2$, the residual variance so that the standard errors of the estimates are:

$$s.e.(m_1) = \sqrt{d_1 s^2} = \text{~~0.0494~~} \; 0.0156$$

$$s.e.(m_2) = \sqrt{d_2 s^2} = \text{~~0.0450~~} \; 0.0142$$

$$s.e.(m_3) = \sqrt{d_3 s^2} = \text{~~0.0885~~} \; 0.0281$$

Example 2.

Nordestinos 2.

Data

q_{01}	q_{02}	q_{03}	q_n
.114(858)	0(1622)	.066(3048)	.083(288)
.081(858)	0(1622)	.068(3048)	.043(288)
.046(900)	0(236)	.017(582)	.019(288)
0(1000)	.201(36)	0(1000)	.038(68)

Using equation (4.53),

iteration

	m_1	m_2	m_3
0	0	0	0
1	.495 072	.189 053	.131 722
2	.505 490	.189 055	.120 336
3	.505 686	.189 055	.120 152
4	.505 690	.189 055	.120 148
5	.505 690	.189 055	.120 148

Using equation (4.54),

iteration

	m_1	m_2	m_3
0	0	0	0
1	.495 072	.189 055	.131 722
2	.505 499	.189 055	.120 322
3	.505 695	.189 055	.120 139
4	.505 699	.189 055	.120 135
5	.505 699	.189 055	.120 135

The results are identical to four decimal places, showing the smallness of the contribution of the second term in equation (4.54).

As in Example 1, the weighted least squares standard errors are obtained.

$$s.e.(m_1) = \sqrt{\{(0.088\ 471\ 99)\ (0.113\ 219\ 06)\}} = 0.3165\ \ 0.1001$$

$$s.e.(m_2) = \sqrt{\{(0.088\ 471\ 99)\ (0.017\ 252\ 91)\}} = 0.1235\ \ 0.0391$$

$$s.e.(m_3) = \sqrt{\{(0.088\ 471\ 99)\ (0.270\ 560\ 77)\}} = 0.4893\ \ 0.1547$$

Applying the modification using lagrangian multipliers to ensure the conservation of proportionate gene flow, the estimates and standard errors for the above examples become:

Nordestino 1.

m_1 =	0.2967	±	0.0494 0.0156	0.3240
m_2 =	0.0393	±	0.0450 0.0142	0.0834
m_3 =	0.6641	±	0.0888 0.0281	0.5926

Nordestino 2.

m_1 =	0.3504	±	0.3165 0.1001	0.6141
m_2 =	0.2380	±	0.1235 0.0391	0.1891
m_3 =	0.4116	±	0.4893 0.1547	0.1969

The values in the right-hand column show for comparison, the linear programming solutions using the method of section 4.1 applied to the above data but taking no account of the sample sizes.

CHAPTER FIVE

CONTINUOUS MIGRATION: EXTENSION OF THEORY TO ANY NUMBER OF POPULATIONS.

By similar arguments to those employed in Chapter Two, the continuous migration model for N populations may be represented completely by a set of differential equations together with a set of initial conditions. To obtain the set of differential equations, write by analogy with equations (2.1) and 2.2), for $q(t)$,

$$q_1(t+\delta t) = (1 - m_{12}(t)\,\delta t)\, q_1(t) + m_{12}(t)\,\delta t\, q_2(t) \qquad (5.1)$$

$$q_2(t+\delta t) = (m_{21}(t)\,\delta t)\, q_1(t) + (1 - m_{21}(t)\,\delta t)\, q_2(t) \qquad (5.2)$$

where the gene frequencies are as defined previously and $m_{12}(t)$ denotes the component of admixture received by population I from population II at time t (or more strictly, $m_{12}(t)\,\delta t$ is received in the interval $(t,\ t+\delta t)$).

In order to generalise the theory presented in Chapter Two, it will be necessary to abandon the sum Σ and difference Δ technique for setting up and solving the differential equations; these symbols do not readily generalise for $N > 2$. Instead, each equation is taken separately to give, by re-arranging, dividing by δt and taking the limit as $\delta t \to 0$, the differential equations:

$$\frac{dq_1(t)}{dt} = - m_{12}(t)\, q_1(t) + m_{12}(t)\, q_2(t) \qquad (5.3)$$

$$\frac{dq_2(t)}{dt} = m_{21}(t)\, q_1(t) - m_{21}(t)\, q_2(t) \qquad (5.4)$$

The set of initial conditions may be stated as,

$$q_1(0) = q_{10} \qquad (5.5)$$

$$q_2(0) = q_{20} \qquad (5.6)$$

and equations (5.3) to (5.6) now completely describe the model.

Now the model may be generalised so that for N populations the differential equations are:

$$\frac{dq_i(t)}{dt} = \sum_{j=1}^{N} m_{ij}(t)\, q_j(t) \qquad (5.7)$$

for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, N$ and where,

$$m_{ii} = - \sum_{\substack{j=1 \\ j \neq i}}^{N} m_{ij} \qquad (5.8)$$

with the initial conditions,

$$q_i(0) = q_{i0} \qquad (5.9)$$

again for $i = 1, 2, \ldots, N$.

Equivalently, the differential equations may be represented in matrix form as:

$$\frac{dq(t)}{dt} = M(t)\, q(t) \tag{5.10}$$

and the initial conditions as:

$$q(0) = q_0 \tag{5.11}$$

The general solution to the equations (5.10) and only simple functions are taken for the matrix elements of $M(t)$ in what follows. A useful result exists in the form of Sauvage's Theorem for the case in which these functions are constants, and this case is now considered.

5.1 Constant Migration: Derivation of Solution.

Where the admixture rates are constant with time, the general differential equation (5.10) simplifies to:

$$\frac{dq_i(t)}{dt} = M q_i(t) \qquad (5.12)$$

in which M is now a constant matrix. In order to obtain a general solution in this case, a theorem is now given, and then the difficulties of satisfying the initial conditions (5.11) are considered.

THEOREM 5.1 (Sauvage (1895)) If the matrix M has latent roots λ_i for $i = 1, 2, \cdots, N$ and characteristic matrix A then a set of particular solutions to (5.12) is given by:

$$q = Ae \qquad (5.13)$$

where $$e = \{e_i\} = \{e^{\lambda_i t}\} \qquad (5.14)$$

(Definition A is the characteristic matrix of M whose latent roots are exhibited on the diagonal of the diagonal matrix Λ if $MA = A\Lambda$.)

Proof This is short and as reference to it is made later, it is given below:

Let B be a square matrix of order N with columns represented by the vectors b_i, and such that:

$$B = MA$$

$$\text{or } b_i = Ma_i = \lambda a_i$$

where a_i is the i^{th} column of A.

From: $q = Ae$, $\frac{dq}{dt} = A\frac{de}{dt} = A\begin{pmatrix}\lambda_1 e_1 \\ \vdots \\ \lambda_N e_N\end{pmatrix} = Be = MAe$

The particular solutions (5.13) are, however, only solutions to the model under consideration if the initial conditions (5.11) are satisfied. The solutions as they stand imply that:

$$q_0 = Au \tag{5.15}$$

because at time $t=0$ the vector e becomes the unit vector u. This last relation will not in general be true, and so the initial conditions are not satisfied. However, the problem may be solved by using a stronger form of Sauvage's Theorem. This is now given.

THEOREM 5.2 The general solution to the model (5.12) together with initial conditions (5.11) is given by:

$$q = AKe \tag{5.16}$$

where K is a diagonal matrix whose elements, k_i, are such that:

$$q_0 = AKu \tag{5.17}$$

Proof This follows the proof of Sauvage's Theorem, noting that A has the property:

$$MA = A\Lambda$$

for Λ the diagonal matrix of latent roots, then:

$$MAK = A\Lambda K = AK\Lambda$$

(since diagonal matrices commute) and hence AK also has this property; the result that (5.16) represents a

particular solution follows.

To show that (5.17) defines K so that the initial conditions will be fulfilled, it is enough to realise that Ku may be replaced by a vector k of elements k_i, and providing that A is non-singular, k is uniquely defined by:

$$k = A^{-1} q_0 \tag{5.18}$$

Example. The above result is now applied to the two population model discussed in Chapter Two. In that case, see equations (2.1) and (2.2),

$$M = \begin{pmatrix} -m^{II} & m^{II} \\ m^{I} & -m^{I} \end{pmatrix} \tag{5.19}$$

with latent roots:

$$0 \quad \text{and} \quad -(m^{I} + m^{II})$$

and latent vectors, proportional to,

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} m^{II} \\ -m^{I} \end{pmatrix}$$

and the characteristic matrix, C, say after omitting the factors of proportionality for ease of calculation is,

$$C = \begin{pmatrix} 1 & m^{II} \\ 1 & -m^{I} \end{pmatrix} \tag{5.20}$$

and
$$C^{-1} = \frac{1}{-(m^{I}+m^{II})}\begin{pmatrix} -m^{I} & -m^{II} \\ -1 & 1 \end{pmatrix}$$

so that
$$k = \frac{1}{m^{I}+m^{II}}\begin{pmatrix} m^{I} & m^{II} \\ 1 & -1 \end{pmatrix}\begin{pmatrix} q_o^{I} \\ q_o^{II} \end{pmatrix}$$

$$= \frac{1}{m^{I}+m^{II}}\begin{pmatrix} m^{I}q_o^{I} + m^{II}q_o^{II} \\ q_o^{I} - q_o^{II} \end{pmatrix} \qquad (5.21)$$

and the solutions are then,

$$q = \begin{pmatrix} 1 & m^{II} \\ 1 & -m^{I} \end{pmatrix}\frac{1}{m^{I}+m^{II}}\begin{pmatrix} m^{I}q_o+m^{II}q_o & 0 \\ 0 & q_o^{I}-q_o^{II} \end{pmatrix}\begin{pmatrix} 1 \\ e^{-(m^{I}+m^{II})t} \end{pmatrix} \qquad (5.22)$$

$$= \begin{pmatrix} \frac{m^{I}q_o + m^{II}q_o}{m^{I}+m^{II}} + \frac{m^{II}(q_o^{I}-q_o^{II})}{m^{I}+m^{II}} e^{-(m^{I}+m^{II})t} \\ \frac{m^{I}q_o + m^{II}q_o}{m^{I}+m^{II}} - \frac{m^{I}(q_o^{I}-q_o^{II})}{m^{I}+m^{II}} e^{-(m^{I}+m^{II})t} \end{pmatrix} \qquad (5.23)$$

which correspond exactly with the results given in Theorem 2.4.

