ON THE MACRO-DYNAMIC STOCHASTIC TREATMENT OF THE SIZE
AND AGE STRUCTURE OF A HUMAN POPULATION

By

Nader Abd-Elmaksoud Fergany

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 666

February 1970
NADER ARD-ELMAKSOUD FERGANY. On the Macro-Dynamic Stochastic Treatment of the Size and Age Structure of a Human Population. (Under the direction of H. B. WELLS.)

Macro-dynamic stochastic formulations of the size and age structure of a human population are considered. No restrictive assumptions are placed on the physical process of population change and demographic variables are assumed to have the simple stochastic structure of a constant depending on time plus a stationary random error component.

Probabilistic characterizations are obtained for the population size at a point in time under discrete and continuous probability structures, and the probability distribution of the number of individuals in two different age groups at a certain point in time is described by its lower moments in general and in the special case of an underlying multivariate normal distribution. The theory developed is illustrated using data for the population of Egypt.

The problem of estimation of a certain function of joint moments of a multivariate distribution using a sample of correlated observations is discussed briefly. It is shown that the usual estimators in the case of a sample of independent observations are biased and that it is possible to arrive at unbiased estimators in the case of \( m \)-dependent processes. The complexity of the variances of estimators considered prevented a more conclusive study based on the mean square error criterion. A Monte Carlo study of a special case in small samples indicates that the degree to which an estimator uses the sample information is the major factor in determining how well the estimator performs in terms of mean square error.
ON THE MACRO-DYNAMIC STOCHASTIC TREATMENT OF THE SIZE AND AGE STRUCTURE OF A HUMAN POPULATION

by

Nader Abd-Elmaksoud Fergany

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics

Chapel Hill

1970

Approved by:

[Signature]

Adviser
BIOGRAPHY

The author was born August 20, 1944, in Gieza, U.A.R. He graduated from the Saidia high school in Gieza in 1959.

He received the Bachelor of Science degree with a major in applied statistics from the Faculty of Economics, Cairo University in 1963. He then joined the Faculty as an instructor. Since October, 1965 he has been on leave from his position pursuing graduate studies in statistics at the University of North Carolina.

The author is married to Moushira Khattab of Cairo, U.A.R.
ACKNOWLEDGMENTS

The author expresses his appreciation to members of his advisory committee, Professors A. Hawley, D. Quade, M. Sheps, H. Van Der Vaart and H. B. Wells.

Professor Wells endured the arduous task of advising the author for more than three years and all members of the Committee gave valuable assistance and counsel.

The financial support of the Population Council and the Carolina Population Center is gratefully acknowledged.

Mrs. Gay Goss typed the dissertation beautifully and efficiently.

Finally, it gives the author great pleasure to acknowledge the significant contributions of the many wonderful people who were associated with the Faculty of Economics at Cairo University during his years of study and work there to his education and edification.
TABLE OF CONTENTS

LIST OF TABLES .......................................................... vi
LIST OF FIGURES ........................................................... vii

Chapter

I. INTRODUCTION AND REVIEW OF LITERATURS ....................... 1

1.1. Introduction ....................................................... 1
1.2. Review of Macro-Analytic Models in Demography ............ 2
  1.2.1. Deterministic Models ........................................ 2
  1.2.2. Stochastic Models ........................................... 8
1.3. Objectives of the Present Study and Approach
    Adopted ............................................................... 13

II. POPULATION SIZE .................................................... 15

2.1. Theoretical Formulation, Discrete Probability
    Structure ............................................................. 15
2.2. A Discrete Error Probability Law ............................. 20
2.3. Theoretical Formulation, Continuous
    Probability Structure ................................................ 27
2.4. An Example ........................................................ 30

III. AGE STRUCTURE .................................................... 37

3.1. Introduction ....................................................... 37
3.2. Statistical Formulation .......................................... 42
  3.2.1. Introduction .................................................. 42
  3.2.2. The Multivariate Normal Case ............................. 45
3.3. A Discrete Approximation to the Human
    Survivorship Function .............................................. 46
3.4. An Example ........................................................ 48

IV. SOME ASPECTS OF THE ESTIMATION PROBLEM ..................... 52

4.1. Estimation of a Certain Function of Joint
    Moments of a Multivariate Distribution Using
    the Same Function of the Joint Moments of a
    Sample of Dependent Observations .............................. 52
Chapter

4.2. A Monte Carlo Study of Some of the Properties of Five Estimators for the Variance of the Product of the Two Components of a Bivariate Normal Time Series in Small Samples. 60
4.2.1. Introduction 60
4.2.2. Procedure 63
4.2.3. Conclusions 66

V. SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH 69
5.1. Summary 69
5.2. Suggestions for Further Research 71

Appendices

A. DATA AND INTERMEDIATE CALCULATIONS FOR SECTION (3.4) 73
B. COMPUTER PROGRAM 78
C. NUMERICAL RESULTS OF THE MONTE CARLO STUDY 87
LIST OF REFERENCES 92
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1. Comparison of the Discrete Error and Symmetric Binomial Distributions for $k = 1,2,3$</td>
<td>26</td>
</tr>
<tr>
<td>2.4.1. Estimated Probability Distribution of the Population of Egypt in Thousands on 9/30/1960</td>
<td>35</td>
</tr>
<tr>
<td>3.4.1. Estimated Male Births in Thousands and Survival Probabilities</td>
<td>50</td>
</tr>
<tr>
<td>A-1. Birth Rate per 1000 Population Corrected for Under-Registration, Egypt (1940-1959)</td>
<td>73</td>
</tr>
<tr>
<td>A-2. Estimated Mid-Year Population in Thousands, Egypt (1940-1959)</td>
<td>74</td>
</tr>
<tr>
<td>A-3. Reported Death Rate per 1000 Males in the Age Group (0-4), Egypt (1940-1959)</td>
<td>75</td>
</tr>
<tr>
<td>A-4. Average Ratio of Single-Year Death Rates to the Death Rate of the Age Group (0-4), Egypt, 1937, 1947 and 1960 Life Tables</td>
<td>76</td>
</tr>
<tr>
<td>A-5. Ratio of Crude Death Rates Corrected for Under-Registration to Reported Rates, Egypt (1940-1959)</td>
<td>77</td>
</tr>
<tr>
<td>C-1. Relative Mean Square Error of the Five Estimators</td>
<td>87</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>2.4.1. Annual Rate of Natural Increase, Egypt, 1947-1959</td>
<td>34</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION AND REVIEW OF LITERATURE

1.1. Introduction

The use of mathematics as a medium of expression in a certain branch of human knowledge is considered by some to be a sign of the dawn of an exact science growing out of that branch. Viewed in that light, the increased use of mathematical models in demography in recent years - as means of both enriching the theory and deriving sound and workable methodologies - is a definite manifestation of the advent of an era of scientific sophistication in this area.

That consideration of stochastic variations in a model improves its resemblance to reality and thus increases its utility is indicated by Mindel C. Sheps [27] in the special case of the study of patterns of human reproduction: "While mean or expected values given by probabilistic models agree with those of deterministic models based on sufficiently similar assumptions, stochastic models offer insight into the effects of chance and into the variations that occur naturally in the biological system." However, we must not lose sight of the fact that the role played by pure chance variations in the problems of human population change tends to be small since the number of individuals considered is usually very large.

Unfortunately, each of the existing macro-analytic stochastic models of population change suffers one or more of the following short-comings:
1. over-simplified assumptions, regarding both the physical process of population change and the stochastic structure associated with it, that widen the gap between reality and the model that abstracts it,

2. complicated mathematical derivations,

3. difficulty of practical applications; gigantic computational efforts are usually required to produce relatively detailed useful results,

4. lack of concern for the problems involved in the estimation of needed parameters from actual data which are both important and challenging.

1.2. Review of Macro-Analytic Models in Demography

1.2.1. Deterministic models. Macro-analytic models of population change are formulated in terms of either discrete time and age variables or continuous time and age variables.

The classic of discrete time deterministic formulations of the problem is that of Leslie [19] and we shall review its basic structure briefly because it serves as the starting point of almost all age-specific discrete time stochastic treatments of the problem of population change.

Leslie considered only the female population and assumed that age-specific rates are independent of time. He also made the following definitions:

\[ n_{xt} = \text{the number of females alive in the age group (} x, x+1) \text{ at time } t, \]

\[ p_x = \text{the probability that a female in the age group (} x, x+1) \]
at time $t$ will be alive in the age group $(x+1, x+2)$

at time $t+1$, $1 \geq P_x \geq 0$,

$F_x$ = the number of daughters born in the interval $(t, t+1)$

per female alive aged $(x, x+1]$ at time $t$, who will be

alive in the age group $(0,1]$ at time $t+1$, $F_x \geq 0$,

and

$$
M = \begin{bmatrix}
F_0 & F_1 & F_2 & \cdots & F_{m-1} & F_m \\
P_0 & 0 & 0 & \cdots & 0 & 0 \\
0 & P_1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & P_{m-1} & 0 \\
\end{bmatrix}
$$

where $(m, m+1]$ is the last age group attainable in the population.

It then follows that

$$
n_{t+1} = M_n_t
$$

where $n_t$ is an $(m+1)$ column vector giving the female age distribution

at time $t$, or

$$
n_t = M^t n_0
$$

where $t = 0$ stands for some initial time period.

Leslie then notes that if $(k, k+1]$ is the last age group in which

reproduction occurs, $F_x = 0$ for $x > k$ and $M$ may be partitioned symmet-

rically at this point as

$$
M = \begin{bmatrix}
A & 0 \\
- & - \\
B & C \\
\end{bmatrix}
$$

whence
\[
N^t = \begin{bmatrix}
- & A^t & - & 0 \\
- & f(A, B, C) & - & C^t \\
\end{bmatrix}
\]

But, since \( C^t = 0 \) for \( t \geq m - k \), he concluded that it is the submatrix \( A \) which is of primary importance and gave his mathematical treatment in terms of it.

The principal continuous time deterministic formulation is due to Alfred J. Lotka [21], [22], [23]. Since it will serve as the basis for our stochastic treatment, some aspects of Lotka's theory that are relevant to this treatment will be reviewed here.

Let \( x \) denote age, \( x \geq 0 \),

\( t \) denote time,

\( N(t) \) be the number of individuals in the population at time \( t \),

\( c(x, t) \) be the proportional age distribution function at time \( t \);

i.e., \( \int_{a_1}^{a_2} c(x, t) \, dx = \) the proportion of individuals in the age group \((a_1, a_2]\) at time \( t \) and \( \int_0^\infty c(x, t) \, dx = 1 \),

\( B(t) \) be the number of births in the population at time \( t \) and \( b(t) = B(t)/N(t) \) be the corresponding birth rate,

\( D(t) \) be the number of deaths in the population at time \( t \) and \( d(t) = D(t)/N(t) \) be the corresponding death rate, and

\( p(x, t) \) be the probability that an individual born at time \( t-x \) will survive to be of age \( x \) at time \( t \).

Considering a closed population, \(^1\) and using the terminology

\(^1\)A closed population is one that increases only through birth and decreases only through death.
presented above, Lotka derived the following relations

\[ N(t) = \int_0^\infty B(t-x) p(x,t) \, dx \quad \ldots (1.2.1.3) \]

\[ c(x,t) = B(t-x) p(x,t) / N(t) \quad \ldots (1.2.1.4) \]

He then remarked that if the survivorship function \( p(x,t) \) and the proportional age distribution function \( c(x,t) \) are mathematically independent of time, both the birth and death rates of the population will be constant and so will be the rate of change in the size of the population. If the birth and growth rates are denoted by \( b \) and \( r \) respectively, it follows that

\[ N(t) = N(t_0) e^{rt} \quad B(t) = B(t_0) e^{rt} \quad D(t) = D(t_0) e^{rt} \quad \ldots (1.2.1.5) \]

where \( t_0 \) is some initial time point. It also follows, from relation (1.2.1.4) above, that

\[ c(x) = b p(x) e^{-rx} \quad \ldots (1.2.1.6) \]

We observe that all functional forms of the age structure presented above depend fundamentally on the survivorship function \( p(x,t) \).²

In the following discussion, we shall be concerned with one generation - real or synthetic - survivorship function, or a \( p(x,t) \) that is constant.

²It should be noted that \( p(x,t) \) is rather complex. It is not a "generation" survivorship function in the sense of a pure death process describing the mortality experience of a cohort of births; nor is it a "period" survivorship function describing the age-specific mortality experience of a certain population at some point in time applied to a synthetic cohort of births. Actually, \( p(x,t) \) at time \( t \) describes segments of the mortality experience of as many birth cohorts as there are age points at time \( t \), with the mortality experience of younger ages represented more often than that of older ages.
over time; hence the argument $t$ will not be displayed.

In addition to the prominent role the survivorship function plays in determining the age structure, it determines a life table completely. For example, the life expectancy at age $x$, $e_x$ say, is given by the expression

$$e_x = \left( \int_{x}^{\infty} p(z) \, dz \right)/p(x) \quad \ldots \quad (1.2.1.7)$$

Consider a pure death process in which a cohort of $N(0)$ births goes through life until the last member of the cohort dies. Let $N(x)$ denote the size of the cohort at age $x$ (in such a process, age and time since the start of the process coincide, of course) then the survivorship function is given by $p(x) = \frac{N(x)}{N(0)}$.

If the instantaneous death rate at age $x$, $\mu(x)$, is a constant, $\mu$ say, then

$$\frac{d \, p(x)}{dx} = -\mu \, p(x)$$

which gives the well-known negative exponential relation

$$p(x) = e^{-\mu x} \quad \ldots \quad (1.2.1.8)$$

on using the initial condition $p(0) = 1$. [7] The last relation defines what might be called a constant-attrition survivorship function.

However, a survivorship function that describes human mortality conditions cannot be expected to follow the constant-attrition survivorship function, for the risk of human mortality is a function of age. The most general approach to a human survivorship function is to consider the probability of death as a function of age, $\mu(x)$, in which case
we have
\[ \frac{d}{dx} p(x) = -\mu(x) p(x) \] giving - on utilizing the initial condition \( p(0) = 1, \)
\[ p(x) = \exp \left( - \int_{0}^{x} \mu(\xi) \, d\xi \right) \quad \ldots (1.2.1.9) \]
It remains, of course, to determine the mathematical form of \( \mu(x) \).

The earliest attempts to find a general mathematical expression that describes the human survivorship function aimed at it directly; \( \text{i.e.}, \) considered \( p(x) \) or \( N(x) \) rather than \( \mu(x) \). De Moivre conjectured in 1725 that \( p(x) \) decreases in an arithmetic progression. In 1825, Gompertz introduced his famous "Laws of Mortality." [8] \( \mu(x) = \beta x^\gamma \)
which was later modified by Makeham in 1860 to \( \mu(x) = \alpha + \beta e^\gamma \). [7]
However, this law was not applicable at all for ages before maturity. Later developments - surveyed by Préchet [13] - started with \( \mu(x) \) or a variation thereof.

Two remarks are in order here. First, most later attempts did not aim at a law of mortality that is based on a general theory regarding the underlying biological process but rather at fitting mathematical expressions to observed age specific probabilities of death. This resulted in expressions that were too complicated formally and/or too specific to yield the analytic advantages that would result from a general mathematical expression. Second, although the reasons are clear - namely, the complexity of a general expression that describes human survivorship for all ages - the shift of the attention of researchers in this area from \( p(x) \) to \( \mu(x) \) and from starting with an
underlying theory to fitting appropriate functional forms to observed μ(x) values is unfortunate and represents a loss in the degree of generality of the resulting expressions. For not much faith can be placed in an empirical fitting that is not backed by an acceptable theory; and - clearly - if μ(x) represents the probability of death in an interval Δx starting at age x, then μ(x) is a monotonically non-decreasing function of Δx and the fitted form - although it may not change basically - will depend on the age units used in the fitting process.

More recently, typical patterns of mortality have been described through the presentation of series of "model" life tables. [8]

1.2.2. Stochastic Models. Although mathematical formulations exist for classical birth and death processes that give the size of the population as a function of time, the systems of equations resulting when birth and death rates are allowed to be general functions of time are not easy to solve and practical applications are cumbersome. [3], [12]

The first age-specific stochastic model of population change is ascribed to Bartlett. [4] Bartlett used discrete time and discrete age groups. He considered females only, assumed fertility and mortality rates to be constant over time and within each age group, and derived asymptotic forms for the linear and quadratic moments of the number of females in an age group when the age groups and time intervals were both made small.

Three years later, Kendall [17] discussed an age-specific continuous time stochastic model of population change. Kendall formulated
the problem using \( N(x,t) \) to describe the state of the population at time \( t \) in the sense that the Stieltjes integral
\[
\int_{x_1}^{x_2} dN(x,t)
\]
the number of individuals in the age group \( (x_1, x_2) \). He admitted that distribution problems are extremely difficult but found it possible to arrive at some of the moments.

