Institute Mimeo Series #8

October 17, 1947

STOCHASTIC PROCESSES

(Notes of a course given at the University of North Carolina in the Fall Quarter, 1946.)

by

M. S. Bartlett

Errata

p. 20: Last expression in sec. 3. For $1/m$ read $1/\sqrt{m}$.

p. 29: The right hand side of equation (1) should be similar to that in (2) (with the ommission of (j)).

p. 33: "reproduction rate" (three lines below equation (3) is

$$s(T)\ r(T) = \Phi(T), \text{ and } \phi(T)dT = d\,\Phi(T).$$

p. 36: Equation (5):

$$C_i = -\alpha(r_i) \Big/ \frac{d}{dr_i}\psi(r_i).$$

p. 38: Equation below (6): for suffix $\begin{cases} k + 2r - 2 & \text{read } k + 2r - 4 \\ k + 2r - 4 & \text{read } k + 2r - 6 \end{cases}$

p. 39: Four lines down in sec. 10: "explicit".

p. 53: Right hand side of equation (22) is $\lambda(n - \mu_t)\mu_t$.

p. 65: For "Wald" read "Wold".

p. 76: Line above equation (12). For "is" read "is in".

p. 82. Top line: $\phi(z) = z^2 + \alpha z + \beta$.

p. 86: Line under equation (8): "on for" for "for".

STOCHASTIC PROCESSES

(Notes of a course given at the University of North Carolina in the Fall Quarter, 1946.)

by

M. S. Bartlett

Prefatory Note

This short course on stochastic processes, given at the University of North Carolina in the fall quarter, 1946, has been intended as an introduction to a subject which both in its theoretical and practical aspects is of extremely wide scope.

On the theoretical side my aim has primarily been to make statistical students familiar with what are on the whole comparatively recent and unfamiliar ideas, and where feasible I have illustrated these ideas in as elementary a way as possible. For example, where there is a duality between processes defined for discrete and continuous time, it seems useful to discuss the former model first, not only (a) in its own right and (b) as an approximate numerical procedure for the latter model, but also as affording some insight into the structure of the theory before the many further problems of rigour arising in the case of continuous time have to be met.

In regard to applications, it was considered advisable to limit these to one or two broad fields of immediate interest to statisticians, and the course fell roughly into two parts:

(i) evolutionary processes of the Markoff type, involving a discrete random variable, and their application to population growth and allied problems;

(ii) stationary processes, involving a random variable with continuous range, and their application to the correlation theory of time-series.

In a previous introductory course given at Cambridge University, a third field was given in place of (i), viz.

(iii) a discussion of the 'random walk' (or 'random migration') problem, and its application in statistics to sequential analysis; but as a separate course on sequential analysis is given in the University of North Carolina, the present choice and limitation of subject matter, which was dictated also by the time allowed for the course, seemed preferable.

In the compilation of these notes, reference has been made to lecture notes taken by Mr. D. N. Nanda. (Any sections added for completeness to the course presented are indicated by an asterisk). Some of the references given in F of the bibliography I owe to Prof. H. Cramér, Prof. M. Kac and Mr. J. E. Moyal. Acknowledgment is also made of the many stimulating discussions on stochastic processes I have had in recent years with Mr. Moyal.

M. S. B.

January, 1947.

CONTENTS

I. General Introduction.

1. Preamble.
2. Classification of stochastic processes.
3. Recommended literature.
4. Random variables and probability distributions.
5. Theoretical specification of stochastic processes.

II. Evolutionary Processes of the Markoff Chain Type, and Their Application to Population Growth, Industrial Renewal Theory, and Allied Problems.

1. Basic equations for Markoff Chains.
2. The equation $p_{k+1} = \underset{\sim}{Q}\underset{\sim}{p}_k$.
3. The case of probability distributions.
4. An example from genetics.
5. Population growth.
6. Industrial renewal theory.
7. Transition to continuous time.

*8. Solution of an integral equation.

9. Fluctuations.
10. Further discussion of the population growth equations.
11. Markoff chains with continuous time.
12. Some examples:
    (i) The negative binomial
    (ii) An elementary 'birth' and 'death' process.
    (iii) The Gaussian frequency curve as the mean of a stochastic model.
    *(iv) The logistic function

*13. Further discussion of the general equations.

III. Stationary Processes, and their Bearing on the Analysis of Time-series.

1. Definition of stationary processes.
2. Stationary processes specified for discrete time.
3. Relation between autocorrelation coefficients and harmonic analysis.
4. Standard errors of sample estimates.
5. Stationary processes specified for continuous time.
6. The continuous linear process.
7. Standard errors in the case of continuous time.
8. Simultaneous time-series.

*9. Further notes on simultaneous time-series.

IV. Selected Bibliography.

I. GENERAL INTRODUCTION

1. Preamble.

In this course we are going to consider a subject which in isolated applications has arisen since the beginnings* of probability theory, but the general theory of which has only recently begun to receive the attention it deserves. Roughly speaking, we may think of this subject as the statistics of 'change', or 'dynamic' theory, in contrast to the 'static' distributional problems with which we have hitherto been mainly preoccupied. In theoretical terms, we have been concerned with a random variable X, or a number of random variables $X_1, X_2, \ldots X_k$, or the distribution of a specified function *f* of the variables $X_1, X_2, \ldots X_k$. We want now to consider a random quantity which depends on a further parameter, which will often denote 'time' and may therefore be denoted by *t*; this leads us to the notion of a random function $X(t)$.

This theoretical picture will represent some actual, e.g. physical, process in the real world, that has some random or stochastic element involved in its structure, and that will therefore be called a random or stochastic process. For convenience, just as probability is used both for the mathematical concept and physical concept, we shall also refer to the mathematical model, represented by $X(t)$, as a stochastic process.

*See the references in F11, especially at the end of Ch. 8.

2. Classification of Stochastic Processes.

Before glancing at the possible applications, we shall find it useful to classify broadly our subject-matter. There is no loss of generality in confining our attention to random quantitative variables, since distributions associated with attributes can always be defined in relation to random variables taking the values 0 to 1. However, as in other applications of probability theory, it is often convenient to consider separately discrete-valued random variables and variables taking a continuous range of values. But we also have now the parameter t, which may take in some problems an enumerable set of values and in others a continuous range of values. This gives us at once four main types of process, which can be exemplified as follows:

(1) X discrete, t discrete e.g. total number of farms in U.S.A. harvesting wheat each year.

(2) X discrete, t continuous e.g. total number of farms in U.S.A.

(3) X continuous, t discrete e.g. total weight of wheat crop in U.S.A. each year.

(4) X continuous, t continuous e.g. total area of cultivated land in U.S.A.

By X continuous in the above classification, we merely mean that it is a random variable with continuous range; the question of the "continuity" of X(t), regarded as a function of t, will be considered later.

So far we have referred only to stochastic processes involving one random quantity X and one parameter t, but obviously we may have to deal with more than one quantity X and/or more than one parameter t. The custom as regards terminology does not seem too constant; for definiteness, we shall refer to processes involving more than one quantity X as multivariate or simultaneous processes, reserving the term multidimensional processes for processes involving more than one parameter t. For example, the changing prices of several commodities considered together would be a simultaneous process in several variables; the fertility of a piece of land would be a function of two spatial cordinates x and y, say, and for certain purposes could be regarded as a random function $F(x,y)$ involving the two spatial dimensions x and y.

Further classifications will be made as the theory is developed, but we may perhaps note in the various applications one further broad division into evolutionary and stationary processes. Thus as examples of evolutionary processes we may cite:

(i) growth and change in biological (including human) populations,

(ii) diffusion problems in physics,

(iii) sequential sampling methods in industry,

whereas in other problems we may be able to think of our processes in a certain sense as stationary. Such stationary processes, which will be more precisely defined later, are important, for example, in the study of:

(i) oscillatory time-series in meteorology, economics, textile research, etc.

(ii) the statistical theory of turbulence,

(iii) electrical 'noise' phenomena.

3. Recommended Literature.

There is as yet no text-book devoted to the subject of stochastic processes, though a number of papers and monographs on particular applications have appeared in recent years, e.g., B 1, 2 and 3 or D 3 (these numbers refer to the selected references given at the end of these notes). For a study of the theory, an advanced knowledge of probability theory is desirable, to be obtained, for example, from A 1 or 2. In addition, A 1 contains a chapter on 'the homogeneous random process'. A 3 is cited for two reasons: its first part includes a useful discussion of probability modes of convergence, and its second part deals with the particular type of stochastic process known as a Markoff chain (see Part II of these notes).

4. Random Variables and Probability Distributions.

Before considering the theoretical specification of stochastic processes, we shall refer briefly to the notation and some of the formulae in the theory of random variables and probability distributions that we shall want to use (cf. A 1 or 2). For mutually exclusive events or classes $A_r (r = 1 \text{ to } k)$, a probability distribution is specified by the symbolic expression

$$E\left\{A\right\} \equiv p_1 A_1 + p_2 A_2 + \dots \quad p_k A_k, \tag{1}$$

or equivalently by the column vector of probability coefficients $p_1, p_2, \dots p_k$. If a number x_r is associated with the event A_r, then the random variable X has a value x_r with probability

$$p\left\{X = x_r\right\} = p_r. \tag{2}$$

When this theory is generalized to cover infinite and even non-denumerable numbers of classes, we define the total or cumulative probability distribution $F(x)$ of X by

$$P\left\{X \leq x\right\} = F(x) \tag{3}$$

where $F(x)$ is a monotonic function, (continuous to the right), increasing from 0 to 1 as x increases. In most cases in practice $F(x)$, which is invariably of the composite Stieltjes integral type (for definiteness we shall exclude the exception with a 'singular component')

$$F(x) = \int_{-\infty}^{x} dF(x) = c_1^2\int_{-\infty}^{x} f(x)dx + c_2^2 \sum_{x_r \leq x} p_r \tag{4}$$

where $c_1^2 + c_2^2 = 1$, is also such that c_1^2 or $c_2^2 = 0$. In the first case, we still have a probability distribution of the type represented by equation (1) (though the number of classes may now be infinite). In the second case, we shall refer to x as a continuous variable. We shall also for convenience sometimes refer to the differential element of the integral in (4), and in the two cases write

$$\begin{aligned} p(x) = dF(x) &= 0, \qquad (x \neq x_r), \\ &= p_r, \qquad (x = x_r), \end{aligned} \tag{5}$$

and $\qquad p(x) = dF(x) = f(x)dx.$

The operation of taking the mean value or expectation of a function $w(X)$ is denoted by the same formal symbol E as in (1), where we now define $E\left\{w(X)\right\}$ by the Lebesgue-Stieltjes integral

$$E\left\{w(X)\right\} = \int_{-\infty}^{\infty} w(x)\, dF(x). \tag{6}$$

Similarly for a vector random variable

$$\underline{R} \equiv (X_1, X_2, \ldots X_e)$$

we define

$$F(\underset{\sim}{r}) \equiv F(x_1, x_2, \ldots x_e) = P(X_i \leq x_i \text{ for all } i, 1,\ldots,e)$$

and

$$E\left\{w(\underset{\sim}{R})\right\} = \int w(\underset{\sim}{r})\, dF(\underset{\sim}{r}). \tag{7}$$

In particular, we have the characteristic function

$$C(\theta) \equiv M(i\theta) \equiv E\left\{\exp i(\theta_1 X_1 + \theta_2 X_2 + \ldots \theta_e X_e)\right\}, \tag{8}$$

defined for all real θ; and, when they exist, the moment-generating function $M(\theta)$, and the moment formulae:

$$\begin{aligned} &\text{Mean of } X_1 && m \equiv E\{X_1\}, \\ &\text{Variance of } X_1 && \sigma^2 \equiv E\left\{(X_1 - m)^2\right\}, \\ &\text{Covariance of } X_i \text{ and } X_j, && \\ & && \rho_{ij}\sigma_i\sigma_j = E\left\{(X_i - m_i)(X_j - m_j)\right\}, \end{aligned} \tag{9}$$

etc. We define as usual the cumulant or semivariant function

$$K(\theta) = \log M(\theta) = \sum_{r=1}^{\infty} \kappa_r \theta^r / r! \tag{10}$$

where $\kappa_1 = m$, $\kappa_2 = \sigma^2$ and for a vector variable, the general second-order coefficient is $\rho_{ij}\sigma_i\sigma_j$. In particular we have for the three most well-known distributions:

$$\begin{aligned} &\text{Poisson,} && K(\theta) = m(e^{\theta} - 1) \\ &\text{Binomial,} && \qquad n \log\left\{1 + p(e^{\theta} - 1)\right\} \\ &\text{Normal} && \qquad m\theta + \tfrac{1}{2}\theta^2\sigma^2 \end{aligned}$$

and for the normal law in several variables, a corresponding expression in the θ_i up to the second degree. We have also for discrete random variables taking the values 0, 1, 2 ... such as the Poisson or binomial, the useful identity

$$\Pi(z) = M(\log z), \tag{11}$$

where $\Pi(z)$ is the probability-generating function. In all cases the total distribution $F(x)$ can always be obtained from $C(\theta)$ (at points of continuity) by Fourier inversion, e.g.

$$F(x) - F(0) = \lim_{T\to\infty} \frac{1}{2\Pi} \int_{-T}^{T} \frac{1-e^{-ix\theta}}{i\theta} C(\theta)\, d\theta, \tag{12}$$

For independent variables X_1, X_2, ... X_e we have

$$C(\theta) = C_1(\theta_1)\; C_2(\theta_2)\; \dots\; C_e(\theta_e), \tag{13}$$

whence in particular for the sum

$$X = X_1 + X_2 \;+ \dots X_e,$$

$$C(\theta) = C_1(\theta)\; C_2(\theta)\; \dots\; C_e(\theta) \tag{14}$$

and if the $K(\theta)$ exist,

$$K(\theta) = K_1(\theta) + K_2(\theta) + \dots K_e(\theta) \tag{15}$$

Finally, we shall require some formulae for conditional random variables. Thus for two random variables X and Y, we define the total distribution function $G(y|x)$ of the conditional variable $Y|x$ by the relation

$$F(x,y) = \int_{-\infty}^{x} G(y|x)\; dF_x(x); \tag{16}$$

equivalently we may write

$$dF(x,y) = p(x,y) = p(y|x)p(x) = dG(y|x)dF_x(x) \tag{17}$$

where in the discrete case

$$p_{rs} = p(x_r,y_s) = p(y_s|x_r)\; p(x_r) = q_{rs}p_r, \text{ say}, \tag{18}$$

and in the case of continuous variation,

$$p(x,y) = f(x,y)dxdy = g(y|x)dy. \quad f_x(x)dx. \tag{19}$$

A function $w(Y)$ has a conditional mean value in relation to $G(y|x)$ given by

$$E\left\{w(Y)\middle| x\right\} = \int w(y)dG(y|x), \tag{20}$$

i.e. in the first case, we have

$$\Sigma_s\; w(y_s)q_{rs} \tag{21}$$

and in the second,

$$\int w(y)g(y|x)dy. \tag{22}$$

In particular the characteristic function of $Y|x$ is

$$E\{e^{i\theta Y}|x\} = \int e^{i\theta y}\, dG(y|x). \tag{23}$$

5. Theoretical Specification of Stochastic Processes.

If a stochastic process $X(t)$ exists from the mathematical standpoint, it follows that for any set of possible values t_r $(r = 1, \ldots, n)$ of the parameter t, we shall have a vector random variable

$$\underset{\sim}{R} \equiv (X_1, X_2, \ldots X_n).$$

This vector random variable $\underset{\sim}{R}$ must of course have a simultaneous distribution $F(\underset{\sim}{r})$. The complete theoretical specification of a stochastic process $X(t)$ must therefore include the specification of $F(\underset{\sim}{r})$ for any finite set of values t_r. It does not immediately follow that such a specification is then <u>sufficient</u> as a definition of the process, which may be a function of t <u>for all t</u>, but under not very restrictive conditions on the regularity of the functions with which we shall be concerned, we may legitimately assume its sufficiency.*

From the existence of the distribution $F(\underset{\sim}{r})$, we may at once deduce some consistency relations which are of the utmost importance in the theory of certain types of stochastic process. We have for two times t_1 and t_2 (by convention we shall take $t_r \geq t_{r-1}$) a distribution $F(x_1, x_2)$, from which we obtain on integration** w.r.t. x_1,

$$F_2(x_2) = \int_{-\infty}^{\infty} G(x_2|x_1)\, dF_1(x_1). \tag{1}$$

Equivalently to (1), it is convenient to write

$$p(x_2) = \int_{x_1} p(x_2|x_1)\, p(x_1), \tag{2}$$

---

*See F 3 and F 7.