5.2 Constant migration: statistical estimation.

In this section, the general problems of estimation are outlined in determining the constant elements of M from sampled values of the gene frequencies in N populations at several points in time. The matrix M of the present chapter has, of course, different properties from M in earlier chapters. Here the solution of the last section will be written as:

$$q = Be + cu \tag{5.24}$$

using the fact that the matrix M has zero row totals, or:

$$Mu = o \tag{5.25}$$

which implies that zero is a latent root and hence that u is a latent vector; in (5.24) the vector e is now short of one element, say the first e°, and B is the characteristic matrix with one row, say the first, missing. Any row of q may then be written:

$$q_i = c + \sum_{j=2}^{N} b_{ij} e^{\lambda_j t} \tag{5.26}$$

Example In the case of two populations, this is equivalent to:

$$\begin{aligned} q_1 &= c + b_{12} e^{\lambda_2 t} \\ q_2 &= c + b_{22} e^{\lambda_2 t} \end{aligned} \tag{5.27}$$

and where the q's obtained from sampling at exactly known times, t_i, are called q_i^*, the statistical model may be represented by:

$$\begin{aligned} q_1^* &= c + b_{12} e^{\lambda_2 t} + \epsilon_1 \\ q_2^* &= c + b_{22} e^{\lambda_2 t} + \epsilon_2 \end{aligned} \tag{5.28}$$

Here ϵ_1 and ϵ_2 represent the errors in the quantities q_1^* and q_2^* and assuming the common constant variance and independence of these errors of one another and from one time to another, the method of least squares may be applied to minimise,

$$\Sigma (q_1^* - c - b_{12} e^{\lambda_2 t})^2 + \Sigma (q_2^* - c - b_{22} e^{\lambda_2 t})^2 \qquad (5.29)$$

with respect to the parameters c, b_{12}, b_{22} and λ_2. The resulting normal equations are by no means trivial to solve and they are given here to illustrate the degree of their difficulty. Suppose that the summation is over n points in time. The equations are:

$$\begin{aligned}
2nc &+ b_{12} \Sigma e^{\lambda_2 t} &+ b_{22} \Sigma e^{\lambda_2 t} &= \Sigma (q_1^* + q_2^*) \\
c \Sigma e^{\lambda_2 t} &+ b_{12} \Sigma e^{2\lambda_2 t} & &= \Sigma q_1^* e^{\lambda_2 t} \\
c \Sigma e^{\lambda_2 t} & & + b_{22} \Sigma e^{2\lambda_2 t} &= \Sigma q_2^* e^{\lambda_2 t} \\
c \Sigma t e^{\lambda_2 t} &+ b_{12} \Sigma t e^{2\lambda_2 t} &+ b_{22} \Sigma t e^{2\lambda_2 t} &= \Sigma (q_1^* + q_2^*) t e^{\lambda_2 t}
\end{aligned} \qquad (5.30)$$

and some simplification may be achieved by applying the transformation of parameters,

$$b = b_{12} + b_{22}$$

The equations then become, taking one of the second and third

equations together with,

$$\begin{aligned} 2nc \quad &+ \quad b\Sigma e^{\lambda_2 t} \quad &&= \quad \Sigma(q_1^* + q_2^*) \\ 2c\Sigma e^{\lambda_2 t} &+ \quad b\Sigma e^{2\lambda_2 t} \quad &&= \quad \Sigma(q_1^* + q_2^*)e^{\lambda_2 t} \\ 2c\Sigma t e^{\lambda_2 t} &+ \quad b\Sigma t e^{2\lambda_2 t} \quad &&= \quad \Sigma(q_1^* + q_2^*)te^{\lambda_2 t} \end{aligned} \tag{5.31}$$

and these are recognisable as the same equations which result from the application of the least squares method to the asymptotic regression model,

$$y^* = a + br^x + \epsilon \tag{5.32}$$

for which methods are considered and developed in the following chapters: to show this write:

$$q_1^* + q_2^* = y$$
$$t = x$$

and $$\epsilon_1 + \epsilon_2 = \epsilon$$

with the parameters corresponding in view of $2c = a$ and $e^{\lambda_2} = r$.

By making use of this correspondence to take advantage of the methods developed for the common asymptotic regression model, it is possible to arrive at separate estimates for b_{12} and b_{22} by using separately the second and third equations of the set (5.30).

Extending the above argument to N populations, the statistical model may be represented by:

$$q^* = Be + cu + \epsilon \tag{5.33}$$

where ϵ is now the vector of errors, $\epsilon = \{\epsilon_i\}$, and, for the purposes of estimating the parameters λ_i and c, the model may first be placed by:

$$u'q^* = u'Be + Nc + u'\epsilon \tag{5.34}$$

which corresponds to the asymptotic model:

$$y^* = a + \sum_{i=1}^{N} b_i r_i^{x} + \epsilon \tag{5.35}$$

and this model is discussed in the following chapters. The separate estimates of the elements of the columns of B may be made afterwards, by using the results of the asymptotic regression substituted into the appropriate least squares equations.

Before considering asymptotic regression problems in detail, a brief discussion of non-constant continuous migrations follows, and the degree of difficulty will be seen to be great in general.

5.3 Non-constant migration.

The general problem of variable continuous migration is now investigated by first assuming that the admixture rates are linear functions of time. Only in special cases are results readily obtainable yet it is possible with sufficient assumptions to consider polynomial functions. The results to be obtained are not entirely satisfactory, in that the admixture matrices to be considered are not close to reality. Bearing this in mind, however, some effort is made to point out a class of solvable problems in this area.

When the admixture matrix $M(t)$, contains elements which are linear functions of time, the matrix can then be resolved into two components, M_0 and M_1, each containing constant elements, such that,

$$M(t) = M_0 + M_1 t \tag{5.36}$$

where the time variable, t, multiplies M_1 in a scalar manner. Substituting this resolution of $M(t)$ into equation (5.10) gives:

$$\frac{dq(t)}{dt} = (M_0 + M_1 t)\, q(t) \tag{5.37}$$

and a simple particular solution of this differential equation is given, in special circumstances, by the result of the following theorem.

THEOREM 5.3 When the matrices M_0 and M_1 share a common characteristic matrix, C, a particular solution to equation (5.37) is given by:

$$q = Ce \tag{5.38}$$

where e is the vector defined by:

$$e = \{ e^{\lambda_i(t)} \} \tag{5.39}$$

in which, $\lambda_i(t) = \lambda_{1i} t + \frac{1}{2}\lambda_{2i} t^2$ (5.40)

and $\{\lambda_{1i}\}$ and $\{\lambda_{2i}\}$ are constant vectors, containing the latent roots of M_0 and M_1 respectively.

Proof Let Λ_1 and Λ_2 be the diagonal matrices with vectors of ~~traces~~ diagonal elements λ_1 and λ_2 respectively. Differentiating (5.38) gives:

$$\frac{dq}{dt} = C(\Lambda_1 + \Lambda_2 t)\, e \tag{5.41}$$

and it is required to show that,

$$C(\Lambda_1 + \Lambda_2 t)e = (M_0 + M_1 t)\, e \tag{5.42}$$

The result follows by comparing coefficients of t on both sides of this last equation so that,

$$C\Lambda_1 = M_0 C$$

and $$C\Lambda_2 = M_1 C$$

The condition that M_0 and M_1 must share a common characteristic matrix seems rather severe in practice. However, there would seem to be no simpler condition for which the differential equation has a particular solution. Perhaps the most likely matrices which might occur as having a common characteristic matrix, are those which are merely scalar multiples of each other. The necessary condition for the latter, in real terms, is that the rates of change of the admixture rates are constant and proportional to the initial admixture rates. Given that this is remotely possible, it is a trivial matter to establish a proper solution to the model so that the initial conditions are satisfied by the same arguments of the previous section.

Where the elements of $M(t)$ are not linear but polynomial functions of time of degree r, the above approach may be extended so that:

$$M(t) = M_0 + M_1 t + \ldots + M_r t^r \qquad (5.43)$$

and the general result is stated in the theorem which follows.

THEOREM 5.4 When $M(t)$ is a polynomial matrix function of time and its component constant matrices M_i for $i = 0, 1, \ldots, r$ share a common characteristic matrix, C, the differential equation:

$$\frac{dq(t)}{dt} = \sum_{i=0}^{r} M_i t^i \, q(t)$$

has a particular solution given by:

$$q = Ce$$

where $$e = \{ e^{\lambda_i(t)} \}$$

and $$\lambda_i(t) = \sum_{j=1}^{r} \frac{1}{j} \lambda_{ij} t^j$$

where $\{\lambda_{ij}\}$ is the matrix whose columns contain the latent roots of $M_0, M_1, \ldots, M_r$.

Proof This follows the argument of the proof of Theorem 5.3.

Although it is not a particularly useful observation in practice, it should be noted that the conditions given in the Theorems 5.3 and 5.4 are both necessary and sufficient for the existence of a particular solution of the form described.

PART II

ASYMPTOTIC REGRESSION METHODS

INTRODUCTION

The estimation of admixture rates in continuous migration models has at its centre a general problem of asymptotic regression. Methodology to cater for asymptotic models in other fields has been the subject of considerable discussion in the literature. From the time of the earliest work of Gompertz (1825) and Mitscherlich (1909, 1930), who derived "laws" or models to describe human mortality and fertiliser response respectively, there has been a need to obtain estimates of non-linear parameters from experimental data. The chapters which follow set out to discuss the more important methods of estimation for single non-linear parameter models and to extend the most reliable general method to cater for many non-linear parameter models; the latter models relate to intermixture between several populations.

In these chapters, the models to be discussed from the theory of intermixture belong to the class of generalised asymptotic regression

in the classification of Turner, Monroe and Lucas (1961) and Turner (1959). The present treatment seeks to be somewhat more detailed and searching then that of these authors but their work offers a perspective to the special regressions considered here.

CHAPTER SIX

GENERAL METHODS FOR A SINGLE NON-LINEAR PARAMETER

First, a single non-linear parameter relation is defined, and then the classical method is given, (called "classical" in view of the use it makes of the established Gauss-Seidel iterative process: the Newton-Raphson process used to obtain least squares estimates of non-linear parameters, see Whittaker and Robinson (1944).). In this section it will become clear that this is not, in general, the best method for the types of special model considered and the reasons for this are given.

It will be convenient to let the equation:

$$\eta = \alpha + \beta f(\rho, x) \tag{6.1}$$

represent the known relationship between the expectation, η , of the dependent variable, y , and the independent variable, x; relations of this form will be called single non-linear parameter (SNLP). In any particular case $f(\rho, x)$ will represent a known function (but for generality it remains unspecified in what follows) and it is subject

only to the condition that its second derivative with respect to ρ is non-zero in order to avoid triviality. In the equation (6.1) α, β and ρ are constants.

The estimation of the constants in the equation using observed pairs of values (x,y) by the principle of least squares, is not a simple matter becuase of the implied non-linearity of f and also as a consequence, of the normal equations. The three normal equations, with ρ known, however, reduce to two simultaneous linear equations in α and β which allow ready solution; in view of this fact, it is natural to hope to establish an iterative method based only on ρ, and this has been successfully achieved by Stevens (1951) as described in the section which follows. Ignoring this feature of the normal equations, however, the classical approach may be adopted. This is now described.

6.1 The classical iterative method.

Given approximate values α_0, β_0 and ρ_0 for the constants in equation (6.1) a Taylor expansion of the function

$$\eta = \eta(\alpha, \beta, \rho ; x)$$

leads to the relation, ignoring second order terms,

$$\eta - \eta(\alpha_0, \beta_0, \rho_0 ; x) = (\alpha-\alpha_0) \frac{\partial \eta}{\partial \alpha}\Big|_{\substack{\alpha=\alpha_0 \\ \text{etc.}}} + (\beta-\beta_0) \frac{\partial \eta}{\partial \beta}\Big|_{\substack{\alpha=\alpha_0 \\ \text{etc}}} + (\rho-\rho_0) \frac{\partial \eta}{\partial \rho}\Big|_{\substack{\alpha=\alpha_0 \\ \text{etc}}}$$

$$= \delta\alpha + \delta\beta f(\rho_0, x) + \beta \delta\rho \frac{\partial f}{\partial \rho}\Big|_{\rho=\rho_0} \qquad (6.2)$$

in which the new "constants" $\delta\alpha$, $\delta\beta$, $\delta\rho$ may be estimated by least squares in view of their linear occurrence in this relation; from them improved values,

$$\alpha_1 = \alpha_0 + \delta\alpha$$

$$\beta_1 = \beta_0 + \delta\beta$$

$$\rho_1 = \rho_0 + \delta\rho$$

may be obtained and the process can be repeated until the corrections become small enough to be negligible. It is clear that these values cannot be expected to converge to zero unless the second order terms which were ignored in forming relation (6.2) really are small; this requires:

$$(\alpha - \alpha_0)^2 \frac{\partial^2 \eta}{\partial \alpha^2} \doteq 0 \qquad (6.3)$$

$$(\beta - \beta_0)^2 \frac{\partial^2 \eta}{\partial \beta^2} \doteq 0 \qquad (6.4)$$

$$(\rho - \rho_0)^2 \frac{\partial^2 \eta}{\partial \rho^2} = (\rho - \rho_0)^2 \beta \frac{\partial^2 f}{\partial \rho^2} \doteq 0 \qquad (6.5)$$

$$(\alpha - \alpha_0)(\beta - \beta_0) \frac{\partial^2 \eta}{\partial \alpha \partial \beta} \doteq 0 \qquad (6.6)$$

$$(\beta - \beta_0)(\rho - \rho_0) \frac{\partial^2 \eta}{\partial \beta \partial \rho} \doteq 0 \qquad (6.7)$$

$$(\alpha - \alpha_0)(\rho - \rho_0) \frac{\partial^2 \eta}{\partial \alpha \partial \rho} \doteq 0 \qquad (6.8)$$

of which equations (6.3), (6.4), (6.6) and (6.8) are clearly satisfied for any SNLP relation, the left-hand sides being zero in each case but equations (6.5) and (6.7) remain to be satisfied. Of these equations

the latter is dependent upon the smallness of $\delta\beta$ (as well as of $\delta\rho$), and it becomes clear that if the approximation β_0 is not good the relation (6.2) is not valid and it is unreasonable to expect the convergence referred to above. The presence of two conditions to be satisfied for any SNLP relation will be seen to be unnecessary in the following section.