Under the following assumptions:

1. one sex
2. the subpopulations generated by two co-existing individuals develop in complete independence of each other
3. the birth rate varies with age but not with time
4. the death rate does not vary, either with age or with time, Kendall developed equations for

\[ E(dN(x,t)), \]
\[ Var(dN(x,t)), \]
\[ Cov(dN(x,t), dN(y,t)). \]

The equations are difficult to solve, except in some unrealistic special cases. For example, in his paper, Kendall solved the case in which both the birth and death rates are independent of age and time. Simpler methods exist for solving this problem now. [12]

More recently, age specific stochastic models of population change were formulated in terms of discrete time and corresponding discrete age grouping. Two formulations are worth mentioning here. The first, due to Pollard [24], revived interest in this type of model and the second, due to Sykes [28], contains one of the most general treatments of the problem available.
Pollard considered females only, restricted his age distribution to prereproductive and reproductive ages and assumed that

1. fertility rates and survival probabilities in Leslie's formulation represent binomial probabilities that are independent of time,

2. a certain individual's death or procreation is independent of other individuals' death or procreation.

Then using Leslie's notation, he derived an expression for

\[ e_t \] the vector of expected number of females in the \((k+1)\) age groups \((0,1], (1,2], \ldots, (k,k+1] \) at time \(t\) and

\[ \mathbf{c}(t) \] the vector that has as its elements the variances and covariances of the number of females in the \(i\)th and \(j\)th age groups at time \(t\), \(c_{ij}(t)\) say, listed in a dictionary order according to the subscripts \(i\) and \(j\) (for \(i \neq j\), \(c_{ij}(t)\) and \(c_{ji}(t)\) are both listed).

This expression is

\[
\begin{pmatrix}
\mathbf{e}_t \\
\mathbf{c}(n)
\end{pmatrix} = \begin{pmatrix}
\mathbf{e}_0 \\
\mathbf{c}(0)
\end{pmatrix} \begin{pmatrix}
\mathbf{A}^n \\
\mathbf{A} \otimes \mathbf{A}^n \\
\mathbf{A} \otimes \mathbf{A}^n \\
\ldots
\end{pmatrix} + \sum_{j=1}^{n-j} (\mathbf{A} \otimes \mathbf{A})^{n-j} B \mathbf{e}_{j-1} \]

\[
\ldots (1.2.2.1)
\]

where \(B\) is a \((k+1)^2 \times (k+1)^2\) matrix whose elements are functions of \(F_x\) and \(P_x, x = 0, 1, 2, \ldots, k\), and \(\mathbf{A} \otimes \mathbf{A}\) is the direct product of the matrix \(\mathbf{A}\) by itself. These results were generalized to higher order moments of \(n_t\).
It is in order here to point to the tremendous computational effort involved in applying (1.2.2.1) for a moderate number of age groups, if we decide that the simplifying assumption on which the above results are based are tolerable.

Sykes also restricted attention to \((k+1)\) age groups at or below the limiting age of reproduction and defined stochastic versions of Leslie's model by assuming respectively that:

1. the deterministic model is subject to additive random errors,

2. the elements of the transition matrix \(A\) are binomial probabilities,

3. the transition matrices are random variables.

If \(A_t\) is the Leslie transition matrix for the time interval \((t,t+1)\), Sykes shows that in all three cases considered

\[
E(n_t) = \left( \sum_{i=0}^{t-1} A_i \right) n_0 = n_t \quad \ldots (1.2.2.2)
\]

We give below some of his further results together with the assumptions required for them to hold in the three above mentioned cases.

\[
\hat{n}_{t+1} = A_t \hat{n}_t + \xi_t \quad \ldots (1.2.2.3)
\]

where \(\{\xi_t\}\) is a sequence of random vectors with

\[E(\xi_t) = 0\text{ and }\text{Cov} (\xi_s, \xi_t) = \Gamma_{s,t} \quad s, t = 0, 1, 2, \ldots\]

whence

\[\text{Var} (n_t) = \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \left( \sum_{k=t-j}^{t-i-1} \Gamma_{t-i-1, t-j-1} \left( \sum_{k=t-j}^{t-i-1} A_k \right)' \right) \Gamma_{t-i-1, t-j-1} (\pi^{t-1} A_k) \quad \ldots (1.2.2.4)\]
Further, if \( \{e_t\} \) is a normal stochastic process,\(^3\) then \( n_t \) is distributed as a multivariate normal with mean \( \mu_t \) and variance \( V_t \). Hence

\[
Q_t = (n_t - \mu_t)' V_t^{-1} (n_t - \mu_t) \text{ is distributed as } \chi_{k+1}^2
\]

\[\ldots (1.2.2.5)\]

which may be used to place confidence intervals on \( \mu_t \).

(2) Assume that individuals in the \( i \)th age group give birth and die with probabilities \( t^{b_i} \) and \( (1 - t^{s_i}) \) respectively at time \( t \) and that all events result from mutually independent binomial trials. This is similar to Pollard's formulation except that Sykes allows the transition matrix to depend on time. The variance matrix of \( n_t \) is given by the relation

\[
V_t = \sum_{i=0}^{t-1} (\pi^{t-1} A_k) \, D_{t-i-1} \, (\pi^{t-1} A_k)', \quad \ldots (1.2.2.6)
\]

where

\[
D_t = \text{Diag} (W_t, \mu_t) \quad \text{and if } A_t = (a_{ij}) \text{ then}
\]

\[
W_t = (w_{ij}) = (a_{ij} (1 - t^{a_{ij}}))
\]

(3) \[n_{t+1} = (A_t + \Lambda_t) n_t \quad \ldots (1.2.2.7)\]

where \( A_t \) is a known Leslie matrix and \( \{\Lambda_t\} \) is a sequence of independent matrix random variables with \( E (\Lambda_t) = 0 \) and \( \text{Var} (\Lambda_t) = S_t \). In this case, Sykes gave the variance matrix of \( n_t \) as

\[^3\text{i.e., for any finite subset} \, (t_1, t_2, \ldots, t_n) \, \text{of indices in the sequence of nonnegative integers, the vector} \, \xi' = (\xi_{t_1}', \xi_{t_2}', \ldots, \xi_{t_n}') \, \text{has the} \, n(k+1) - \text{variate normal distribution.}\]
\[ V_t = \sum_{\ell=0}^{t-1} \left( \prod_{i=0}^{t-\ell-1} A_i \right) P_\ell \left( \prod_{i=0}^{t-\ell-1} A_i \right)' \]...(1.2.2.8)

where

\[ P_\ell \text{ is a } (k+1) \times (k+1) \text{ matrix whose } i^j \text{th element is given by} \]

\[ \chi_{ij}^\ell = \sum_r \sum_s \text{Cov}[\chi_{ir}^{\delta}, \chi_{js}^{\delta}] (\chi_{rs}^{\nu} + \chi_{rs}^{\mu}) \]

and

\[ \Lambda_t = (\delta_{ij}), \ V_t = (v_{rs}) \text{ and } \mu_t = (\mu_r). \]

Again, we can easily see that, even if we accept the various restrictive assumptions of the above formulations, it is very difficult to put their results to reasonably detailed practical applications.

As a case in point, in the example Sykes gave in his paper, only three age groups were considered and the transition matrix did not vary with time.

1.3. Objectives of the Present Study and Approach Adopted

The purpose of this study is to investigate some aspects of the mode of change in the size of a human population and its age structure under simple and realistic stochastic set-ups with the hope of producing practical methodologies for the demographer who wishes to incorporate stochastic variations in his calculations. The line of development of the stochastic formulations employed, which is a stochastic extension of some of Alfred Lotka's work [23], is not directly related to any of the existing macro-analytic stochastic models and the resulting formulations do not suffer from most of the shortcomings of the previous ones summarized in (1.1).

There is some conceptual difficulty in a stochastic treatment of
human population change in a certain society since in reality we have only one such society and it is difficult to imagine how that society could be subject to chance variations. However, we may consider the observed pattern of population change as a sample of one observation from a universe of possible patterns. This is the approach adopted here.

The unit of analysis considered in this treatment is a human population possessing certain characteristics and subject to vital change continually affecting the number and characteristics of its members. This vital change is effected through two processes, incrementation by the birth of new individuals and immigration, and decrementation by the death of living individuals and emigration. Such a population may be described by three types of random variables expressing forces of incrementation, decrementation, and the resulting population as a function of time. This way of describing the dynamics of human population change is most natural and could be considered as an attempt to theorize the underlying concepts of population change in classical (numerical) demographic analysis.

In the present formulation, demographic variables are assumed to have the stochastic structure of a constant depending on time plus a stationary random error component. Although the exposition of the present formulation is set forth entirely in terms of human populations, the principles involved are applicable to all other types of populations, and so are the results.
CHAPTER II

POPULATION SIZE

2.1. Theoretical Formulation, Discrete Probability Structure

Consider the probability space \((\Omega, \mathcal{P}, P)\) where \(\Omega = \{\omega_1, \omega_2, \ldots\}\)

is a set of a countable number of elements, \(\mathcal{P}\) is the power set of \(\Omega\)
and \(P\) is a probability measure on \(\mathcal{P}\).

Elements of \(\Omega\) are identified here with conglomerations of relative socio-cultural conditions - in the broadest sense of the word - whose effect on the factors affecting change in demographic variables is to introduce stochastic variations as a component of their total variation. Although the theory presented here does not depend on the nature of the probability measure \(P\), stochastic variations are treated as normal-type\(^1\) discrete errors around an expected value determined by overall socio-cultural change expressed in terms of time trends in demographic variables. Accordingly, elements of \(\Omega\) are conceived of as having a one-to-one correspondence with the class of integers - or a symmetric subset of it - such that they could be arranged as

\[ \Omega = \{\ldots, \omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, \ldots\} \]

where \(\omega_i\), \(i = 0, \pm 1, \pm 2, \ldots\),

denotes an \(i\)-unit deviation from the expected value; and the probability measure \(P\) is conceived of as assigning probabilities to elements

\(^1\)i.e., symmetric around the expected value which is also the mode of the distribution.
of \( P \) in the fashion of an error probability distribution about some expected value.

It is probably more realistic to think of relative socio-cultural conditions as forming a continuum, but the case where \( \Omega \) is countable has a simplified probability structure and for this reason is presented first. A continuous probability structure version will be considered in (2.3) below.

Let \( N_d(\tau, \cdot) : \Omega \rightarrow \mathcal{N} \) where \( 0 < \tau \leq t \) and \( \mathcal{N} \) is the set of non-negative integers such that \( N_d(\tau, \omega) = n_d(\tau) \) is the size of the population at time \( \tau \) corresponding to \( \omega \in \Omega \), which is a discrete function of time. For the sake of mathematical convenience, we shall approximate \( N_d(\tau, \cdot) \) by the differentiable function of time \( N(\tau, \cdot) \).

Let \( R(\tau, \cdot) : \Omega \rightarrow \Lambda \), where \( 0 < \tau \leq t \) and \( \Lambda \) is a countable set of points on a finite real interval such that \( R(\tau, \omega) = r_\omega(\tau) \) is the rate of change in the size of the population of time \( \tau \) corresponding to \( \omega \in \Omega \) defined by

\[
\frac{d n_\omega(\tau)}{d\tau} = r_\omega(\tau) n_\omega(\tau) \forall \omega \in \Omega \quad \ldots (2.1.1)
\]

such that \( R^{-1}(\tau, \cdot) \) exists.

From (2.1.1), it follows that

\[
n_\omega(t) = n_\omega(0) \exp \left( \int_0^t r_\omega(\nu) d\nu \right) \forall \omega \in \Omega \quad \ldots (2.1.2)
\]

where \( t = 0 \) stands for some initial time point.

In order to consider the probability structure while allowing the relative socio-cultural conditions to change with time we define

\[
\mathcal{F}(\cdot) : (0, t] \rightarrow \Omega
\]
such that $K(\tau) = \omega_\tau$ is the set of relative socio-cultural conditions existing at time $\tau \in (0,t]$. Let $L(\tau)$ be the set of all possible such functions $\omega_\tau$. We may now think of $R(\tau, \cdot) \in (0,t]$ as mapping $L(\tau)$ into a space of functions $\mathcal{F}(\tau, \cdot) \in (0,t]$ whose typical element $R(\tau, \omega_\tau) = r_{\omega_\tau}(\tau)$ represents the functional dependence of the rate of change in the population size on time and the path of relative socio-cultural conditions prevailing in the time interval $(0,t]$. $N(\tau, \cdot)$ may also be redefined in a similar manner. Note that $r_{\omega_\tau}(\tau)$ and $n_{\omega_\tau}(\tau)$ are "sample functions" in the ordinary sense of stochastic processes if we take our underlying sample space to be $\prod_{0<\tau<t} \Omega$.

Now, equation (2.1.2) may be written as

$$n^*(\tau) = n_{\omega_\tau}(\tau) = n_\omega(0) \exp \left( \int_{0}^{t} r_{\omega_\tau}(\tau) \, d\tau \right) \forall \omega_\tau \in L(\tau) \ldots (2.1.3)$$

for relation (2.1.1) holds for any path $\omega_\tau \in L(\tau) \in (0,t]$ with $\omega_\tau$ replacing $\omega$ at $\tau$. (Note that $r_{\omega_\tau}(\tau)$ is a function of $\tau$ only).

Clearly, the sample space to consider in discussing the probability structure of $r_{\omega_\tau}(\tau)$ and $n_{\omega_\tau}(\tau)$ is $L(\tau)$. However, any probability measure on $\mathcal{A}(L(\tau))$ - the $\sigma$-field on $L(\tau)$ - is of limited practical interest in the problem of human population change of one society since we would not be able to estimate it. All the data normally available represent only one element in $L(\tau)$.

A more useful probability structure results from considering equation (2.1.2) in conjunction with the probability measure $P$ on $P$ when it is assumed to remain stationary over time. Given the initial
population size at \( t = 0 \), the form of the time-trend in \( R(t, \cdot) \) - its expected value as a function of time - and the probability law \( P \), the probability distribution of \( N(t, \cdot) \) could be determined - or approximated - making use of remarks (1) and (2) below.

Remarks:

(1) Let \( X : \Omega \rightarrow \Lambda_1 \ni X(\omega) = x \),
\[ Y : \Lambda_1 \rightarrow \Lambda_2 \ni Y(x) = y \]
where \( \Omega, \Lambda_1, \) and \( \Lambda_2 \) are countable sets elements such that \( P(X^{-1}) \) and \( P(X^{-1} Y^{-1}) \) are defined.

In other words,
\[ X : (\Omega, \mathcal{A}, P) \rightarrow (\Lambda_1, \mathcal{A}_1, P X^{-1}), \]
\[ Y : (\Lambda_1, \mathcal{A}_1, P X^{-1}) \rightarrow (\Lambda_2, \mathcal{A}_2, P X^{-1} Y^{-1}) \]
where \( P \) is a probability measure on the \( \sigma \)-field \( \mathcal{A} \), \( (P X^{-1}) \) is the probability measure on \( \mathcal{A}_1 \) induced by \( X \), and \( (P X^{-1} Y^{-1}) \) is the probability measure on \( \mathcal{A}_2 \) induced by \( Y \).

The probability law of the random variable \( Y \) is determined completely by the probability law of \( X \), or by \( P \) on \( \mathcal{A} \), in a straightforward fashion. For, clearly,
\[
Pr\{Y = y\} = P X^{-1} Y^{-1} \{y\} \\
= P (X^{-1} Y^{-1} \{y\}) = P (X^{-1} \{x\}) \\
= P \{x\}
\]

(2) A discrete probability distribution may be regarded as an approximation to a continuous (or a finer discrete) one with the probability mass points of the approximation located at some well-defined
position within intervals of the continuous distribution (or groups of mass points for the finer discrete one) and carrying the probability measure associated with the corresponding intervals (or group of mass points).

(3) In practice, annual rates of natural increase would be the most reasonable estimates of rates of change in the population size for they are usually available on a yearly basis and hence would provide a sample of reasonable size over a relatively short period of time. But the rate of natural increase is not the same as the rate of change defined by (2.1.1).

For \( \omega \in \Omega \), let \( n(t+1) = n(t) \ e^r \). (Note that the subscript \( \omega \) is suppressed); i.e., the rate of population change is assumed to be constant throughout the time unit \( (t, t+1] \). Assuming a closed population, the rate of natural increase for the same time period is defined by

\[
r' = \frac{n(t+1) - n(t)}{n(t+1)} \quad \ldots (2.1.4)
\]

\[
= [n(t) \ e^r - n(t)]/[n(t) \ e^{r/2}]
\]

\[
= e^{r/2} - e^{-r/2}
\]

\[
= 2\left\{\left(\frac{r}{2}\right) + \frac{1}{3!} \left(\frac{r}{2}\right)^3 + \frac{1}{5!} \left(\frac{r}{2}\right)^5 + \ldots\right\}
\]

\[
= r + 2\left\{\frac{1}{3!} \left(\frac{r}{2}\right)^3 + \frac{1}{5!} \left(\frac{r}{2}\right)^5 + \ldots\right\} \quad \ldots (2.1.5)
\]

i.e. \( r' > r \) for \( r > 0 \)

and \( r' < r \) for \( r < 0 \)
and

\[
\begin{align*}
    r' &= 2 \sinh\left(\frac{X}{2}\right) \\
    r &= 2 \sinh^{-1}\left(\frac{r'}{2}\right) \\
    &= 2 \ln\left(\frac{r'}{2} + \sqrt{\frac{r'^2}{4} + 1}\right)
\end{align*}
\] 

.(2.1.6)

The absolute difference between \( r \) and \( r' \) is very small in the customary range of values for these variables, but when we consider the application of these rates to relatively large populations for a large number of intervals such that \( r \) is constant within each interval but may vary from one to another, we might do well to make the necessary adjustment. Tables of hyperbolic functions would be of use in numerical conversions. \( ^[14] \)

If the population is open to migration, the above-mentioned relations hold with the only difference that \( r' \) represents the balance of natural increase and net migration.