**It should be noted that while $F(x_1, x_2)$ will in general depend on t_1 and t_2, the distribution $F_2(x_2)$ obtained after integration must depend only on t_2.

(where the integral sign is to denote summation if x is discrete).

$$p(x_3) = \int_{x_2} p(x_3|x_2)\, p(x_2)$$
$$= \int_{x_2} p(x_3|x_2) \left\{ \int_{x_1} p(x_2\ x_1)\, p(x_1) \right\}$$

or, interchanging the order of integration, we have

$$\int_{x_1} p(x_3|x_1)\, p(x_1) = \int_{x_1} \left\{ \int_{x_2} p(x_3|x_2) p(x_2|x_1) \right\} p(x_1). \quad (3)$$

Thus (3) is formally true for all processes; it must not be inferred from it, however, that the expressions inside the integral sign w.r.t. x_1 are equal. In fact,

$$p(x_3|x_1) = \int_{x_2} p(x_3|x_2,\ x_1)\, p(x_2|x_1) \quad (4)$$

and only equals $\int_{x_2} p(x_3|x_2)\, p(x_2|x_1)$

<u>if</u> $p(x_3|x_2,\ x_1) = p(x_3|x_2)$.

This last condition is a special case of the general condition that for any two sets t_i, t_j (all $j >$ all i)

$$p(x_j \text{ for all } j | x_i \text{ for all } i) = p(x_j \text{ for all } j | x_o), \quad (5)$$

where t_o represents the greatest of the set t_i. This general condition implies that when the value of X at the last available value of t is known, no previous history adds anything further to the probable future history of X, and is the condition characterizing what is known as a <u>Markoff process</u>. Markoff processes in which the variable X is discrete are also sometimes called <u>Markoff chains</u>. In the applications in Part II, we shall need to develop further the theory of some particular types of Markoff chain.

II. EVOLUTIONARY PROCESSES OF THE MARKOFF CHAIN TYPE, AND THEIR APPLICATION TO POPULATION GROWTH, INDUSTRIAL RENEWAL THEORY, AND ALLIED PROBLEMS.

1. Basic equations for Markoff chains.

For a Markoff process we had

$$p(x_3|x_1) = \int_{x_2} p(x_3|x_2)\, p(x_2|x_1). \tag{1}$$

Since we have also for such a process

$$\begin{aligned} p(x_1,x_2,\ \ldots,x_n) &= p(x_1)p(x_2|x_1)p(x_3|x_1,x_2)\ \ldots \\ &= p(x_1)\prod_{r=1}^{n-1} p(x_{r+1}|x_r), \end{aligned} \tag{2}$$

we note that a Markoff process is completely characterized by its 'initial distribution' $p(x_1)$ and the conditional distribution $p(x_{r+1}|x_r)$.

It will often be convenient to use vector and matrix notation for discrete-valued variables, or for corresponding distributions of attributes. Thus for such cases, the equation

$$p(x_3) = \int_{x_2} p(x_3|x_2)\, p(x_2)$$

may be written algebraically

$$p_{i,3} = \Sigma_j\, q_{ij,32}\, p_{j,2}, \tag{3}$$

where p_i denotes the probability of the $\underline{i}$th classification, or $\underline{i}$th value x_i, or in matrix notation,

$$\underset{\sim}{p}_3 = \underset{\sim}{Q}_{32}\, \underset{\sim}{p}_2, \tag{4}$$

where $\underset{\sim}{p}$ is the column vector (p_i), and $\underset{\sim}{Q}$ the matrix (q_{ij}). We have then, by a repetition of (4),

$$\underset{\sim}{p}_3 = \underset{\sim}{Q}_{32}\, \underset{\sim}{Q}_{21}\, \underset{\sim}{p}_1; \tag{5}$$

but only for a Markoff chain do we also have

$$\underset{\sim}{Q}_{31} = \underset{\sim}{Q}_{32}\, \underset{\sim}{Q}_{21}. \tag{6}$$

As an elementary example suppose the time t_k corresponds to the $\underline{k}$-th independent trial of an event with probability p. Then the number of times the event has occurred at time t_k is given by the ordinary binomial distribution with index k. Equation (3) in this case becomes

$$p_{i,k+1} = p\, p_{i-1,k} + (1-p)p_{i,k},\ (p_{-1} = 0), \qquad (7)$$

or in the matrix notation of (4),

$$\begin{pmatrix} p_0 \\ p \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p_{k+1} \end{pmatrix}_{k+1} = \begin{pmatrix} 1-p & 0 & 0 & \cdots \\ p & 1-p & 0 & \cdots \\ 0 & p & 1-p & \cdot \\ 0 & 0 & p & \cdots \\ \cdot & & & \ddots \\ \cdot & & & \cdot \end{pmatrix} \begin{pmatrix} p_0 \\ p \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p_k \\ 0 \end{pmatrix}_k \qquad (8)$$

where $p_{i,o} = \delta_{i,o}$, (1 for i = 0, 0 for i > 0).

A simple way of solving this system of difference equations is to form the probability-generating function, or equivalently the moment-generating function, for the variable X_k. We have

$$\Pi_{k+1}(z) = (1\ z\ z^2 \ldots) \begin{pmatrix} 1-p & 0 & 0 & \cdots \\ p & 1-p & 0 & \cdots \\ 0 & p & 1-p & \cdot \\ 0 & 0 & p & \ddots \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \end{pmatrix}_k \qquad (9)$$

$$= \left[1 + p(z-1)\right] \Pi_k(z)$$

An even simpler derivation of equation (9) is obtained by noting that the total value X_{k+1} consists of the value X_k plus the <u>independent</u> score 1 or 0 at the k+1-th trial. This is a very special type of Markoff process and implies that

$$\Pi_{k+1}(z) = \Pi_1(z)\ \Pi_k(z),$$

where in this case

$$\Pi_1(z) = 1 - p + pz.$$

We have of course from (9)

$$\Pi_k(z) = \left[1 + p(z-1)\right]^k, \qquad (10)$$

the well-known generating function for the binomial distribution.

The above example indicates the frequent value of generating functions and transforms, as we shall see again in many other problems. At present, however, we shall find it convenient to study further the solution of linear difference equations of the type (3) or (4), when the matrix of coefficients $(q_{ij}) = \underset{\sim}{Q}$ is constant, and the number of values over which _i_ or _j_ runs is finite, and equal to m.

2. The equation $\underset{\sim}{p}_{k+1} = \underset{\sim}{Q}\underset{\sim}{p}_k$.

In the last section $\underset{\sim}{Q}$ denoted a conditional probability distribution, so that, as is the binomial example, the sum of the elements of each column represents the sum of the conditional probabilities of X_{k+1} for given x_k and is consequently unity. However, since the equation

$$\underset{\sim}{p}_{k+1} = \underset{\sim}{Q}\underset{\sim}{p}_k \tag{1}$$

will often arise without $\underset{\sim}{p}_k$ necessarily denoting a probability distribution, we shall not yet impose this condition. When $\underset{\sim}{Q}$ is constant and given, equation (1) has the obvious but important solution

$$\underset{\sim}{p}_k = \underset{\sim}{Q}^k\, \underset{\sim}{p}_o, \tag{2}$$

this being the special case of the solution

$$\underset{\sim}{p}_k = \prod_{i=1}^{k} \underset{\sim}{Q}_{i,\,i-1}\, \underset{\sim}{p}_o \tag{3}$$

when $\underset{\sim}{Q}$ is not constant, the matrix product denoting the _ordered_ product of the $\underset{\sim}{Q}$'s. To examine further the structure of the solution (2), we shall recapitulate the main features of the 'spectral resolution' of the matrix $\underset{\sim}{Q}$. For simplicity we shall omit the degenerate case when the characteristic equation

$$\left| Q - \lambda \right| = 0, \tag{4}$$

where $\left| Q - \lambda \right|$ is the determinant of $(q_{ij} - \lambda\delta_{ij})$, has zero or equal latent roots (for the case of repeated roots, see, for example C 2).

Corresponding to the root λ_i of (4), we define latent column and row vectors s_i, t_i' respectively, where

$$Q\, s_i = \lambda_i s_i, \qquad t_i'\, Q = t_i' \lambda_i \tag{5}$$

or if

$$S \equiv (s_1, s_2, \dots s_m),$$
$$T \equiv (t_1, t_2, \dots t_m),$$

where t_i is the transpose of t_i',

$$Q\, S = S \Lambda, \qquad T'\, Q = \Lambda T', \tag{6}$$

where Λ is the diagonal matrix of roots λ_i. Hence

$$Q = S \Lambda S^{-1} = (T')^{-1} \Lambda\, T'. \tag{7}$$

Since

$$t_j'\, Q\, s_i = t_j'(\lambda_i s_i) = (t_j' \lambda_j)\, s_i,$$
$$t_j'\, s_i = 0 \text{ for } j \neq i.$$

If we choose also the scale of t_j' and s_i so that

$$t_j'\, s_j = 1,$$

we have $T'S = 1$, and (7) may be written

$$Q = S \Lambda T' = \sum_{i=1}^{m} \lambda_i\, s_i\, t_i' \tag{8}$$

From the properties of the matrices

$$A_i \equiv s_i\, t_i',$$

namely,

$$A_i A_j = \begin{cases} 0 & (i \neq j), \\ A_i, & (i = j), \end{cases} \quad \text{and} \quad \sum_{i=1}^{m} A_i = 1,$$

we have

$$Q^k = \sum_{i=1}^{m} \lambda_i^{\,k}\, A_i. \tag{9}$$

Now the s_i form a complete linearly independent set, i.e., there is no relation

$$c_1\underset{\sim}{s}_1 + c_2\underset{\sim}{s}_2 + \ldots c_m\underset{\sim}{s}_m = 0,$$

for if there were, we should obtain by multiplying through by any $\underset{\sim}{t}_j'$, that $c_j = 0$ for all j. It is thus always possible to express the initial vector $\underset{\sim}{p}_o$ in terms of the vectors $\underset{\sim}{s}_i$; in fact

$$\underset{\sim}{p}_o = \sum_{i=1}^{m}(\underset{\sim}{t}_i' \underset{\sim}{p}_o)\underset{\sim}{s}_i = \sum_{i=1}^{m}\alpha_i\underset{\sim}{s}_i, \text{ say.}$$

Hence our solution $\underset{\sim}{p}_k$ has the structure

$$\underset{\sim}{p}_k = \sum_{i=1}^{m}\alpha_i\lambda_i^k\underset{\sim}{s}_i. \qquad (10)$$

To obtain formulae for $\underset{\sim}{s}_i$ or $\underset{\sim}{t}_i'$, we may note that, if the adjoint of a matrix $\underset{\sim}{M}$ is written adj $\underset{\sim}{M}$, we have

$$(\underset{\sim}{Q} - \lambda_i)\ \mathrm{adj}(\underset{\sim}{Q} - \lambda_i) = |\underset{\sim}{Q} - \lambda_i|\ (\delta_{ij}) = 0 \qquad (11)$$

Alternatively, it is known that any matrix $\underset{\sim}{Q}$ satisfies its own characteristic equation $\phi(\underset{\sim}{Q}) = 0$, or

$$\prod_{j=1}^{m}(\underset{\sim}{Q} - \lambda_j) = (\underset{\sim}{Q} - \lambda_i)\prod_{j\neq i}(\underset{\sim}{Q} - \lambda_j) = 0. \qquad (12)$$

Comparing (11) or (12) with the equation

$$(\underset{\sim}{Q} - \lambda_i)\ \underset{\sim}{s}_i = 0,$$

we note that a solution s_i can be obtained from any column of $\mathrm{adj}(\underset{\sim}{Q} - \lambda_i)$ or $\prod_{j\neq i}(Q - \lambda_j)$. Similarly $\underset{\sim}{t}_i'$ can be taken proportional to any row of either matrix, with the further adjustment

$$\underset{\sim}{t}_i'\ \underset{\sim}{s}_i = 1.$$

From the solution (10) we see that in general, when there is one root, λ_1 say, with largest absolute value, and $\alpha_1 \neq 0$, we have ultimately the asymptotic solution

$$p_k \rightarrow \alpha_1\lambda_1^k\ \underset{\sim}{s}_1 \qquad \text{as } k \rightarrow \infty. \qquad (13)$$

<u>Example:</u>

Consider the equation

$$\underset{\sim}{p}_k = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 2 & 2 \\ 1 & 1 & 2 \end{pmatrix}^k \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

We easily find $\lambda_1 = 4$, $\lambda_2 = 2$, $\lambda_3 = 1$.

We also find

$$(\underset{\sim}{Q} - 2)(\underset{\sim}{Q} - 1), \qquad (\underset{\sim}{Q} - 4)(\underset{\sim}{Q} - 1), \qquad (\underset{\sim}{Q} - 4)(\underset{\sim}{Q} - 2)$$

$$\begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}, \quad \begin{pmatrix} -2 & 0 & 2 \\ 2 & 0 & -2 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & -1 & 2 \\ 2 & 2 & -4 \\ -1 & -1 & 2 \end{pmatrix}.$$

Hence we may take

$$\underset{\sim}{s}_1 = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}, \qquad \underset{\sim}{s}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \qquad \underset{\sim}{s}_3 = \begin{pmatrix} -1/3 \\ 2/3 \\ -1/3 \end{pmatrix},$$

$$\underset{\sim}{t}_1' = (1/3, \quad 1/3, \quad 1/3)$$

$$\underset{\sim}{t}_2' = (-1, \quad 0, \quad 1)$$

$$\underset{\sim}{t}_3' = (1/3, \quad 1/3, \quad -2/3).$$

We obtain, since

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1/3 \\ 2/3 \\ -1/3 \end{pmatrix}$$

the solution

$$\underset{\sim}{p}_k = 4^k \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} - 2^k \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1/3 \\ 2/3 \\ -1/3 \end{pmatrix}$$

$$\to 4^k \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \quad \text{asymptotically as } k \to \infty.$$

It will be seen that the existence of a limiting set of ratios for the elements of $\underset{\sim}{p}_k$ no longer holds if more than one root has the same largest modulus (or perhaps if $\alpha_1 = 0$).

3. The Case of Probability Distributions.

In the case when $\underset{\sim}{Q}$ represents a conditional probability distribution and $\underset{\sim}{p}_k$ a probability distribution, we have already noted that the sum of the elements in each column of $\underset{\sim}{Q}$, and the sum of the elements of $\underset{\sim}{p}_k$, are necessarily unity. Consider the latent roots of $\underset{\sim}{Q}$ for this case. If we add all the rows of the characteristic equation determinant to form a new first row, we obtain for each element of the first row, $1 - \lambda$, whence $\lambda = 1$ is a latent root. It is obvious that no root can have modulus greater than 1 if $\underset{\sim}{p}_k$ is to remain a probability distribution; if all other roots have modulus less than 1, the latent vector corresponding to the root 1 will give the limiting distribution $\underset{\sim}{p}_k$ as $k \to \infty$.