THEOREM 6.1 The classical iterative method applied to SNLP relations requires two conditions concerning the first approximations β_0 and ρ_0 to be satisfied.

Proof It is only necessary to consider whether it is possible for the two equations (6.5) and (6.7) to become one for certain special functions $f(\rho, x)$; this is true for $f(\rho, x)$ which satisfy the second order linear differential equation:

$$(\rho - \rho_0)^2 \beta \frac{\partial^2 f}{\partial \rho^2} = (\beta - \beta_0)(\rho - \rho_0) \frac{\partial f}{\partial \rho} \qquad (6.9)$$

To obtain the general solution to this equation write:

$$(\rho - \rho_0) = u, \qquad z = \frac{\partial f}{\partial \rho} \qquad (6.10)$$

so that (6.9) is equivalent to:

$$u^2\beta \frac{\partial z}{\partial u} = (\beta-\beta_0)uz$$

or

$$\frac{\partial z}{z} = \frac{\beta-\beta_0}{\beta}\frac{\partial u}{u}$$

so that integrating gives:

$$\log z = \frac{\beta-\beta_0}{\beta} \log u + \text{constant}$$

or

$$z = cu^k$$

where c is an arbitrary constant and $k = (\beta-\beta_0)/\beta$

or again,

$$\frac{\partial f}{\partial u} = cu^k$$

and integrating once more:

$$f = \frac{c}{k}u^{k-1} + d$$

where d is an arbitrary constant. Substituting back the general solution of (6.9) is seen to be:

$$f = c'(\rho-\rho_0)^{-\beta_0/\beta} + d \qquad (6.11)$$

in which c' and d are arbitrary constants, with respect to ρ, but at least one of which is a function of x in order that f shall be a function of x. But this function f is inadmissible as a special case of the function in an SNLP relation because it is a function of the approximations β_0 and ρ_0. The required result follows.

6.2 Stevens' Method.

The fitting of the relationship (6.1) was considered by Stevens (1951) for the special case in which,

$$f(\rho, x) = \rho^x \tag{6.12}$$

and he arrived at an iterative process dependent only upon a first approximation ρ_0 to the constant ρ ; he observed that a single parameter iterative method could be applied to any general function $f(\rho, x)$ but he did not indicate the full generality of this result and his work is extended in Chapter Nine.

To show how Stevens arrived at this method, the observation made in the last section is now taken up again. This is that two of the three normal equations of least squares are linear in two of the constants, α and β , and provide a pair of simultaneous linear equations if the third constant ρ is known. Stevens realised that ρ need not be known but merely needed to be the subject of an iterative process, the solution of the pair of equations at each stage producing automatically estimates of the first two constants.

More formally the above method is justified by showing that the $\delta\alpha$ and $\delta\beta$ may be eliminated from the three normal equations which result from the application of the principle of least squares to the

relation (6.2). The result of the elimination, as Stevens showed, is a set of three linear equations in α, β and $\beta\delta\rho$ instead of in $\delta\alpha$, $\delta\beta$ and $\delta\rho$. Iterating in only one constant, ρ, provides a second order iterative process with the same power as the Newton-Raphson process for solving equations containing functions of a single variable.

Stevens provided tables of the inverse matrix for solving the three equations on the basis of equally spaced x-values $x = 0,1,2,\ldots,(n-1)$ and for various first approximations to ; the ranges of his tabulation ~~as~~ are $n = 5(1)7$ and $\rho_0 = 0.25(0.01)0.70$, though he suggests that these ranges might be extended to advantage. Several authors have found it worthwhile to extend these tables at Stevens' suggestion but only one of these has been published - that carried out as part of the present research. The full details of the extension which covers the ranges $n = 5(1)30$, $\rho = 0.10(.01)0.90$, published as Hiorns (1964) are given in Chapter Eight.

6.3 Schneider's Method.

Approaching the normal equations of least squares for the fitting of the relation:

$$E(y) = \eta = \alpha + \beta\rho^{x} \qquad (6.13)$$

from the same point of view as Stevens (1951), Schneider (1963) realised that if the estimation of ρ can be carried out separately, the determination of $\hat{a}$ and $\hat{b}$, the least squares estimates of α and β, is a trivial matter requiring only the solution of a pair of simultaneous linear equations.

The least squares estimates $\hat{a}$, $\hat{b}$ and $\hat{r}$ are those values of the estimates a, b and r which satisfy the equations:

$$na + b\Sigma r^{x} = \Sigma y \qquad (6.14)$$

$$a\Sigma r^{x} + b\Sigma r^{2x} = \Sigma y r^{x} \qquad (6.15)$$

$$a\Sigma x r^{x-1} + b\Sigma x r^{2x-1} = \Sigma y x r^{x-1} \qquad (6.16)$$

Between these equations, the elimination of a is straightforward, leading to the pair of equations in b and r:

$$b\left(\Sigma r^{2x} - (\Sigma r^{x})^{2}/n\right) = \Sigma y r^{x} - \Sigma y \Sigma r^{x}/n \qquad (6.17)$$

$$b\left(\Sigma x r^{2x-1} - \Sigma r^{x}\Sigma x r^{x-1}/n\right) = \Sigma y x r^{x-1} - \Sigma y \Sigma x r^{x-1}/n \qquad (6.18)$$

and the elimination of b between these two equations leaves a single equation in r ; this is

$$\frac{\Sigma y r^{x} - \Sigma y \Sigma r^{x}/n}{\Sigma r^{2x} - (\Sigma r^{x})^{2}/n} - \frac{\Sigma y x r^{x-1} - \Sigma y \Sigma x r^{x-1}/n}{\Sigma x r^{2x-1} - \Sigma r^{x} \Sigma x r^{x-1}/n} = 0 \qquad (6.19)$$

making use of the assumption that the denominators in this equation do not vanish (this will only be so for special and trivial values of r and x). This single equation may now be rewritten:

$$(\Sigma y r^{x} - \Sigma y \Sigma r^{x}/n)\, A(r) - (\Sigma y x r^{x-1} - \Sigma y \Sigma x r^{x-1}/n)\, B(r) = 0$$

where

$$A(r) = \Sigma x r^{2x-1} - \Sigma r^{x} \Sigma x r^{x-1}/n$$

$$B(r) = \Sigma r^{2x} - (\Sigma r^{x})^{2}/n$$

and re-arranging,

$$\Sigma y r^{x-1} \{ r A(r) - x B(r) \} - \{ (\Sigma r^{x}/n) A(r) + (\Sigma x r^{x-1}/n) B(r) \} \Sigma y = 0$$

or

$$F(r) = \Sigma y r^{x-1} (G(r) + x) - H(r) \Sigma y = 0 \qquad (6.20)$$

in which,

$$G(r) = -r A(r) / B(r) \qquad (6.21)$$

$$H(r) = -(\Sigma r^{x}/n) A(r) / B(r) - \Sigma x r^{x-1} \qquad (6.22)$$

Schneider (1963) solves the single equation (6.20) iteratively and to facilitate this in general, he provides tables of the functions $G(r)$ and $H(r)$ for $n = 4(1)7$ and $r = 0.01(0.01)0.99$ with the

independent variable x taking the values $x = 0,1,2,\ldots(n-1)$.

In his paper, Schneider (1963) proposes the evaluation of the function $F(r)$ for three values of r and obtains the minimum of this function by parabolic interpolation. Here interpolation with $F(r)$ is simpler than the interpolation possible with the sum of squares itself proposed by Will (1936); Will's suggestion requires the evaluation of the sum of squares for trial values of r using linear regression on a and b for each. Where a computer is available, however, it is probably more satisfactory to solve equations of this type by the Newton-Raphson iterative process.

Let r_o be an approximate solution to equation (6.20) then,

$$\delta r = - \frac{F(r_o)}{F'(r_o)} \qquad (6.23)$$

where $$F'(r_o) = \left.\frac{dF(r)}{dr}\right|_{r=r_o}$$

provides a correction to r_o so that,

$$r_1 = r_o + \delta r$$

is a better approximation to a solution of the equation. This process is repeated until successive values of r differ only by a negligible amount.

It follows from the above derivation that the solution to

equation (6.20) may be substituted into any pair of the equations (6.14) - (6.16) which may then be solved to provide solutions $\hat{a}$ and $\hat{b}$; the latter together with $\hat{r}$ complete the set of least squares estimates.

6.4 Hartley - Gauss - Newton.

Hartley (1961) proposed a modification of the classical approach described in section 6.1. An application of this modification to the relation (6.1) is now given. The essence of the new method is that the corrections $\delta\alpha$, $\delta\beta$ and $\delta\rho$ are not added directly to the current best approximate values to provide improved solutions. Instead the following values are formed:

$$\begin{aligned} \alpha_1 &= \alpha_0 + \lambda\,\delta\alpha \\ \beta_1 &= \beta_0 + \lambda\,\delta\beta \\ \rho_1 &= \rho_0 + \lambda\,\delta\rho \end{aligned} \tag{6.24}$$

where λ is chosen so that $0 \le \lambda \le 1$ and for which,

$$\Sigma\,(y - \alpha_1 - \beta_1 f(\rho_1, x))^2 \tag{6.25}$$

is a minimum.

The minimisation to find λ, if carried out analytically, leads to a difficult algebraic equation in λ and a more practical method is given in the example in Hartley's paper. This alternative consists of setting $\lambda = 0$, $\frac{1}{2}$ and 1 and carrying out a parabolic interpolation using the three values of the quantity (6.25) which result from the substitution of λ.

It is at once clear that this method is at least as powerful as the classical method for $\lambda = 1$ above provides the values of the

latter. The relatively small amount of work involved in carrying out this modification suggests that it is a worthwhile one in that it is likely to improve the speed of convergence. Hartley's method is of the same type as that proposed by Booth (1957) requiring only a little extra computation at each iteration. Other general methods for function minimisation are proposed by Levenberg (1944) and, using conjugate gradients Hestens and Stiefel (1952), Davidon (1959) Powell (1962) and Fletcher and Powell (1963), and these prove powerful for difficult and pathological non-linear functions but require more computation than is usually necessary for the relations under present discussion. Consideration of the convergence properties is not made here, however, but the more fundamental question of whether Hartley's method is applicable is taken up in the following section.

An important observation not made by Hartley, is that if this modification were applied not to the classical method but to the Stevens' method, then the effect for SNLP relations should prove more effective.