2.2. A Discrete Error Probability Law

From the properties of the probability measure \( P \) described at the beginning of (2.1), it seems natural that the probability distribution of a variable like \( R(T,*) \) at a certain point in time would take the form of an error probability distribution around an expected value determined by the time trend in \( R(T,*) \).

An appropriately-labeled normal distribution or a binomial distribution with probability parameter \( \frac{1}{2} \) would serve this purpose. However, the data we usually encounter in demographic estimation are not commensurate with high degrees of accuracy in the sense of probability
mass concentration around the mean, which makes an alternative to the normal distribution in this regard desirable. The same observation applies to the symmetric binomial distribution.

The purpose of this section is, taking the normal distribution as a standard, to devise a simple discrete error probability distribution that does not display a high degree of accuracy in the sense described above. It is well known that the sum of independent continuous uniform random variables approaches the normal distribution as the number of components increases, hence it was reasonable to expect that the sum of independent discrete uniform random variables would approach the form of a discrete error distribution as the number of components increases. The probability mass function (p.m.f.) of the sum of three such random variables was judged to achieve a balance between simplicity and closeness to the desired form. This p.m.f. has, in addition, the advantage of simple calculations of probabilities and the disadvantage of a restricted number of mass points.

Consider the probability distribution

\[
\Pr\{X = x\} = \frac{1}{2k+1} \quad x = -k, -k + 1, \ldots, 0, \ldots, k - 1, k
\]

\[\ldots(2.2.1)\]

Lemma (2.2.1)

It follows from the relation (2.2.1) that the probability generating function (p.g.f.) of \(X\) is

\[
G_X(s) = E(s^X) = \frac{s^{-k}}{2k+1} \left( \sum_{x=0}^{2k} s^x \right)
\]

\[\ldots(2.2.2)\]
clearly, \( E(X) = 0 \) and
\[
\text{Var}(X) = \frac{k^2 + k}{3}
\]
may be calculated directly or as \( C_X''(1) \).

Lemma (2.2.2)

Let \( X_1, X_2 \) be independently distributed according to (2.2.1),
\[
Y = X_1 + X_2, \text{ then}
\]
\[
\Pr(Y = y) = \frac{-|y| + 2k + 1}{(2k + 1)^2} \quad y = -2k, -2k+1, \ldots, 0, \ldots, 2k-1, 2k
\]

The proof is immediate from first principles. This gives the p.m.f. of a discrete triangular distribution and may be used as the building block for the next step.

Theorem (2.2.1)

Let \( X_1, X_2, X_3 \) be independently distributed according to (2.2.1),
\[
Z = X_1 + X_2 + X_3
\]

\[
m_1 = -|z| + 3k + 1, \quad \text{and}
\]
\[
m_2 = -|z| + k, \text{ then}
\]
\[
\Pr(Z = z) = \phi(z)/(2k+1)^3 \quad \text{where}
\]
\[
\phi(z) = \begin{cases} 
  m_1(m_1 + 1)/2, & |z| = k, k+1, \ldots, 3k \\
  m_1(m_1 + 1)/2 - 3m_2(m_2 + 1)/2, & |z| = 1, 2, \ldots, k-1 \\
  3k^2 + 3k + 1 & z = 0
\end{cases}
\]

\[\ldots(2.2.4)\]
Proof

Let \( \phi(z) \) = number of ways in which \( x_1 + x_2 + x_3 = z \), from first principles – using Lemma (2.2.2), or expanding the p.g.f. for \( z \),

\[
G_z(s) = (G_X(s))^3 = \left( \frac{s}{2k+1} \right)^3 \sum_{x=0}^{\infty} s^x
\]

\[
= \left( \frac{s}{2k+1} \right)^3 \left( \sum_{x=0}^{\infty} (x+1) s^x + \sum_{x=2k+1}^{\infty} (4k-x+1) s^x \right) \sum_{x=0}^{\infty} s^x
\]

we get the following table

| \( |z| \) | \( \phi(z) \) |
|------|----------------|
| 0    | \((k+1) + (k+2) + \ldots + (2k+1) + \ldots + (k+2) + (k+1)\) |
| 1    | \(k + (k+1) + \ldots + (2k+1) + \ldots + (k+3) + (k+2)\) |
| 2    | \((k-1) + k + \ldots + (2k+1) + \ldots + (k+4) + (k+3)\) |
| \vdots | \vdots |
| k-2  | \(3 + 4 + \ldots + (2k+1) + (2k) + (2k-1)\) |
| k-1  | \(2 + 3 + \ldots + (2k) + (2k+1) + (2k)\) |
| k    | \(1 + 2 + \ldots + (2k-1) + (2k)\) |
| k+1  | \(1 + 2 + \ldots + (2k-1) + (2k)\) |
| \vdots | \vdots |
| 3k-1 | 1 + 2 |
| 3k   | 1 |

Clearly, for \( Z = 0 \)

\[
\phi(0) = (2k+1) + (2k^2) + 2 \frac{k(k+1)}{2}
\]

\[
= 3k^2 + 3k + 1
\]
and for \( |z| = k, k+1, \ldots, 3k \)

\[
\phi(z) = m_1 (m_1+1)/2
\]

Finally, for \( |z| = 1, 2, \ldots, k-1 \), after some algebraic manipulation we get

\[
\phi(z) = m_1 (m_1+1)/2 - \{m_2 + (m_2+1) + \ldots + (m_2+m_2)\}
\]

\[
= m_1 (m_1+1)/2 - m_2 (m_2+1) - (1 + 2 + \ldots + m_2)
\]

\[
= m_1 (m_1+1)/2 - 3 m_2 (m_2+1)/2.
\]

Further,

\[
3k \sum \phi(z) = (3k^2+3k+1) + 2 \sum_{z=1}^{3k} \frac{(3k-z+1)(3k-z+2)}{2}
\]

\[
- 2 \sum_{z=1}^{k-1} \frac{3}{2} (k-z) (k-z+1)
\]

\[
= 8k^3 + 12k^2 + 6k + 1 = (2k+1)^3
\]

i.e., (2.2.4) is a proper p.m.f.

Q.E.D.

We note the following:

(1) For \( Z = 0 \),

\[
m_1 (m_1+1)/2 - 3 m_2 (m_2 + 1)/2 = 3k^2 + 3k + 1
\]

and hence \( \Pr(Z = z) \) may be thought of as composed of three second degree polynomials in \( Z \) as it is the case with the continuous uniform variates.

(2) It is clear, from the symmetry of the p.m.f. around the mean, that all the odd moments of \( Z \) vanish. Using the p.m.f. or the p.g.f. to evaluate the even moments directly is straightforward but tedious. However, it is easy to arrive at the variance as
\[ \text{Var} (Z) = \sum_{i=1}^{3} \text{Var} (\bar{x}_i) \text{ by virtue of independence} \]
\[ = k^2 + k \]

(3) The constant \(k\) determines both the number of probability mass points, \((6k+1)\), and the variance of the distribution

\[ \sqrt{\text{Var}(Z)} = \sigma_Z = \sqrt{k^2 + k} \]
\[ = k \sqrt{1 + \frac{1}{k}} \]

i.e., \(k < \sigma_Z < k + \frac{1}{2}\) since \(k^2 + k < (k + \frac{1}{2})^2\) which means that all the probability mass of the distribution of \(Z\) is concentrated in the range \((- 3\sigma_Z, 3\sigma_Z)\)

(5) The probability mass concentrated in the interval \((-\sigma_Z, \sigma_Z)\) cannot exceed

\[ P_c = (2k+1)^{-3} \left\{ (3k^2 + 3k + 1) + 2 \sum_{z=1}^{k-1} \left[ \frac{(3k-z+1)(3k-z+2) - 3(k-z)(k-z+1)}{2} \right] + \sum_{z=k+1}^{k+1} \frac{(3k-z+1)(3k-z+2)}{2} \right\} \]
\[ = \frac{16k^3 + 14k^2 - 2k - 3}{3(2k+1)^3} < \frac{2}{3} \]

...(2.2.5)

which means that the p.m.f. for \(Z\) allows for more uncertainty around the mean than the normal distribution.

Alternatively, we may compute the kurtosis of \(Z\)

\[ K = \frac{\text{E}(Z^4)}{(\text{E}(Z^2))^2} \]

\[ = \frac{13k^2 + 13k - 1}{5k(k+1)} \text{ which lies in the interval} \]
\[ [2.5, 2.6] \text{ compared to the normal kurtosis of 3.0.} \]
(6) Table 2.2.1 compares the probability masses associated with \((6k+1)\) mass points for the symmetric binomial and discrete error distributions for \(k = 1, 2, 3\) with the \((6k+1)\) mass points numbered 1, 2, ..., 6k+1 in order, showing that the discrete error distribution allows for more uncertainty than the symmetric binomial.

**TABLE 2.2.1**

**COMPARISON OF THE DISCRETE ERROR AND SYMMETRIC BINOMIAL DISTRIBUTIONS FOR \(k = 1, 2, 3\)**

<table>
<thead>
<tr>
<th>(k)</th>
<th>Mass Point</th>
<th>Probability Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Discrete Error</td>
</tr>
<tr>
<td>1</td>
<td>1 or 7</td>
<td>.037</td>
</tr>
<tr>
<td></td>
<td>2 or 6</td>
<td>.111</td>
</tr>
<tr>
<td></td>
<td>3 or 5</td>
<td>.222</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.259</td>
</tr>
<tr>
<td>2</td>
<td>1 or 13</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>2 or 12</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td>3 or 11</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td>4 or 10</td>
<td>.080</td>
</tr>
<tr>
<td></td>
<td>5 or 9</td>
<td>.120</td>
</tr>
<tr>
<td></td>
<td>6 or 8</td>
<td>.144</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>.152</td>
</tr>
<tr>
<td>3</td>
<td>1 or 19</td>
<td>.0029</td>
</tr>
<tr>
<td></td>
<td>2 or 18</td>
<td>.0087</td>
</tr>
<tr>
<td></td>
<td>3 or 17</td>
<td>.0175</td>
</tr>
<tr>
<td></td>
<td>4 or 16</td>
<td>.0292</td>
</tr>
<tr>
<td></td>
<td>5 or 15</td>
<td>.0437</td>
</tr>
<tr>
<td></td>
<td>6 or 14</td>
<td>.0612</td>
</tr>
<tr>
<td></td>
<td>7 or 13</td>
<td>.0816</td>
</tr>
<tr>
<td></td>
<td>8 or 12</td>
<td>.0962</td>
</tr>
<tr>
<td></td>
<td>9 or 11</td>
<td>.1050</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>.1079</td>
</tr>
</tbody>
</table>
2.3. **Theoretical Formulation, Continuous Probability Structure**

Consider the probability space \((\mathcal{R}, \mathcal{B}, P)\) where \(\mathcal{R}\) is the real line, \(\mathcal{B}\) is the Borel field on \(\mathcal{R}\), and \(P\) is an absolutely continuous probability measure on \(\mathcal{B}\).

Although the theory presented here does not depend on the nature of the probability measure \(P\), we think of points on \(\mathcal{R}\) as having the same physical meaning as elements of \(\Omega\) in (2.1) above with \(x \in \mathcal{R}\) expressing an \(x\)-unit deviation from the expected value of the random variable \(X\) defined by \(P\) on \(\mathcal{B}\) with probability density function (p.d.f.) \(p(x)\); and of \(P\) as an error-type absolutely continuous probability measure, the prime example being the normal probability measure on the Borel line.

For the purposes of this section we consider \(R(\tau, x, \tau \in (0, t), x \in \mathcal{R}\) to be a measurable stochastic process, and require that any two distribution functions \(F(x_{\tau_1}(x))\) and \(F(x_{\tau_2}(x))\) corresponding to \(R(\tau_1, x)\) and \(R(\tau_2, x)\), \(\tau_1, \tau_2 \in (0, t]\) differ only in location.

It is well known that if \(A \in (0, t]\) is a Lebesgue measurable parameter set and if \(\int_A E|R(\tau, x)|d\tau < \infty\) then - since the value of an absolutely convergent iterated integral is independent of the order of integration -

\[
\mathcal{E}\{ \int_A R(\tau, x) \, d\tau \} = \int_A \mathcal{E}[R(\tau, x)] \, d\tau, \quad [10]
\]

Hence, if we define \(V(t, x) = \int_0^t x(\tau) \, d\tau\), and if \(\int_0^t |E(R(\tau, x))|d\tau < \infty\),

we have

\[
\mathcal{E}[V(t, x)] = \int_0^t \mathcal{E}[R(\tau, x)]d\tau \quad \ldots (2.3.1)
\]
where, we may note, \( E(R(\tau, x)) = r_0(\tau) = \mu(\tau) \) say.

Also, we have

\[
\begin{align*}
\text{Var} \, (V(t, x)) &= E((V(t, x) - E(V(t, x)))^2) \\
&= E(\int_0^t (r_x(\tau) - r_0(\tau)) \, d\tau)^2 \\
&= E(\psi(x) \, t)^2
\end{align*}
\]

where \( \psi(x) \) is the difference between \( r_x(\tau) \) and \( r_0(\tau) \) which is constant for all \( \tau \in (0, t] \). Further, \( \psi(x) \) is symmetric around zero and has the simple form \( \psi(x) = \alpha x, \alpha > 0 \). Hence

\[ \text{Var} \, (V(t, \cdot)) = (ut \sigma_x^2) \] where \( \sigma_x^2 \) is the variance of the random variable \( X \) with p.d.f. \( p(x) \).

Now, let \( a(\tau) \) and \( b(\tau) \) have continuous derivatives in \((0, t] \), then

\[
\pi_t = \Pr\{a(t) \leq V(t, x) \leq b(t)\}
\]

\[
= \Pr\left\{ \int_0^t a'(\tau) \, d\tau \leq \int_0^t r_x(\tau) \, d\tau \leq \int_0^t b'(\tau) \, d\tau \right\}
\]

\[
= \Pr\{a'(\tau) \leq r_x(\tau) \leq b'(\tau) \quad \forall \tau \in (0, t]\}
\]

since, because of the parallelism of \( R(\tau, x_1) \) and \( R(\tau, x_2) \) \( x_1, x_2 \in \mathcal{R}, \) \( \tau \in (0, t] \) resulting from the properties of \( R(\tau, x) \) given earlier, \( a(t) \leq V(t, x) \leq b(t) \) holds if and only if \( a'(\tau) \leq r_x(\tau) \leq b'(\tau) \) \( \forall \tau \in (0, t]. \)

Further, if we choose

\[ a'(\tau) = \mu(\tau) - a \quad \text{and} \quad b'(\tau) = \mu(\tau) + b \quad \text{where} \quad a, b \in \mathcal{R} \),

\( a, b > 0 \), then
and because of the fact that \( F(r_{\tau_1}(x)) \) and \( F(r_{\tau_2}(x)) \) differ only in location, we find that

\[
\Pi_{b}^a(\tau) = \Pr\{a'(\tau) \leq r_x(\tau) \leq b'(\tau)\} \\
= \int_{a'(\tau)}^{b'(\tau)} dF(r_{\tau}(x)) \\
= \Pr\{-a \leq x \leq b\} = \int_{-a}^{b} p(x) \, dx \\
= \Pi_{b}^a
\]

does not depend on \( \tau \).

To establish that \( \Pi_{t} = \Pi_{b}^a \) we define the probability space \((R(\mathcal{R}, t), R(\mathcal{B}, t), \mathcal{P}(\mathcal{P}))\) by the following one-to-one correspondence relations.

\[ x \in \mathcal{R} \Rightarrow \{R(\tau, x) \forall \tau \in (0, t] \in R(\mathcal{R}, t) \] 

\[ B \in \mathcal{B} \Rightarrow \{R(\tau, x) \forall x \in B, \forall \tau \in (0, t] \in R(\mathcal{B}, t) \] 

\[ R(\mathcal{P})(R(\mathcal{B}, t)) = \Pr\{R(\tau, x) \forall x \in B, \forall \tau \in (0, t]\} \\
= \Pr\{R(\tau, x) : x \in B, \tau \in (0, t]\} \\
= \int_{B} p(x) \, dx 
\]

Finally it should be clear that the transformation \( V(t, x) \) preserves probability structure in the sense that

\[ \Pr\{V(t, x) : x \in B\} = \Pr\{R(\tau, x) : x \in B, \tau \in (0, t]\} \]

or, more specifically,
\[ \pi_t = \Pr[a(t) \leq V(t, x) \leq b(t)] = \pi_a \pi_b \] 

(2.3.3)

For example, if \( p(x) \) is the standard normal p.d.f., then \( R(t, \cdot) \) is distributed as normal \((r_0(t), \alpha^2)\).\(^2\) It then follows that \( V(t, \cdot) \) is distributed according to normal \((E(V(t, \cdot)), \alpha^2 t^2)\), and, finally, \( N(t, \cdot) \) is distributed lognormally with density

\[ f(n(t)) = \frac{1}{n(t)\sqrt{2\pi \alpha^2 t}} \exp\left\{ \frac{-\frac{1}{2\alpha^2} \left( \ln n(t) - \mu_v(t) - \ln n(0) \right)^2}{} \right\} \]

and

\[ \begin{align*}
\text{mean} &= n(0) \exp\left( \mu_v(t) + \frac{1}{2} \alpha^2 t^2 \right) \\
\text{variance} &= n^2(0) \exp\left( 2 \mu_v(t) + \alpha^2 t^2 \right) [e^{\alpha^2 t^2} - 1] \\
\text{median} &= n(0) \exp\left( \mu_v(t) \right) \\
\text{mode} &= n(0) \exp\left( \mu_v(t) - \alpha^2 t^2 \right)
\end{align*} \]

(2.3.4)

where \( \mu_v(t) = E(V(t, \cdot)) \).