We may illustrate that no limiting distribution need exist by considering the important case of the purely deterministic or permuting type* of matrix $\underset{\sim}{Q}$. We shall suppose that the matrix $\underset{\sim}{Q}$ has one non-zero element in each column, so that these non-zero elements are necessarily unity. We shall suppose for simplicity that $\underset{\sim}{Q}$ does not break up into closed groups of permutations, and it is then always possible to define the order of the 'states' 0, 1, 2, ... m of the random variable X so that the non-zero elements occur immediately to the right of the diagonal elements. Thus

---

*The particular case illustrated in class was for $m = 2$, $\lambda = \pm 1$.

$$\underset{\sim}{Q} = \begin{pmatrix} 0 & 1 & 0 & 0 \dots \\ 0 & 0 & 1 & 0 \dots \\ 0 & 0 & 0 & 1 \dots \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix}$$

and the characteristic equation becomes $\lambda^m = 1$, with the m roots of unity,

$$w_j = \exp 2\pi ji/m \qquad (j = 1 \dots m).$$

Thus

$$\underset{\sim}{p}_k = \sum_{j=1}^{m} \alpha_j \, w_j^{\,k} \, s_j$$

permanently, representing a process in which the 'path' followed out by each element of the initial probability distribution is deterministic, moving cyclically through the different 'states' until after m permutations it returns to its initial state, after which the whole cycle is repeated.

By evaluating adj$(\underset{\sim}{Q} - w_k)$ we readily find that $\underset{\sim}{s}_j{}'$ is proportional to $(1 \; w_j \; w_j^{\,2} \dots w_j^{\,m-1})$, and $\underset{\sim}{t}_j{}'$ to $(1 \;\; w_j^{\,-1} \; w_j^{\,-2} \dots w_j^{\,-(m-1)})$, or what is the same thing, to $(1 \;\; \bar{w}_j \; \bar{w}_j^{\,2} \dots \bar{w}_j^{\,m-1})$, where $\bar{w}_j$ is the complex conjugate of w_j. Thus $\underset{\sim}{t}_j{}'$ is the complex conjugate of $\underset{\sim}{s}_j{}'$, provided each is adjusted in scale by the same factor 1/m.

4. An Example From Genetics.

As a simple illustration of problems involving the equation

$$\underset{\sim}{p}_k = \underset{\sim}{Q}^k \underset{\sim}{p}_o$$

consider the selection and inbreeding for pure lines in genetics. For a single pair of allelomorphs A and a, the most powerful form of inbreeding from the heterozygote Aa, possible in certain types of plant, is by self-fertilization. Since the mating Aa x Aa gives rise to

$$(\tfrac{1}{2}A + \tfrac{1}{2}a)(\tfrac{1}{2}A + \tfrac{1}{2}a) = \tfrac{1}{2}Aa + (\tfrac{1}{4}AA + \tfrac{1}{4}aa), \qquad (1)$$

or 50 per cent heterozygotes, we have the proportion of heterozygotes reduced by one half at each generation. For one of the next most powerful methods, practiced with some animals, of brother-sister mating, we shall classify the different possible situations in terms of the possible matings. From a mating of the progeny represented by (1), we obtain

$$(\tfrac{1}{4}AA + \tfrac{1}{2}Aa + \tfrac{1}{4}aa) \times (\tfrac{1}{4}AA + \tfrac{1}{2}Aa + \tfrac{1}{4}aa).$$

Similarly from Aa x aa, we obtain matings in the next generation of

$$(\tfrac{1}{2}A + \tfrac{1}{2}a)a \times (\tfrac{1}{2}A + \tfrac{1}{2}a)a.$$

In this way our classification becomes:

Progeny mating / Parent mating	aa x aa AA x AA	aa x Aa AA x Aa	Aa x Aa	AA x aa
aa x aa AA x AA	1	1/4	1/8	0
aa x Aa AA x Aa	0	1/2	1/2	0
Aa x Aa	0	1/4	1/4	1
AA x aa	0	0	1/8	0

Treating the table as our $\underset{\sim}{Q}$ matrix, we find for its characteristic or latent roots

$$(\lambda - 1)(\lambda - \tfrac{1}{4})(\lambda^2 - \tfrac{1}{2}\lambda - \tfrac{1}{4}) = 0$$

or $$\lambda_1 = 1, \quad \lambda_{2,3} = \frac{1 \pm \sqrt{5}}{4}, \quad \lambda_4 = \tfrac{1}{4}.$$

In this problem we are most interested in the proportion of heterozygotes in any generation. If we call the probabilities for the above classes of mating in the table p, q, r, and s (in order), the proportion of heterozygotes will be

$$k = r + \tfrac{1}{2}q.$$

Hence, without finding the complete solution for p, q, r, and s, we may write for h alone,

$$h_k = A\left(\frac{1 + \sqrt{5}}{4}\right)^k + B\left(\frac{1 - \sqrt{5}}{4}\right)^k + C\left(\tfrac{1}{4}\right)^k \qquad (2)$$

and determine A, B and C from the values h_1, h_2, h_3, which are readily found to be 1, $\frac{1}{2}$, $\frac{1}{2}$ if $r_1 = 1$. (We have taken advantage in (2) of the obvious fact that there will be no term $D(1)^k$ in h_k, but of course this need not be assumed).

5. Population Growth.

As another application of our basic difference equation, we shall make use of it in the theory of population growth, where it no longer refers to probabilities, but to expected numbers. A fuller explanation of the meaning of the equation in this case, and its place in a more complete stochastic process theory of population growth, will be given later. At present we shall content ourselves with setting up the usual equation, discussing it first of all in discrete terms (see C 2).

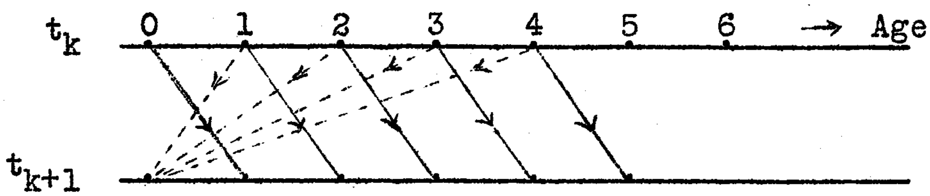
We consider the change in numbers from one unit of time to the next, for simplicity, considering the female population alone and using age-groups corresponding to our time-unit. We adopt the following notation:

$m_{i,k}$ -- the expected number of females in age-group i, i + 1 at time t_k.

p_i -- the probability that a female aged i to i+1 at time t_k will be alive at time t_{k+1}.

F_i -- the expected number of daughters born in the interval t_k to t_{k+1} per female aged i to i+1 at time t_k, and alive in age-group 0 to 1 at time t_{k+1}.

We assume that the mortality and fertility rates remain constant; (this does not strictly imply that the rates over finite age-groups will be constant, owing to the changing composition of the population, but we shall also assume that our time-unit is fine enough for these rates to be effectively constant). A schematic picture of changes in the population is then as follows:



(The continuous sloping lines represent the natural aging of the population, the dotted lines represent births).

We have the equations:

$$\left.\begin{aligned} m_{0,k+1} &= \Sigma_i \; F_i \; m_{i,k}, \\ m_{i+1,k+1} &= P_i \; m_{i,k}, \quad (i \geqslant 0), \end{aligned}\right\} \qquad (1)$$

or in our matrix notation,

$$\mathbf{m}_{k+1} = \mathbf{Q} \, \mathbf{m}_k \qquad (2)$$

where

$$\mathbf{Q} = \begin{pmatrix} F_0 & F_1 & F_2 & \cdots \\ P_0 & 0 & 0 & \\ 0 & P_1 & 0 & \\ \vdots & & \ddots & \\ \vdots & \vdots & & \ddots \\ \vdots & \vdots & \vdots & \end{pmatrix}. \qquad (3)$$

The matrix will be closed owing to a maximum age limit; if we are mainly interested in the number of births, we may effectively close it earlier at the end of the reproductive age r, for we shall have corresponding to this limit a partition of $\mathbf{Q}$ of the form

$$\mathbf{Q} \equiv \left(\begin{array}{c|c} \mathbf{A} & 0 \\ \hline \mathbf{B} & \mathbf{C} \end{array}\right)$$

and it follows that

$$\mathbf{Q}^2 = \left(\begin{array}{c|c} \mathbf{A}^2 & 0 \\ \hline \mathbf{BA}+\mathbf{CB} & \mathbf{C}^2 \end{array}\right)$$

etc, so that if we denote the expected female population up to the reproductive age limit by $\mathbf{n}$, we have

$$\mathbf{n}_{k+1} = \mathbf{A}\,\mathbf{n}_k. \tag{4}$$

The solution

$$\mathbf{n}_k = \mathbf{A}^k\,\mathbf{n}_o \tag{5}$$

may be obtained by direct computational methods, or by the investigation of the characteristic roots of $\mathbf{A}$. It is, however, convenient to transform from the reference system $\mathbf{n}$, which given the age-distribution at any time, to a new reference system $\mathbf{l}$, where

$$\mathbf{l} = \mathbf{H}\,\mathbf{n}, \tag{6}$$

and

$$\mathbf{H} = \begin{pmatrix} P_oP_1\ldots P_{r-1} & 0 & \ldots\ldots & 0 \\ 0 & P_1P_2\ldots P_{r-1} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \cdot\; P_{r-1} & 0 \\ 0 & 0\;\ldots & 0 & 1 \end{pmatrix}. \tag{7}$$

For any non-singular transformation, we have

$$\mathbf{l}_{k+1} = \mathbf{H}\,\mathbf{n}_{k+1} = \mathbf{HA}\,\mathbf{n}_k = \mathbf{HAH}^{-1}\mathbf{l}_k = \mathbf{L}\,\mathbf{l}_k,$$

where $\mathbf{L} = \mathbf{HAH}^{-1}$.

We have here

$$\mathbf{L} = \begin{pmatrix} F_o & F_oF_1 & P_oP_1F_2 & \ldots & P_oP_1\;.\;.\;P_{r-1}F_r \\ 1 & 0 & 0 & & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & \ldots & \ddots & \vdots \\ 0 & 0 & 0 & & 1 \end{pmatrix}. \tag{8}$$

The new reference system has computational advantages over the original system. To interpret it, we notice that

$$\underset{\sim}{l}_k = \begin{pmatrix} P_oP_1\ldots P_{r-1}m_{o,k} \\ P_1\ \ldots P_{r-1}m_{1,k} \\ \vdots \\ m_{r,k} \end{pmatrix} = \begin{pmatrix} m_{r,k+1} \\ m_{r,k+r-1} \\ \vdots \\ m_{r,k} \end{pmatrix} = P_oP_1\ldots P_{r-1}\begin{pmatrix} m_{o,k} \\ m_{o,k-1} \\ \vdots \\ m_{o,k-r} \end{pmatrix}$$

so that the new system of equations simply corresponds to the relation

$$m_{o,k+1} = \sum_{i=o}^{r}(P_oP_1\ldots P_{i-1}F_i)m_{o,k-i} \tag{9}$$

connecting the expected births at time t_{k+1}, with the expected births at preceding times.

The latent roots for the $\underset{\sim}{L}$ (or $\underset{\sim}{A}$) matrix are given by the characteristic equation

$$\lambda^{r+1} - F_o\lambda^r - P_oF_1\lambda^{r-1}\ldots. - P_oP_1\ldots.P_{r-1}F_r = 0, \tag{10}$$

or equivalently, by

$$\sum_{i=o}^{r}(P_oP_1\ldots P_{i-1}F_i)\lambda^{-(i+1)} = 1. \tag{11}$$

Since there is only one change in sign in (10), there is only one real positive root, and this will usually in practice be the root with largest modulus. It will obviously from (11) be $\gtreqless 1$, according as the sum of elements

$$\sum_{i=o}^{r}(P_oP_1\ \ldots\ P_{i-1}F_i) \gtreqless 1.$$

If

$$\underset{\sim}{L} = \sum_{i=1}^{r+1}\lambda_i\ \underset{\sim}{s}_i\ \underset{\sim}{t}_i{}',$$

we have seen that if there is a single root λ_1 with largest modulus, then asymptotically

$$\underset{\sim}{l}_k \to \lambda_1{}^k\alpha_1\ \underset{\sim}{s}_1, \tag{12}$$

where $$\underset{\sim}{l}_o = \sum_{i=1}^{r+1}\alpha_i\ \underset{\sim}{s}_i.$$

We note also that

$$\underset{\sim}{L}\begin{pmatrix}\lambda_i^{\,r}\\ \lambda_i^{\,r-1}\\ \cdot\\ \cdot\\ \cdot\\ 1\end{pmatrix} = \begin{pmatrix}F_o\lambda_i^{\,r}+P_oF_1\lambda_i^{\,r-1}\ldots\ P_oP_1\ldots\ P_{r-1}F_r\\ \lambda_i^{\,r}\\ \cdot\\ \cdot\\ \cdot\\ \lambda_i\end{pmatrix} = \begin{pmatrix}\lambda_i^{\,r+1}\\ \lambda_i^{\,r}\\ \cdot\\ \cdot\\ \cdot\\ \lambda_i\end{pmatrix}$$

$$= \lambda_i\begin{pmatrix}\lambda_i^{\,r}\\ \lambda_i^{\,r-1}\\ \cdot\\ \cdot\\ \cdot\\ 1\end{pmatrix},$$

whence $\underset{\sim}{s_i}'$ is proportional to $(\lambda_i^{\,r},\lambda_i^{\,r-1}\ldots\ 1)$.

To illustrate the type of data which might be handled by these methods, we quote the following table from C 2, which should be consulted for further details. It is suggested by Leslie as a possible table of mortality and fertility rates for the brown rat under optimum breeding conditions. The largest root corresponding to these figures is 1.56246, indicating the limiting rate of increase.

i	A		L
(30 day unit)	P_i	F_i	1st row
0 –	0.94697	–	–
1 –	0.99665	–	–
2 –	0.99926	0.3964	0.3741
3 –	0.99899	1.4939	1.4089
4 –	0.99863	2.1777	2.0517
5 –	0.99817	2.5250	2.3756
6 –	0.99753	2.6282	2.4682
7 –	0.99667	2.6749	2.5059
8 –	0.99553	2.6018	2.4293
9 –	0.99399	2.4419	2.2698
10 –	0.99196	2.1865	2.0202
11 –	0.98926	1.9044	1.7454
12 –	0.98572	1.7259	1.5648
13 –	0.98107	1.4918	1.3332
14 –	0.97511	1.2415	1.0885
15 –	0.96748	0.9522	0.8141
16 –	0.95797	0.7141	0.5907
17 –	0.94631	0.4618	0.3659
18 –	0.93247	0.2518	0.1888
19 –	0.91649	0.0901	0.0630
20 –		0.0035	0.0022

6. Industrial Renewal Theory.

In the related application to industrial renewal theory the basic equation we have been using refers again directly to probabilities. We suppose that a stock of industrial articles have a life-time subject to random variation, each article as it wears out being replaced. The 'death distribution' giving the probability of a new article wearing out after a time t_k to t_{k+1} is supposed known. Lotka (C 3) quotes the following example from Kurtz.

Age Interval	Survival Chance to Beginning of Period	Frequency Distribution of 'deaths'
0 - 1	1	-
1 - 2	1	-
2 - 3	1	0.003
3 - 4	0.997	0.009
4 - 5	0.988	0.018
5 - 6	0.970	0.030
6 - 7	0.940	0.057
7 - 8	0.883	0.103
8 - 9	0.780	0.141
9 - 10	0.639	0.139
10 - 11	0.500	0.138
11 - 12	0.362	0.132
12 - 13	0.230	0.104
13 - 14	0.126	0.063
14 - 15	0.063	0.037
15 - 16	0.026	0.022
16 - 17	0.004	0.004
17 - 18	-	-

In this problem the P's and F's of the last section are defined in terms of data of the type quoted by the relations:

$P_o P_1 \ldots P_i$ is the survival chance to the beginning of the period $t_i - t_{i+1}$, $(i = 0, 1, 2, \ldots)$.

$F_i = 1 - P_i$ is the expected replacement due to 'deaths' in the period $t_i - t_{i+1}$.