6.5 Discussion.

In order to make some comparison between the methods described in the earlier sections, it will be assumed that the independent variable, x, takes n equally spaced values, scaled so that:

$$x = 0, 1, 2, \ldots, (n-1) \tag{6.26}$$

Furthermore, the special SNLP relation treated earlier is reconsidered as it is an important practical one and it does enable detailed conclusions to be reached; this relation was:

$$E(y) = \eta = \alpha + \beta\rho^{x} \tag{6.27}$$

Conditions (6.3) - (6.8) were obtained in section 6.1 which must hold for the application of the classical method. Of these conditions only two remain for SNLP relations. These concern first approximations β_0 and ρ_0 to the estimates of β and ρ. In terms of estimates replace β_0 and ρ_0 by b_0 and r_0 so that the conditions become:

$$(r-r_0)^2 b \frac{\partial^2}{\partial r^2}(r^x) \doteq 0 \tag{6.28}$$

and

$$(b-b_0)(r-r_0)\frac{\partial^2}{\partial b \partial r}(br^x) \doteq 0 \tag{6.29}$$

A useful and reasonable measure of the size of these two quantities would seem to be the magnitude of the perturbation which they introduce into the normal equations if they are kept in the

expansion (6.2). Initially, we consider their combined effect upon the first normal equation. The perturbation introduced into this equation is:

$$\pi = \tfrac{1}{2} \Sigma (\delta r)^2 b x(x-1) r^{x-2} + \Sigma (\delta b)(\delta r) x r^{x-1} \tag{6.30}$$

and for any given positive value δ, a condition is now derived for which,

$$|\pi| < \delta \tag{6.31}$$

Two known series sums will be of use in establishing this condition, These are:

$$\sum_{0}^{n-1} x r^{x-1} = -\frac{n r^{n-1}}{1-r} + \frac{1-r^n}{(1-r)^2} \tag{6.32}$$

$$\Sigma x(x-1) r^{x-2} = -\frac{n(n-1) r^{n-2}}{1-r} - \frac{2n r^{n-1}}{(1-r)^2} + \frac{1-r^n}{(1-r)^3} \tag{6.33}$$

More particularly it is of some interest to consider the case where $0 < r < 1$ and the number of points n, is large. Then the following results will be of use:

$$\Sigma x r^{x-1} \doteq \frac{1}{(1-r)^2} \tag{6.34}$$

$$\Sigma x(x-1) r^{x-2} \doteq \frac{1}{(1-r)^3} \tag{6.35}$$

From (6.30),

$$\pi = [\, b(\delta r)^2 f_1(r) + (\delta b)(\delta r) f_2(r) \,]$$

where,

$$f_1(r) = \tfrac{1}{2}\{-n(n-1)r^{n-2}(1-r)^2 - 2nr^{n-1}(1-r) + (1-r^n)\}/(1-r)^3$$

and,

$$f_2(r) = \{-nr^{n-1}(1-r) + (1-r^n)\}/(1-r)^2 \qquad (6.36)$$

so that (6.31) holds if,

$$|\delta r| \,.\, |\, b(\delta r) f_1(r) + \delta b\, f_2(r)| < \delta \qquad (6.37)$$

A weaker condition than (6.37) is given by the following set of inequalities:

$$|b|(\delta r)^2 < \delta_1 / f_1(r) \qquad (6.38)$$

$$|\delta r| \,.\, |\delta b| < \delta_2 / f_2(r) \qquad (6.39)$$

and

$$\delta_1 + \delta_2 < \delta$$

For large samples from (6.34) and (6.35),

$$\pi = \tfrac{1}{2}\frac{b(\delta r)^2}{(1-r)^2} + \frac{(\delta b)(\delta r)}{(1-r)^3}$$

$$= \{\tfrac{1}{2} b(\delta r)^2 + (\delta b)(\delta r)(1-r)\}/(1-r)^3$$

For Stevens' Method the corresponding perturbation is obtained from this by putting $\delta b = 0$, and the comparison of these perturbations may, therefore, be made by using the ratio:

$$\frac{\tfrac{1}{2} b(\delta r)^2 + (\delta b)(\delta r)(1-r)}{\tfrac{1}{2} b(\delta r)^2} = 1 + \frac{\delta b}{2b\,\delta r}(1-r)$$

and it follows that Stevens' Method produces a lower absolute perturbation if:

$$\left| 1 + \frac{\delta b}{2b\,\delta r}(1-r) \right| > 1 \qquad (6.40)$$

i.e. either if:

$$1 + \frac{\delta b}{2b\,\delta r}(1-r) > 1$$

which implies

$$\frac{\delta b}{2b\,\delta r} > 0 \qquad (6.41)$$

(for growth curves and certain other curves $b < 0$ so that this merely requires δb and δr to be of opposite sign).

or if,

$$-1 - \frac{\delta b}{2b\,\delta r}(1-r) > 1$$

which implies

$$- \frac{\delta b}{2b\,\delta r}(1-r) > 2$$

i.e.

$$- \frac{\delta b}{4b} > \frac{\delta r}{1-r} \qquad (6.42)$$

In general b is larger in magnitude than r, and this is usually true of the errors of approximation δb and δr. This fact makes inequality (6.42) a possible condition in practice. Where the second term in the modulus in (6.40) is small (< 1), however, the inequality is as likely to hold as not because of the general uncertainty of the signs of δb and δr. Further discussion of this point is difficult but it would seem that there are more cases in practice when the perturbation is smaller by Stevens' Method than otherwise. In view of the difficulty met with here, the argument applied above is not repeated for the remaining Normal Equations.

CHAPTER SEVEN

SPECIAL METHODS.

7.1 Small sample techniques and grouping.

In practical uses made of the SNLP model (6.13), small samples are often encountered and apparently four points commonly describe the curve adequately. When this is the case, considerable simplification of the computation required follows. This is specially true when the four points represent equally spaced values of the independent variable, for it is then possible to carry out some algebraic analysis preparatory to the computation. So great is the simplification that there is some temptation to reduce quite large sets of data to smaller sets by grouping. The merits of grouping exist entirely in the fact that it reduces Computational Labour: a convincing argument of Stevens (1951) shows that the loss of information is large and that the estimates are of low efficiency.

7.2 Gomes' Method.

In a paper following upon work with W.L. Stevens and E. Malavolta, P. Gomes (1949, 1953) makes use of the elimination of the parameters α and β from the least squares equations for the SNLP relation (6.13) to derive a polynomial equation for an estimate of the non-linear parameter ρ. This polynomial equation takes the form:

$$\sum_{i=1}^{n} y_i J_i(r) = 0 \qquad (7.1)$$

and the polynomial functions $J_i(r)$ are tabulated for $n = 4$ and 5 so that interpolation for a root is possible. The method is limited by the larger number of tabulated polynomials which would be needed for larger values of n. Pathological examples given by Gomes, which show the advantage over Stevens' Method as originally presented with limited tabulation, may be discounted now that a wider use is made of electronic computers and because of the extended range of tabulation of Stevens' matrix.

There is some affinity between Gomes' Method and that newly proposed by Schneider, as described in Chapter Six. The important difference, however, is the improved formulation of the polynomial equation (7.1) which allows a more satisfactory solution without the limitation to small samples.

7.3 Patterson's Method.

Patterson (1956) proposed a simple ratio estimator for the non-linear parameter ρ in the relation (6.13). The simplicity of form of this estimator makes it an attractive one in practice but it would be unreasonable to expect too much in the way of optimal properties. Investigations into simple ratio estimators are numerous and the relevant works are White (1956) who independently obtained some of Patterson's formulae, Patterson (1958, 1960), Finney (1958) and Patterson and Lipton (1959); these authors also discuss quadratic ratio estimators which are considered in the following section.

Suppose that an estimator r for ρ is required where r is a ratio of two linear functions of the observed y-values. For equally spaced integer values of x let $x = 0,1,2,\ldots,(n-1)$. Such a ratio estimator may be written

$$r = \frac{\sum_{i=0}^{n-1} b_i y_i}{\sum_{i=0}^{n-1} q_i y_i} \qquad (7.2)$$

and a simple condition for r to be a consistent estimator of ρ is that

$$\frac{E\{\sum b_i y_i\}}{E\{\sum q_i y_i\}} = \rho \qquad (7.3)$$

By substituting for the y, rearranging and comparing powers of ρ it

quickly follows that (7.3) is satisfied for all values of α, β and ρ only if

$$\Sigma p_i = \Sigma q_i = 0$$
$$p_0 = q_{n-1} = 0$$

and
$$p_i = q_{i-1} \qquad (i = 1, 2, \ldots, n-1) \tag{7.4}$$

The freedom of choice for values of coefficients in the linear functions is still considerable and it would seem sensible to seek to require the minimisation of the variance of this estimator. However, although the result of this minimisation produces the maximum likelihood estimator, r, as Patterson proved, the resulting values of the coefficients are themselves functions of ρ and therefore not very useful in practice. Instead, values of the coefficients are chosen which give small variance and consequent high efficiency over practical ranges of ρ.

A disturbing feature of the ratio estimators is their inherent bias. This was investigated for four points by Finney (1958) while Patterson (1958) gave expressions for the bias when using from four to seven points. From the relations given above it follows that for four points the ratio estimator r may be written

$$r = \frac{y_4 + (\lambda-1)y_3 - \lambda y_2}{y_3 + (\lambda-1)y_2 - \lambda y_1} \tag{7.5}$$

where λ is a constant, and from this Finney obtained the bias as

$$\text{bias}(r) = \frac{\sigma^2}{\beta^2\rho^2(1-\rho^2)} \frac{(\lambda-1)^2 + 2\rho(\lambda^2-\lambda+1)}{(\lambda+\rho)^2} \qquad (7.6)$$

where each y is independent with variance σ^2. The numerator vanishes only for a pair of imaginary values of λ and the bias would appear to be very large for ρ near to zero or unity. Nevertheless the size of the bias in linear estimators may be better for larger n than in more complicated quadratic estimators: this point was adequately proved by Patterson (1958) with reference to the Taylor estimator to be described in the next section.

In the present context it would seem reasonable to consider the possibility of obtaining ratio estimators for the parameters in many non-linear parameter (MNLP) models e.g.

$$E(y) = \alpha + \sum_{j=1}^{p} \beta_j \rho_j^{x} \qquad (7.7)$$

The unsatisfactory nature of some of the properties of the SNLP estimators as described above and in the next section, implies that for more complicated models the situation may well be much worse. This is indeed the case, but it is probably worthwhile to indicate, in passing, the extent of the difficulties involved. To do this, let $p = 2$ in the model (7.7) and define the linear ratio estimator given by (7.2) as r_1, the estimator of ρ_1. For consistency of this estimator, the

simplest condition would seem to be of the form of (7.3) and upon substitution this becomes

$$\frac{F(p)}{F(q)} = \rho_1 \qquad (7.8)$$

where $F(p) = (\Sigma p_i)\alpha + \beta_1 \Sigma p_i \rho_1^i + \beta_2 \Sigma p_i \rho_2^i$

and p and q represent the vectors of coefficients $\{p_i\}$ and $\{q_i\}$ respectively. An inspection of the coefficients of the powers of ρ_1, and ρ_2 in the rearranged form of (7.8) is enough to show that it is not possible to obtain values for the coefficients which are independent of α, β, ρ_1 and ρ_2. This is a considerable obstacle in the attempt to find a general estimator and it is apparent that the difficulties which arise in seeking reasonable efficiency or unbiassedness will be greater than in the previous case. No doubt for known approximate values of the parameters a satisfactory estimator could be achieved which would be local to these values. As this is not thought to be a useful possibility in the present context, it is not explored further. Instead, quadratic ratio estimators are described briefly before proceeding to deal with developments connected with Stevens' Method which are believed to have some practical value.

7.4 Internal regression and general quadratic ratio estimators.

Some account is now given of the use of "internal regression" in estimating non-linear parameters. This method is based on the linearity which may exist when the model is re-stated in terms of differences of adjacent y-values observed at equally spaced values of x. Of course, the error in the linearised model is no longer independent and this leads to bias and inefficiency in the estimates. Two of the methods which have been suggested of this type are now described and reference will once more be made to the valuable study by Finney (1958).

Hartley (1948) proposed the estimation of the parameters in the SNLP asymptotic regression model (6.13) by the "internal regression" of $y_{i+1} - y_i$ upon $y_{i+1} + y_i$. The linearity is expressed in a simplified form for any equally spaced integer values for the x_i, by the relation

$$\eta_{i+1} - \eta_i = 2\alpha \frac{1-\rho}{1+\rho} - \frac{1-\rho}{1+\rho}(\eta_{i+1} + \eta_i) \qquad (7.9)$$

where the η-values are used to represent the expectation of the y-values. It is clear that on substitution of the y-values, the error in the statistical model would not be independent, and proceeding to obtain estimates of α and ρ by linear regression introduces a bias. This bias was investigated by Finney (1958) who showed that for

$n = 4$, the bias in the estimate of ρ is given by

$$\text{bias}(r) = \frac{\sigma^2}{\beta^2\rho^2(1-\rho)^2} \frac{3+4\rho+4\rho^2+4\rho^3+3\rho^4}{2(1+\rho+\rho^2)^2} \tag{7.10}$$

where σ^2 is the variance of each observed y-value. The above is a simpler form than Hartley's original proposal which required various partial sums to be formed. However, both the above and the original method are no more than ratio estimators with quadratic functions of the y-values as numerator and denominator. For the original Hartley method the bias, again for four observed values, is given by

$$\text{bias}(r) = \frac{\sigma^2}{\beta^2\rho^2(1-\rho)^2} \frac{2(3+4\rho+20\rho^2+16\rho^3+7\rho^4)}{(1+\rho)(3+4\rho+3\rho^2)^2} \tag{7.11}$$

and this is only slightly less than the bias given by (7.10), as ρ varies between 0 and 1, being equal to it for $\rho = 1$.