Also, \( f \) is skewed to the right and has positive kurtosis compared to the normal distribution. \([1],[16],[18]\)

2.4. An Example

Data on the population of Egypt were used to illustrate some aspects of the previous theory.

The 1947 census gave the country's population as of the 31st of

\(^2\)We should actually use a standard normal variate truncated to a finite interval, e.g., \((-4,4)\) or \((-5,5)\), if \( R(t, \cdot) \) is to be considered finite.
March as 18967 thousands. This figure was considered to have suffered from an overall over-enumeration stimulated by a rationing census which was taken in 1945, and a sizeable under-enumeration of children less than ten years of age. A correction for over-enumeration carried out by El-Badry [11] resulted in an estimated population of 17907 thousands. Another correction for under-enumeration of children less than ten years of age - using the standard demographic technique of surviving estimated births in the ten years preceding the census date - inflated that estimate to 19619 thousands. This last figure was used here for illustration purposes.

Egypt's is virtually a closed population - i.e., external migration is a negligible component of population change - hence the annual rate of natural increase is a proper measure of population change. The series of rates of natural increase in the period (1947-1959) was taken to estimate the rate of population change defined by (2.1.1). The conversion of rates of natural increase to rate of change in accordance with remark (2) at the end of section (2.1) above was not deemed necessary since the difference did not exceed $1 \times 10^{-6}$ in any year (it is implied in this comparison, of course, that the rates of change remain constant for a whole year). El-Badry [11], gave the following series of rates of natural increase, adjusted for under-registration of births and deaths, per 1000 population.

<table>
<thead>
<tr>
<th>Year</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947</td>
<td>22.5</td>
</tr>
<tr>
<td>1948</td>
<td>21.3</td>
</tr>
<tr>
<td>1949</td>
<td>20.4</td>
</tr>
<tr>
<td>1950</td>
<td>23.8</td>
</tr>
<tr>
<td>1951</td>
<td>24.6</td>
</tr>
<tr>
<td>1952</td>
<td>27.0</td>
</tr>
<tr>
<td>1953</td>
<td>22.5</td>
</tr>
<tr>
<td>1954</td>
<td>24.4</td>
</tr>
<tr>
<td>1955</td>
<td>23.8</td>
</tr>
<tr>
<td>1956</td>
<td>23.7</td>
</tr>
<tr>
<td>1957</td>
<td>19.1</td>
</tr>
<tr>
<td>1958</td>
<td>23.3</td>
</tr>
<tr>
<td>1959</td>
<td>25.0</td>
</tr>
</tbody>
</table>
A polynomial in time was used to estimate the time trend in the rates of natural increase. The least squares method was employed to estimate the coefficients of this polynomial. An assumption needed for the optimum properties of least squares to hold, namely the independence of the stochastic elements of $R(\tau, \cdot)$ over time, would not, in general, be met by the data of the type considered here. These random variables are, in general, serially correlated. In this case, least squares estimators are still unbiased but not guaranteed to be minimum variance.

It was decided to approximate the time trend with a quadratic function of time. The estimated least squares function was

$$
E(\hat{R}(\tau, \cdot)) = 0.021531 + 0.0006466 \tau - 0.00004501 \tau^2
$$

with $\tau = 0$ at the 1947 census date; i.e., 1946.25 \hspace{1cm} \ldots (2.4.1)

The mean accounted for 99.3% of the total sum of squares. Although the linear and quadratic terms contributed only .025 and .079 respectively to the square of the multiple correlation coefficient, it was decided to keep them in the prediction equation. Higher degree polynomials were excluded from consideration because the resulting time patterns were judged to be inappropriate in terms of projecting the trend even to the near future; the cubic and quartic coefficients contributed .093 and .355 to $R^2$ respectively. Although the choice of the second degree polynomial is rather subjective, it was deemed to both improve the predictive value of the prediction equation and provide a more interesting example than the case of a constant trend.

We consider next the probability structure of the rate of growth of the population in the time period under consideration.

Under the discrete probability structure, a constant multiple of
a random variable that has a probability distribution of the type described in section (2.2) above was superposed on the variation of the growth rates that was not explained by the quadratic function of time. A 19-interval probability law, i.e., $k = 3$, was used, this only determines the degree of detail desired in the probability distribution.

The sample variance about the regression line, $\pi^2 = \sum_{i=1}^{13} u_i^2$ where $u_i$ is the $i^{th}$ residual, was taken to estimate the variance of the stochastic component of $R(\tau, \cdot) = \alpha^2 (3^2 + 3) = 12\alpha^2$; and the step size for the translated random variable $R(\tau, \cdot)$ was estimated by equating $\pi^2 = 45956 \times 10^{-10}$ to $12\alpha^2$ as $\hat{\alpha} = .000619$.

The fitted regression curve was taken as the center of a grid of 19 such parallel curves each .000619 units above or below the next in either direction (if there is one) with associated probability masses determined according to the probability law considered. It should be noted that it is implied here that the random mechanism introducing the stochastic variation around the mean does not change with time, although the mean itself does and that this probability structure amounts to approximating the structure resulting from considering $(P, r_\omega(\tau))$ over the time period $(0, t]$ by continuous grouping of elements of $\Omega$ into 19 groups in a manner consistent with the structural characteristics of $P$.

If we are willing to project the time pattern of growth rates and the estimated probability structure into the future, it could be used to determine an approximate probability distribution for future population estimates making use of remark (1) in (2.1). On the other hand, we could stipulate future changes in the functional form of the time
pattern of growth rates and retain the estimated probability structure. As an example of the first case, the population of Egypt at the time of the 1960 census was estimated using the 1947 adjusted census figure and the information provided by the rates of natural increase. Making use of the fact that the 1960 census is referred to the 30th of September, equation (2.4.1), and the above theory and discussion, the following results were obtained.

### TABLE 2.4.1

**ESTIMATED PROBABILITY DISTRIBUTION OF THE POPULATION OF EGYPT IN THOUSANDS ON 9/30/1960**

<table>
<thead>
<tr>
<th>Interval</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24774 - 24981</td>
<td>.002 915</td>
</tr>
<tr>
<td>24982 - 25190</td>
<td>.008 746</td>
</tr>
<tr>
<td>25191 - 25402</td>
<td>.017 493</td>
</tr>
<tr>
<td>25403 - 25615</td>
<td>.029 155</td>
</tr>
<tr>
<td>25616 - 25830</td>
<td>.043 732</td>
</tr>
<tr>
<td>25831 - 26047</td>
<td>.061 224</td>
</tr>
<tr>
<td>26048 - 26265</td>
<td>.081 633</td>
</tr>
<tr>
<td>26266 - 26486</td>
<td>.096 210</td>
</tr>
<tr>
<td>26487 - 26708</td>
<td>.104 956</td>
</tr>
<tr>
<td>26709 - 26932</td>
<td>.107 872</td>
</tr>
<tr>
<td>26933 - 27158</td>
<td>.104 956</td>
</tr>
<tr>
<td>27159 - 27386</td>
<td>.096 210</td>
</tr>
<tr>
<td>27387 - 27616</td>
<td>.081 633</td>
</tr>
<tr>
<td>27617 - 27847</td>
<td>.061 224</td>
</tr>
<tr>
<td>27848 - 28081</td>
<td>.043 732</td>
</tr>
<tr>
<td>28082 - 28317</td>
<td>.029 155</td>
</tr>
<tr>
<td>28318 - 28554</td>
<td>.017 493</td>
</tr>
<tr>
<td>28555 - 28794</td>
<td>.008 746</td>
</tr>
<tr>
<td>28795 - 29037</td>
<td>.002 915</td>
</tr>
</tbody>
</table>
with \[\text{mean} = 26832 \text{ thousands and}\]
\[\text{standard deviation} = 777 \text{ thousands}\]

Under the continuous probability formulation, the following results were obtained, in the normal case

\[\mu_\nu(13.5) = 0.312 676\]

\[E(N(13.5,\cdot)) = 26832 \text{ thousands}\]

\[\sqrt{\text{Var}(N(13.5,\cdot))} = 734 \text{ thousands}\]

Also, properties of the normal p.d.f. may be used to make approximate probability statements about \(V(t,\cdot)\) and, consequently, \(N(t,\cdot)\).
In our example,

\[\Pr(25312 \leq N(13.5,\cdot) \leq 28419) = 0.95\]

The 1960 census count was 25984 thousands. When corrected for under-enumeration of children less than ten years of age, the population figure rose to an estimated 26896 thousands.
CHAPTER III

AGE STRUCTURE

3.1. **Introduction**

Let $N(t)$ and $r(t)$ have the same definitions as $N(t,\cdot)$ and $R(t,\cdot)$ respectively in (2.1.1),

$B(t), b(t)$ and $c(x,t)$ have the same definitions as in Lotka's formulation in (1.2.1),

$N(x,t) = c(x,t) N(t),$

$m(x,t)$ be the birth rate for individuals of age $x$ at the time epoch $t$, and

$$p(x,t) = \frac{\text{number of individuals of age } x \text{ at time } t \text{ in the population}}{\text{number of births at time } (t-x) \text{ in the population}}$$

This generalized formulation of survivorship incorporates the balance of forces of migration affecting the population in addition to mortality conditions. In this case, $p(x,t)$ may, conceivably, exceed 1, and "generalized" death rates computed in such a situation may become negative. However, such oddities do not introduce much logical difficulty and it is in this generalized sense that death and survivorship will be considered here. The roots of this formulation are ascribed to Hyrenius. [29]

---

1All variables considered in this chapter are random variables. The explicit dependence of these functions on elements of the underlying sample space is suppressed in the interest of clarity because it does not serve a significant purpose.
Using the above notation, the following relations describe the age structure of the population at time $t$:

$$N(x,t) = B(t-x) \ p(x,t) \quad \ldots (3.1.1)$$

$$c(x,t) = B(t-x) \ p(x,t) / \int_{0}^{t} B(t-x) \ p(x,t) \ dx$$

$$= b(t-x) \ p(x,t) \ \exp\left(- \int_{t-x}^{t} r(\tau) \ d\tau\right) \quad \ldots (3.1.2)$$

making use of the relation $N(t) = N(0) \ \exp\left(\int_{0}^{t} r(\tau) \ d\tau\right)$.

Equation (3.1.1) is fundamentally simpler, conceptually easier and more amenable to statistical manipulation than (3.1.2) which is a generalization of the familiar formulation of Lotka and probably more advantageous from the demographic analysis point of view since for its purposes it is both easier and more meaningful to speak in terms of rates of occurrence rather than the number of occurrences. Of course, if the total age span possible at time $t$ is considered, then $c(x,t)$ is only a constant multiple of $N(x,t)$.

The last observation points out the major source of difficulty in determining the complete age structure of a population at a point in time under the present formulation if all the demographic variables are allowed to be general functions of time. If $\omega_{t}$ is the highest attainable age at time $t$, then to determine the complete age structure at this point in time, the history of birth and survival in the population under consideration must be available for the time period $((t-\omega_{t}), t]$. This ambitious requirement - as well as mathematical convenience - has led students of demography pursuing this formulation to consider special cases in which some of the demographic variables are
mathematically independent of time.

If \( p(x,t) \) and \( b(t) \) are mathematically independent of time in such a way that \( r(t) \) does not depend on time in the interval \((t-b-a),(t-a)\) \( a \geq 0, b \geq 0 \), then \( c(x,t) \) must also be mathematically independent of time for \( a \leq x \leq (a+b) \) and we have

\[
c(x,t) = c(x) = bp(x) e^{-rx} \quad a \leq x \leq (a+b)
\]

which is Alfred Lotka's celebrated equation. The same form may be shown to hold at \( t \to \infty \) if \( p(x,t) \) and \( m(x,t) \) do not depend on time. [21]

A slightly modified version of this formulation requires knowledge of the age distribution function at a point in time prior to \( t \), \( N(x,t-a), a > 0 \) say, and the course of birth and survival in the population in the intervening time period \((t-a), t] \) only in order to determine the complete age structure at time \( t \). For this formulation, define

\[
p^*(x,a,t) = \frac{\text{number of individuals of age } x \text{ at time } t \text{ in the population}}{\text{number of individuals of age } (x-a) \text{ at time } (t-a) \text{ in the population}}
\]

\[ x \geq a > 0 \]

and note that \( p^*(x,a,t) = p(x,t)/p(x-a,t-a) \)

and \( p^*(x,x,t) = p(x,t) \).

Then, we can write

\[
N(x,t) = N(x-a, t-a) p^*(x,a,t) \quad x \geq a \\
= B(t-x) p(x,t) \quad x < a
\]

(3.1.3)

The last relation is exactly the same as in the previous formulation and is written in the usual demographic practice as
= \left( \int_0^\infty m(\xi, t-x) N(\xi, t-x) \, d\xi \right) p(x,t).

This is the model of component projections which is clearly quite practical in a deterministic set-up. This is so because it does not require hard-to-get historical data. But, on the other hand, since estimates of the age structure are available only through rather infrequent surveys and censuses, the estimation of the underlying stochastic mechanism - in a stochastic treatment of the age structure - is severely hindered. Consequently, the distribution of $N(x,t)$ is usually given conditional on the initial $N(x,t-a)$. Data on birth and survival are available much more frequently.

In actual applications, irrespective of the formulation used, we are usually interested in age groups rather than age points; i.e., in the age distribution function as a set - (rather than a point) - function of age. Hence we consider

$$b^c N_c(t) = \left\{ \begin{array}{l}
\int_b^c N(x,t) \, dx \\
\int_b^c B(t-x) \, p(x,t) \, dx \quad \text{if } c \leq a \\
\int_b^c N(x-a, t-a) p^*(x,a,t) \, dx \quad \text{if } b \geq a
\end{array} \right\}$$

or

$$b^c N_c(t) = \left\{ \begin{array}{l}
\int_b^c B(t-x) \, dx \\
\int_b^c N(x-a, t-a) p^*(x,a,t) \, dx
\end{array} \right\}$$

Since all the functions involved in the above relations are always non-negative, then - by the second mean value theorem - we have

$$b^c N_c(t) = K_1 \left\{ \begin{array}{l}
\int_b^c B(t-x) \, dx \\
\int_b^c N(x-a, t-a) p^*(x,a,t) \, dx
\end{array} \right\}$$

.. .(3.1.4)

$$\text{or}$$

$$b^c N_c(t) = K_1 \left\{ \begin{array}{l}
\int_b^c B(t-x) \, dx \\
\int_b^c N(x-a, t-a) p^*(x,a,t) \, dx
\end{array} \right\}$$

.. .(3.1.5)
where
\[
\min_{b \leq x \leq c} p(x, t) \leq K_1 \leq \max_{b \leq x \leq c} p(x, t) \quad \text{and} \quad B \quad \text{is the number of births in the time interval } ((t-c), (t-b)); \text{ or}
\]
\[
b'_{c}(t) = K_2 \int_{b}^{c} N(x-a, t-a) \, dx \]
\[
= K_2 \frac{N(t-a)}{(b-a)(c-a)} \quad \ldots (3.1.6)
\]

where
\[
\min_{b \leq x \leq c} p^*(x, a, t) \leq K_2 \leq \max_{b \leq x \leq c} p^*(x, a, t) \quad \text{and} \quad \frac{N(t-a)}{(b-a)(c-a)}
\]
is the number of individuals in the age group \((b-a), (c-a)\) in the population at time \((t-a)\).

\(K_1\) and \(K_2\) are usually approximated by some sort of average value. For example, if survivorship conditions are constant in the time period \(((t-a), t] \) for birth cohorts of the time period \(((t-a-c), (t-a-b)] \) and survivorship probabilities from age \((x-a)\) at time \((t-a)\) to age \(x\) at time \(t\) are weighted by the cohort size at time \((t-a)\), we have
\[
K_2 = \int_{b}^{c} p^*(x, a) \frac{p(x-a)}{p(x-a)} \, dx / \int_{b}^{c} p(x-a) \, dx
\]
\[
= \int_{b}^{c} p(x) \, dx / \int_{b}^{c} p(x-a) \, dx
\]
\[
= \frac{L}{(c-b)} / \frac{L}{(c-b)(b-a)} \quad \ldots (3.1.7)
\]
where
\[
\frac{L}{n_x} = \int_{x}^{x+n} p(x) \, dx
\]

which is the regular demographic practice.
3.2. Statistical Formulation

3.2.1. Introduction

In this section we discuss the nature of the probability distribution of the age structure at a point in time.