If we transform as before to the $\underset{\sim}{L}$ and $\underset{\sim}{l}$ system, we obtain for the first row of $\underset{\sim}{L}$ the quantity

$P_o P_1 \ldots P_{i-1} F_i = P_o P_1 \ldots P_{i-1}(1 - P_i), = f_i$, say, which represents the chance of surviving to the beginning of the period $t_{i-1} - t_i$ and then 'dying' (wearing out) during the $t_i - t_{i+1}$ interval. This is just the 'death distribution' illustrated in the last colums of the tabulated data. We obtain as our equation for the replacements, (assuming for definiteness these are made at the end of each time-interval),

$$l_{o,k+1} = \sum_{i=o}^{r} f_i l_{o,k+1}. \tag{1}$$

As an alternative to handling this equation by the methods previously outlined, we may note that if we define the first 'generation' of replacements as the replacements which came in directly for the original articles, the second generation as those coming in to replace the first generation and so on, then for the j+1-th generation we have a similar equation to (1), but with the jth generation on the right hand side, i.e.,

$$l_{o,k+1}(j + 1) = \sum_{i=o}^{r} f_i l_{o,k-i}(j). \tag{2}$$

Now the right hand side of this equation is mathematically the 'convolution' of f and $l_o(j)$, ($l_{o,k-i}(j) = 0$ for $k - i < 0$), that is, if we form the corresponding moment-generating function transform of both sides,

$$M_{j+1}(\theta),\text{ say, } = M(\theta)M_j(\theta),$$

where $M(\theta)$ is the m.g.f. of the distribution f_i; and hence by repeating the argument, we obtain finally

$$M_j(\theta) = M^j(\theta)M_0 = M^j(\theta) \tag{3}$$

if $M_0 = 1$, i.e., the original stock was new at time zero. This result (3) shows that the distribution of the jth generation of replacements increases its cumulants proportionately to j, since

$$K_j(\theta) = jK(\theta), \tag{4}$$

and will tend to normality as j increases.

Example.

To take a simpler numerical example than the data quoted, suppose the f_i distribution is

0 – 1	1 – 2	2 – 3	3 – 4	4 –
0.2	0.4	0.2	0.2	–

with mean 2.4 (taking the time values at the ends of the intervals) and variance 1.04.

Proceeding by direct computational methods, we construct the A matrix as

$$\begin{pmatrix} 0.2 & 0.5 & 0.5 & 1.0 \\ 0.8 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{pmatrix}$$

and operate repeatedly on the initial vector with first component unity. To obtain in addition for interest the contributions from the different generations, we may shift the first component of the new vector to another column to indicate that it denotes a new generation. In this way the following complete enumeration of the age-distributions for each generation, and the corresponding age-distribution for all articles, was rapidly obtained. The expected total replacements at each time are underlined.

AGE DISTRIBUTION BY GENERATIONS

	Total	Zero	1st	2nd	3rd	4th	5th
t_0	0-1: 1 1-2: 0 2-3: 0 3-4: 0	1 0 0 0					
t_1	$\underline{.2}$ .8 0 0	0 .8 0	.2				
t_2	$\underline{.44}$ .16 .40 0	0 0 .4	.40 .16	.04			
t_3	$\underline{.368}$ .352 .080 .200	0 0 0 .2	.20 .32 .08	.160 .032	.008		
t_4	$\underline{.4896}$ .2944 .1760 .0400		.20 .16 .16 .04	.240 .128 .016	.0480 .0064	.0016	
t_5	$\underline{.37312}$ .39168 .14720 .08800		 .16 .08 .08	.240 .192 .064 .008	.1200 .0384 .0032	.01280 .00128	.00032

After we have four periods, the single equation (1) (or (2) for individual generations) may be used to obtain further values. For example,

$$\underline{.37312} = .2(.4896) + .4(.368) + .2(.44) + .2(.2).$$

Completing the corresponding values for the second generation we obtain also the figures completing the distribution of second generation replacements viz

t_6	t_7	t_8
.200	.080	.040 ,

whence we may check that the mean and variance for the entire distribution of the second generation are 4.8 and 2.08, just twice the values for the distribution of the first generation.

This direct method is more useful here than the characteristic root technique, though of course this may be studied if desired. Thus we find the characteristic equation

$$(\lambda - 1)(\lambda^3 + 0.8\lambda^2 + 0.4\lambda + 0.2) = 0$$

with, besides the unit root, one about -2/3 and two complex.

Since there is only one unit root, the limiting expected rate of replacement will tend to a constant value which must obviously be the reciprocal of the mean life-time of a new article, i.e., 1/4.8.

7. Transition to Continuous Time.

It may be noted with the single equation we obtained connecting births at different times in the theory of population growth, and with the corresponding equation in the theory of industrial renewal, that if we let the time interval approach zero, the summation on the right will become equivalent to an integration. Equation (2) of the last section, for example, becomes replaced by

$$l_o(t,\ j+1) = \int_o l_o(t - T, j)\ dF(T), \tag{1}$$

where $l_o(t)$ is the rate* of renewal, and $l_o(t,j)$ the component rate of renewal corresponding to the jth generation. From (1) we still have, however,

$$M_j(\theta) = M^j(\theta). \tag{2}$$

For example, consider the special case when the chance of 'death' is independent of age. The distribution function $F(t)$ is well-known in this case to be given by

*More generally, where this rate does not remain finite, we may consider the integrated rate or total renewal.

$$1 - F(T) = \underset{h \to 0}{Lt}\,(1 - \lambda h)^{T/h} = e^{-\lambda T} \qquad (T \geq 0).$$

and

$$M(\theta) = \int_0^\infty \lambda e^{\theta T} e^{-\lambda T}\, dT = (1 - \theta/\lambda)^{-1}.$$

Hence

$$M_j(\theta) = (1 - \theta/\lambda)^{-j},$$

which corresponds to the frequency function

$$(\lambda T)^{j-1} e^{-\lambda T} \lambda dT/\Gamma(j).$$

The total rate of replacements at time t will be

$$\sum_{j=1} l_0(t,j) = \sum_{j-1=0}^{\infty} (\lambda T)^j\, e^{-\lambda T}/j!$$
$$= \lambda,$$

which is the reciprocal of the mean value of the distribution $F(T)$. In this case, we thus check the obvious solution that the expected rate of replacement is constant from the start, in contrast with the usual limiting stability after initial oscillations.

Similarly in the theory of population growth we now obtain the integral equation

$$m_0(t) = \int_0 m_0(t - T)d\Phi(T), \qquad (3)$$

where if $s(T)$ is the survival chance of females up to time T, and $r(T)$ the reproduction rate at time T, the effective 'reproduction rate' from such females will be $s(T)r(T) = \ (T)$, and $\ (T)dT = d\Phi(T)$.

If $r(T)$ is zero for $T > T'$, we may write (3) in the form

$$m_0(t) = \int_t^{T'} m_0(t - T)d\Phi(T) + \int_0^t m_0(t - T)d\Phi(T)$$

$$= \mu(t) + \int_0^t m_0(t - T)d\Phi(T) \qquad (4)$$

where $\mu(t)$ is defined by values of $m_0(t)$ for $t < 0$.

This integral equation is of standard form, and its solution in connection with the present problem has been discussed by several writers (see, for example, Lotka (C 3), Feller (C 1)). A common method of solution represents a direct analogue of the characteristic root method with the discrete form. For example, equation (11) of section 5, if we write $r = \log \lambda$, goes over into the equation

$$\int_0^\infty e^{-rt} d\Phi(T) = 1, \qquad (5)$$

and we would anticipate that the asymptotic behaviour of our solution would depend on the root r_1 with the largest modulus. Since $\Phi(T)$ is an increasing function, equation (5) has only one real root, which is $\gtreqless 0$ according as

$$R = \int_0^\infty d\Phi(T) \gtreqless 1,$$

and which in general determines the ultimate behaviour of the solution. In order to investigate the roots of (5), it is customary to fit a frequency function to $\Phi(T)/R$, and further express the solution as a series of the type $\Sigma_i \, a_i e^{r_i t}$, where r_i are the roots of (5) (cf. C 3), but the direct computational methods available in the discrete case are often preferable; there are moreover dangers in deriving the present series type of solution, and before it is used in any problem the student should consult Feller's rigorous discussion (C 1).

*8. Solution to an Integral Equation.

For reference it is convenient to quote here (without proof) an abbreviated form of some of Feller's theorems. We shall consider the solution here only in the case when the functions $\mu(t)$ and $\phi(t)$ in the integral equation

$$m(t) = (t) + \int_0^t m(t - T)\, \phi(T) dT, \qquad (1)$$

are bounded (and non-negative).

(i) Let

$$\begin{cases} \Psi(r) = \int_0^\infty e^{-rT}\, \phi(T)\, dT \\ \alpha(r) = \int_0^\infty e^{-rT}\, \mu(T)\, dT \end{cases}$$

converge for $r > \sigma$. Then there is a unique (non-negative) solution for $m(t)$ of (1), bounded in every finite interval. If further $\lim_{r \to \sigma+0} \Psi(r) > 1$, let $\sigma' > \sigma$ be the (real) root of $\Psi(r) = 1$; otherwise, when $\lim_{r\to\sigma+0}\Psi(r) \leq 1$, write $\sigma' = \sigma$. Then for $r > \sigma'$, we have

$$\beta(r) = \frac{\alpha(r)}{1 - \Psi(r)}, \tag{2}$$

where $\beta(r)$ is the Laplace integral

$$\beta(r) = \int_0^\infty e^{-rT} m(T)\, dT.$$

(ii) To consider the asymptotic behaviour of the solution we assume further that

$$\begin{cases} \int_0^\infty \mu(T) dT = a, \\ \int_0^\infty \phi(T) dT = R. \end{cases}$$

It is noted that the 'mean value' $t^{-1}\int_0^t m(T)dT$ may converge to a limit as t increases even if m(t) does not. Feller shows that for a 'mean value' limit C to exist, it is necessary and sufficient that R = 1, and

$$\int_0^\infty T\, \phi(T)dT = \nu_1 \text{ (finite)};$$

then $C = a/\nu_1$.

For $R \neq 1$, we have the results

$$\int_0^\infty m(T)dT = a/(1 - R) \text{ if } R < 1.$$

If $R > 1$, let σ' be the positive root of $\Psi(r) = 1$. Then the 'mean value' limit of n(t) is a/ν_1', where

$$n(t) = e^{-\sigma' t} m(t); \quad \int_0^\infty e^{-\sigma' T}\, T\phi(T)dT = \nu_1'.$$

(iii) When R = 1, if further

$$\int_0^\infty T^2\, \phi(T)dT = \nu_2 \text{ (finite)},$$

$\phi(t)$ is of bounded total variation in $(0,\infty)$, and $\lim_{t\to\infty}\mu(t) = 0$, then

$$\lim_{t \to \infty} m(t) = a/\nu_1 \tag{3}$$

(iv) Finally, in order that m(t) (defined by (2)) can be represented by the series solution

$$m(t) = \Sigma_i \; c_i e^{r_i t} ,$$

the series converging absolutely for $t = 0$, and r_i being the roots of $\psi(r) = 1$, it is necessary and sufficient that

$$\beta(r) = \frac{\alpha(r)}{1 - \psi(r)} = \Sigma_i \frac{c_i}{r - r_i} \tag{4}$$

and that $\Sigma_i |c_i|$ converge absolutely. The coefficients c_i are given by

$$c_i = \frac{-d\alpha(r_i)}{dr_i} \quad \psi'(r_i). \tag{5}$$

In particular it is necessary that $\beta(r)$ is a one-valued function.

9. Fluctuations.

A further caution in interpreting the solution of the integral equation (1) of the previous section is, however, advisable, especially before any excessive attention to its limiting form as $t \to \infty$ is given for our particular applications. It must not be regarded as more than an expectation equation, and does not tell us how many individuals will ultimately exist, starting from any finite number of individuals. This will become more apparent when we set up complete stochastic equations for situations of this kind, but it may be as well if we illustrate it first by a very simple example, returning for convenience to a discrete model.*

*Another example, relating to the type of problem discussed in sec. 4, is given in my paper in J. Genetics, 35(1937).

Suppose the population growth equation for a simple organism takes the form

$$\begin{pmatrix} m_{o,k+1} \\ m_{1,k+1} \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} m_{o,k} \\ m_{1,k} \end{pmatrix}, \qquad (1)$$

corresponding to an expectation of two offspring in the second time-interval, and a chance of survival to the second (reproductive) time-interval of $\frac{1}{2}$. Such a population is obviously non-increasing as far as m is concerned, the solution in the case $(m_{o,o}\ m_{1,o}) = (\frac{1}{2}n, \frac{1}{2}n)$ being easily found to be

$$\begin{pmatrix} m_{o,k} \\ m_{1,k} \end{pmatrix} = \tfrac{3}{4}n \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix} - \tfrac{1}{4}n(-1)^k \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix}. \qquad (2)$$

Suppose for our complete stochastic model we assume that the number of offspring is indeed fixed at two, but that the chance of survival to the reproductive period is an independent chance for each individual, then if the corresponding <u>actual</u> numbers of individuals corresponding to m are N, we have for the joint moment-generating function of N_o and N_1,

$$\begin{aligned} M_{k+1}(\theta_o,\theta_1) &\equiv E\left\{\exp(N_{o,k+1}\ \theta_o + N_{1.k+1}\ \theta_1)\right\} \\ &= E\left\{\exp(2N_{1,k}\ \theta_o)\cdot(\tfrac{1}{2} + \tfrac{1}{2}e\ \theta_1)N_{o,k}\right\} \\ &= M_k(\log\left[\tfrac{1}{2} + \tfrac{1}{2}e^{\theta_1}\right],\ 2\theta_o), \qquad (3) \end{aligned}$$

a functional equation for M_k, or alternatively, by writing $e^{\theta} = z$, for Π_k. Taking logarithms, we have

$$K_{k+1}(\theta_o,\theta_1) = K_k(\log\left[\tfrac{1}{2} + \tfrac{1}{2}e^{\theta_1}\right],\ 2\theta_o), \qquad (4)$$

whence by expansion,

$$\begin{aligned} &m_{o,k+1}\theta_o + m_{1,k+1}\theta_1 + \tfrac{1}{2}v_{oo,k+1}\theta_o^{\,2} + v_{o1,k+1}\theta_o\theta_1 + \\ &\tfrac{1}{2}v_{11,k+1}\theta_1^{\,2} + \dots \\ &\quad = m_{o,k}\log\left[1 + \tfrac{1}{2}(e^{\theta_1} - 1)\right] + 2m_{1,k}\theta_o + \\ &\quad \tfrac{1}{2}v_{oo,k}\log^2\left[1 + \tfrac{1}{2}(e^{\theta_1} - 1)\right] + 2v_{o1,k}\theta_o\log\left[1+\tfrac{1}{2}(e^{\theta_1}-1)\right] \\ &\quad + 2v_{11,k}\theta_o^{\,2} + \dots \end{aligned}$$

or by equating coefficients of powers of θ_o and θ_1, we obtain

$$\begin{pmatrix} m_{o,k+1} \\ m_{1,k+1} \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} m_{o,k} \\ m_{1,k} \end{pmatrix},$$

reproducing again equation (1); and also for the variances,

$$\begin{pmatrix} v_{oo,k+1} \\ v_{11,k+1} \end{pmatrix} = \begin{pmatrix} 0 & 4 \\ \frac{1}{4} & 0 \end{pmatrix} \begin{pmatrix} v_{oo,k} \\ v_{11,k} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \frac{1}{4} & 0 \end{pmatrix} \begin{pmatrix} m_{o,k} \\ m_{1,k} \end{pmatrix}. \quad (5)$$

From (5), we have

$$\begin{aligned} v_{oo,k+2} &= v_{oo,k} + m_{o,k} \\ &= v_{oo,k-2} + m_{o,k} + m_{o,k-2} \\ &\quad \cdots \end{aligned}$$

or since $m_{o,k-2r}$ is constant, v_{oo} <u>increases indefinitely with</u> k.