A computationally simpler method of the same type is due to Dr. St. C.S. Taylor and this appeared in the literature for the first time in the paper of Finney (1958). The method consists of regressing y_{i+1} upon y_i and the relation

$$\eta_{i+1} = \alpha(1-\rho) + \rho\eta_i \tag{7.12}$$

expresses the linearity which makes it easy to obtain estimates of some sort. Again these estimates are biassed and Finney proved in

this case that, for $n = 4$,

$$\text{bias}(\rho) = -\frac{\sigma^2}{\beta^2\rho^2(1-\rho^2)} \cdot \frac{1 + 4\rho + \rho^2}{2(1+\rho+\rho^2)^2} \qquad (7.13)$$

which is much smaller than either of the above and is of opposite sign.

In view of the fact that the "internal regression" estimators of ρ are all in the form of a ratio of quadratic functions of the y-values, Finney attempted to find the general quadratic estimator with least variance. This he partly succeeded in doing by providing, in the face of heavy analysis, an estimator which is fully efficient at the ends of the range of values of ρ and with as small a variance as possible in between. The "general" estimator deduced is now given, for $n = 4$, to indicate the complexity and difficulty of extending this result.

$$r_a = \frac{4y_4^2 - 4y_4y_3 - 3y_4y_2 - y_4y_1 - y_3^2 + 7y_3y_2 - y_3y_1 - 3y_2^2 + 2y_2y_1}{4y_4y_2 - 4y_4y_1 - 4y_3^2 + 4y_3y_1 + 2y_2^2 - 4y_2y_1 + 2y_1^2} \qquad (7.14)$$

By direct evaluation, Finney showed that this estimator had slightly smaller variance over the range of values of ρ but in general recommended Patterson's linear estimator in preference to any quadratic estimator. In a similar manner, Patterson and Lipton (1959)

investigated quadratic estimators, indicating that Hartley's method maintained high efficiency and relatively low bias generally and would be preferable to any other quadratic ratio estimator, except perhaps in the special case of four point regression, when the Taylor estimator has some merit, as described above.

CHAPTER EIGHT

TABLES FOR THE APPLICATION OF STEVENS' METHOD

8.1 Introduction.

In Chapter Six Stevens' Method was briefly considered; here the practical aspects of the Method are discussed and the computational procedure outlined in full. This section is intended to serve as a guide to the extended Stevens' tables and to offer a procedure which may be adopted in practice for many diverse applications of the Method.

As was stated earlier, the Method deals with asymptotic regression curves of the form:

$$\eta = E(y) = \alpha + \beta\rho^{x} \qquad (8.1)$$

where it is assumed that x is a fixed independent variable taking the value $0,1,2,\ldots.(N-1)$ and that a corresponding single value of y is available for each of the N values of x; α, β and ρ are constants to be determined in the course of the fitting. The Method remains useful for certain other curves and other restrictions upon the x-values;

these are discussed in section 8.4.

By an ingenious simplification of the maximum likelihood equations, W.L. Stevens (1951) showed that the estimation of the three constants reduces merely to the iterative estimation of ρ, from a first approximate r_0 with the α and β values being generated automatically by the process. In his paper Stevens provided tables mainly for the ranges $N = 5(1)7$ and $r_0 = .25(.01).70$; he suggested that his tables might profitably be extended and the aim of the present work is to make available such an extended set of tables.

The ranges considered here should cover the values which are most likely to occur in practice; these are $N = 5(1)30$ and $r_0 = .10(.01).90$. A discussion of how the tables may be used for values outside these ranges is given in section 8.5.

8.2 Description of the Method.

Forming the normal equations of least squares, or equivalently of maximum likelihood, and the information matrix I, defined by R.A. Fisher (1956), leads to a modified information matrix I_o such that:

$$I_o \begin{pmatrix} a \\ b \\ b\,\delta r \end{pmatrix} = \begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \end{pmatrix}$$

where a and b are the estimated values of the parameters α and β where δr is a correction to an approximation, r_o, to r, the estimate of ρ, and where $Y_0 = \Sigma y$, $Y_1 = \Sigma y r^x$ and $Y_2 = \Sigma y x r^{x-1}$ these sums being over $x = 0,1,2,\ldots.(N-1)$.

It can be shown that I_o is independent of a and b and it is, in fact, I with the second row and column each divided by the maximum likelihood estimate of b.

The iterative process is then defined by:

$$\begin{pmatrix} a \\ b \\ b\,\delta r \end{pmatrix} = I_o^{-1} \begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \end{pmatrix}$$

or

$$\begin{pmatrix} a \\ b \\ b\,\delta r \end{pmatrix} = \begin{pmatrix} F_{AA} & F_{AB} & F_{AR} \\ F_{AB} & F_{BB} & F_{BR} \\ F_{AR} & F_{BR} & F_{RR} \end{pmatrix} \begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \end{pmatrix} \qquad (8.2)$$

or equivalently, without using matrices, by:

$$\begin{aligned} a &= F_{AA}Y_0 + F_{AB}Y_1 + F_{AR}Y_2 \\ b &= F_{AB}Y_0 + F_{BB}Y_1 + F_{BR}Y_2 \\ b\delta r &= F_{AR}Y_0 + F_{BR}Y_1 + F_{RR}Y_2 \end{aligned} \qquad (8.3)$$

where the F-values are the elements of the matrix I_0^{-1}. These values are functions of r_0 only, since the elements of I_0 were independent of a and b.

In practice, δr is obtained by dividing the right-hand side of the last equation by that of the preceding equation of (8.3).

This iterative process has several advantages as well as the primary one that it deals in the estimation of the single parameter ρ; perhaps in order of importance these may be stated as:

(i) the process is a second order one since it is based on the second order derivative of the likelihood function occurring in a Taylor expansion of the function,

(ii) the estimates are of maximum efficiency,

(iii) the estimates are asymptotically unbiassed, and

(iv) large sample standard errors are available for the estimates.

Of these (ii) and (iii) are a direct consequence of the estimates being those which satisfy the maximum likelihood equations. This is, of course, only true under conditions where the residuals are

independent and of equal weight. Consideration of (iv) may now be extended by stating the formulae for the standard errors of the estimates:

$$\begin{aligned} s.e.(a) &= \sqrt{(F_{AA}\ s^2)} \\ s.e.(b) &= \sqrt{(F_{BB}\ s^2)} \\ s.e.(r) &= \{\sqrt{(F_{RR}\ s^2)}\}/b \end{aligned} \qquad (8.4)$$

where s^2 is the sum of squared deviations about the fitted curve divided by $(N-3)$.

8.3 Computational procedure.

It is first necessary to obtain an approximate value r_0, for the parameter ρ. If no other knowledge upon the size of this value is available, the following method may be used effectively when the fit is expected to be at all a good one.

By sketching a curve by eye through the points (x_i, y_i) choose a convenient numerical value, ξ_0, of x near to the centre of the range of fitting. Equidistant from ξ_0 take two other values, one near each end of the range and let these be called $\xi_0 - h$ and $\xi_0 + h$ respectively. Read off the values η_{-1}, η_0 and η_1 from the sketched curve as the values of y which correspond to $x = \xi_0 - h$, ξ_0 and $\xi_0 + h$ form the small difference table:

$\xi_0 - h$	η_{-1}	
		$\eta_0 - \eta_{-1}$
ξ_0	η_0	
		$\eta_1 - \eta_0$
$\xi_0 + h$	η_1	

The approximation required is then:

$$r_0 = \left(\frac{\eta_1 - \eta_0}{\eta_0 - \eta_{-1}} \right)^{\frac{1}{h}} \qquad (8.5)$$

or,

$$r_0 = \text{antilog} \left[\frac{1}{h} \left\{ \log (\eta_1 - \eta_0) - \log (\eta_0 - \eta_{-1}) \right\} \right]$$

where the base of logarithms is arbitrary.

Using this value of r_0 or a value arrived at in some other way, the iteration routine may now be entered. Before doing so, however, it is useful to compute $Y_0 = \Sigma y$ for once and all at the beginning, as this quantity remains unchanged throughout the calculation.

It is important to fix at the outset the size of the correction δr to r_0 which will be considered negligible.

(1) Compute $Y_1 = \Sigma y r_0^x$ and $Y_2 = \Sigma y x r_0^{x-1}$ where the summations are over all values of x, i.e. $x = 0,1,2,\ldots,(N-1)$.

(2) Using the appropriate row of F-values from the tables for r_0 and N calculate a, b and $b\delta r$ using the relation (8.2) or (8.3), and deduce the value of δr.

(3) If δr is not negligible, replace r_0 by $r_0 + \delta r$ and go back to (1); if δr is negligible the estimation is complete, the current values of a, b and r_0 being the final estimates.

(4) Using the fitted relationship, the expected y-values may be obtained for $x = 0,1,2,\ldots,(N-1)$ and the sums of deviations and of squared deviations computed. Of these sums, the former has expected value zero and the latter is divided by $(N-3)$ to give the estimated error variance, s^2.

(5) Where appropriate, large sample standard errors may be formed using the relations (8.4).

In unpublished notes, H. Linhart (1959) has suggested that the determination of the final estimates may be checked by converging to the final estimate r_0 from both sides. To do this, choose a starting value r_0 on the other side of the final value from that used above and repeat the estimation comparing both the final estimate r_0 and the sum of squared deviations with those obtained above.

8.4 Use of the tables for other curves.

Alternative models for which Stevens' Method may be applicable are now listed:

(1) $$y = \alpha + \beta e^{-\gamma x} \qquad (8.6)$$

the growth curve in physiology and biology, e.g. Harrison, Hiorns and Weiner (1964), or the learning curve in psychology; this becomes identical to the present model if $e^{-\gamma}$ is replaced by ρ.

(2) $$y = y_0 + d\left(1 - 10^{-kx}\right) \qquad (8.7)$$

the Mitscherlich's law, Mitscherlich (1909, 1930), describing fertiliser response in agriculture; here $y_0 + d = \alpha$, $-d = \beta$ and $10^{-k} = \rho$. An alternative form of this law is:

$$y = A\left\{1 - 10^{-c'(x+b')}\right\} \qquad (8.8)$$

and this becomes the present model if,

$A = \alpha$, $-A\left(10^{-c'b'}\right) = \beta$ and $10^{-c'} = \rho$

(3) $$z = \exp(\alpha + \beta\rho^x) \qquad (8.9)$$

Gompertz law, is used for graduating life tables in actuarial work and for predicting price changes in economics; this is equivalent if z is replaced by $\exp(y)$.

(4) $$z = 1/(\alpha + \beta\rho^x) \qquad (8.10)$$

the logistic curve occurs in demography to describe population growth e.g. Yule (1925); here write for equivalence, $z = \frac{1}{y}$.

This list is not, of course, exhaustive but it indicates the wide applicability of Stevens' Method in many fields. By way of warning, it should be added that although a law or function may be transformable into the general model considered here, the method should not be applied unless the x-variable is independent or is relatively free from error, whilst the y-variable contains error and is dependent upon x.

Example. Mice are weighted at weekly intervals between the ages 3 and 12 weeks. The following table gives the means over several litters:

age (wks.)	3	4	5	6	7	8	9	10	11	12
wt. (g.)	8.3	12.6	14.6	15.3	16.8	18.2	19.3	20.7	21.7	21.4

In the course of the fitting the weights will be represented by the variable y and the ages by $(x+3)$ so that $x = 0,1,2,\ldots,9$.