For a fixed age point and a fixed time point, \( x_0, t_0 \) respectively - under both the original Lotkian formulation and the component projections set-up - the problem of determining the probability distribution of \( N(x_0, t_0) \) is one of either finding the distribution of the product of two random variables that cannot, in general, be assumed stochastically independent, or finding the probability distribution of a constant multiple of a random variable.\(^2\)

Further, the probability distribution of \( N(x, t_0) \) is the joint distribution of \( N(x_0, t_0) \) for all \( x_0 \), and, in general, the marginal distributions - for fixed age and time points - are not expected to be those of independent random variables.

Also, implementing the approximations discussed in (3.1), it is clear that the stochastic treatment of the age distribution function as a set-function of age is essentially the same as its stochastic treatment as a point function of age.

We shall consider the joint distribution of the number of individuals in two different age groups in the population at some point in time. Demographic variables will be assumed to have the simple stochastic structure of a constant depending on time (a deterministic component)

\(^2\)Note that in the case of using \( m(x, t_0-x_0) \) to arrive at \( B(t_0-x_0) \), the probability distribution of \( B(t_0-x_0) \) is a transformation of that of \( m(x, t_0-x_0) \) since it is usually not possible to treat \( N(x, t_0-x_0) \) as a random variable.
representing the expected value of the variable plus a stationary - in
time - stochastic error component. This restricted form will not be
required in the theoretical development but is suggested as the basis
for estimation in practical application.

The joint distribution referred to in the preceding paragraph is
a function of the joint distribution of a number of demographic vari-
ables that depends on the underlying formulation (Lotkian, component
projections or a combination of both). This number of component vari-
ables ranges from two to four depending on the situation. In the most
general case, Lotkian formulation, namely,

\[ N(x_1,t) = B(t-x_1) \ p(x_1,t) \]
\[ N(x_2,t) = B(t-x_2) \ p(x_2,t) \]

we have an initial joint distribution of four random variables expressing
demographic phenomena, two for number of births and two for survival
probabilities. These four random variables are not - in general -
stochastically independent. This set of four components is divided into
two subsets of two random variables each, (one for the number of births
and the other for the corresponding survival probability), and the final
probability distribution is the joint distribution of the products of
the two random variables in each of the two subsets. All other cases
may be deduced from this by considering degenerate random variables.

The distribution of the product of two stochastically dependent
random variables is usually difficult to obtain, let alone the bivariate
distribution of the two two-variate products of four stochastically
dependent random variables. Because of this difficulty, we shall be:
satisfied with describing the last distribution by its lower moments; i.e., the means, variances and covariance of its two components. The covariance is probably of the least practical importance. These moments are functions of the joint moments of the original four-variate distribution.

An interesting and important statistical problem arises in the estimation of the joint moments of the original four-variate distribution. The usual methods of estimation are not strictly applicable here since the basic requirement of a sample of n independent observations is not met in time series data. However, we shall use the same functions of sample moments \( \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{4} \pi_{i j} r_{j} \) to estimate functions of population moments \( \pi \frac{1}{n} \frac{1}{j} \pi_{i j} \) and the usual unbiased estimators of covariances in the normal case.

In general, let

\[
\begin{bmatrix}
Y_1 \\
Y_2
\end{bmatrix} = \begin{bmatrix}
X_{11} \\
X_{21} \\
X_{11} \\
X_{22}
\end{bmatrix}
\]

Then

\[
\begin{align*}
E(Y_{1i}) &= E(X_{11} \cdot X_{12}) \\
\text{Var}(Y_{1}) &= \text{Var}(X_{11} \cdot X_{12}) = E((X_{11} X_{12})^2) - E^2(X_{11} X_{12}) \\
\text{Cov}(Y_{1}, Y_{2}) &= E(X_{11} X_{12} X_{21} X_{22}) - E(X_{11} X_{12}) E(X_{21} X_{22})
\end{align*}
\]

\[\ldots (3.2.1.1)\]

In actual applications, we may set out to estimate these quantities without specifying a parent distribution. Alternatively, we may
stipulate an underlying probability distribution as a basis for the stochastic variation in the demographic variables, and use its properties to arrive at expressions for the required moments and then estimate these, hoping that the latter procedure, being more specific, may prove to be easier. One such parent distribution, the multivariate normal,\(^3\) is considered here as providing the random error component in the variation of the demographic variables.

3.2.2. The Multivariate Normal Case

If \(X_1\) and \(X_2\) are distributed identically and independently as the standard normal variate, then the distribution function of their product is \(\frac{1}{\pi} K_0(Z)\) where \(K_0(Z)\) is the Bessel function of the second kind of a purely imaginary argument of zero order. Further, if \(X_1\) and \(X_2\) are not stochastically independent, the distribution function of their product is available in a closed form. [9] Unfortunately, this form is too complicated to be of much analytical use. Limited tabulations of the distribution function are available in the case of independence. [2]

Consider the four dimensional normal variable \(X' = (X_1, X_2, X_3, X_4)\) with mean vector \(\mu' = (\mu_1, \mu_2, \mu_3, \mu_4)\) and covariance matrix \(V = (\sigma_{ij})\) \(i, j = 1, 2, 3, 4.\) Let \(t' = (t_1, t_2, t_3, t_4)\) be a vector of mathematical variables. Then the moment generating function of \(X\) is given by

\[
\phi_X(t) = \exp (\mu' t + \frac{1}{2} t' V t).
\]

Making use of the properties of \(\phi_X(t)\), we get

\(^3\)In an exact treatment, the multinormal distribution would have to be truncated to a finite rectangular region in the four dimensional space.
\[ E (X_i X_j) = \sigma_{ij} \mu_i \mu_j, \]
\[ \text{Var} (X_i X_j) = \sigma^2_{ij} + \sigma_{ii} \mu_j^2 + 2\sigma_{ij} \mu_i \mu_j + \sigma_{jj} \mu_i^2 \]
\[ + \sigma_{ii} \sigma_{jj}, \quad i, j = 1, 2, 3, 4 \] 
\[ \{ \text{..(3.2.2.1)} \}
\]

and
\[
\text{Cov} (X_1 X_2, X_3 X_4) = \sigma_{23} \sigma_{14} + \sigma_{13} \sigma_{24} + \sum_{i=1}^{2} \sum_{j=3}^{4} \mu_i \mu_j \sigma \]
\[ \{ (3-1), (7-j) \}
\]

which are sufficient to determine the lower moments of the distribution of \([X_1 X_2, X_3 X_4]\).

3.3. A Discrete Approximation to the Human Survivorship Function

An approximation to the survivorship function \(p(x)\) based on a finite number of age groups is motivated by the following consideration. In practice we can only observe a finite approximation to \(p(x)\) and for most practical purposes we do not need more than that. One such approximation may be arrived at as follows:

Divide the possible age span \([0, \omega]\) into \(n\) intervals of the form \((a_i, a_{i+1}], \quad i = 1, 2, \ldots, n\) where \(a_1 = 0\) and \(a_{n+1} = \omega\) such that \(\mu(x)\) can be considered approximately constant within each of the \(n\) age groups. Let the probability of death per age unit within the \(i^{th}\) age interval be approximated by \(q_i\) say.

Clearly, the assumptions needed for a constant - attrition survivorship function are approximately satisfied within each of the \(n\) age intervals considered and we have the original \(p(x)\) function approximated by \(n\) negative exponentials for
\[ \mu(x) = \mu_n(x) = \sum_{i=1}^{n} q_i 1(a_i, a_{i+1}] \quad \ldots (3.3.1) \]

where \[ 1(a_i, a_{i+1}] = 1 \quad a_i < x \leq a_{i+1} \]

= 0 otherwise is the indicator function of the interval \((a_i, a_{i+1}].\) It then follows — from (1.2.1.8) — that

\[
p(x) = p_n(x) = \exp\left\{ -\left( \sum_{i=1}^{m-1} q_i (a_{i+1} - a_i) + (x - a_m)q_m \right) \right\}
\]

\[
= \prod_{i=1}^{m-1} \exp\left\{ -q_i \frac{x-a_i}{\lambda_i} \right\} \cdot e^{-q_m (x-a_m)}
\]

\[
= k_m e^{-q_m (x-a_m)} \quad \ldots (3.3.2)
\]

where \(a_m < x \leq a_{m+1}, \ \lambda_i = a_{i+1} - a_i, \) and \(k_m\) is the probability of survival from birth to the beginning of the \(m^{th}\) age group.

Note that the conditional probability of death in the \(i^{th}\) age interval given survival to the beginning of the interval is equal to

\[
(p(a_i) - p(a_{i+1}))/p(a_i) = 1 - e^{-\lambda_i q_i} \quad \ldots (3.3.3)
\]

which bears a strong resemblance to the Reed and Merrell formula resulting from the direct assumption that \(\ln p(x) = \alpha + \beta x\) within the age interval. [25]

Further, it turns out that \(q_i\) is equal to the central death rate \(^4\) of the \(i^{th}\) age group, \(d_i\) say, which is the measure of age-specific mortality usually available, since

---

[^4]: Relating the number of deaths in a certain age group over a certain period of time in a population to the number of person-years lived in this population by individuals in those age groups and time period.
\[ d_i = \frac{p(a_i) - p(a_{i+1})}{\int_{a_i}^{a_{i+1}} p(x)dx} = \frac{k_i - k_i e^{-q_i l_i}}{k_i \frac{1}{q_i} (1 - e^{-q_i l_i})} = q_i \quad \ldots (3.3.4) \]

The approximation (3.3.2) may be used to calculate life expectancies rather easily, for

\[ e(a_i) = \frac{1}{k_i} \int_{a_i}^{\infty} p_n(x) \, dx \]

\[ = \frac{1}{k_i} \sum_{j=i}^{n} \int_{a_j}^{a_{j+1}} e^{-q_j(x-a_j)} \, dx \]

\[ = \frac{1}{k_i} \sum_{j=i}^{n} (k_j/q_j)(1 - e^{-q_j l_j}) \]

\[ = \frac{1}{k_i} \sum_{j=i}^{n} (k_j - k_{j+1})/q_i \quad \ldots (3.3.5) \]

3.4. An Example

In this section we apply some aspects of the previous theory to estimate the number of males in the two age groups (0) and (1-4) — using the conventional census age grouping — in Egypt on January 1, 1965. Males in these two age groups are the survivors of male births in the period January 1, 1960 — December 31, 1964.

Yearly male births were estimated using a series of crude birth rates corrected for under-registration by El-Badry [11], a sex ratio at birth of 105 males per 100 females and a series of estimated mid-year population figures — using census figures adjusted for under-enumeration of children less than ten years of age and assuming exponential growth between census counts.
Approximate single-year central death rates were derived from the reported death rate for the age group (0-4) using divisors computed from complete life tables for the years 1937, 1947 and 1960. [6] These rates were multiplied by correction factors computed by relating crude death rates adjusted for under-registration to reported crude death rates. [11] Finally, the resulting rates were transformed to survival probabilities using the method described in (3.3) and separation factors of 0.75, 0.7, 0.65, 0.6 and 0.55. Appendix (A) includes the original data and some of the intermediate calculations. Table 3.3.1 shows the resulting births and survival probabilities estimates.

The next step is to estimate the time trend in the four demographic variables under consideration and then use deviations from the trend to estimate the moments of the underlying distribution. We shall confine ourselves to polynomials and use least squares to get unbiased estimates of their coefficients.

For $X_1$, $X_2$ and $X_4$, we fitted straight lines. The square of the multiple correlation coefficient was 0.845, 0.907 and 0.991 for $X_1$, $X_2$ and $X_4$ respectively, and the inclusion of a quadratic term in the regression equation did not contribute to $R^2$ more than .01 in any of the three cases. The fitted regression lines were

$$
\begin{align*}
\hat{E}(X_1) &= 409.86 + 8.915t \\
\hat{E}(X_2) &= 0.8585 + 0.00386t \\
\hat{E}(X_4) &= 0.5939 + 0.01057t
\end{align*}
$$

with $t = 0$ at mid-year 1940; i.e., 1939.5.
<table>
<thead>
<tr>
<th>Calendar year</th>
<th>Male births in thousands in year t</th>
<th>Probability of survival from birth to be in the age group (0) at the start of (t+1)</th>
<th>Male births in thousands in the four-year period starting with year t.</th>
<th>Probability of survival from birth in a four-year period starting with year t to be in the age group (1-4) at the start of year (t+5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940 1</td>
<td>432</td>
<td>.861</td>
<td>1656</td>
<td>.597</td>
</tr>
<tr>
<td>1</td>
<td>426</td>
<td>.864</td>
<td>1656</td>
<td>.605</td>
</tr>
<tr>
<td>2</td>
<td>391</td>
<td>.856</td>
<td>1702</td>
<td>.617</td>
</tr>
<tr>
<td>3</td>
<td>407</td>
<td>.876</td>
<td>1764</td>
<td>.635</td>
</tr>
<tr>
<td>4</td>
<td>432</td>
<td>.877</td>
<td>1852</td>
<td>.638</td>
</tr>
<tr>
<td>5</td>
<td>472</td>
<td>.871</td>
<td>1899</td>
<td>.641</td>
</tr>
<tr>
<td>6</td>
<td>453</td>
<td>.889</td>
<td>1907</td>
<td>.654</td>
</tr>
<tr>
<td>7</td>
<td>495</td>
<td>.897</td>
<td>1964</td>
<td>.661</td>
</tr>
<tr>
<td>8</td>
<td>479</td>
<td>.870</td>
<td>2002</td>
<td>.672</td>
</tr>
<tr>
<td>9</td>
<td>480</td>
<td>.894</td>
<td>2071</td>
<td>.685</td>
</tr>
<tr>
<td>1950 1</td>
<td>510</td>
<td>.901</td>
<td>2125</td>
<td>.697</td>
</tr>
<tr>
<td>1</td>
<td>533</td>
<td>.896</td>
<td>2159</td>
<td>.708</td>
</tr>
<tr>
<td>2</td>
<td>548</td>
<td>.906</td>
<td>2185</td>
<td>.723</td>
</tr>
<tr>
<td>3</td>
<td>534</td>
<td>.905</td>
<td>2165</td>
<td>.738</td>
</tr>
<tr>
<td>4</td>
<td>544</td>
<td>.918</td>
<td>2142</td>
<td>.746</td>
</tr>
<tr>
<td>5</td>
<td>559</td>
<td>.912</td>
<td>2157</td>
<td>.754</td>
</tr>
<tr>
<td>6</td>
<td>528</td>
<td>.932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>511</td>
<td>.921</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>559</td>
<td>.927</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>598</td>
<td>.930</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For $X_3$, it is important to realize that yearly births determine period births completely, and since we have expressed $X_1$ as

$$X_{1i} = \alpha + \beta i + \epsilon_i \quad i = 0, 1, \ldots, 19$$

where $E(\epsilon_i) = 0$. It follows that

$$X_{3j} = \sum_{i=j}^{j+3} X_{1i} \quad j = 0, 1, \ldots, 15$$

$$= (4\alpha + 6\beta) + (4\beta) j + \sum_{i=j}^{j+3} \epsilon_i$$

$$= \alpha' + \beta'j + \epsilon'_j$$

where $E(\epsilon'_j) = 0$, therefore (3.4.1) implies

$$E(X_3) = 1692.93 + 35.66t \quad \ldots \quad (3.4.2)$$

If we assume that the observed time trends will continue for a period of five years (which we do not have to do if we want to stipulate different trends), and that the underlying stochastic structure about these trends is stationary, and using the expected values of the four demographic random variables on 1/1/1965 predicted from the estimated regression lines in conjunction with the deviations from the lines, we get the following results for the number of males in Egypt as of January 1, 1965.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Method</th>
<th>Mean (*)</th>
<th>Standard Deviation (*)</th>
<th>Covariance (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>distribution free</td>
<td>593.3</td>
<td>21.5</td>
<td>distribution -120.2</td>
</tr>
<tr>
<td></td>
<td>normal</td>
<td>593.3</td>
<td>22.1</td>
<td>free</td>
</tr>
<tr>
<td>1-4</td>
<td>distribution free</td>
<td>1939.5</td>
<td>36.3</td>
<td>normal -132.6</td>
</tr>
<tr>
<td></td>
<td>normal</td>
<td>1939.5</td>
<td>37.6</td>
<td></td>
</tr>
</tbody>
</table>

(*) in thousands.
CHAPTER IV

SOME ASPECTS OF THE ESTIMATION PROBLEM

4.1. Estimation of a Certain Function of Joint Moments of a Multivariate Distribution Using the Same Function of Joint Moments of a Sample of Dependent Observations

Consider the $k$-variate random variable $Y' = (Y_1, Y_2, \ldots, Y_k)$ with c.d.f. $F_{Y'}(\cdot)$. Suppose we have a sample of $n$ observations from this parent distribution, $Y_1, Y_2, \ldots, Y_n$ where

$$Y_{i1}' = (Y_{i11}, Y_{i21}, \ldots, Y_{ik1}) \quad i = 1, 2, \ldots, n$$

such that the $n$ observations are not independent.

Let $X_{r1}$ denote the product \( \prod_{l=1}^{k} Y_{r1l}^{r_l} \), in chapters 2 and 3, we used \( m_r = \frac{1}{n} \sum_{i=1}^{n} X_{r1}^{r_1} \), say, where $r_l = 0, 1, 2, \ldots$ and $l = 1, 2, \ldots, k$ as an estimator for $\mu_r = E(\prod_{l=1}^{k} Y_{r1l}^{r_l}) = E(X_{r1}^{r_1})$. We also used $(m_r - m_s)$ to estimate $(\mu_r - \mu_s)$. In this chapter, we investigate some of the properties of these estimators.