We might note further, for the probability-generating function,

$$\Pi_{k+1}(z_o, z_1) = \Pi_k(\tfrac{1}{2} + \tfrac{1}{2}z_1, z_o^2), \quad (6)$$

whence

$$\begin{aligned} \Pi_{k+2r}(0,0) &= \Pi_{k+2r-2}(\tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2})^2,\ (\tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2})^2)^2) \\ &= \Pi_{k+2r-4}(\tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2})^2)^2, \\ &\qquad (\tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2})^2)^2)^2) \quad \cdots \end{aligned}$$

or since the limit of the steadily increasing sequence

$$u_1 = \tfrac{1}{2} + \tfrac{1}{2}(\tfrac{1}{2})^2, \quad u_2 = \tfrac{1}{2} + \tfrac{1}{2}u_1^2, \quad u_3 = \tfrac{1}{2} + \tfrac{1}{2}u_2^2, \ \ldots$$

is given by

$$x = \tfrac{1}{2} + \tfrac{1}{2}x^2,$$

or $x = 1$, we obtain

$$\mathop{\mathrm{Lt}}_{r \to \infty} \Pi_{k-2r}(0,0) = \Pi_k(1,1) = 1,$$

<u>which gives a probability of unity for entire extinction of the population after a long enough time</u>. These results

(cf. also section 10, and example (iii), section 11) show that it cannot be assumed that the mean of expected number in a finite population undergoing stochastic changes remains representative of the actual numbers that exist after a long time.

10. Further Discussion of the Population Growth Equations.*

From the results of the last section it will evidently be advisable to try to set up the complete stochastic equations for a population growth process. This will require a rather more explicity and therefore less general specification of the stochastic process assumed than was sufficient to obtain the integral equation of II 7. The linearity of this equation in $m_{o,t}$ ensures its validity as an expectation equation under fairly general conditions; however, it will be instructive to demonstrate this under particular assumptions and at the same time to stress again its limitations as a complete representation of the actual history of a finite population.

We shall consider the changes during a small interval Δt, and make the following assumptions:

(a) the chance of an individual, born in the interval $(t_r - \Delta t, t_r)$, and alive at time $t = t_s$, surviving to time $t + \Delta t - t_s + 1$ is $1 - \mu(T_r)\Delta t + o(\Delta t)$, where $\mu(T_r)$ is a function of $T_r = t_s - t_r$.

(b) the chance of an individual, born in the interval $(t_r - \Delta t, t_r)$ and alive at time $t = t_s$, giving birth to a new individual in the interval t_s, $t_s + \Delta t$ surviving at time $t_s + \Delta t$, is $\lambda(T_r)\ t + o(\Delta t)$.

*This discussion may alternatively (as in class), be taken after sections 11 and 12, but since the equations have been left in approximate 'discrete' form, the present order seems preferable.

(c) the number of individuals at $t = t_s$, born in the interval $(t_r - \Delta t, t_r)$, is denoted by $N_r(t_s)$.

The chance of multiple births, or the effect of gestation period, is neglected; all probabilities are assumed to act independently. The chance of a birth or a death during Δt for finite N and small Δt corresponding to existing individuals of particular age then becomes approximately proportional to the existing number N_r corresponding to that age, and the chance of more than one event (birth or death) is of higher order in Δt. Thus we may write, tracing the changes from t_s to t_{s+1},

$$M_{t_s} + \Delta t(\theta_r) = E_{N_r}\left\{\exp \sum_{r=-\infty}^{s+1} \theta_r N_r(t_s)\right\}$$

$$= E_{N_r}\left\{ \left(\sum_{r=-\infty}^{s} \left[N_r(t_s)\mu(T_r)\Delta t e^{-\theta_r} + N_r(t_s)\lambda(T_r)\Delta t e^{\theta_{s+1}}\right] + 1 - \sum_{r=-\infty}^{s} N_r(t_s)\left[\mu(T_r)\Delta t + \lambda(T_r)\Delta t\right] + o(\Delta t)\right) \times \exp \sum_{r=-\infty}^{s} \theta_r N_r(t_s)\right\}$$

or

$$\Delta M_{t_s} = \Delta t\left\{\sum_{r=-\infty}^{s} \left[\mu(T_r)(e^{-\theta_r} - 1) + \lambda(T_r)(e^{\theta_{s+1}} - 1)\right] N_r(t_s) \times \exp \sum_{r=-\infty}^{s} \theta_r N_r(t_s)\right\} + o(\Delta t)$$

$$= \Delta t\left\{\sum_{r=-\infty}^{s} \left[\mu(T_r)(e^{-\theta_r} - 1) + \lambda(T_r)(e^{\theta_{s+1}} - 1)\right] \frac{\partial}{\partial \theta_r}\right\} M_{t_s} + o(\Delta t) \qquad (1)$$

or to the same order of approximation

$$\Delta K_{t_s} \sim \Delta t\left\{\sum_{r=-\infty}^{s} \left[\mu(T_r)(e^{-\theta_r} - 1) + \lambda(T_r)(e^{\theta_{s+1}} - 1)\right] \frac{\partial}{\partial \theta_r}\right\} K_{t_s}. \qquad (2)$$

Expanding this equation in powers of θ_r, we obtain up to the terms of second degree;

$$\sum_{r=-\infty}^{s}(\Delta m_r)\theta_r + m_{s+1}\theta_{s+1} + \sum_{r=-\infty}^{s}(\Delta v_{rr})\tfrac{1}{2}\theta_r^2 + v_{s+1,s+1}\,\tfrac{1}{2}\theta_{s+1}^2$$

$$+\sum_{r=-\infty}^{s} v_{r,s+1}\,\theta_r\theta_{s+1} + \sum_{r\neq q}^{s}(\Delta v_{qr})\theta_q\theta_r$$

$$\sim t\sum_{r=-\infty}^{s}\Big\{\big[\mu(T_r)(-\theta_r + \tfrac{1}{2}\theta_r^2) + \lambda(T_r)(\theta_{s+1} + \tfrac{1}{2}\theta_{s+1}^2)\big]$$

$$\Big[m_r + \sum_{q=-\infty}^{s}\theta_q\, v_{rq}\Big]\Big\}$$

whence

$$\left.\begin{aligned} m_{s+1} &\sim \sum_{r=-\infty}^{s}\lambda(T_r)m_r\Delta t,\\ \Delta m_r &\sim -\mu(T_r)m_r\Delta t; \end{aligned}\right\} \tag{3}$$

$$\left.\begin{aligned} v_{s+1,s+1} &\sim \sum_{r=-\infty}^{s}\lambda(T_r)m_r\Delta t,\\ v_{s+1,q} &\sim \sum_{r=-\infty}^{s}\lambda(T_r)v_{qr}\Delta t, \quad (q \leqq s),\\ \Delta v_{qq} &\sim \big[\mu(T_q)m_q - 2\mu(T_q)v_{qq}\big]\Delta t, \quad (q \leqq s),\\ \Delta v_{pq} &\sim \big[-\mu(T_p) - \mu(T_q)\big]\, v_{pq}\Delta t, \quad (p \neq q, \leqq s). \end{aligned}\right\} \tag{4}$$

For the total population at time t_{s+1},

$$N = \sum_{r=-\infty}^{s+1} N_r = N_{s+1} + \sum_{r=-\infty}^{s} N_r, \tag{5}$$

we easily find from the above equations

$$\left.\begin{aligned} \Delta m &\sim \sum_{r=-\infty}^{s}\big[\lambda(T_r) - \mu(T_r)\big]\, m_r\Delta t,\\ \Delta v &\sim \sum_{r=-\infty}^{s}\big[\lambda(T_r) + \mu(T_r)\big]\, m_r\Delta t\\ &\quad + 2\sum_{r=-\infty}^{s}\sum_{q=-\infty}^{s}\big[\lambda(T_r) - \mu(T_r)\big]\, v_{qr}\Delta t. \end{aligned}\right\} \tag{6}$$

These equations in the limit include (set (3)) the expectation equations corresponding to the integral equation of sections II, 7 and 8. (They also include the particular case of $\lambda(T)$

and $\mu(T)$ constant considered more precisely in example (ii) of the next section). Without attempting here to solve the equations in the general case, we may note that we may expect the process to exhibit somewhat similar properties to those obtained in simpler cases. For example, if the mean of the population is stationary (for all age-groups), we may show that its variance still increases indefinitely. For from equation (5) we have for fluctuations during the interval t_s, $t_s + \Delta t$

$$\delta N = \delta N_{s+1} + \sum_{r=-\infty}^{s} \delta N_r$$

where the deviation δN_{s+1} is independent of the deviations δN_r, $(r \leq s)$. Hence the addition to the variance during this time interval is greater than the addition to the variance arising from δN_{s+1}. But from equations (3) and (4) this variance $v_{s+1,s+1}$ is equal to m_{s+1} (in fact under the assumptions made the increment δN_{s+1} is easily seen to have a Poisson distribution), and hence is constant if m_{s+1} is constant. It follows that the total variance increases without limit although the mean of the population is stationary.

Of course, this does not imply that the expected numbers are not sufficiently representative of the actual numbers for a <u>finite</u> time. Under the above assumptions the equations could be derived on the basis of one initial individual, and the cumulant function for N_0 individuals (for definiteness of the same age) is then simply N_0 times the cumulant function for one individual. Hence for any finite time we have $\sqrt{v}/m$ of order $1/\sqrt{N_0}$, i.e., the coefficient of variation about the mean can be made as small as we wish by taking a large enough initial population. But for small populations it is important to consider the size of fluctuations even for moderate values of t, e.g., for small biological populations or for small populations of physical particles. The contrast of this case of industrial renewal, for which the total number of 'individuals' is necessarily fixed, is apparent, in spite of the possibility that the integral equations for the mean rate of replacement may be identical. (There are, of course, fluctuations in the number of renewals required at any time, but since the expected renewals on the basis of one initial article are actual probabilities, the fluctuation at any particular time about the expected number for N_0 initial articles is given by the binomial distribution).

11. Markoff Chains with Continuous Time.

We shall now formulate more precisely the transition to continuous time of our general basic equations for Markoff chains. If there is no limit to the subdivision of the time-intervals, we may allow $t_3 \to t_2$ in the equations of section II, 1. Let $t_3 = t_2 + \Delta t$. Suppose the probability of a change in the interval $t, t + \Delta t$ is represented by the diagonal matrix

$$\underset{\sim}{P}(t)\Delta t + o(\Delta t),$$

and further that the asymptotic conditional distribution given that such a change occurs is $\underset{\sim}{S}(t)$, so that

$$\begin{aligned} \underset{\sim}{Q}(t_2 + \Delta t, t_1) &= \underset{\sim}{Q}(t_2 + \Delta t, t_2)\underset{\sim}{Q}(t_2, t_1) \\ &= \left[1 - \underset{\sim}{P}(t_2)\Delta t\right] \underset{\sim}{Q}(t_2, t_1) \\ &\quad + \left[\underset{\sim}{S}(t_2)\underset{\sim}{P}(t_2)\Delta t\right] \underset{\sim}{Q}(t_2, t_1) + o(\Delta t), \end{aligned}$$

whence

$$\underset{\Delta t \to 0}{\mathrm{Lt}} \frac{\underset{\sim}{Q}(t_2 + \Delta t, t_1) - \underset{\sim}{Q}(t_2, t_1)}{\Delta t}$$

$$= (\underset{\sim}{S}(t_2) - 1)\underset{\sim}{P}(t_2)\underset{\sim}{Q}(t_2, t_1)$$

or
$$\frac{\partial \underset{\sim}{Q}(t_2, t_1)}{\partial t_2} = \underset{\sim}{R}(t_2)\underset{\sim}{Q}(t_2, t_1), \tag{1}$$

where

$$\underset{\sim}{R}(t_2) = (\underset{\sim}{S}(t_2) - 1)\ \underset{\sim}{P}(t_2). \tag{2}$$

Multiplying equation (1) to the right by $\underset{\sim}{p}(t_1)$, we have also

$$\frac{\partial p(t_2)}{\partial t_2} = \underset{\sim}{R}(t_2)\ \underset{\sim}{p}(t_2). \tag{3}$$

So far we have been considering the case of finite distributions, but there is no restriction in our notation to prevent our considering also the case when the number of probability terms represented by $\underset{\sim}{p}$ becomes denumerably infinite. Sufficient conditions for the existence of a unique solution to our equations (1) and (3) under such a more general formulation have been given by Arley (B, 1).

We shall also find it useful here to make frequent use of operational methods in conjunction with the probability and moment-generating functions. These methods (see section 13) will be valid under Arley's conditions, (which ensure the existence and differentiability of moments of all orders), and in particular will be valid for the examples we shall consider.

We earlier considered the binomial distribution as an example of a stochastic process. Here we shall treat first the Poisson distribution. We assume that the chance of an event occurring in Δt, independently of what has previously occurred, is $\lambda\Delta t + o(\Delta t)$, and that the chance of more than one event occurring is $o(\Delta t)$. Then our basic equation before we proceed to the limit would be

$$p_i(t + \Delta t) = (1 - \lambda\Delta t)p_i(t) + \lambda\Delta t\, p_{i-1}(t) + o(\Delta t)$$

or in matrix notation, if we neglect $o(\Delta t)$,

$$\begin{pmatrix} p_o \\ p_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}_{t+\Delta t} = \begin{pmatrix} 1-\lambda\Delta t & 0 & 0 & \cdot\ \cdot \\ \lambda\Delta t & 1-\lambda\Delta t & 0 & \\ 0 & \lambda\Delta t & 1-\lambda\Delta t & \\ 0 & 0 & \lambda\Delta t & \\ \cdot & \cdot & \cdot & \end{pmatrix} \begin{pmatrix} p_o \\ p_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}_t$$

In the limit as $\Delta t \rightarrow o$, the matrices $\underset{\sim}{P}$ and $\underset{\sim}{S}$ defined above are easily seen to be

$$\underset{\sim}{P} = \lambda \begin{pmatrix} 1 & 0 & 0 & & \\ 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \cdot \end{pmatrix}, \quad \underset{\sim}{S} = \begin{pmatrix} 0 & 0 & 0 & & \\ 1 & 0 & 0 & & \\ 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \cdot \end{pmatrix},$$

and

$$\underset{\sim}{R} = \lambda \begin{pmatrix} -1 & 0 & 0 & \\ 1 & -1 & 0 & \\ 0 & 1 & -1 & \\ 0 & 0 & 1 & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Forming the probability-generating function $\Pi(z)$, we find

$$\frac{\partial \Pi_t(z)}{\partial t} = \lambda(1 \quad z \quad z^2 \ \ldots) \begin{pmatrix} -p_o \\ p_o - p_1 \\ p_1 - p_2 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}_t$$

$$= \lambda(z - 1)\Pi_t(z), \qquad (4)$$

or equivalently

$$\frac{\partial M_t(\theta)}{\partial t} = \lambda(e^\theta - 1)M_t(\theta), \tag{5}$$

$$\frac{\partial K_t(\theta)}{\partial t} = \lambda(e^\theta - 1). \tag{6}$$

From (6), we have

$$\begin{aligned} K_t(\theta) &= \lambda t(e^\theta - 1) + K_o(\theta) \\ &= \lambda t(e^\theta - 1) \quad \text{if } K_o = 0, \end{aligned} \tag{7}$$

which represents the Poisson distribution with mean λt and general term

$$p_i(t) = \frac{(\lambda t)^i e^{-\lambda t}}{i!} \tag{8}$$

As in the case of the binomial, this result is obtained most simply from the multiplication of m.g.f.'s corresponding to independent components. For we have

$$M_{t+\Delta t}(\theta) = \left[(1 - \lambda\Delta t) + \lambda\Delta t e^\theta + o(\Delta t)\right] M_t(\theta),$$

whence equations (5) and (6) follow at once.

12. Some Examples.

We shall now work out a few further examples which, for reasons either of principle or of practice, are of importance. Operational methods will be used.

(i) The negative binomial.