From a rough sketch of the curve, the mid-point of the range of x, $x = 4.5$ is taken and two end values, 0.5 and 8.5; the corresponding y values from this graph are given by the table:

x	y	Δ
0.5	10.5	
		7.5
4.5	18	
		3.3
8.5	21.3	

$$r_0 = \left(\frac{3.3}{7.5}\right)^{\frac{1}{4}} = 0.81$$

To determine r to two decimal places, proceed with the following scheme:

$$Y_0 = \Sigma y = 168.4$$

x	r_0^x	xr_0^{x-1}	y	y'	$y-y'$
0	1.0000 0000	0	8.3	8.94	-.64
1	.8100 0000	1.0000 0000	12.6	11.71	.89
2	.6561 0000	1.6200 0000	14.6	13.96	.64
3	.5314 4100	1.9683 0000	15.3	15.80	-.50
4	.4304 6721	2.1257 6400	16.8	17.29	-.49
5	.3486 7844	2.1523 3605	18.2	18.50	-.30
6	.2824 2954	2.0920 7064	19.3	19.49	-.19
7	.2287 6792	1.9770 0678	20.7	20.29	.41
8	.1853 0202	1.8301 4336	21.2	20.95	.25
9	.1500 9464	1.6677 1818	21.4	21.48	-.08
				Sum Check	.01

$$s^2 = \frac{2.4645}{7}$$

$$= \underline{\underline{0.35207}}$$

Matrix from tables, $r_0 = 0.81$, $N = 10$.

$$\begin{pmatrix} 4.80585 & -3.95752 & -1.75021 \\ -3.95752 & 3.87596 & 1.31778 \\ -1.75021 & 1.31778 & 0.69430 \end{pmatrix} \begin{pmatrix} 168.4 \\ 67.120718 \\ 297.041553 \end{pmatrix} = \begin{pmatrix} 23.7885 \\ -14.8537 \\ -0.049074 \end{pmatrix}$$

Using $Y_1 = \Sigma y r_0^x = 67.120718$, $Y_2 = \Sigma y x r^{x-1} = 297.041553$

$$\delta r = \frac{-0.049074}{-14.8537} = 0.003304 \qquad \therefore r = r_0 + \delta r = 0.8133.$$

The iteration is complete because $\delta r < 0.005$

$F_{AA}s^2 =$	1.6920	$\therefore$ s.e.(a) =	1.30	$a = 23.79$	$\pm$	1.30
$F_{BB}s^2 =$	1.3646	$\therefore$ s.e.(b) =	1.17	$b = -14.85$	$\pm$	1.17
$\frac{F_{RR}s^2}{b^2} =$	0.00110791	$\therefore$ s.e.(r) =	0.0333	$r = 0.8133$	$\pm$	0.0333

These standard errors are asymptotic in the sense explained in section 8.2.

8.5 Data for which the ranges of tabulation are exceeded.

Large values of r.

When r is large, becoming nearly unity, there is a good case for fitting the parabola, as Stevens illustrated with an example.

If $\tau = 1 - \rho$ is small so that τ^3 is everywhere negligible compared with ϵ_i, it is clear that (8.1) can be rewritten as:

$$E(y) = \alpha + \beta(1-\tau)^x \qquad (8.11)$$

or expanding binomially, as

$$E(y) = (\alpha+\beta) - \beta\tau x + \tfrac{1}{2}\beta x(x-1)\tau^2 \qquad (8.12)$$

from which the asymptotic regression can be equated with the quadratic regression

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 \qquad (8.13)$$

and this may be fitted by orthogonal polynomials (e.g. Fisher and Yates, p.30 (1957)). The asymptotic parameters correspond to the polynomial parameters and are given in terms of them by

$$\rho = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}, \quad \alpha = \beta_0 - \frac{(\beta_1+\beta_2)^2}{2\beta_2^2} \quad \text{and} \quad \beta = -\frac{(\beta_1+\beta_2)^2}{2\beta_2^2} \qquad (8.14)$$

Small values of r.

With r^2 negligible, consideration of the inverse of the modified information matrix, I_0^{-1}, (defined in section 8.2), leads

to the approximate relation:

$$I_o^{-1} = \frac{1}{N-2-4r}\begin{pmatrix} 1 & -1 & -1-r \\ -1 & N-1-4r & -(N-3)r+1 \\ -1-r & -(N-3)r+1 & N-1-2r \end{pmatrix} \qquad (8.15)$$

whose elements can be used in place of the F-values, provided the errors, ϵ_i, are everywhere large compared with r^2.

Values of N outside the given range.

From (8.15), if I_o^{-1} is replaced in an obvious notation by F_N^r then,

$$(N-2-4r)\,F_N^r = \begin{pmatrix} 1 & -1 & -1-r \\ -1 & -1-4r & 3r+1 \\ -1-r & 3r+1 & -1-2r \end{pmatrix} + N\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -r \\ 0 & -r & 1 \end{pmatrix}$$

$$= R_1 + N R_2 \quad , \text{ say,}$$

and $(N-1-4r)\,F_{N+1}^r = R_1 + (N+1)R_2$

so that by subtraction,

$$(N-1-4r)\,F_{N+1}^r - (N-2-4r)\,F_N^r = R_2$$

and hence the recurrence relations:

$$F_{N+1}^r = \left(\frac{N-2-4r}{N-1-4r}\right)F_N^r + \left(\frac{1}{N-1-4r}\right)R_2 \qquad (8.16)$$

$$F_N^r = \left(\frac{N-1-4r}{N-2-4r}\right)F_{N+1}^r - \left(\frac{1}{N-1-4r}\right)R_2 \qquad (8.17)$$

may be used to extend the range of tabulation for small values of r, again where r^2 is negligible by comparison with the error term, ϵ_i^2.

8.6 The checking of the tabulation.

The tables were computed in rows in the order in which they appear so that the usual visual graduation check by columns would fail to detect any "build-up" error. This check was applied successfully with respect to gross errors and none of these were located.

A more satisfactory check was available due to a fortunate property of the matrix inverses and this check is independent of the method of calculation. To show this, a recurrence relation may be derived.

The modified information matrix I_o can be shown to be, (omitting the lower triangle elements of the symmetric matrices which follow)

$$I_o = \begin{pmatrix} N & \Sigma r^{x} & \Sigma x r^{x-1} \\ & \Sigma r^{2x} & \Sigma x r^{2x-1} \\ & & \Sigma x^2 r^{2x-2} \end{pmatrix}$$

and this matrix will be called A_N^r indicating the summations over $x = 0,1,2,\ldots,(N-1)$. Defining:

$$\Delta_N^r = \begin{pmatrix} 1 & r^N & N r^{N-1} \\ & r^{2N} & N r^{2N-1} \\ & & N^2 r^{2N-2} \end{pmatrix}$$

and if A_{N+1}^r is defined as I_o with summations over $x = 0,1,2,\ldots,N$

then:

$$A^r_{N+1} = A^r_N + \Delta^r_N$$

For $N > 2$ the A matrices are non-singular so that multiplying from the left by $(A^r_{N+1})^{-1}$ and from the right by $(A^r_N)^{-1}$ shows that:

$$(A^r_N)^{-1} = (A^r_{N+1})^{-1} + (A^r_{N+1})^{-1} \Delta^r_N (A^r_N)^{-1}$$

or,

$$(A^r_N)^{-1} - (A^r_{N+1})^{-1} = (A^r_{N+1})^{-1} \Delta^r_N (A^r_N)^{-1}$$

The matrix Δ^r_N is always singular since its rows are proportional and it has, therefore, rank zero. By the ~~theorem on the rank of a product,~~ product rule for determinants, the determinant of the right-hand side product is zero since the determinant of Δ^r_N is zero. ~~matrix cannot have rank greater than zero and is thus singular.~~ It follows directly from this that the determinant of the left-hand side matrix is zero, i.e.

$$\left|(A^r_N)^{-1} - (A^r_{N+1})^{-1}\right| = 0$$

which is the required check. By forming the differences of rows of F-values which have the same r-value but adjacent N-values, a matrix is obtained whose determinant should be zero. This check was applied by the computer, indicating that nowhere was the value of this determinant greater than 10^{-6}.

In addition, another valuable check was used, particularly during the development of the tabulation programme, and this was a direct comparison with existing independent tabulations over smaller

ranges by W.L. Stevens (1951) for $N = 5(1)7$ and S. Lipton (private communication) for $N = 5(1)12$.

Finally, the most positive check (but this was not applied extensively as it takes longer than other checks) is to use the tables for fitting a set of points which lie exactly on the curve; if the true value of r is used as starting value, the expected correction δr must be zero.

CHAPTER NINE

EXTENSION OF STEVENS' METHOD.

Stevens showed that the likelihood equations allow the elimination of all but the non-linear parameter to provide an iterative procedure. He suggested that his method, applied to:

$$\eta = E(y) = \alpha + \beta\rho^{x} \tag{9.1}$$

could be applied to general models of the form:

$$\eta = \alpha + \beta f(\rho, x) \tag{9.2}$$

where $f(\rho, x)$ is any function.

Whereas this is true, the method extends much further. It is not, of course, restricted to a single non-linear function containing a parameter, and neither is it only valid for a single independent variable x. The most general function for which Stevens' elimination is possible, appears to be:

$$\eta = \alpha + \sum_{j=1}^{p} \beta_j f_j(\rho_j; x_1, \ldots, x_q) \tag{9.3}$$

where f_j are any functions of q independent variables $x_1, \ldots, x_q$ and the parameter vectors ρ_j whose elements are exclusive to the function

f_j . The proof which follows assumes that each r_j has one element only r_j and the case where these vectors are of higher orders is a straightforward extension.

Introduce $f_0 = 1$, always and define $b_0 = a$ so that:

$$\eta = \sum_{j=0}^{p} \beta_j f_j(\rho_j ; x_1, \ldots, x_q) \qquad (9.4)$$

and, assuming normality for the independent errors in the y-values, the usual maximum likelihood equations (MLE) with β_j replaced by b_j , ρ_j replaced by r_j are $(p+1)$ equations for $k = 0,1,2,\ldots p$:

$$-\sum_j b_j \sum_i f_k f_j + \sum y_i f_k = 0 \qquad (9.5)$$

and p equations for $k = 1,2,\ldots p$:

$$-b_k \sum b_j \sum \frac{\partial f_k}{\partial r_k} f_j + b_k \sum y_i \frac{\partial f_k}{\partial r_k} = 0 \qquad (9.6)$$

where f_j represents the general function $f_j(r_j ; x_1, \ldots, x_q)$ and the suffix i is used to refer to repeated observed values of the variables. For non-linear functions f_j the above equations are difficult to solve and R.A. Fisher's general method, Fisher (1925, 1956), is relevant. This involves the use of the information matrix, I , which is readily obtained by differentiating the left-hand sides of the equations above with respect to each parameter in turn, and replacing each y by its expected value, η . The latter is found by taking the mean of y over all repeated sets of observations.

In view of the structure of the likelihood equations I, in this case, is a partitioned matrix given by:

$$I = \begin{pmatrix} A & B \\ B' & C \end{pmatrix} \tag{9.7}$$

where,

$$A = \{A_{ij}\} = \Sigma f_i f_j$$

$$B = \{B_{ij}\} = b_j \Sigma f_i \frac{\partial f_j}{\partial r_j}$$

$$C = \{C_{ij}\} = b_i b_j \Sigma \frac{\partial f_i}{\partial r_i} \frac{\partial f_j}{\partial r_j}$$

Here A and C are square matrices of orders $(p+1)$ and p respectively, B is a $p \times (p+1)$ matrix and B' is the transpose of B.

R.A. Fisher (1925) proposed that if $\hat{v}$ represents the vector of parameters which satisfy the likelihood equations and v_i' is an approximation to $\hat{v}$ then:

$$w = I\,(v' - \hat{v}) \tag{9.8}$$

where w is the vector of residuals upon substituting v' into the MLE. This result holds for v' close to $\hat{v}$ because of the smallness of squares of elements $(v' - \hat{v})$ by Taylor's Theorem. Further, it is not dependent in any way upon the form of model which determines I. Here the linear combination model form can be applied to advantage.

It will be convenient to define the vector, δ by:

$$\delta = v' - \hat{v} \tag{9.9}$$

so that (9.8) can be written:

$$w = I\delta \tag{9.10}$$

Suppose that I_o is defined:

$$I_o = \begin{pmatrix} A & B^* \\ B^{*\prime} & C^* \end{pmatrix} \tag{9.11}$$

where:

$$B^* = \{(B_{ij} / b_j)\} = \Sigma f_i \frac{\partial f_j}{\partial r_j}$$

and,

$$C^* = \{ C_{ij} / b_i b_j \} = \Sigma \frac{\partial f_i}{\partial r_i} \frac{\partial f_j}{\partial r_j}$$

with A as defined previously.