In the course of such an investigation, we may want to consider the $kn$-variate distribution of the sample; i.e., $Z' = (Y_1', Y_2', \ldots, Y_n')$ with c.d.f. $F_{Z'}(\cdot)$. Unfortunately, however, we end up with a sample of one observation only from $F_{Z'}(\cdot)$.

It is easy to show that $m_r$ is an unbiased estimator for $\mu_r$ since
\[ E(m_r) = \frac{1}{n} \sum_{i=1}^{n} \int \ldots \int x_i^{r} \, dF_i \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \int \ldots \int x_i^{r} \, dF_i \]
\[ \text{except} \]
\[ = \frac{1}{n} \sum_{i=1}^{n} E(x_i^{r}) = \frac{1}{n} \sum_{i=1}^{n} \mu_r = \mu_r. \]

Also, the calculation of the covariance of two such estimators \( m_r \) and \( m_s \) is straightforward. We have

\[
\text{Cov}(m_r, m_s) = E\left[ \left( \frac{1}{n} \sum_{i=1}^{n} (x_i^{r} - E(x_i^{r})) \right) \left( \frac{1}{n} \sum_{i=1}^{n} (x_i^{s} - E(x_i^{s})) \right) \right]
\]
\[ = \frac{1}{n^2} \left\{ \left[ \sum_{i \neq j}^{n} \text{Cov}(x_i^{r}, x_j^{s}) \right] + \frac{1}{n} \sum_{i \neq j}^{n} \text{Cov}(x_i^{r}, x_j^{s}) \right\}
\]
\[ = \frac{1}{n} \text{Cov}(x_1^{r}, x_1^{s}) + \frac{1}{n} \sum_{i \neq j}^{n} \text{Cov}(x_i^{r}, x_j^{s})
\]
\[ = \frac{1}{n} \text{Cov}(x_1^{r}, x_1^{s}) + (1 - \frac{1}{n}) \overline{\text{Cov}}(x_i^{r}, x_j^{s}) \]
\[
\text{where} \quad \overline{\text{Cov}}(x_i^{r}, x_j^{s}) = \frac{1}{ij} \overline{\sigma}_{rs} \text{denotes the arithmetic mean of the } n(n-1) \text{ covariances.}
\]
If the $n$ observations are independent, the second term of the right hand side of (4.1.1) vanishes, and if any two observations have the same dependence structure, hence the same distribution, then $i_j \sigma_{rs}$ is equal to the common covariance, $i_j \sigma_{rs}$ say.

As a special case of (4.1.1) when $r^*_l = s^*_l$, $l = 1, 2, \ldots, k$ we get

$$
\text{Var} \ (m_r) = \frac{1}{n} \text{Var} \ (X^*_i) + (1 - \frac{1}{n}) \ \overline{\text{Cov}} \ (X^*_i X^*_j) = \frac{1}{n} \sigma_{rr} + (1 - \frac{1}{n}) \ i_j \sigma_{rr} \quad \ldots(4.1.2)
$$

This is a restatement of the fact that if a set of $n$ random variables are distributed such that they all have the same variance, then their average internal covariance must not be less than $\left( \frac{\sigma^2}{n-1} \right)$ where $\sigma^2$ is their common variance. It follows that, in the limit, $i_j \sigma_{rr}$ is non-negative.

The situation gets more complicated when we consider $(m_s' m_t)$ as an estimator of $(u_s' u_t)$. To start with, the estimator is - in general - biased for

$$
E(m_s' m_t) = u_s' u_t + \left\{ \frac{1}{n} \sigma_{st} + (1 - \frac{1}{n}) i_j \sigma_{st} \right\} \quad \ldots(4.1.3)
$$

as may be shown either directly or as a corollary of (4.1.1) in conjunction with the relation $E(XY) = \text{Cov} \ (X, Y) + E(X) \ E(Y)$. As a matter of fact, $(m_s' m_t)$ cannot even be asymptotically unbiased for $(u_s' u_t)$ unless $i_j \sigma_{st} = 0$.

Further, the bias term in (4.1.3) is not estimable by the usual estimators of sample covariances. It is easy to see that
\[
\sum_{i=1}^{n} (X_i^s - \overline{X}^s)(X_i^t - \overline{X}^t) + \sum_{i \neq j=1}^{n} (X_i^s - \overline{X}^s)(X_j^t - \overline{X}^t) = 0
\]

Hence

\[
E(\sum_{i=1}^{n} (X_i^s - \overline{X}^s)(X_i^t - \overline{X}^t)) = -E(\sum_{i \neq j=1}^{n} (X_i^s - \overline{X}^s)(X_j^t - \overline{X}^t))
\]

\[\ldots (4.1.4)\]

and

\[
E(\sum_{i=1}^{n} (X_i^s - \overline{X}^s)(X_i^t - \overline{X}^t))
\]

\[
= E\{\sum_{i=1}^{n} (X_i^s - \mu_s - \overline{X}^s - \mu_s)(X_i^t - \mu_t - \overline{X}^t - \mu_t)\}
\]

\[
= \sum_{i=1}^{n} \{\sigma_{st} + \frac{1}{n}\left(\sum_{j=1}^{n} \sigma_{st} + \sum_{j' \neq 1}^{n} ij' \sigma_{st}\right) - \frac{2}{n} (\sigma_{st} + \sum_{j=1}^{n} \sigma_{st})\}
\]

\[
= (n-1) (\sigma_{st} - \overline{\sigma}_{st})
\]

\[\ldots (4.1.5)\]

Clearly, the bias term in (4.1.3) cannot be expressed as a function of \((\sigma_{st} - \overline{\sigma}_{st})\).

We also observe that

\[
E(\sum_{i \neq j=1}^{n} X_i^s X_j^t) = n(n-1)(\overline{\sigma}_{st} + \mu_s \mu_t)
\]

\[\ldots (4.1.6)\]

and

\[
E(\sum_{i=1}^{n} X_i^s X_i^t) = n(\sigma_{st} + \mu_s \mu_t)
\]

\[\ldots (4.1.7)\]
Hence we may consider
\[ e_1 = \frac{1}{n(n-1)} \sum_{i \neq j=1}^{n} X_i^{s} X_j^{t} \quad \text{and} \]
\[ e_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^{s} X_i^{t} \quad \text{as well as} \quad e_3 = m_s m_t \quad \text{as biased estimators for} \quad (\mu_s \mu_t). \]

An alternative approach to what we have been considering so far stipulates certain dependence patterns among the \( n \) observations of the sample. A particularly appealing pattern was considered by P. K. Sen. [26] Sen extended the U-statistics method of estimation of functionals of distribution functions [15] to the case of \( m \)-dependent stationary stochastic processes.

Let \( \{X_n\} \) be a sequence of vector-valued random variables, Sen called such a sequence \( m \)-dependent if the random vector \((X_1, X_2, \ldots, X_t)\) is stochastically independent of the random vector \((X_s, X_{s+1}, \ldots)\) whenever \( s-r > m \) and it is said to be stationary if the joint distribution of \( X_i, \ldots, X_{i+r} \) is independent of \( i \) for all \( r \).

Let \( f(X^\alpha_1, X^\alpha_2, \ldots, X^\alpha_k) \) be a statistic symmetric in its arguments \( X^\alpha_1, \ldots, X^\alpha_k \), \( \alpha_1 < \alpha_2 < \ldots < \alpha_k \). If \( E(f(X^\alpha_1, \ldots, X^\alpha_k)) = g \), we call \( f \) a kernel for \( g \).

Define \( h = E(f(X^\alpha_1, \ldots, X^\alpha_k) | \alpha_{j+1} - \alpha_j > m \) for \( j = 1, 2, \ldots, k-1) \) and the symmetric estimator based on a sample of \( n \) observations, \( n \geq m \) (k-1) + k,
\[ U(X_1, \ldots, X_n) = \binom{n-m(k-1)}{k}^{-1} \sum_{S} f(X^\alpha_1, \ldots, X^\alpha_k) \quad \ldots (4.1.8) \]

where the summation extends over all possible \( \binom{n-m(k-1)}{k} \) sets of
(α₁, ..., αₖ) satisfying (αₗ₊₁ - αₗ) > m.

The estimate \( \hat{U}(X_1, ..., X_n) \) is unbiased for \( h \). Sen showed that it is also consistent and, under the conditions of existence of the second and third order moments of \( f(X_1, ..., X_n) \), that \( \hat{U}(X_1, ..., X_n) \) is asymptotically normally distributed with mean \( h \) and variance \( k^2 \zeta \) and gave a consistent estimator for \( \zeta \).

Under the m-dependence set-up, we may formally generalize \( c_1 \) to get an unbiased estimator for \( (\mu_\sigma \mu_t) \). Consider

\[
\begin{align*}
m e_1 &= \binom{n-m}{2}^{-1} \sum_{i<j=1}^{n} \frac{1}{2} (X_i^t X_j^s + X_i^s X_j^t) \\
&= \frac{1}{(n-m)(n-m-1)} \sum_{i\neq j=1}^{n} X_i^s X_j^t \\
&\quad \text{for } |i-j|>m
\end{align*}
\]

Clearly \( m e_1 \) is unbiased for \( (\mu_\sigma \mu_t) \) and \( 0 e_1 = e_1 \). We may also want to consider the estimator

\[
(m)e_1 = \frac{1}{n(2n-m-1)} \sum_{i\neq j=1}^{n} X_i^s X_j^t \\
\quad \text{for } |i-j|<m
\]

Recapitulating, we have five estimators that we may use to estimate \( (\mu_\sigma \mu_t) \). These are \( me_1 \), \( me_1 \), \( e_1 \), \( e_2 \), and \( e_3 \). Let

\[
c_1 = \frac{(n-m)(n-m-1)}{n^2}, \quad c_2 = \frac{m(2n-m-1)}{n^2}, \quad c_3 = \frac{1}{n}
\]

\[
c_1' = \frac{(n-m)(n-m-1)}{n(n-1)}, \quad c_2' = \frac{m(2n-m-1)}{n(n-1)}, \text{ then these estimators are related by the following relations.}
\]
\[ e_1 = c_1^m e_1^m + c_2^m (m)e_1^m \]
\[ e_3 = (1 - \frac{1}{n}) e_1^m + \frac{1}{n} e_2^m \]
\[ = c_1^m e_1^m + c_2^m (m)e_1^m + c_3^m e_2^m \]
and, in general,
\[ E(m_e_1) = \mu_s \mu_t + \sigma_{ij}^{(m)} \]
where \( \sigma_{ij}^{(m)} \) is the average of the \((n-m)(n-m-1)\)
covariances of \((x_i^s, x_j^t)\) such that \(|i - j| > m\)
\[ E(m_e_1) = \mu_s \mu_t + \sigma_{ij}^{(m)} \]
where \( \sigma_{ij}^{(m)} \) is the average of the \(m(2n-m-1)\)
covariances of \((x_i^s, x_j^t)\) such that \(|i - j| \leq m\)
\[ E(e_1) = \mu_s \mu_t + c_1^m \sigma_{ij}^{(m)} + c_2^m \sigma_{ij}^{(m)} \]
\[ E(e_2) = \mu_s \mu_t + \sigma_{st} \]
\[ E(e_3) = \mu_s \mu_t + c_1 \sigma_{ij}^{(m)} + c_2 \sigma_{ij}^{(m)} + c_3 \sigma_{st} \]
under \(m\)-dependence \( \sigma_{ij}^{(m)} = 0 \) and asymptotically \( c_1 \rightarrow 1 \)
and \( c_2, c_2', c_3 \rightarrow 0 \), and we can distinguish two subsets of five estimators accordingly. On one hand, we have a fundamental estimator, namely \( m_{e_1} \),
on which two other estimators, \( e_1 \) and \( e_3 \), are based. These three
estimators are asymptotically equivalent. Since \( m_{e_1} \) is unbiased under
\(m\)-dependence, \( e_1 \) and \( e_3 \) are asymptotically unbiased under \(m\)-dependence
and have the same variance as \(e_1\). On the other hand, we have \((m)e_1\) and \(e_2\) which have the undesirable property that they are always biased; i.e., for all sample sizes, unless – of course – \(\sigma_{st}^{(m)} = 0\) or \(\sigma_{st} = 0\) respectively which is very unlikely.

Clearly we should be more interested in small sample results rather than asymptotic results because the length of time series of demographic data is usually restricted. Unfortunately the nature of the bias terms of the five estimators does not allow precise comparisons. Only general tendencies can be indicated.

For the first subset of the five estimators, \(m e_1\) seems to be the choice estimator for \(m\)-dependent processes and \(n > m(k - 1) + k\), otherwise \(m e_1 = 0\). However, we need to either know \(m\) or estimate it from the sample. Even if \(m\)-dependence does not describe the dependence structure of the process under consideration adequately, \(m e_1\) may still be preferred to \((m)e_1\) or \(e_1\) if there is reason to expect that

\[
|\sigma_{st}^{(m)}| > |\sigma_{st}| \quad \text{or} \quad |\sigma_{st}^{(m)}| > |\sigma_{st}|
\]

respectively, as is usually the case with time series. Between the two groups of estimators, \(m e_1\) may be preferred to \((m)e_1\) for the same reasons that is may be preferred to \(e_1\), and it would probably be less biased than \(e_2\) since it is reasonable to expect that \(|\sigma_{st}| > |\sigma_{st}^{(m)}|\). Finally, within the second subset of estimators, \((m)e_1\) would be better than \(e_2\) if \(|\sigma_{st}^{(m)}| < |\sigma_{st}|\) which is likely to be the case.

However, we must remember that in estimation from small samples, the relative merit of an estimator is determined to a great extent by how much use it makes of the sample information, and the five estimators
under consideration here use various portions of the sample. For example $e_1$, $e_3$ make increasingly more use of the sample information, in that order. Also, the amount of relative use of the sample information for $e_1$ and $(m) e_1$ depends on the relation of $m$ to $n$.

A still more conclusive comparison among the five estimators considered would be based on a more decisive criterion than unbiasedness only, like the mean square error of an estimator. All indications are, however, that such a criterion would not be feasible analytically extrapolating from the experience with unbiasedness. As a matter of fact the variance terms needed for a study based on M.S.E. are too general and too complicated to be useful analytically.

Obviously, the problem gets to be considerably more complex when we consider estimators for $\hat{\mu}_r - \mu$ based on $m$ and the five estimators for $\mu_r$ discussed above.

4.2. A Monte Carlo Study of Some of the Properties of Five Estimators for the Variance of the Product of the Two Components of a Bivariate-Normal Time Series in Small Samples

In this section we use the Monte Carlo method of statistical experimentation to estimate the mean square error of five estimators for a special case of the estimation problem discussed in (4.1).

4.2.1. Introduction

Let $\{x_t\}$ be a $(p \times 1)$ vector-valued normal time series$^2$ with

$$E(x_t) = 0 \quad (p \times 1)$$

$^2_i.e., for any subset of indices $(t_1, t_2, \ldots, t_n)$ the vectors $(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$ have the np-multivariate normal distribution.
\[ \Sigma_h = \text{Cov}(X_t, X_{t+h}) = (h \sigma_{ij}) \]

\[ i, j = 1, 2, \ldots, p, \forall t \]

where
\[
\Sigma_0 = \begin{bmatrix}
1 & \rho_{12} & \cdots & \rho_{1p} \\
\rho_{21} & 1 & \cdots & \rho_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{p1} & \rho_{p2} & \cdots & 1
\end{bmatrix}
\]

...(4.2.1)

Further, let
\[ \Sigma_h = e^{-\alpha h} \Sigma_0 \quad \alpha > 0 \]

such that \( e^{-\alpha(m+1)} = 0 \). This defines a special case of an \( m \)-dependent multi-normal time series.

Then a subsequence of \( n \) consecutive elements of \( \{X_t\} \) has the \( np \)-variate normal with

\[ \mathbb{E}(X'_{t+1} \cdots X'_{t+n}) = \begin{bmatrix} 0' \\ 1 \times np \end{bmatrix} \]

and
\[
\text{Var} \begin{bmatrix}
X_{t+1} \\
X_{t+2} \\
\vdots \\
X_{t+n}
\end{bmatrix} = \begin{bmatrix}
\Sigma_0 & \Sigma_1 & \cdots & \Sigma_{n-1} \\
\Sigma_1 & \Sigma_0 & \cdots & \Sigma_{n-2} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{n-1} & \Sigma_{n-2} & \cdots & \Sigma_0
\end{bmatrix} = \Sigma_{(np \times np)}
\]
\[
\Sigma = \begin{bmatrix}
1 & e^{-\alpha} & \cdots & e^{-(n-1)\alpha} \\
-\alpha & e & \cdots & e^{-(n-2)\alpha} \\
\vdots & \vdots & \ddots & \vdots \\
-\alpha & e & \cdots & 1
\end{bmatrix} \otimes \Sigma_0 \text{ for all } t. \tag{4.2.1.3}
\]

In the most general problem considered in Chapters II and III, we had \( p = 4 \) for the estimation of

\[
(E(X_1X_2X_3X_4) - E(X_1X_2)E(X_3X_4)) \text{ where } X_1, X_2, X_3 \text{ and } X_4 \text{ are the components of } X_t, \text{ note that this parameter is independent of } t. \text{ This function of moments represents the covariance of the number of individuals in two different age groups. Unfortunately, a complete investigation of this case using Monte Carlo methods would be both cumbersome and of questionable utility simply because of the large number of correlation patterns that we have to consider among the four components of } X_t. \text{ Furthermore, as indicated earlier, it is likely that this covariance will not be of much practical importance.}

A special case that is of considerable practical importance refers to the problem of estimation of the variance of the number of individuals in a certain age group. In this case, \( p = 2 \) and the parameter is

\[
V = E(X_1^2X_2^2) - E^2(X_1X_2).
\]

This case will be considered here.