Suppose the probability of a new occurrence is still equal to $\lambda\Delta t$ on the average, but is to some extent linearly dependent on the number existing at time t; explicitly, we assume for the probability of a new occurrence

$$\lambda\Delta t\left(\frac{1 + \mu N_t}{1 + \mu\lambda t}\right) + o(\Delta t).$$

We obtain

$$M_{t+\Delta t}(\theta) = E_N\left\{\left[1 + \lambda\Delta t\left(\frac{1+\mu N_t}{1+\mu\lambda t}\right)\left(e^{\theta} - 1\right)\right] e^{N_t\theta}\right\} + o(\Delta t).$$

whence

$$\frac{\partial M_t(\theta)}{\partial t} = \frac{\lambda}{1+\mu\lambda t}\,(e^{\theta} - 1)\left[M_t(\theta) + \mu\frac{\partial}{\partial\theta} M_t(\theta)\right] \tag{1}$$

or equivalently

$$\frac{\partial K_t(\theta)}{\partial t} = \frac{\lambda}{1+\mu\lambda t}\,(e^{\theta} - 1)\left[1 + \mu\frac{\partial}{\partial\theta} K_t(\theta)\right] \tag{2}$$

To solve this partial differential equation, we employ the standard method of solution of an equation of the type

$$A(x,y,z)\frac{\partial z(x,y)}{\partial x} + B(x,y,z)\frac{\partial z(x,y)}{\partial x} = C(x,y,z) \tag{3}$$

viz., we find two independent solutions u = constant, and v = constant, of the subsidiary equations

$$\frac{dx}{A} = \frac{dy}{B} = \frac{dz}{C}\,, \tag{4}$$

and hence obtain the completely general solution

$$\zeta(u,v) = 0 \tag{5}$$

or equivalently

$$u = \psi(v) \tag{6}$$

where ζ (or ψ) is an arbitrary function.

Solving (2) in this way, we obtain the subsidiary equations

$$\frac{dT}{1+T} = \frac{-d\phi}{\phi(1+\phi)} = \frac{dL}{\phi}\,,$$

where for convenience we have written

$$\lambda\mu t = T, \qquad e^{\theta} - 1 = \phi\,, \qquad K = L/\mu\,.$$

We have solutions

(a) $\log(1+T) = \int\left(\frac{1}{1+\phi} - \frac{1}{\phi}\right) d\phi = \log(1+\frac{1}{\phi}) + \log C$

or $$(1 + T) = C(1 + \frac{1}{\phi})$$

and

$$e^L = D/(1 + \phi).$$

Hence the general solution is

$$e^L(1 + \phi) = \psi(\frac{(1 + T)\phi}{1 + \phi}). \tag{7}$$

For $t = T = 0$, we must have, (if $K_0 = 0$),

$$(1 + \phi) = \psi\left(\frac{\phi}{1 + \phi}\right),$$

whence $$\psi(x) = (1 - x)^{-1}.$$

Thus for any finite $t > 0$,

$$L = \log\left\{\frac{1}{1 + \phi}\left(1 - \frac{\phi(1 + T)}{1 + \phi}\right)^{-1}\right\}$$

$$= - \log(1 - T\phi),$$

or $$K(\theta) = -1/\mu \quad \log\left\{1 - \lambda\mu t(e^{\theta} - 1)\right\}, \tag{8}$$

which is the cumulant function of the well-known negative binomial distribution with probability-generating function

$$\Pi(z) = (1 + \lambda\mu t - \lambda\mu tz)^{-1/\mu}. \tag{9}$$

(ii) <u>An elementary 'birth' and 'death' process</u>.

We assume that the chance of an individual 'dying' in t, $t + \Delta t$ is $\mu\Delta t + o(\Delta t)$, independently of past history or of other individuals. We also assume thatthe chance of a particular individual giving rise to a new individual, i.e., a 'birth', is $\lambda\Delta t + o(\Delta t)$. This type of model has obvious possibilities both in the theory of growth of biological populations, and also in physical applications (cf. B, 1); although somewhat too simplified to be of immediate practical use, it is of considerable theoretical interest in illustrating principles.

Again we consider $M_{t+\Delta t}(\theta)$, and obtain

$$M_{t+\Delta t}(\theta) = E_N\left\{1 + N_t\left[\lambda(e^{\theta}-1) + \mu(e^{-\theta}-1)\right]e^{N_t\theta}\Delta t\right\} + o(\Delta t),$$

whence

$$\frac{\partial M_t(\theta)}{\partial t} = \left[\lambda(e^{\theta}-1) + \mu(e^{-\theta}-1)\right]\frac{\partial M_t(\theta)}{\partial \theta}, \tag{10}$$

or equivalently

$$\frac{\partial K_t(\theta)}{\partial t} = \left[\lambda(e^{\theta}-1) + \mu(e^{-\theta}-1)\right]\frac{\partial K_t(\theta)}{\partial \theta}. \tag{11}$$

The subsidiary equations here are

$$\frac{dt}{1} = \frac{-\,d\theta}{\lambda(e^{\theta}-1) + \mu(e^{-\theta}-1)} = \frac{dK}{0}, \tag{12}$$

where $K = C$ is a solution, and so is

$$t = D - \int \frac{d(e^{\theta})}{(e^{\theta}-1)(\lambda e^{\theta}-\mu)},$$

or

$$t - D = \begin{cases} -\dfrac{1}{\lambda-\mu}\log\left(\dfrac{e^{\theta}-1}{\lambda e^{\theta}-\mu}\right), & (\lambda \neq \mu), \\ \dfrac{1}{\lambda(e^{\theta}-1)}, & (\lambda = \mu). \end{cases}$$

Hence $K = \psi(\Theta)$, where

$$\Theta(t) = \begin{cases} \dfrac{e^{(\lambda-\mu)t}(e^{\theta}-1)}{(\lambda e^{\theta}-\mu)}, & (\lambda \neq \mu), \\ \lambda t - 1/(e^{\theta}-1), & (\lambda = \mu). \end{cases}$$

Putting $t = 0$, we have, when $\lambda \neq \mu$,

$$(\lambda e^{\theta} - \mu)\Theta_0 = e^{\theta} - 1,$$

or

$$(\lambda\Theta_0 - 1)e^{\theta} = \mu\Theta_0 - 1,$$

and hence if $K_0 = N_0\theta$,

$$N_0\theta = N_0\log\left(\frac{\mu\Theta_0 - 1}{\lambda\Theta_0 - 1}\right) = \psi(\Theta_0),$$

which determines the function ψ. We finally obtain

$$K_t(\theta) = N_o \log\left\{\left[\frac{\mu e^{(\lambda-\mu)t}(e^\theta-1)}{\lambda e^\theta - \mu} - 1\right]\Big/\left[\frac{e^{(\lambda-\mu)t}(e^\theta-1)}{\lambda e^\theta - \mu} - 1\right]\right\}. \quad (13)$$

In the case $\lambda = \mu$, $\Theta_o = -1/(e^\theta - 1)$ or $e^\theta = 1 - 1/\Theta_o$, whence

$$N_o\theta = N_o \log\,(1 - 1/\Theta_o) = \psi(\Theta_o),$$

and

$$K_t(\theta) = N_o \log\left\{\left[1 - (\lambda t - 1)(e^\theta - 1)\right]\Big/\left[1 - \lambda t(e^\theta - 1)\right]\right\}. \quad (14)$$

If we expand (13) or (14) in powers of θ, we obtain the cumulants of N_t. In particular, it may be shown that the mean and variance are given by

$$m_t = N_o\, e^{(\lambda-\mu)t},$$

$$v_t = \begin{cases} N_o \dfrac{\lambda+\mu}{\lambda-\mu} e^{(\lambda-\mu)t}\left[e^{(\lambda-\mu)t} - 1\right], & (\lambda \neq \mu), \\ 2\lambda t N_o, & (\lambda = \mu), \end{cases} \quad (15)$$

so that in the case $\lambda = \mu$ we again have an example of a process with stationary or oscillating mean (here exactly constant) but with ever-increasing variance.

To examine the probabilities for the above process, we simply write $\Pi_t(z) = M_t(\log z)$. For example, when $\lambda = \mu$, and $N_o = 1$, we obtain

$$\begin{aligned}\Pi_t(z) &= 1 - (\lambda t - \frac{1}{z-1})^{-1} \\ &= 1 - \frac{1-z}{1+\lambda t}(1 - \frac{\lambda t}{1+\lambda t} z)^{-1} \\ &= 1 - \frac{1-z}{1+\lambda t}\sum_{r=0}^{\infty}\left(\frac{\lambda t}{1+\lambda t}\right)^r z^r,\end{aligned}$$

whence

$$\left.\begin{aligned}p_r(t) &= \left\{\left(\frac{\lambda t}{1+\lambda t}\right)^{r-1} - \left(\frac{\lambda t}{1+\lambda t}\right)^r\right\}\frac{1}{1+\lambda t} \\ &= \frac{1}{(1+\lambda t)^2}\left(\frac{\lambda t}{1+\lambda t}\right)^{r-1}, \qquad (r > 0), \\ p_o(t) &= \frac{\lambda t}{1+\lambda t}, \quad \to 1 \text{ as } t \to \infty.\end{aligned}\right\} \quad (16)$$

<u>Again we see that the probability of extinction tends to one in this example in spite of a constant mean.</u> This conclusion would not be altered for any other (finite) value of N_o, for more generally we find for arbitrary N_o and μ,

$$\Pi_t(0) = \begin{cases} \left(\dfrac{e^{(\lambda-\mu)t} - 1}{\frac{\lambda}{\mu} e^{(\lambda-\mu)t} - 1}\right)^{N_o} \to \begin{cases} \left(\frac{\mu}{\lambda}\right)^{N_o} \\ 1 \end{cases} \text{as } t \to \infty, & (\lambda > \mu), \\ & (\lambda < \mu), \\ \left(\dfrac{\lambda t}{1 + \lambda t}\right)^{N_o} \to 1 & (\lambda = \mu). \end{cases}$$

(iii) <u>The Gaussian frequency curve as the mean of a stochastic model</u>.

In the last example the mechanism of the process was independent of time, so that if the constant μ was less than λ, the number of individuals was likely to increase indefinitely. Such processes, where the evolution in time depends on constant quantities might be called <u>temporally homogeneous processes</u>.* If we wish to represent a process with a characteristic rise and fall, like an epidemic, some modification must be made. One useful modification is as follows. For convenience we adopt a definite terminology appropriate to the epidemiological field.

We assume as before that the chance of an event, this time a new 'infection', is given in a small interval Δt by $\lambda \Delta t$ for each already infected individual, but that the chance of an already infected individual being no longer a 'case', i.e., recovering, increases linearly with the time, and is given by $\mu t \Delta t$. Then our equation for K becomes

*This expression is, however, reserved by some writers for processes which are stationary.

modified to

$$\frac{\partial K}{\partial t} = \left\{\mu t(e^{-\theta} - 1) + \lambda(e^{\theta} - 1)\right\}\frac{\partial K}{\partial \theta}. \tag{17}$$

While this equation does not appear* now to be soluble in finite terms, we may investigate the mean and variance as usual. We obtain for the mean, from the coefficient of θ in (17),

$$\frac{\partial m_t}{\partial t} = (\lambda - \mu t)m_t,$$

whence

$$m_t = m_o e^{\lambda t - \frac{1}{2}\mu t^2}, \tag{18}$$

which rises to a maximum at $t = \lambda/\mu$ and then drops to zero. Similarly for the variance we find

$$\frac{\partial v_t}{\partial t} = m_t(\lambda + \mu t) + 2v_t(\lambda - \mu t),$$

whence

$$v_t e^{-2(\lambda t - \frac{1}{2}\mu t^2)} = \int_o^t m_o(\lambda + \mu t)e^{-(\lambda t - \frac{1}{2}\mu t^2)}\, dt$$

or

$$v_t = m_o e^{\lambda t - \frac{1}{2}\mu t^2}(1 - e^{\lambda t - \frac{1}{2}\mu t^2}) + 2m_o\lambda\, e^{2(\lambda t - \frac{1}{2}\mu t^2)}\int_o^t e^{-(\lambda t - \frac{1}{2}\mu t^2)}\, dt. \tag{19}$$

*(iv) The logistic function.

A different kind of restriction on population growth occurs when the chance of an increase is restricted by a maximum size for the population. Considered as an epidemic model, we here assume that the chance of infection is proportional not only to the number already infected, but also to the number of non-infected 'susceptibles' remaining in the

*Mr. David G. Kendall has since pointed out to me that equation (17) does in fact lead to a complete solution.

population. This gives the chance of a new infection in $(t, t + \Delta t)$ proportional to $N_t(n - N_t)\Delta t$, where n is the total possible number of infections. We obtain the partial differential equation for $M_t(\theta)$:

$$\frac{\partial M_t(\theta)}{\partial t} = \lambda(e^{\theta} - 1)\left[n\frac{\partial M_t(\theta)}{\partial \theta} - \frac{\partial^2 M_t(\theta)}{\partial \theta^2}\right]. \qquad (20)$$

In particular, the coefficient of θ gives

$$\frac{\partial m_t}{\partial t} = \lambda(nm_t - \nu), \qquad (21)$$

where $\nu_t \equiv E\left\{N_t^2\right\}$ is the second moment of N_t. We notice that, owing to the non-linearity of the chance of an increase in total number on the existing number N_t, equation (21) is no longer an equation for n_t identical with the standard equation for this problem leading to the logistic function (cf. Feller, F 5). This standard equation is

$$\frac{\partial \mu_t}{\partial t} = \lambda(n - \mu_t)\mu_t, \qquad (22)$$

whence

$$\mu_t = \frac{n}{1 + e^{-\lambda n(t-t_o)}}, \qquad (23)$$

$$\frac{\partial \mu_t}{\partial t} = \frac{1}{4}\lambda n^2 \operatorname{sech}^2\left[\frac{1}{2}\lambda n(t - t_o)\right], \qquad (24)$$

equation (23) giving the total 'population of infections' and equation (24) the rate of new cases.

For large n, the difference between (21) and (22) is unimportant, since ν_t will not differ effectively from m_t^2. Thus, although in this case the stochastic model affects even the expected value of N_t compared with the standard solution, the latter will adequately represent the number present for large n.

*13. Further Discussion of the General Equations.

It is evident that the operational method used in the preceding sections may be generalized. In general, if a conditional change from N_t to $N_{t+\Delta t}$ is represented by the moment-generating function factor multiplying $e^{N_t\theta}$,

$$1 + \psi(\theta, t, N_t)\Delta t + o(\Delta t) \tag{1}$$

we obtain the operational equation

$$\frac{\partial M_t(\theta)}{\partial t} = \psi(\theta,\ t,\ \frac{\partial}{\partial\theta})M_t(\theta); \tag{2}$$

and similarly in the case of more than one random variable N_t. In particular, if the function ψ is independent of t, we have the operational solution

$$M_t(\theta) = \exp\left\{\psi(\theta,\ \frac{\partial}{\partial\theta})t\right\}M_o(\theta). \tag{3}$$

Returning to the type for which ψ in (1) is proportional to N_t or in (3) is proportional to $\partial/\partial\theta$, we have (3) expressible in the form

$$M_t(\theta) = \exp\left\{t\,\frac{\partial}{\partial\eta(\theta)}\right\}M_o(\theta), \tag{4}$$

say, or equivalently for $\Pi(z)$,

$$\Pi_t(z) = \exp\left\{t\,\frac{\partial}{\partial\zeta(z)}\right\}z^{n_o} \tag{5}$$

if $\zeta(z) \equiv \eta(\theta)$ for $z = e^{\theta}$, then $\Pi_o(z) = z^{n_o}$.