Now ~~if~~ the vector of residuals w is defined as:

$$w = (w_{B0}, w_{B1}, \cdots, w_{Bp}, w_{R1}, \cdots, w_{Rp})$$

in an obvious way, and:

$$w_o = (w_{B0}, w_{B1}, \cdots, w_{Bp}, w_{R1}/b_1, \cdots, w_{Rp}/b_p)$$

Given a set of approximate values, the small corrections to these values will be denoted by:

$$\delta = (\delta b_0, \delta b_1, \cdots, \delta b_p, \delta r_1, \cdots, \delta r_p)$$

Let,

$$\delta_o = (\delta b_0, \delta b_1, \cdots, \delta b_p, \delta r_1/b_1, \cdots, \delta r_p/b_p)$$

so that from (9.10):

$$w_o = I_o \delta_o$$

or alternatively,

$$\delta_o = I_o^{-1} w_o$$

and by substituting from (9.8) and (9.9):

$$\delta_o = v_o + I_o^{-1} z$$

where

$$v_o = (b_o', b_1', \ldots, b_p', 0, \ldots, 0)$$

and

$$z = (\Sigma y, \Sigma y f_1', \ldots, \Sigma y f_p', \Sigma y \frac{\partial f_1'}{\partial r_1}, \ldots, \Sigma y \frac{\partial f_p'}{\partial r_p'})$$

in which,

$$f_j' = f_j(r_j', x_j)$$

and,

$$\frac{\partial f_j'}{\partial r_j} = \frac{\partial f_j}{\partial r_j}\Big|_{r=r_j'}$$

for $j = 1,2,\ldots, p$. The iterative process is now defined by:

$$\delta_o - v_o = I_o^{-1} z$$

in which the right-hand side is a function only of the non-linear parameters. More concisely, if:

$$t = (b_o, b_1, \ldots, b_p, b_1\delta r_1, \ldots, b_p \delta r_p)$$

then the iterative process is:

$$t = I_o^{-1} z$$

The converged values of the estimates obtained by use of this iterative process will satisfy the maximum likelihood equations. ~~and~~ They

have the usual properties of full efficiency and ~~asymptotic unbiasedness.~~ are asymptotically unbiased. Further, for estimation from large samples, the estimates are asymptotically normally distributed and standard errors may be attached to them using the modified information matrix:

$$\text{var}(\hat{b}_i) = s^2 F_{ii} \qquad i = 1, 2, \ldots, p$$

$$\text{var}(\hat{r}_i) = s^2 F_{jj} / \hat{b}_i^2 \qquad j = p+i$$

where F_{ii} is the i^{th} diagonal element of I_o^{-1} and s^2 is the estimate of the variance of the error in each observed y-value.

It is now clear that in place of r_j in each function there may be a vector of several non-linear parameters. The above iterative method still applies provided that no non-linear parameter occurs in more than one function.

9.1 Confidence limits and regions for non-linear parameters.

It is a feature of the classical iterative approach, or again of Stevens' Method, that approximate confidence intervals may be obtained immediately from the information matrix used during the course of the estimation. Whether the approximation is a useful one or not depends entirely upon the degree of non-linearity in the MNLP relation. This point is discussed fully by Beale (1960) who derives a measure of non-linearity and obtains confidence regions for models which become almost linear under some transformation. The computational difficulties encountered in obtaining these regions, leaves their general usefulness open to some doubt, particularly in view of the fact that the interpretation of a region is so difficult for three or more parameters; more acceptable would be a simultaneous set of confidence intervals. Stone (1960) shows some concern for the need for more confidence intervals and he indicates that simultaneous intervals are obtainable from the usual approximate confidence region. This approach is now considered for the asymptotic models under present discussion and then extended to obtain limits for new regions proposed by Halperin (1963) following Williams (1962).

From the last section, let $\hat{\theta}$ represent the vector of maximum likelihood estimates of k parameters so that

$$\hat{\theta} = \{ \hat{b}, \hat{r} \} \qquad (9.12)$$

where $\hat{b}$ and $\hat{r}$ are the similar vectors for the linear and non-linear parameters separately. If $F_{k,n-k}(\alpha)$ is the 100α % value of the probability function for the F-distribution, and if, to correspond with the previous notation, s^2 is the error variance estimate, I is the information matrix, then

$$(\theta - \hat{\theta})' I (\theta - \hat{\theta}) \leq k s^2 F_{k,n-k}(\alpha) \qquad (9.13)$$

represents the approximate confidence region within which the true value of θ falls with probability α. This region represents the interior of an ellipsoid and Stone (1960) proposes to use instead the circumscribing polyhedron defined by the set of bounding hyperplanes

$$\theta - \hat{\theta} = \pm \sqrt{\{ k s^2 F_{k,n-k}(\alpha) \, diag(I^{-1}) \}}$$

and d may be introduced to represent the quantity whose square root is taken so that the simultaneous confidence intervals are given by the vector equation:

$$\hat{\theta} - d \leq \theta \leq \hat{\theta} + d$$

These intervals define a conservative confidence region, and the probability that <u>all</u> of the k inequalities are satisfied is <u>at least</u> α. The asymptotic normality of maximum likelihood estimates allows an application of the more general theorem by Scheffé (1953) on simultaneous confidence intervals for a linear combination of random

variables. It is a straightforward matter to show that the Scheffé intervals for the above simplified case are those used by Stone.

In a recent paper, Williams (1962) approaches the (SNLP) relation (6.1) as a specific type of non-linear relation in an attempt to obtain an exact confidence interval for the non-linear parameter, ρ. This he proceeds to do by showing that Stevens' modified normal equations for the model are a consequence of applying a Taylor expansion to it using only ρ for this purpose and the consequent linearity of the regression. The reduction in size of the correction term in the linear regression during the iterative process which follows, suggests that this term could be tested for significance at any stage. When this term fails to reach significance at a given probability level, according to the appropriate sums of squares in an analysis of variance, the corresponding confidence limit is presumed to have been passed. A close study of the linear regression sums of squares with and without the error term, allows Williams to obtain in this way an exact interval for ρ.

A broader approach is taken by Halperin (1963) who considers the effect of regression on the given non-linear variables together with a set of general variables: showing that the latter may be replaced by corrections to the non-linear parameters, a result previously given by Hiorns (1962b). This author follows the work of Williams and proposes

general confidence regions for MNLP relations. The results are given below, in equation (9.17) and (9.18).

Let X be the matrix of observations containing the p terms in the p_1 non-linear parameters and D represent a matrix of p_2, for the moment, unspecified variables and each of these matrices is supposed to have rows to correspond with the sample size but X has p columns and D has p_2 columns. Suppose that the model is specified by

$$\eta = E(y) = X\beta_1$$

where β_1 is the vector of p parameters occurring linearly in the model and consider the supplemented regression

$$\eta = (X, D)\beta \tag{9.15}$$

where $\beta' = (\beta_1', \beta_2')$

and β_2 contains those p_2 parameters which occur linearly with the unspecified variables in D. Following the usual regression theory, the asymptotically unbiassed estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$ are contained in $\hat{\beta}$ where

$$\hat{\beta} = \begin{pmatrix} X'X & X'D \\ D'X & D'D \end{pmatrix}^{-1} \begin{pmatrix} X' \\ D' \end{pmatrix} y \tag{9.16}$$

Furthermore, the significance of the whole or part of the regression

may be treated by the variance ratios:

$$F_{p_1+p_2,\, n-p_1-p_2} = \frac{n-p_1-p_2}{p_1+p_2} \; \frac{(\hat{\beta}-\beta)'(X,D)'(X,D)(\hat{\beta}-\beta)}{y'y - \hat{\beta}'(X,D)'(X,D)\hat{\beta}} \tag{9.17}$$

$$F_{p_2,\, n-p_1-p_2} = \frac{n-p_1-p_2}{p_2} \; \frac{\hat{\beta}_2' U'U \hat{\beta}_2}{y'y - \hat{\beta}'(X,D)'(X,D)\hat{\beta}} \tag{9.18}$$

where $$U = D'\left(I - X(X'X)^{-1}X'\right) \tag{9.19}$$

and (9.18) may be used to define a confidence region for β_2 in the normal way. Such a region may be disjoint or infinite but this feature will be discussed later.

The choice of the variables in D is now clear, for the matrix inverted in (9.16) can be made to correspond with the modified information matrix, I_0, defined by the last section in equation (9.11) if

$$D = \{\, \partial f_j(x_i) / \partial r_j \,\} \tag{9.20}$$

With this definition of D it follows that the parameters in β_2 are the elements at the lower end of the vector t of the last section and are defined by

$$\beta_2 = \{\beta_{2j}\} = \{\beta_{1i}\,\delta\rho_i\} \tag{9.21}$$

so that the maximum likelihood estimates of these parameters are zero.

For many non-linear paremeters there is considerable difficulty both in obtaining the confidence region from (9.18) and in interpreting it. The calculation will be simplified if the appropriate matrix product can be replaced by an alternative estimate of the variance in the dependent variable and this can be quickly achieved if replications of this variable are available. It remains then only to explore the matrix function in the numerator of the right-hand side of (9.18) for different trial values for the non-linear parameters. For one such parameter this was carried out by Williams (1962) in one example, but in another he shows that considerable effort may be needed to study this matrix function even for a single parameter.

The interpretation of confidence regions can only be a simple matter when the boundaries of the regions are simply defined and closed. This is the case for linear estimates when the boundary is an ellipsoid with its centre at the maximum likelihood point in the parameter space, as indicated by Hotelling (1931). Of course, the proposal by Stone, quoted above is based on the assumption that the regions have ellipsoidal boundaries and a further assumption is made that these ellipsoids are not rotated about the parameter axes. The additive effect of these two assumptions would prove drastic on all but the most moderate of non-linear regressions. Taking the assumptions separately, it is not difficult to find a set of

conservative simultaneous intervals for a rotated ellipsoid. If (9.13) represents the interior of an ellipsoid then the lengths of the semi-axes are proportional to the square roots of the latent roots of I^{-1}. The hypersphere of radius proportional to the largest of these square roots, therefore, is a conservative region in that it encloses all points in the parameter space which satisfy (9.13). In practice, the largest latent root is easiest to obtain, as is well known, and therefore the equation of this hypersphere may be determined readily. Another attractive feature is that simultaneous intervals may be obtained directly from the intercepts of the parameter axes with this hypersphere, taking no further account of rotation. Against this, however, is the overlooking of information contained about the variances of all but one of the estimates. The importance of this is great only if heterogeneity of these variances exists: it would seem reasonable to test for this in the usual way with Bartlett's test and the only variances available will be the asymptotic values obtained from I^{-1}.

A more exact measure of the importance of choosing a hypersphere may be obtained as follows. The volume, V_E, of the ellipsoidal region given by (9.13) is

$$V_E = \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)} \frac{c^k}{\sqrt{|I|}} \qquad (9.22)$$

where c^2 denotes the right-hand side of (9.13). (This result follows

directly from an orthogonal transformation of the positive definite matrix I and a simple substitution, or equivalently by using the result which is sometimes called Aitken's integral, (Turnbull and Aitken (1932)), i.e.

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}\theta' A \theta + it'\theta} \, d\theta_1 \, d\theta_2 \cdots d\theta_k$$

putting $t = 0$, and the gamma function results from the Dirichlet integration over the hypersphere or transformed ellipsoid. The proof is given by Cramér (1946, pp 118-20) and elsewhere.) From this the volume of the hypersphere, V_S, with radius proportional to ρ, the square root of the largest latent root of I^{-1}, is readily deduced by replacing I by the unit matrix divided scalarly by ρ^2. A measure describing how conservative the region described by the hypersphere would be, is then

$$\delta = \frac{V_S - V_E}{V_E} = \rho^k \sqrt{|I|} - 1 \qquad (9.23)$$

Here $100\delta\%$ can be defined as the percentage increase in the probability α that the overall probability that the hypersphere rather than the ellipsoid (9.13) contains the true point in the parameter space, but this will not be true for corresponding simultaneous intervals. In the above, of course the fact that the inverse information matrix is proportional to the asymptotic covariance matrix for the parameters means that δ is analogous to a measure of the proportional increase in the

generalized standard deviation, if the latter is defined to be the square root of the generalized variance originally derived by Wilks (1932). Another analogous quantity is the scatter coefficient as given by Frisch (1929) which is the reciprocal of one plus δ given above. In the extreme, when all parameter estimates have the same asymptotic variance and zero covariances, this coefficient is unity and δ is zero.

The corresponding volume of the hypercube, V_c whose sides are the hyperplanes representing the simultaneous confidence intervals, is now considered. This hypercube has edges of length $2\rho c$, the length of each interval and its volume is, therefore,

$$V_c = (2\rho c)^k \qquad (9.24)$$

For $k=1$, the hypercube degenerates to the line segment $\theta_1 = \pm \rho c$ and the length of this single interval is the "volume" of the hypercube; in this case the volumes of the ellipsoid and hypersphere given above both correspond with this value, as is expected from the coincidence of the degenerate cases of these regions. For higher values of k, the volumes diverge as is shown in the table below. In the table $Y = 1/\sqrt{|I|}$.