Equation (3.2.2.1) gives

\[
V = \sigma_{12}^2 + \sigma_{11}\mu_2^2 + 2\sigma_{12}\mu_1\mu_2 + \sigma_{22}\mu_1^2 + \sigma_{11}\sigma_{22}
\]
in our special case, using the conditions included in (4.2.1.1), this reduces to

\[ V = \rho_{12}^2 + 1 \]  \hspace{1cm} (4.2.1.4)

Let \( e_0 = \frac{1}{n} \sum_{i=1}^{n} (X_{1i}X_{2i})^2 \) be the unbiased estimator for \( E(X_1^2X_2^2) \), and \( m^e_1, (m)^e_1, e_2 \) and \( e_3 \) be defined from the corresponding formulae in (4.1) by putting \( t_\ell = s_\ell = 1, \ell = 1, 2, \) and \( X_i = X_{1i}X_{2i} \). The following estimators for \( V \) were considered

\[
\begin{align*}
S_1 & = e_0 - m^e_1 \\
S_2 & = e_0 - (m)^e_1 \\
S_3 & = e_0 - e_1 \\
S_4 & = e_0 - e_3 \\
S_5 & = \hat{\sigma}_{12}^2 + \hat{\sigma}_{11}^2 + \hat{\mu}_2^2 + 2\hat{\sigma}_{12}\hat{\mu}_1\hat{\mu}_2 + \hat{\sigma}_{22}^2 + \hat{\mu}_1^2 + \hat{\sigma}_{11}^1\hat{\sigma}_{22}^1
\end{align*}
\]  \hspace{1cm} (4.2.1.5)

where \( \hat{\cdot} \) denotes the usual unbiased estimator of the parameter in the case of a sample of independent observations.

We note that \( (e_0 - e_2) = 0 \), which does not make it a useful estimator for \( V \) although \( e_0 \) and \( e_2 \) may be - under certain conditions - reasonable estimates of \( E(X_1^2X_2^2) \) and \( E^2(X_1X_2) \) respectively.

### 4.2.2. Procedure

The following values of \( n, m \) and \( \rho_{12} = \rho \), say, were considered

\[
\begin{align*}
n & = 5, 10, 15 \\
m & = 0, 1, 3, 5, 10, 15 \\
\rho & = -.75, (.25), .75 \ \footnote{3} \\ \footnote{3} \rho = \pm 1 \text{ were not included to keep } \Sigma \text{ positive definite.}
\end{align*}
\]
For each combination of \( n \), \( m \) and \( p \), 1000 samples of \( 2n \) independent normal deviates were generated using the IBM subroutine RANDU to generate uniform random numbers [30] and J. Bell's modification of the Box-Muller routine to transform them to independent normal deviates [5].

This sample of \( 2n \) observations was transformed to a single random observation from a \( 2n \)-variate normal with mean \( \mathbf{0} \) and variance matrix \( \Sigma \) determined by \( n \), \( m \), \( p \) and the covariance structure described in (4.2.1).

The following two remarks were used in the transformation.

(1) Let \( \mathbf{Y} \sim N(\mathbf{0}, \mathbf{I}) \) and

\[
\mathbf{Z} = \mathbf{C} \mathbf{Y} \quad \text{then} \quad \mathbf{Z} \sim N(\mathbf{0}, \mathbf{C} \mathbf{C}')
\]

(2) Choleski's Factorization

If \( \Sigma \) is a symmetric positive definite matrix, then

\[
\Sigma = \mathbf{T} \mathbf{T}' \quad \text{where} \quad \mathbf{T} \quad \text{is a lower triangular matrix. The elements of} \quad \mathbf{T} \quad \text{are determined by the following relations.}
\]

\[
t_{ij} = 0 \quad i < j
\]

\[
t_{11} = \sqrt{\sigma_{11}}
\]

\[
t_{ii} = \sigma_{ii}/t_{11}
\]

\[
t_{ii} = \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} t_{ik}^2} \quad i = 1, 2, 3, \ldots, p
\]

\[
t_{ij} = (\sigma_{ij} - \sum_{k=1}^{j-1} t_{ik} t_{jk})/t_{jj} \quad i \neq j, i,j = 2, 3, \ldots, p
\]
Obviously, the one observation from the 2n-variate normal may be considered as a sample of n correlated observations from a bivariate normal distribution determined by $\rho$, whose correlation structure is determined by $m$, which is what we actually need. For each of the 1000 samples, the five estimators $S_1$, $S_2$, $S_3$, $S_4$ and $S_5$ were computed and then their mean and variance over the 1000 resulting values were used to estimate their mean square error (M.S.E.).

In computing the estimators use was made of the following forms.

Let $Y_i = X_{1i} X_{21}$, $i = 1, 2, \ldots, n$, then

$$m_1 = \frac{(Y' E_1 Y)/((n-m)(n-m-1))}{(m)}$$

$$e_1 = \frac{(Y' E_2 Y)}/(n(2n-m-1)) \quad n \geq m + 2$$

$$e_1 = \frac{(Y' E_2 Y)}/(n(n-1)) \quad n < m + 2$$

$$e_3 = \frac{(Y' J Y)}{n^2}$$

(4.2.2.1)

where $J$ is an $(n \times n)$ matrix of 1's,

$$E_{1ij} = \begin{cases} 1 & |i-j| > m \\ 0 & |i-j| \leq m \end{cases}$$

$$E_{2ij} = \begin{cases} 0 & i=j \\ 0 & |i-j| > m \\ 1 & |i-j| \leq m, i \neq j \end{cases}$$

We may point out here that it would have been possible - but very arduous - to avoid the sampling experiment performed to arrive at numerical values for the M.S.E. of the first four estimators for the
combinations of \( n, m \) and \( p \) considered here using the general form that can be laboriously obtained. This procedure would have the advantage of the absence of sampling errors from the M.S.E. estimates. However, such a procedure was considered prohibitively tedious for the fifth estimator. Consequently, in order to enhance comparability and save time and effort, we decided on the Monte Carlo simulation procedure for the five estimators.

4.2.3. Conclusions

The variance was invariably the major component of M.S.E. of the five estimators. Table C.1 summarizes the results of the Monte Carlo study. It gives the Relative Mean Square Error, defined as the ratio of the M.S.E. to the parameter value being estimated and henceforth denoted by R.M.S.E. This will be our criterion for comparing the five estimators.

The main conclusion to be drawn out of the results of the study is that for small sample sizes (up to 15 observations) making maximum use of the sample information is more important than avoiding correlation structures by basing estimators on the subsets of uncorrelated observations in the sample that form kernels for the parameter being estimated. This is indicated by the facts that \( S_2 \) had, uniformly, the lowest R.M.S.E. of all estimators and \( S_2 \) had lower R.M.S.E. than \( S_1 \) when \( m \) was relatively large compared to \( n \).

Naturally, the performance of the five estimators, in terms of R.M.S.E. - improved as the sample size increased; also, the differences among the estimators decreased considerably. The range of the R.M.S.E. was 14.6, 3.8 and 2.7 for \( n = 5 \), 10 and 15 respectively. In particular,
the magnitude of the relative superiority of \( S_4 \) with respect to the 
rest of the estimators declined as \( n \) increased. Further, the uniform 
ordering pattern of the five estimators that held for \( n = 5 \) over all 
values of the correlation coefficient and time lag - with the exception 
of \( S_1 \) and \( S_2 \) exchanging relative positions as \( m \) increased - was replaced 
by a criss-crossing pattern that was much in effect for \( n = 10, 15 \).

The fifth estimator, based on normal theory, had - consistently - 
the highest R.M.S.E. - and with a wide margin - except for \( m = 15 \) when 
\( S_1 \) came close to it, for \( n = 5 \). Evidently this was due to the large 
number of component parameters that had to be estimated in the course 
of computing \( S_5 \) and the non-linear form in which they were used. However, it showed the largest proportional decline in R.M.S.E. as \( n \) 
increased. Further, it was not uniformly the worst estimator starting 
with \( n = 10 \). Actually, it competed with \( S_4 \) for the lowest R.M.S.E. 
for \( n = 15 \), moderate correlation and intermediate time lags. We may 
even speculate that for larger samples, \( S_5 \) will probably compare more 
favourably or have the lowest R.M.S.E. as the disadvantages of a large 
number of component parameters and non-linear forms get overwhelmed by 
sample size as the present results indicate.

For \( m = 0, (m)^{e_1} = 0 \) and for \( m \) relatively small compared to \( n \), 
\( (m)^{e_1} \) makes relatively little use of the sample information; hence, 
\( (m)^{e_1} \) had a relatively high R.M.S.E. as may be expected. Similarly, 
for \( m + 2 > n \) \( e_1 = 0 \) and for \( m \) relatively large compared to \( n \) \( e_1 \) 
makes relatively little use of the sample information. Clearly, \( S_3 \) 
tended to be dominated by \( S_1 \) or \( S_2 \) depending on the relative magnitude 
of \( m \) to \( n \). For \( n = 5 \), with \( S_5 \) being uniformly worse than all other
estimates, $S_2$ and $S_1$ exchanged the position of the second-worst estimator as $m$ increased; and for $n = 10, 15$ they also exchanged their relative position - which was the worst for $n = 15$. The R.M.S.E. of $S_1$ and $S_2$ tended to increase drastically in response to increases in $|\rho|$. 
CHAPTER V

SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

5.1. Summary

Stochastic models of human population change are valuable theoretical constructs and useful tools of practical applications. However, available stochastic formulations are of limited use because of over-simplified assumptions, complicated mathematics and difficulty of applications.

In the present study, simple macro-analytic stochastic formulations of the problem of human population change over time are considered on two levels of detail, namely, total population size and age structure of the population. In the development of these formulations, no conditions were imposed on the physical mechanism producing the process of population change and the resulting methodologies are simpler than what is presently available.

Demographic variables are assumed to have the simple stochastic structure of a constant depending on time plus a stationary random error component; i.e., the stochastic mechanism producing random variations as a component of the total variation in the population is assumed to remain stationary over time.

In Chapter II, a basic probability space is defined whose elements are identified with relative conglomerations of socio-cultural conditions that are responsible for the random component in the change of
the population.

A discrete error-type probability law that allows for more uncertainty - in the sense of less probability mass concentration around the mean - relative to the normal and symmetric binomial distributions is discussed. Such a distribution would be of value for types of data that are not commensurate with high degrees of accuracy as is the case with demographic data for developing countries for example.

Using characteristics of the above-mentioned probability space, probabilistic characterizations of the distribution of the population at a certain point in time are arrived at under both discrete and continuous probability structures. The theory discussed is illustrated by data for the population of Egypt.

The problem of the probability distribution of the age distribution function is considered in Chapter III. A generalized formulation of survivorship from birth incorporating net migration is used to present the age distribution function formulations of Lotka and the model of component projections.

The probability distribution of the number of individuals in two different age groups at a certain point in time is described by its lower moments, i.e., the means, variances and covariance, in general and in the special case of an underlying multivariate normal distribution. A numerical example is given using Egyptian data.

Because demographic data needed for the application of stochastic models of population change are usually in the form of time series, classical methods of estimation are not strictly applicable for the samples available are not - in general - composed of independent observations.
An investigation of some aspects of this estimation problem is carried out in Chapter IV for the case of a certain function of joint moments of a multivariate distribution. It is found that estimators usually considered are biased and that it is possible to define unbiased estimators for \( m \)-dependent processes by basing estimators on subsamples of independent observations that can be formed from the original sample. Because the variances of estimators considered are too general and too complex to be of much analytical use, a special case was investigated using the Monte Carlo method of statistical experimentation to study the mean square error of five estimators for the variance of the number of individuals in a certain age group at a certain point in time, assuming an underlying normal distribution, in small samples.

The main conclusions to be drawn out of the study are that maximum use of sample information outweighs all other considerations in comparing the performance of estimators in terms of M.S.E., and that an estimator based on normal theory does not compare favorably with other estimators that do not assume an underlying normal distribution, presumably because it depends on a large number of parameters in a non-linear fashion.

5.2. Suggestions for Further Research

Two main directions for further useful research that are indicated by this study are:

(1) investigation of probabilistic characterizations for the distribution of the age distribution function that describe the distribution more fully than its moments,

(2) further examination of the problem of parameter estimation
from time series data in general, and the estimation of functions of joint moments of multivariate distributions from serially correlated observations in particular.
APPENDICES

A. DATA AND INTERMEDIATE CALCULATIONS FOR SECTION (3.4)

TABLE A.1

BIRTH RATE PER 1000 POPULATION CORRECTED FOR UNDER-REGISTRATION
EGYPT (1940-1959)

<table>
<thead>
<tr>
<th>Year</th>
<th>1940</th>
<th>1941</th>
<th>1942</th>
<th>1943</th>
<th>1944</th>
<th>1945</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth rate</td>
<td>46.2</td>
<td>45.1</td>
<td>41.0</td>
<td>42.2</td>
<td>44.3</td>
<td>47.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1946</th>
<th>1947</th>
<th>1948</th>
<th>1949</th>
<th>1950</th>
<th>1951</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth rate</td>
<td>45.5</td>
<td>49.1</td>
<td>46.3</td>
<td>45.3</td>
<td>47.1</td>
<td>48.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1952</th>
<th>1953</th>
<th>1954</th>
<th>1955</th>
<th>1956</th>
<th>1957</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth rate</td>
<td>48.2</td>
<td>45.9</td>
<td>45.7</td>
<td>45.9</td>
<td>42.3</td>
<td>40.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1958</th>
<th>1959</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth rate</td>
<td>42.8</td>
<td>44.8</td>
</tr>
</tbody>
</table>

**TABLE A.2**

**ESTIMATED MID-YEAR POPULATION IN THOUSANDS, EGYPT (1940-1959)**

<table>
<thead>
<tr>
<th>Year</th>
<th>1940</th>
<th>1941</th>
<th>1942</th>
<th>1943</th>
<th>1944</th>
<th>1945</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>18236</td>
<td>18434</td>
<td>18635</td>
<td>18838</td>
<td>19043</td>
<td>19250</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1946</th>
<th>1947</th>
<th>1948</th>
<th>1949</th>
<th>1950</th>
<th>1951</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>19460</td>
<td>19734</td>
<td>20201</td>
<td>20678</td>
<td>21167</td>
<td>21668</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1952</th>
<th>1953</th>
<th>1954</th>
<th>1955</th>
<th>1956</th>
<th>1957</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>22180</td>
<td>22704</td>
<td>23241</td>
<td>23791</td>
<td>24353</td>
<td>24929</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1958</th>
<th>1959</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>25518</td>
<td>26121</td>
</tr>
</tbody>
</table>

*Using census figures corrected for under-enumeration of children less than ten years of age and assuming exponential population growth between censuses.*

**Source:** Compiled from "Population Trends in the U.A.R." [6]
<table>
<thead>
<tr>
<th>Year</th>
<th>1940</th>
<th>1941</th>
<th>1942</th>
<th>1943</th>
<th>1944</th>
<th>1945</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death rate</td>
<td>117.4</td>
<td>112.1</td>
<td>116.5</td>
<td>100.1</td>
<td>99.5</td>
<td>106.9</td>
</tr>
<tr>
<td>Year</td>
<td>1946</td>
<td>1947</td>
<td>1948</td>
<td>1949</td>
<td>1950</td>
<td>1951</td>
</tr>
<tr>
<td>Death rate</td>
<td>93.1</td>
<td>83.6</td>
<td>108.3</td>
<td>88.3</td>
<td>81.6</td>
<td>86.2</td>
</tr>
<tr>
<td>Year</td>
<td>1952</td>
<td>1953</td>
<td>1954</td>
<td>1955</td>
<td>1956</td>
<td>1957</td>
</tr>
<tr>
<td>Death rate</td>
<td>78.7</td>
<td>79.4</td>
<td>68.5</td>
<td>70.2</td>
<td>59.0</td>
<td>66.6</td>
</tr>
<tr>
<td>Year</td>
<td>1958</td>
<td>1959</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death rate</td>
<td>61.6</td>
<td>57.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of life</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>2.100</td>
<td>1.300</td>
<td>0.725</td>
<td>0.325</td>
<td>0.200</td>
</tr>
</tbody>
</table>