Write

$$z = \zeta^{-1}\left[\zeta(z)\right],$$

then our operational solution becomes

$$\Pi_t(z) = \left\{\zeta^{-1}\left[\zeta(z) + t\right]\right\}^{n_o}. \tag{6}$$

These equations are of course alternative or particular forms of the general equations (1) and (3) of section 10. The general equation was

$$\frac{\partial \underset{\sim}{Q}(t,t_o)}{\partial t} = \underset{\sim}{R}(t)\ \underset{\sim}{Q}(t,t_o) \tag{7}$$

and was derived from the limit of

$$\underset{\sim}{Q}(t + \Delta t,\ t_o) = (1 + \underset{\sim}{R}(t)\Delta t)\underset{\sim}{Q}(t,\ t_o) + o(\Delta t)$$

as $\Delta t \to 0$; it thus has the formal solution

$$\underset{\sim}{Q}(t,t_o) = P\int_{t_o}^{t}(1 + \underset{\sim}{R}(t)dt) \tag{8}$$

where $P\int_{t_o}^{t}(1 + \underset{\sim}{R}(t)dt)$ represents the 'product integral'

$$\lim_{\Delta t \to o} \prod_{i=o}^{n-1}\left[1 + \underset{\sim}{R}(t_o + i\Delta t)\right]$$

and $\underset{\sim}{Q}(t_o,t_o)$ is simply the diagonal unit matrix and may be dropped.

To obtain a general solution equivalent to (8), we integrate (7) to

$$\underset{\sim}{Q}(t,t_o) = 1 + \int_{t_o}^{t} \underset{\sim}{R}(t)\underset{\sim}{Q}(t,t_o)dt, \tag{9}$$

an integral equation with a formal series solution

$$\underset{\sim}{Q}(t,t_o) = \sum_{i=o}^{\infty} \underset{\sim}{T}_i(t,t_o), \tag{10}$$

where

$$\begin{cases} \underset{\sim}{T}_o(t,t_o) = 1, \\ \underset{\sim}{T}_{i+1}(t,t_o) = \int_{t_o}^{t} \underset{\sim}{R}(t)\underset{\sim}{T}_i(t,t_o)dt. \end{cases}$$

It may be shown that such a formal series is a valid solution (see Arley B 1) if the matrix $\underset{\sim}{R}(t)$ is <u>absolutely exponentiable</u> in (t_o,t), i.e.

$$\exp\left\{(t - t_o)\underset{\sim}{K}\right\} \equiv \sum_{i=o}^{\infty}\frac{(t - t_o)^i}{i!}\,\underset{\sim}{K}^i$$

exists, where

$$\underset{\sim}{K} = \max_{t_o \le T \le t}\left|\underset{\sim}{R}(T)\right|.$$

A closed expression for the solution (10) is possible if we have further that $\underset{\sim}{R}(t)$ and $\underset{\sim}{T}_1(t,t_o)$ commute. For in this case

$$\frac{d}{dt}\left[\underset{\sim}{T}_1(t,t_o)\right]^i = i\left[\underset{\sim}{T}_1(t,t_o)\right]^{i-1}\underset{\sim}{R}(t), \qquad (11)$$

and

$$\begin{cases} \underset{\sim}{T}_2(t,t_o) = \int_{t_o}^{t} \frac{1}{2}\frac{d}{dt}\left[\underset{\sim}{T}_1(t,t_o)\right]^2 dt \\ \qquad\qquad = \frac{1}{2}\left[\underset{\sim}{T}_1(t,t_o)\right]^2 \\ \quad \cdots \\ \underset{\sim}{T}_i(t,t_o) = \frac{1}{i!}\left[\underset{\sim}{T}_1(t,t_o)\right]^i \end{cases}$$

whence

$$\underset{\sim}{Q}(t,t_o) = \exp\left\{\underset{\sim}{T}_1(t,t_o)\right\}$$

$$= \exp\left\{\int_{t_o}^{t} R(t)dt\right\}. \qquad (12)$$

Finally, we may quote from Arley sufficient conditions for $\underset{\sim}{R}(t)$ to be absolutely exponentiable. We consider matrices of the column semi-diagonal type, i.e., the non-zero elements in each column extend at most only a finite and fixed number 1 of places beyond the diagonal. We suppose further that the sum of the absolute values of the elements in the q-th column is bounded by the expression q f(t), where f(t) is further bounded by C in the interval (t_o,t). Then $\underset{\sim}{R}(t)$ is absolutely exponentiable, and a unique solution of (7) exists, in the interval (t_o,t) such that

$$|t - t_o|\ 1\ C < 1, \qquad (13)$$

and may always be extended to further times by stages. Further, the moments of all orders of $\underset{\sim}{p}(t) = \underset{\sim}{Q}(t,t_o)\ \underset{\sim}{p}(t_o)$ exist and are differentiable with continuous differential coefficients under these conditions, which are also generalized to cover processes in more than one variable.

III. STATIONARY PROCESSES, AND THEIR BEARING ON THE ANALYSIS OF TIME-SERIES.

1. Definition of Stationary Processes.

We have seen in the case of population growth how the mean of a process may be constant, but the variance changing. Such a process we might term <u>stationary to the first order</u>, in contrast with a <u>completely stationary process</u>. By the latter we shall (cf. Khintchine D 2) mean a process whose simultaneous distribution at $t_1 \ldots t_n$ (or equivalently the simultaneous characteristic function) depends only on the n - 1 differences $t_2 - t_1, \ldots t_n - t_1$. When the moments of all orders exist, we shall then have also

$$E\left\{X_1^{\alpha_1} X_2^{\alpha_2} \ldots X_n^{\alpha_n}\right\} = \psi(t_2 - t_1, \ldots t_n - t_1). \quad (1)$$

In particular,

$$E\left\{X^{\alpha}\right\} = \mu_{\alpha}$$

independently of t, and

$$E\left\{X_1^{\alpha} X_2^{\beta}\right\} = \mu_{\alpha\beta}(t_2 - t_1),$$

a function of the difference or interval $T = t_2 - t_1$. It is sometimes useful to consider processes for which (1) is satisfied for $\alpha_1 + \alpha_2 + \ldots \alpha_n \leq r$; such processes may be called <u>stationary to the r-th order</u>. Thus a process <u>stationary to the second order</u> has its moments of the form

$$\begin{aligned} E\left\{X\right\} &= m, \\ E\left\{(X_1 - m)(X_2 - m)\right\} &= w(t_2 - t_1) \quad\quad (2) \\ &= \sigma^2 \rho(t_2 - t_1), \end{aligned}$$

where $\rho(o) = 1$. The function $\rho(T)$ is called the <u>autocorrelation coefficient</u>, being the correlation coefficient between any two values of X an interval T apart. (For real X, $\rho(T)$ is a symmetric function of T).

We note that the distribution or characteristic functions of <u>normal processes</u>, i.e., processes whose distributions for any set $t_1 \ldots t_n$ are multivariate normal, involve only the first and second-order moments. It at once follows that a <u>normal process, stationary to the second order, is completely stationary</u>.

In this part, we shall assume throughout that any stochastic process discussed is completely stationary. However, we shall be mainly concerned with the <u>correlational</u> properties of stationary processes, and many of these properties will depend only on the structure of the process as specified by its first and second moments, and will evidently still hold for processes stationary only to the second order. The possible values of X will usually be assumed to have a continuous range.

2. <u>Stationary Processes Specified for Discrete Time</u>.

We shall first consider processes defined only for values of t_r spaced at regular intervals. The simplest case of a stationary stochastic process is then: --

(i) <u>The completely random process</u>, defined by the property

$$p(X_1 \ldots X_n) = \prod_{r=1}^{n} p(X_r),$$

and illustrated by a sequence of independent observations, such as the score 0 to 1 for 'heads' or 'tails' when tossing a coin. In this case we obviously have

$$\rho_{r-s} = \rho(t_r - t_s) = 0, \quad (r \neq s). \tag{1}$$

As a second example, we define a particular case of a stationary Markoff process by

(ii) <u>The linear Markoff process</u>, where

$$X_{s+v} = \lambda_v X_s + Y_{s+v}, \tag{2}$$

where X_s and Y_{s+v} are independent for any $v > 0$. Averaging, we have for the means

$$m_x = \lambda_v m_x + m_{y,v}$$

or $$m_{y,v} = m_x(1 - \lambda_v) \tag{3}$$

for the mean of Y_{s+v} in terms of the mean of X. Taking for convenience the mean of X to be zero, we have on multiplying by X_s before averaging,

$$\sigma^2 \rho_v = \lambda_v \sigma^2,$$

or

$$\lambda_v = \rho_v. \tag{4}$$

We have further, since the partial correlation of X_{s+v} and X_{s-u} for given X_s $(v,u > 0)$ is zero,

$$\rho_{u+v} - \rho_u \rho_v = 0$$

whence

$$\rho_v = \rho_1^{\,v}. \tag{5}$$

As a somewhat more general <u>linear process</u>, we consider next

(iii) <u>The linear autoregressive process</u>.

$$X_{s+v} = \lambda_1 X_{s+v-1} + \lambda_2 X_{s+v-2} + \dots + \lambda_v X_s + Y_{s+v}, \tag{6}$$

where Y_{s+v} is independent of $X_{s+v-1}, X_{s+v-2}, \dots$. In particular we consider such a scheme with two non-zero coefficients λ_1 and λ_2; or writing $a = -\lambda_1$, $b = -\lambda_2$, we have

$$X_{s+2} + aX_{s+1} + bX_s = Y_{s+2}, \tag{7}$$

where Y_{s+2} is independent of $X_{s+1}, X_s, \dots$.

This equation (7) was introduced by Yule (E 7) as a model of an oscillating system subject to random shocks (for example, an oscillating pendulum bombarded irregularly by boys equipped with peashooters). As such we shall see later that it is only approximate; nevertheless, (7), or the more general equation (6), may be studied as a useful empirical model for series of discrete observations.

From its form it is an example of a <u>stochastic linear difference equation</u>, and may be solved by methods appropriate to ordinary difference equations. Thus, using operational methods of solution, we obtain as the 'particular solution' of (7)

$$X_s = \frac{1}{E^2 + aE + b} Y_{s+2},$$

where E is the usual finite difference operator $1 + \Delta$, or

$$\begin{aligned} X_s &= \frac{1}{\mu_1 - \mu_2}\left\{\frac{1}{E - \mu_1} - \frac{1}{E - \mu_2}\right\} Y_{s+2} \\ &= \frac{1}{\mu_1 - \mu_2}\left\{E^{-1}(1 - \mu_1 E^{-1})^{-1} - E^{-1}(1 - \mu_2 E^{-1})^{-1}\right\} Y_{s+2} \\ &= \sum_{r=0}^{\infty}\left\{\frac{\mu_1^{r+1} - \mu_2^{r+1}}{\mu_1 - \mu_2} Y_{s-r}\right\}, \end{aligned} \tag{8}$$

where μ_1 and μ_2 are the roots of the equation

$$x^2 + ax + b = 0.$$

The complementary function is of the form

$$A\mu_1^{\,s} + B\mu_1^{\,s}, \tag{9}$$

but, since this is a function of s, it must in general be zero for a stationary series. The physical picture is that we imagine the process 'started-up' some time ago, and X_s then steadies down to the solution (8) when $|\mu| < 1$, while the complementary function sinks to zero.*

*An exception occurs when $|\mu| = 1$; this leads to the case of strictly harmonic series, which are referred to in section III, 3.

From (7) we may also obtain a difference equation for ρ_s. Multiplying by X_o and averaging, we have after dividing through by σ^2,

$$\rho_{s+2} + a\rho_{s+1} + b\rho_s = 0,$$

with a solution of the form (9). Here s refers to the interval, not to the absolute time, and (9) thus represents a valid solution. To obtain the constants A and B, we note on multiplying (7) by X_{s+1} that

$$\rho_1(1 + b) + a = 0 \tag{10}$$

whence, with $\rho_o = 1$, we obtain

$$A + B = 1,$$

$$A\mu_1 + B\mu_2 = -\frac{a}{1 + b} = \frac{\mu_1 + \mu_2}{1 + \mu_1\mu_2},$$

whence

$$A = \frac{\mu_1(1 - \mu_2^2)}{(\mu_1 - \mu_2)(1 + \mu_1\mu_2)}, \quad B = \frac{-\mu_2(1 - \mu_1^2)}{(\mu_1 - \mu_2)(1 + \mu_1\mu_2)},$$

and

$$\rho_s = \frac{\mu_1^{s+1}(1 - \mu_2^2) - \mu_2^{s+1}(1 - \mu_1^2)}{(\mu_1 - \mu_2)(1 + \mu_1\mu_2)} . \tag{11}$$

If we write $\mu_1 = \sqrt{b}\, e^{i\theta}$, $\mu_2 = \sqrt{b}\, e^{-i\theta}$, $(0 < b < 1)$, and $2\sqrt{b} \operatorname{Cos}\theta = -a$, $(a^2 < 4b)$, we obtain

$$\rho_s = b^{\frac{1}{2}s}\left\{\frac{\operatorname{Sin}(s + 1)\theta - b \operatorname{Sin}(s - 1)\theta}{(1 + b)\operatorname{Sin}\theta}\right\}$$

$$= b^{\frac{1}{2}s} \frac{\operatorname{Sin}(s\theta + \Psi)}{\operatorname{Sin}\theta}, \tag{12}$$

where $\tan\Psi = \tan\theta . (1 + b)/(1 - b)$; (cf. Kendall E 1).

The solution (8) suggests the more general type of linear process

$$X_s = \sum_{r=o}^{\infty} f(r)\, Y_{s-r} \tag{13}$$

or if for convenience we define $f(r) = 0$ for $r < 0$,

$$X_s = \sum_{u=-\infty}^{\infty} f(s - v) Y_v, \tag{14}$$

where as before Y_v is a completely random series. It is assumed that $E\{Y_v\} = 0$, $E\{Y_j^{\ r}\}$ finite for all $r > 0$, and that $\sum_{r=0}^{\infty} f^2(r)$ converges. Under these conditions it may be shown that the infinite sum in (14) is well-defined in the stochastic sense as a limit 'in the mean' (see section 5) of the finite sum, and has distributional properties given by the limiting distributional properties of the finite sum. Thus from the independence of Y_v we may write

$$K(\theta_1, \theta_2) = \log E\left\{\exp(\theta_1 X_s + \theta_2 X_{s+t})\right\}$$

$$= \sum_{v=-\infty}^{\infty} K_r(\theta_1 \ f(s-v) + \theta_2 \ f(s+t-v))$$

where $K_y(\theta)$ is the cumulant function of Y, or

$$K(\theta_1, \theta_2) = \sum_{u=-\infty}^{\infty} K_y(\theta_1 \ f(u) + \theta_2 \ f(u+t)). \tag{15}$$

whence

$$\sigma^2(X) = \sigma^2(Y) \ \Sigma \ f^2(u),$$

$$\operatorname{cov}(X_s, X_{s+t}) = \sigma^2(Y) \ \Sigma \ f(u) \ f(u+t). \tag{16}$$

We may also note from the more general cumulant function connecting four instants of time, viz.

$$K(\theta_1, \theta_2, \theta_3, \theta_4) = \Sigma \ K_y(\theta_1 f(u) + \theta_2 f(u+t) + \theta_3 f(u+t+v) + \theta_4 f(u+t+v+w))$$

that

$$\left.\begin{aligned} &K_4(X) = K_4(Y) \ \Sigma \ f^4(u), \\ &K_{1111}(X_s, X_{s+t}, X_{s+t+v}, X_{s+t+v+w}) \\ &\qquad = K_4(Y) \ \Sigma \ f(u) \ f(u+t) \ f(u+t+v) \ f(u+t+v+w) \end{aligned}\right\} . \tag{17}$$

To illustrate formulae (16), let us consider the particular linear process we had in (8). Here we had

$$f(u) = \frac{\mu_1^{u+1} - \mu_2^{u+1}}{\mu_1 - \mu_2}.$$

we obtain

$$\Sigma f^2(u) = \frac{1}{(\mu_1-\mu_2)^2}\left[\frac{\mu_1^2}{1-\mu_1^2} - \frac{2\mu_1\mu_2}{1-\mu_1\mu_2} + \frac{\mu_2^2}{1-\mu_2^2}\right]$$

$$= \frac{1}{a^2-4b}\left[\frac{a^2 - 2b(1+b)}{1 - a^2 + 2b + b^2} - \frac{2b}{1 - b}\right]$$

or

$$\frac{\sigma^2(X)}{\sigma^2(Y)} = \frac{1 + b}{(1 - b)\{(1 + b)^2 - a^2\}}, \tag{18}$$

and

$$\Sigma\ f(u)\ f(u+t) = \frac{\mu_1}{(\mu_1-\mu_2)^2}\left[\frac{\mu_1^{t+2}}{1-\mu_1^2} - \frac{\mu_1\mu_2^{t+1}}{1-\mu_1\mu_2} - \frac{\mu_2\mu_1^{t+1}}{1-\mu_1\mu_2} + \frac{\mu_2^{t+2}}{1-\mu_2^2}\right]$$

$$= \frac{\mu_1^{t+1}(1-\mu_2^2) - \mu_2^{t+1}(1-\mu_1^2)}{(1-\mu_1^2)(1-\mu_2^2)(1-\mu_1\mu_2)(\mu_1-\mu_2)},$$

whence as before

$$\rho_t = \frac{\Sigma\ f(u)\ f(u+t)}{\Sigma\ f^2(u)}$$

$$= \frac{\mu_1^{t+1}(1-\mu_2^2) - \mu_2^{t+1}(1-\mu_1^2)}{(\mu_1-\mu_2)(1+\mu_1\mu_2)}.$$

3. Relation Between Autocorrelation Coefficients and Harmonic Analysis.

So far we have considered some particular cases of stationary processes defined for discrete t. Before going on to consider the general correlational properties of such processes, we shall make one or two preliminary remarks.