TABLE 9.1

Region	No of parameters 1	2	3	4	k
Ellipsoid, V_E	$2\gamma c$	$\pi\gamma c^2$	$\frac{4}{3}\pi\gamma c^3$	$\frac{1}{2}\pi^2\gamma c^4$	$\frac{\pi^{k/2}\gamma c^k}{\Gamma(\frac{k}{2}+1)}$
Hypersphere, V_S	$2\rho c$	$\pi\rho^2 c^2$	$\frac{4}{3}\pi\rho^3 c^3$	$\frac{1}{2}\pi^2\rho^4 c^4$	$\frac{\pi^{k/2}\rho^k c^k}{\Gamma(\frac{k}{2}+1)}$
Hypercube, V_C	$2\rho c$	$4\rho^2 c^2$	$8\rho^3 c^3$	$16\rho^4 c^4$	$2^k\rho^k c^k$
Ratio V_S/V_E	1	ρ^2/γ	ρ^3/γ	ρ^4/γ	ρ^k/γ
Ratio V_C/V_S	1	$4/\pi$	$6/\pi$	$32/\pi^2$	$\frac{2^k\Gamma(\frac{k}{2}+1)}{\pi^{k/2}}$

Some consideration is now due of non-ellipsoidal regions as these are the kind most likely to be encountered when applying (9.18). An approach is given which caters for open and closed regions though in the former case the existence of points at infinity in the set of allowed parameter values is assumed. Let B_j be the set of values of the parameter β_{2j} for which (9.18) is satisfied and call β_{2Sj} that member of the set which has largest magnitude and β_{2Ij} that member with smallest magnitude so that

$$\beta_{2Sj} = \sup_{B_j}(\beta_{2j}) \qquad (9.25)$$

and

$$\beta_{2Ij} = \inf_{B_j}(\beta_{2j})$$

If the set B_j is empty for any set these two values are taken as plus and minus infinity respectively. For all j, let β_{2S} and β_{2I}

be the vectors of elements defined; then it follows that

$$\beta_{2I} \leq \beta_2 \leq \beta_{2S} \tag{9.26}$$

defines a set of simultaneous conservative confidence intervals.

The exploration of these intervals further is difficult and promises to be lengthy and even particular cases of the general MNLP models must extend beyond the scope of the present research. However, it is possible to state this problem more precisely for the MNLP models in which

$$f_j = \rho_j^{\alpha} \tag{9.27}$$

It is straightforward to show that, following (9.20) from the quadratic form in the numerator of (9.18)

$$u'u = x'T^2x - x'Tx(x'x)^{-1}(x'Tx) \tag{9.28}$$

where T is a diagonal matrix with elements equal to the values of α in the sample, and β_2 is defined by

$$\beta_2 = (u'u)^{-1}u'y$$

so that

$$\beta_2'(u'u)\beta_2 = y'u(u'u)^{-1}u'y \tag{9.29}$$

and

$$u = \{I - x(x'x)^{-1}x'\}Tx$$

In the above x is a function of a current value of the vector ρ_0.

For given sample values for x and y, together with an estimate s^2, values of ρ, or equivalently deviations $\delta\rho$ from ρ may be found which satisfy (9.18) from (9.28) and (9.29) above.

REFERENCES

Aitken, A.C. (1956) Determinants and Matrices. Edinburgh: Oliver and Boyd. 9th edition.

Aitken, A.C. (1935) On least squares and linear combinations of observations. Proc. Roy. Soc. Edin. 55, 42.

Aitken, A.C. (1937) The evaluation of the latent roots and vectors of matrix. Proc.Roy. Soc. Edin. 57, 269-304.

Bartlett, M.S. (1956) Stochastic Processes. Cambridge University Press.

Beale, E.M.L. (1960) Confidence regions in non-linear estimation. Jour. Roy. Statist. Soc. Ser. B, 22, 41-88.

Booth, A.D. (1957) Numerical Methods. London: Butterworths (p.154)

Charnes, A., Cooper, W.W. and Ferguson, R.O. (1955) Optimal estimation of executive compensation by linear programming. Management Science, 1, 138-50.

Cramér, H. (1946) Mathematical Methods of statistics. Princeton: Princeton University Press.

Dantzig, G.B. (1949) Maximisation of a linear function of variables subject to linear inequalities. Activity Analysis of Production and Allocation. (Ed. T.C. Koopmans) New York: John Wiley.

Davidon, W.C. (1959) Variable Metric method for minimisation. A.E.C. Res. and Dev. report ANL-5990.

Feller, W. (1958) Probability Theory. New York: John Wiley 2nd edition.

Finney, D.J. (1958) The efficiencies of alternative estimators for an asymptotic regression equation. Biometrika 45, 370-388.

Fisher, R.A. (1925) Theory of statistical estimation. Proc. Camb. Phil. Soc. 22, 700.

Fisher, R.A. (1930) The genetical theory of natural selection. London: Oxford University Press.

Fisher, R.A. (1956) Statistical Methods and Scientific Inference Edinburgh: Oliver and Boyd.

Fisher, R.A. and Yates, F. (1957) Statistical Tables. Edinburgh: Oliver and Boyd.

Fisher, W.D. (1961) A note on curve fitting with minimum diviations by linear programming. Jour. Amer. Statist. Assoc. 56, 359-62.

Fletcher, R. and Powell, M.J.D. (1963) A rapidly convergent descent method for minimisation. Computer Journal, 6, 163-8.

Fourier, J.B.J. (1888-90) Oeuvres de Fourier (Ed. G. Darboux) Paris: Gauthier Villars. Vol. 2, 334-5.

Fraser, R.A., Duncan, W.J. and Collar, A.R. (1946) Elementary Matrices. Cambridge University Press.

Fréchet, (1937-8) Recherches theoriques modernes sur la théorie des probabilités. Traité du Calcul des Probabilités (ed Borel) 1, 3. Paris.

Frisch, R. (1929) Correlation and scatter in statistical variables. Nord. Statist. Tidskr., 8, 36.

Glass, B. and Li, C.C. (1953) The dynamics of racial intermixture. Amer. Jour. Hum. Genetics, 5, 1-20.

Gomes, F.P. (1953) The use of Mitscherlich's regression law in the analysis of experiments with fertilizers. Biometrics, 9, 498-516.

Gomes, F.P. and Malavolta, E. (1949) Aspectos matemáticos e estatísticos da lei de Mitscherlich. Escola de Agricultura, Piracicaba, Sao Paulo, Brasil (mimeograph).

Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philos. Trans. Roy. Soc. London, Ser. A, 115, 513-585.

Guest, P.G. (1961) Numerical methods of Curve fitting. Cambridge University Press.

Haldane, J.B.S. (1924-32) A mathematical theory of natural selection. Proc. Camb. Phil. Soc., 23, 19, 158, 363, 838; 26, 220; 27, 131; 28, 244.

Halperin, M. (1963) Confidence interval estimation in non-linear regression. Jour. Roy. Statist. Soc. Ser. B, 25, 330-3.

Harrison, G.A., Hiorns, R.W. and Weiner, J.S. (1964) The growth of mice in a fluctuating environment. Proc. Roy. Soc. Ser. B, 160, 137-48.

Hartley, H.O. (1948) The estimation of non-linear parameters by 'internal least squares'. Biometrika, 35, 32-45.

Hartley, H.O. (1961) The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares. Technometrics, 3, 269-380.

Hestenes, M.R. and Stiefel, E. (1952) Methods of conjugate gradients for solving linear systems. Jour. Res. N.B.S., 49, 409.

Hiorns, R.W. (1962a) A further note on the extension of Stevens' tables for asymptotic regression. Biometrics, 18, 123.

Hiorns, R.W. (1962b) Statistical estimation for linear compound models. Int. Cong. Mathematicians, Stockholm. Abstracts.

Hiorns, R.W. (1964) The fitting of growth and allied curves of the asymptotic regression type by Stevens' Method. London: Department of Statistics, University College. Cambridge University Press. Tracts for Computers Series Vol. XVIII

Hotelling, H. (1931) The generalization of Student's ratio. Ann. Math. Statist, 2, 360-78.

Levenberg, K.A. (1944) A method for the solution of certain non-linear problems in least squares. Quart. Jour. Appl. Math. II, 164-168.

Linhart, H. (1959) Unpublished notes on Stevens' Method.

Lipton, S. Tables for Stevens' method. Private communication.

Mitscherlich, E.A. (1909) Das Gesetz des Minimums und das Gesetz des abnehmenden Bodersertrags. Landw. Jahrb. 38, 537-52.

Mitscherlich, E.A. (1930) Die Bestimmung des Dungerbedurfnisses des Bodens. Berlin: Paul Parey.

Owen, A.R.G. (1959) Mathematical models for selection. (in Natural selection in human populations: D.F. Roberts and G.A. Harrison, eds.) London: Pergamon Press.

Patterson, H.D. (1956) A simple method for fitting an asymptotic regression curve. Biometrics, 12, 323-9.

Patterson, H.D. (1958) The use of autoregression in fitting an exponential curve. Biometrika, 45, 389-400.

Patterson, H.D. and Lipton, S. (1959) An investigation of Hartley's method. Biometrika, 46, 281-92.

Patterson, H.D. (1960) A further note on a simple method for fitting an exponential curve. Biometrika, 47, 177-80.

Powell, M.J.D. (1962) An iterative method for finding stationary values of a function of several variables. Computer Journal, 5, 147.

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

Roberts, D.F. and Hiorns, R.W. (1962) The dynamics of Racial Intermixture. Amer. Jour. Hum. Genetics, 14, 261-77.

Roberts, D.F. and Hiorns, R.W. (1963) Population mixture and gene frequency change. Proc. Second. Int. Cong. Human Genetics, Rome, 151.

Roberts, D.F. (1955) The dynamics of racial intermixture in the American Negro. Amer. Jour. Hum. Genetics, 7, 361-7.

Sauvage, L. (1895) Théorie generale des systèmes d'equations differentielles linéaires et homogènes. Annales de Toulouse Vol VIII.

Scheffé, H. (1953) A method for judging all contrasts in the analysis of variance. Biometrika, 40, 87-104.

Schneider, B. (1963) Die Bestimmung der Parameter im Ertragsgesetz von E.A. Mitscherlich. Sonderdruck aus Biometrische Zeitschrift, Band 5, Heft 2.

Stevens, W.L. (1951) Asymptotic Regression. Biometrics, 7, 247-67.

Stone, M. (1960) Contribution to discussion of paper by E.M.L. Beale. Jour. Roy. Statist. Soc. Ser. B, 22, 84-5.

Sylvester, J.J. (1883) Phil. Mag. (5) 16, 267.

Turnbull, H.W. and Aitken, A.C. (1932) An introduction to the theory of canonical matrices. London: Blackie.

Turner, M.E. (1959) The single process law: a study in non-linear regression. Dissertation no. 59-6571 University Microfilms, Inc. Ann. Arbor, Mich.

Turner, M.E., Monroe, R.J. and Lucas, H.L. (1961) Generalised asymptotic regression and non-linear path analysis. Biometrics, 17, 120-143.

Wagner, H.M. (1959) Linear programming techniques for regression analysis. Jour. Amer. Statist. Assoc., 54, 206-12.

Wagner, H.M. (1962) Non-linear regression with minimal assumptions. Jour. Amer. Statist. Assoc., 57, 572-8.

Whittaker, E. and Robinson, G. (1944) The calculus of observations. 4th edition London: Blackie.

Wilks, S.S. (1932) Certain generalizations in the analysis of variance. Biometrika, 24, 471-94.

Wilks, S.S. (1962) Mathematical Statistics. New York: John Wiley.

Will, H.S. (1936) On a general solution for the parameters of any function with application to the theory of organic growth. Ann. Math. Statist. 7, 165-90.

Williams, E.J. (1959) Regression Analysis. New York: John Wiley.

Williams, E.J. (1962) Exact fiducial limits in non-linear estimation. Jour. Roy. Statist. Soc. Ser. B, 24, 125-39.

Wright, S. (1931) Evolution in mendelian populations. Genetics, 16, 97.

Yule, U. (1925) The growth of populations and the factors which control it. Jour. Roy. Statist. Soc. 88, 1-58.