### TABLE A.5

**RATIO OF CRUDE DEATH RATES CORRECTED FOR UNDER-REGISTRATION TO REPORTED RATES, EGYPT (1940-1959)**

<table>
<thead>
<tr>
<th>Year</th>
<th>1940</th>
<th>1941</th>
<th>1942</th>
<th>1943</th>
<th>1944</th>
<th>1945</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>1.221</td>
<td>1.249</td>
<td>1.276</td>
<td>1.267</td>
<td>1.258</td>
<td>1.235</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1946</th>
<th>1947</th>
<th>1948</th>
<th>1949</th>
<th>1950</th>
<th>1951</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>1.204</td>
<td>1.243</td>
<td>1.225</td>
<td>1.209</td>
<td>1.220</td>
<td>1.212</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1952</th>
<th>1953</th>
<th>1954</th>
<th>1955</th>
<th>1956</th>
<th>1957</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>1.198</td>
<td>1.200</td>
<td>1.197</td>
<td>1.256</td>
<td>1.141</td>
<td>1.174</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1958</th>
<th>1959</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>1.175</td>
<td>1.222</td>
</tr>
</tbody>
</table>

**Source:** Compiled from El-Badry's paper. [11]
B. COMPUTER PROGRAM

The FORTRAN IV computer program used in the Monte Carlo study described in (4.2) is reproduced below:

```
DIMENSION X(50,50),T(50,50),S(50,50),SD(50),R(7),M(8),A(8),B(25)
1,Y1(50),Y2(50),Y(25),E1(25,25),E2(25,25),ES(5),ESM(5),ESV(5),
2ESMES(5),ESRMSE(5),T1(2500),E11(625),E21(625)

C BUILDING LAG AND CORRELATION ARRAYS

M(1)=0
M(2)=1
M(3)=3
DO 2 J=4,6
2 M(J)=(J-3)*5
DO 3 J=1,6
3 A(J)=15./(M(J)+1.)
R(1)=-.75
DO 4 I=2,7
4 R(I)=R(I-1)+.25
DO 1 N=10,30,10
N2=N/2
WRITE(3,9)N2
9 FORMAT(1H1,13HSAMPLE SIZE =,5X,I5)
DO 1 J=1,6
WRITE(3,19)M(J)
19 FORMAT(//,1H0,4HLAG=,14X,I5)
```
C BUILDING CORRELATION DECAY FACTORS ARRAY

DO 5 K=1,25
5 B(K)=0.0
MJ=M(J)
IF(MJ.EQ.0)GO TO 91
DO 6 K=1,MJ
B(K)=EXP(-A(J)*K)
IF(ABS(B(K)).LT..1E-10)B(K)=0.0
6 CONTINUE
91 MJ2=M(J)+2

C BUILDING MATRICES REQUIRED FOR COMPUTING ESTIMATORS

IF(N2.LT.MJ2)GO TO 15
DO 17 K=1,N2
DO 17 L=1,N2
IF(INABS(K-L)-MJ)117,117,18
18 E1(K,L)=1.
E2(K,L)=0.0
GO TO 17
117 E1(K,L)=0.0
E2(K,L)=1.
17 CONTINUE
121 DO 49 K=1,N2
49 E2(K,K)=0.0
GO TO 16
15 DO 20 K=1,N2
DO 20 L=1,N2
E1(K,L)=0.0
IF(K-L)21,200,21
21 E2(K,L)=1.
GO TO 20
200 E2(K,L)=0.0
20 CONTINUE
16 DO 1 I=1,7
   WRITE(3,29)R(I)
29 FORMAT(/'H0,24HCORRELATION COEFFICIENT=',F10.6)

C BUILDING VARIANCE MATRIX

   DO 7 K=1,N
5   X(K,K)=1.
   DO 8 K=2,N
6   X(K,K-1)=R(I)
   DO 59 K=1,N,2
7   K2=K+2
   DO 59 L=K2,N,2
8   LB=(L-K)/2
9   X(L,K)=B(LB)
10  X(L+1,K)=X(L,K)*R(I)
   DO 10 K=2,N,2
11  K2=K+2
   DO 10 L=K2,N,2
12  LB=(L-K)/2
13  X(L,K)=B(LB)
14  X(L-1,K)=X(L,K)*R(I)
DO 81 K=1,N
DO 81 L=1,K
IF(ABS(X(K,L)).LT..1E-10)X(K,L)=0.0
81 CONTINUE
DO 101 K=1,N
DO 101 L=1,K
101 X(L,K)=X(K,L)
C GENERATING MULTINORMAL OBSERVATIONS
CALL FCTRIZ(X,N,T)
W=0.0
IX=I*MJ2*N+1
DO 26 L=1,5
ESM(L)=0.0
26 ESV(L)=0.0
NS=1000
DO 11 NN=1,NS
DO 12 L=1,N
CALL NORDEV(IX,W,Z)
IF(ABS(Z).LT..1E-10)Z=0.0
12 Y1(L)=Z
CALL PACKR(T,50,50,T1,N,N)
CALL MULT(T1,N,N,Y1,N,1,Y2)
DO 13 L=1,N,2
L1=L+1
L2=L1/2
13 Y(L2)=Y2(L)*Y2(L1)
C COMPUTATION OF ESTIMATES

ES01=0.0

DO 14 L=1,N2

14 ES01=ES01+Y(L)*Y(L)
ES0=ES01/N2

CALL PACKLS(E1,25,25,E11,N2,N2)
CALL ATRBA(Y,N2,1,E11,N2,N2,ES(1))
CALL PACKLS(E2,25,25,E21,N2,N2)
CALL ATRBA(Y,N2,1,E21,N2,N2,ES(2))

ES(3)=ES(1)+ES(2)
ES(4)=ES01+ES(3)

IF(N2.LT.N32)GO TO 22

ES(1)=ES(1)/((N2-M(J))*(N2-M(J)-1))
ES(2)=ES(2)/(N2*(2*N2-M(J)-1))

GO TO 23

22 ES(2)=ES(2)/(N2*(N2-1))

23 ES(3)=ES(3)/(N2*(N2-1))

ES(4)=ES(4)/(N2*N2)

DO 24 L=1,4

24 ES(L)=ES0-ES(L)

YM1=0.0
YM2=0.0
V1=0.0
V2=0.0
V12=0.0

DO 25 L=1,N,2
L1 = L + 1
YM1 = YM1 + Y2(L)/N2
YM2 = YM2 + Y2(LJ)/N2
V1 = V1 + Y2(L) * Y2(L)/(N2-1)
V2 = V2 + Y2(L1) * Y2(L1)/(N2-1)

25 V12 = V12 + Y2(L) * Y2(L1)/(N2-1)
V1 = V1 - N2 * YM1 * YM1/(N2-1)
V2 = V2 - N2 * YM2 * YM2/(N2-1)
V12 = V12 - N2 * YM1 * YM2/(N2-1)
ES(5) = V12 * V12 + V1 * YM2 * YM2 + V2 * YM1 * YM1 + 2 * V12 * YM1 * YM2 + V1 * V2

DO 11 L = 1, 5
ESM(L) = ESM(L) + ES(L)/NS

11 ESV(L) = ESV(L) + ES(L) * ES(L)/NS
P = R(I) * R(I) + 1.

DO 27 L = 1, 5
ESV(L) = ESV(L) - ESM(L) * ESM(L)
ESMSE(L) = ESV(L) + (ESM(L) - P)**2

27 ESRMSE(L) = ESMSE(L)/P

1 WRITE(3, 39) P, (L, ESM(L), ESV(L), ESMSE(L), ESRMSE(L), L=1, 5)

39 FORMAT(1HO, 16HPARAMETER VALUE=, 4X, F12.6, //3X, 9HESTIMATOR, 12X,
14HMEAN, 12X, 8HVARIANCE, 12X, 6HM.S.E., 14X, 8HR.M.S.E. // (6X, 15, 6X, 4
2(F14.6, 5X)))

STOP
END
SUBROUTINE FCTRIZ(X,N,T)

DIMENSION X(50,50),T(50,50),S(50,50),SD(50)

DO 14 K=1,N
  DO 14 L=K,N
  14 T(K,L)=0.0
    T(1,1)=SQRT(X(1,1))
    IF(N-1)11,12,2
  2 DO 3 I=2,N
  3 T(I,1)=X(I,1)/T(1,1)
     N1=N-1
     IF(N1-1)12,1,4
  4 DO 5 K=2,N1
     K1=K+1
     DO 5 I=K1,N
  5 S(I,K)=T(I,1)*T(K,1)
     DO 6 J=2,N
  6 SD(J)=T(J,1)*T(J,1)
     DO 10 K=2,N1
     T(K,K)=SQRT(X(K,K)-SD(K))
     K1=K+1
     DO 9 I=K1,N
  9 T(I,K)=(X(I,K)-S(I,K))/T(K,K)
     IF(N1-K)12,1,7
  7 DO 8 J=K1,N
  8 SD(J)=SD(J)+T(J,K)*T(J,K)
     DO 10 L=K1,N1
  10 CONTINUE
L1=L+1

DO 10 M=L1,N

10 S(M,L)=S(M,L)+T(M,K)*T(L,K)

SD(N)=SD(N)+T(N,N1)*T(N,N1)

T(N,N)=SQRT(X(N,N)-SD(N))

RETURN

11 WRITE(3,13)

13 FORMAT(1H0,33H VARIANCE MATRIX DIMENSION ERROR)

12 RETURN

END

SUBROUTINE NORDEV(IX,W,Z)

IF(W)1,1,2

2 Z=2.*W*V*U

W=0.0

RETURN

1 CALL RANDU(IX,IX,V)

IX=IX

CALL RANDU(IX,IX,U)

IX=IX

U=2.*U-1.

W=V*U+U*U

IF(1.-W)1,3,3

3 CALL RANDU(IX,IX,X)

IX=IX

W=SQRT(-2.*ALOG(X))/W
\[ Z = (V^T V - U^T U)^T \]

RETURN

END

The subroutines:

PACKR : packs a rectangular matrix as a column vector
MULT : multiplies two rectangular matrices
PACKLS: packs a lower symmetric matrix as a column vector
ATRBA : performs the matrix multiplication A'BA

were written by Dr. T. G. Donnelly for the Matrix Subroutines Package of the Department of Biostatistics, University of North Carolina at Chapel Hill, North Carolina.
C. NUMERICAL RESULTS OF THE MONTE CARLO STUDY

TABLE C.1

RELATIVE MEAN SQUARE ERROR OF THE FIVE ESTIMATORS

<table>
<thead>
<tr>
<th>Parameter Value</th>
<th>n</th>
<th>m</th>
<th>Estimator</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0</td>
<td></td>
<td>1.5625</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.225</td>
<td>2.655</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.061</td>
<td>3.390</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.225</td>
<td>2.655</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.743</td>
<td>1.769</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.820</td>
<td>5.776</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>5.925</td>
<td>3.527</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.635</td>
<td>4.111</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.879</td>
<td>3.222</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.779</td>
<td>2.107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.594</td>
<td>3.169</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.634</td>
<td>2.721</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.390</td>
<td>1.782</td>
</tr>
<tr>
<td>n</td>
<td>m</td>
<td>n Estimator</td>
<td>Correlation Coefficient</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---</td>
<td>-------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.75 -0.5 -0.25 0.0 0.25 0.5 0.75</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6.580</td>
<td>4.013 4.369 1.527 2.617 3.524 7.746</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.353</td>
<td>2.666 3.834 1.621 2.188 2.507 4.745</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.353</td>
<td>2.666 3.834 1.621 2.188 2.507 4.745</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.215</td>
<td>1.748 2.497 1.085 1.438 1.652 3.098</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>11.539</td>
<td>7.879 4.508 2.834 5.221 5.682 10.853</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>7.260</td>
<td>4.586 2.308 1.392 2.139 2.809 9.668</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.114</td>
<td>2.910 1.818 1.373 1.641 1.710 6.095</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.114</td>
<td>2.910 1.818 1.373 1.641 1.710 6.095</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.709</td>
<td>1.919 1.231 0.934 1.121 1.200 4.017</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>15.252</td>
<td>5.525 2.242 1.655 2.244 4.770 9.756</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.771</td>
<td>3.136 1.776 1.537 1.679 2.612 3.573</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.771</td>
<td>3.136 1.776 1.537 1.679 2.612 3.573</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.761</td>
<td>2.098 1.191 1.051 1.154 1.753 2.350</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>15.560</td>
<td>6.890 2.972 2.387 3.802 5.318 10.238</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>2.459</td>
<td>1.200 0.992 1.067 1.021 1.266 1.726</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.811</td>
<td>1.784 1.188 1.031 1.147 1.896 2.989</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.459</td>
<td>1.200 0.992 1.067 1.021 1.266 1.726</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.006</td>
<td>0.997 0.813 0.863 0.826 1.043 1.431</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.666</td>
<td>1.839 1.302 0.974 1.047 1.994 2.696</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>m</td>
<td>Estimator</td>
<td>Correlation Coefficient</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---</td>
<td>-----------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.75</td>
<td>-0.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2.137</td>
<td>1.482</td>
<td>1.030</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.353</td>
<td>2.152</td>
<td>1.173</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.088</td>
<td>1.498</td>
<td>1.014</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.707</td>
<td>1.224</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.988</td>
<td>2.156</td>
<td>1.354</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.709</td>
<td>1.495</td>
<td>1.064</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.354</td>
<td>1.933</td>
<td>1.106</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.665</td>
<td>1.464</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.373</td>
<td>1.186</td>
<td>0.815</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.397</td>
<td>1.931</td>
<td>1.193</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2.397</td>
<td>2.120</td>
<td>1.167</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.872</td>
<td>2.197</td>
<td>1.153</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.180</td>
<td>1.907</td>
<td>1.078</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.768</td>
<td>1.547</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.118</td>
<td>1.780</td>
<td>1.274</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>3.824</td>
<td>2.404</td>
<td>1.022</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.280</td>
<td>1.501</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.280</td>
<td>1.501</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.864</td>
<td>1.228</td>
<td>0.710</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.566</td>
<td>2.179</td>
<td>1.220</td>
</tr>
</tbody>
</table>
\begin{table}
\centering
\begin{tabular}{cccccccc}
\hline
n & m & Estimator & Correlation Coefficient & \multicolumn{7}{c}{-0.75} \\
& & & & -0.5 & -0.25 & 0.0 & 0.25 & 0.5 & 0.75 \\
\hline
15 & 1 & 4.265 & 3.045 & 1.405 & 0.865 & 1.154 & 2.636 & 4.451 \\
2 & 2.282 & 1.946 & 1.031 & 0.792 & 0.906 & 1.433 & 2.237 \\
3 & 2.282 & 1.946 & 1.031 & 0.792 & 0.906 & 1.433 & 2.237 \\
4 & 1.880 & 1.590 & 0.864 & 0.663 & 0.762 & 1.178 & 1.832 \\
5 & 3.290 & 2.182 & 1.303 & 0.948 & 1.175 & 2.748 & 3.557 \\
\hline
15 & 0 & 1 & 1.343 & 0.957 & 0.600 & 0.537 & 0.671 & 0.790 & 1.075 \\
2 & 2.328 & 1.396 & 0.716 & 0.528 & 0.780 & 1.169 & 2.014 \\
3 & 1.343 & 0.957 & 0.600 & 0.537 & 0.671 & 0.790 & 1.075 \\
4 & 1.178 & 0.839 & 0.527 & 0.472 & 0.592 & 0.702 & 0.944 \\
5 & 1.457 & 0.968 & 0.620 & 0.569 & 0.655 & 0.980 & 1.438 \\
\hline
1 & 1 & 1.484 & 0.907 & 0.691 & 0.463 & 0.776 & 1.043 & 1.432 \\
2 & 2.478 & 1.312 & 0.795 & 0.466 & 0.843 & 1.502 & 2.292 \\
3 & 1.487 & 0.903 & 0.697 & 0.461 & 0.775 & 1.047 & 1.423 \\
4 & 1.294 & 0.791 & 0.612 & 0.412 & 0.676 & 0.915 & 1.250 \\
5 & 1.549 & 0.977 & 0.561 & 0.504 & 0.568 & 1.135 & 1.447 \\
\hline
3 & 1 & 1.655 & 0.967 & 0.676 & 0.453 & 0.633 & 1.099 & 1.465 \\
2 & 2.412 & 1.274 & 0.760 & 0.445 & 0.689 & 1.493 & 2.343 \\
3 & 1.625 & 0.939 & 0.681 & 0.449 & 0.625 & 1.094 & 1.491 \\
4 & 1.417 & 0.817 & 0.597 & 0.398 & 0.548 & 0.956 & 1.304 \\
5 & 1.450 & 0.942 & 0.573 & 0.381 & 0.590 & 1.157 & 1.545 \\
\hline
\end{tabular}
\caption{Continued}
\end{table}
<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>Estimator</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-0.75</td>
<td>-0.5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2.285</td>
<td>1.038</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.060</td>
<td>1.311</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.225</td>
<td>1.022</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.934</td>
<td>0.895</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.698</td>
<td>1.090</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1.774</td>
<td>1.143</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.602</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.341</td>
<td>0.843</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.183</td>
<td>0.748</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.449</td>
<td>1.059</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>3.053</td>
<td>2.007</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.621</td>
<td>1.195</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.621</td>
<td>1.195</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.417</td>
<td>1.050</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.864</td>
<td>1.493</td>
</tr>
</tbody>
</table>
LIST OF REFERENCES


4. Bartlett, M. S. Stochastic Processes, Notes of a course given at the University of North Carolina in the Fall quarter, 1946, pp. 28-42.