Firstly, we have not yet indicated how we may measure or estimate ρ_t, which is the true correlation between X_s and X_{s+t}.

Since this correlation for a stationary process is independent of s, we may hope to estimate it by observing the relation between x_s and x_{s+t} in an observed series for a number of values x_s $(s = 1 \ldots n)$, but such a procedure must be kept distinct from the theoretical averaging, over the joint distribution of X_s and X_{s+t}, which determines ρ_t. For example, consider the observed series (defined for <u>all</u> real s)

$$x_s = \operatorname{Cos} \lambda s. \tag{1}$$

We have here a simple harmonic series for which the 'time-average'

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \operatorname{Cos} \lambda s \, ds = 0$$

and further

$$\frac{1}{T} \int_0^T \operatorname{Cos} \lambda s \operatorname{Cos} \lambda (s+t) \, ds$$

$$= \frac{1}{T} \int_0^T (\operatorname{Cos}^2 \lambda s \operatorname{Cos} \lambda t - \operatorname{Cos} \lambda s \operatorname{Sin} \lambda s \operatorname{Sin} \lambda t) ds$$

$$\to \tfrac{1}{2} \operatorname{Cos} \lambda t,$$

or since the same expression with $t = 0$ tends to $\frac{1}{2}$, the observed correlation obtained between x_s and x_{s+t} for an increasing continuous range of s would be $\cos \lambda t$. But since we have not yet specified the stochastic process for which (1) represents an observed or 'realized' value, the correlation ρ_s is also not yet specified. Suppose, however, we specify this process by

$$X_s = \operatorname{Cos} (\lambda s + Y), \tag{2}$$

where Y is a random variable uniformly distributed in $(0, 2\pi)$. Then we readily find

$$E\left\{X_s\right\} = \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Cos}(\lambda s + y) dy = 0$$

$$E\left\{X_s X_{s+t}\right\} = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{Cos}(\lambda s + y)\ \operatorname{Cos}(\lambda s + y + t)dy$$

$$= \frac{1}{2}\operatorname{Cos}\lambda t,$$

so that the stochastic process we have defined is stationary at least to the second order, and its moments are those obtained by the 'time-averaging' procedure. For a process like (2) we note also that the autocorrelation ρ_t continually oscillates like the series X_s itself, and never dies down to zero.

Secondly, the existence of these different types of autocorrelation coefficients leads to the query: "Are there any restrictions on the types of autocorrelations that are possible?" It is easy to see that there are. Let us work out the partial correlation between X_t and X_{t+2s} when X_{t+s} is given. We must have

$$-1 \leq \frac{\rho_{2s} - \rho_s^2}{1 - \rho_s^2} \leq 1,$$

whence we obtain not only the trivial relation $\rho_{2s} \leq 1$, but also

$$\rho_{2s} - 2\rho_s^2 + 1 = 0. \qquad (3)$$

By taking $s < a$, $2s > a$, we find, for example, that the autocorrelation $\rho_t = 1$ $(t \leq a)$, $\rho_t = 0$ $(t > a)$, is not a valid possibility. Such consistency conditions can all be included in a single important theorem, which we shall now consider.

We give the theorem here for the case of discrete time, in the form given by Wald (D 3). The theorem is an analogue of the corresponding theorem for continuous time (Khintchine, D 2), which will be given in due course.

The necessary and sufficient condition that ρ_s $(s = 0, \pm 1, \pm 2, \dots)$ *is the autocorrelation coefficient of some stationary process* (*defined for discrete* t) *is that*

$$\rho_s = \int_0^{\pi} \cos sw\, d\, F(w), \tag{4}$$

<u>where</u> F(w) <u>is a distribution function defined between</u> 0 <u>and</u> π.

<u>Necessity</u>: The necessity follows from the existence of the non-negative definite Hermitian form

$$\sigma^2 \sum_{u=1}^{n}\sum_{v=1}^{n} a_u \bar{a}_v \rho_{u-v} = E\left\{\left|\sum_{u=1}^{n} a_u X_u\right|^2\right\} \geqq 0, \tag{5}$$

whence it follows from a theorem by Bochner* that the equation

$$\tfrac{1}{2}\int_0^{2\pi} e^{isT} d\, H(T) = \bar{\rho}_s \tag{6}$$

has a non-decreasing solution with H(0) = 0. Putting s = 0, we see further that $H(2\pi) = 2$, and if we write

$$\begin{cases} w = T, & F(w) = H(T), \qquad (0 \leqq T < \pi), \\ w = 2\pi - T, & F(w) = 2 - H(T), \quad (\pi \leqq T \leqq 2\pi), \end{cases}$$

then it is easily seen that (4) and (6) are equivalent.

<u>Sufficiency</u>: Write $X_t = \sqrt{2} \cos (tZ + Y)$, where the distribution of Z is F(w), and the distribution of Y is independent of Z, and uniformly distributed in $(0, 2\pi)$. Then $E\{X_t\} = 0$ and

$$E\{X_t X_{t+s}\} = E\{\cos sZ\}$$

$$= \int_0^{\pi} \cos sw\, dF(w) = \rho_s$$

It was noted in section I, 4 that any distribution function we assume to be of the form

$$F(w) = c_1^2 F_1(w) + c_2^2 F_2(w)$$

where $F_1(w)$ represents a distribution function with a frequency function $f_1(w)$, and $F_2(w)$ is a step-function. It follows that ρ_s is then also of the form

*The corresponding theorem for continuous s is discussed in the Lectures by S. Bochner on Fourier Analysis (Princeton University, 1936-37), Chapter 16.

$$\rho_s = c_1^2(\rho_s)_1 + c_2^2(\rho_s)_2,$$

where

$$(\rho_s)_1 = \int_0^{\pi} \text{Cos } sw\ f_1(w)\ dw, \tag{7}$$

and

$$(\rho_s)_2 = \Sigma_r\ p_r\ \text{Cos } sw_r. \tag{8}$$

As a special case of (8), we have the autocorrelation coefficient Cos λs considered earlier in this section. Equation (7) has the usual Fourier series inversion formula,

$$f_1(w) = \frac{1}{\pi}\sum_{s=-\infty}^{\infty}(\rho_s)_1\ \text{Cos } sw. \tag{9}$$

The formal integration of this equation gives the inversion formula which is in fact valid for any distribution function F(w),

$$F(w) = \frac{1}{\pi}\left\{w + 2\sum_{s=1}^{\infty}\rho_s\ \frac{\text{Sin } sw}{s}\right\}. \tag{10}$$

This analysis of ρ_s is of fundamental importance in interpreting the harmonic analysis of a time-series set of observations $x_1, x_2, \ldots, x_n$. Let us consider the usual 'periodogram' analysis of such observations, where we compute

$$\begin{cases} A_p = \sqrt{\frac{2}{n}}\ \Sigma\ X_r\ \text{Cos}\left[(2\pi p/n)r\right] \\ B_p = \sqrt{\frac{2}{n}}\ \Sigma\ X_r\ \text{Sin}\left[(2\pi p/n)r\right] \end{cases}$$

and then the 'intensity' $I_p = A_p^2 + B_p^2$. The factor $\sqrt{(2/n)}$ has been inserted to make $E\{I_p\} = 2\sigma^2$ for a completely random process. We have the identity

$$I_p = \frac{2}{n}\sum_{r=1}^{n}\sum_{u=1}^{n} X_r X_u\ \text{Cos } k(r-u), \tag{11}$$

where $k = 2\pi p/n$, or if we write

$$C_s = \frac{1}{n-|s|}\sum_{r=1}^{n-|s|} X_r X_{r+s} \tag{12}$$

$$I_p = 2\sum_{s=n+1}^{n-1}\frac{(n-|s|)}{n}\ C_s\ \text{Cos } ks. \tag{13}$$

For convenience we assume $E\{X\} = 0$; we then have on taking the expectation of both sides of the identity (13),

$$E\{I_p\} = 2\sigma^2\left[1 + 2\sum_{s=1}^{n-1}\left(1 - \frac{s}{n}\right)\rho_s \operatorname{Cos} ks\right]. \qquad (14)$$

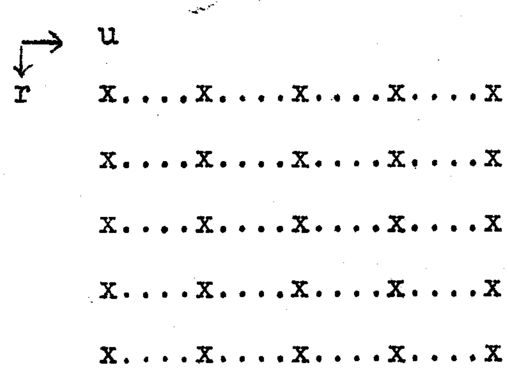
We may now consider $E\{I_p\}$ not only for a completely random series, but for possible forms of ρ_s. The usual periodogram argument considers what happens when $\rho_s = \operatorname{Cos}\lambda s$. We may write (14) in the alternative form

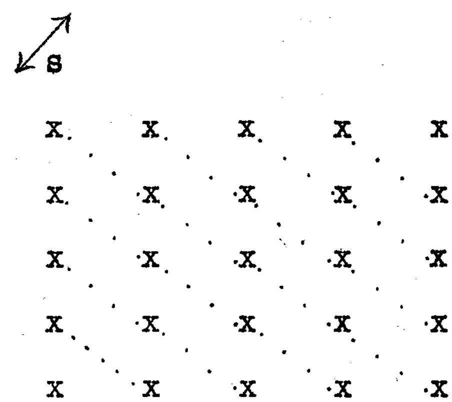
$$E\{I_p\} = \frac{2\sigma^2}{n}\sum_{v=o}^{n-1}\sum_{s=-v}^{v}\rho_s \operatorname{Cos} ks; \qquad (15)$$

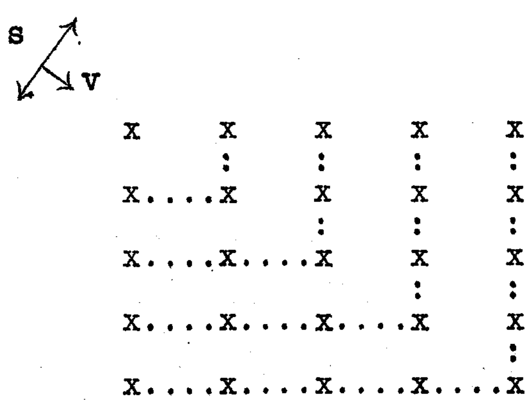
(the identity of the sums

$$\sum_{r=1}^{n}\sum_{u=1}^{n} f(r-u) = \sum_{s=-n+1}^{n-1}(n - |s|)f(s) = \sum_{v=o}^{n-1}\sum_{s=-v}^{v} f(s)$$

follows from the equivalence of the following counting procedures:






).

Hence on substituting Cox λs for ρ_s, we obtain

$$E\{I_p\} = \frac{2\sigma^2}{n}\sum_{r=o}^{n-1}\sum_{s=-v}^{v}\operatorname{Cos}\lambda s \operatorname{Cos} ks.$$

Using the formula

$$\sum_{s=-v}^{v} \cos \mu s = \sin(v + \tfrac{1}{2})\mu / \sin \tfrac{1}{2}\mu , \qquad (16)$$

we obtain

$$E\left\{I_p\right\} = \frac{\sigma^2}{n} \sum_{v=0}^{n-1} \left\{ \frac{\sin[(v+\frac{1}{2})(k+\lambda)]}{\sin[\frac{1}{2}(k+\lambda)]} + \frac{\sin[(v+\frac{1}{2})(k-\lambda)]}{\sin[\frac{1}{2}(k-\lambda)]} \right\},$$

or from the further formula

$$\sum_{v=0}^{n-1} \sin(v + \tfrac{1}{2})\mu = \sin^2 \tfrac{1}{2}n\mu / \sin \tfrac{1}{2}\mu , \qquad (17)$$

we finally obtain

$$E\left\{I_p\right\} = \frac{\sigma^2}{n} \left\{ \frac{\sin^2[\frac{1}{2}n(k+\lambda)]}{\sin^2[\frac{1}{2}(k+\lambda)]} + \frac{\sin^2[\frac{1}{2}n(k-\lambda)]}{\sin^2[\frac{1}{2}(k-\lambda)]} \right\}; \qquad (18)$$

(the formulae (16) and (17) are most simply derived by taking real or imaginary parts of suitable exponential sums).

In general this will be $O(\frac{1}{n})$, but as $k \to \lambda$, the second term $\to$ $n\sigma^2$, and $E\left\{I_p\right\}$ becomes large, a result which is used to detect the presence of harmonic components in the original series.

However, it is stressed that if a component of the autocorrelation ρ_s of the type $(\rho_s)_1$ exists, we have a corresponding component of $E\left\{I_p\right\}$ given by (14), which as n becomes large tends (cf. equation (9)) to

$$2\pi\sigma^2 f_1(k). \qquad (19)$$

This component, although finite, may also be rather large compared with $2\sigma^2$ for some values of k; so that for a large but finite value of n it is not so simple to separate the strictly harmonic components when the original series is of the general type we have considered in this section.

As an example, consider the linear Markoff process with autocorrelation coefficient $\rho_s = \rho_1^{|s|}$. This process contains no strictly harmonic components, and its transform F(w) corresponding to ρ_s is given by

$$f(w) = \frac{1}{\pi} \sum_{s=-\infty}^{\infty} \rho_1^{|s|} \operatorname{Cos} sw = \frac{1 - \rho_1^2}{\pi(1 + \rho_1^2 - 2\rho_1 \operatorname{Cos} w)}. \quad (20)$$

This function has a maximum in $(0,\pi)$ at $w = 0$, given by $f(w) = (1 + \rho_1)/[\pi(1 - \rho_1)]$, so that for large ρ_1 this function becomes large at w near zero. As a check on result (18), we re-obtain ρ_s from

$$\int_0^{\pi} \frac{\operatorname{Cos} sw\,(1-\rho_1^2)dw}{\pi(1+\rho_1^2-2\rho_1 \operatorname{Cos} w)} = \frac{1}{2}\int_0^{2\pi} \frac{e^{isw}(1-\rho_1^2)dw}{\pi(1-\rho_1 e^{iw})(1-\rho_1 e^{-iw})},$$

or on writing $z = e^{iw}$, and integrating round the unit circle, we evaluate the residue at the single pole inside the circle, $z = \rho_1$, (for $s > 0$), obtaining the result we started from, $\rho_s = \rho_1^s$, $(s > 0)$.

4. Standard Errors of Sample Estimates.

We shall now consider further the problem of estimating correlation or regression coefficients from a sample of data in the form of a time-series, i.e., n ordered observations $x_1, \ldots, x_n$. We confine ourselves as before to the case of stationary processes.

Suppose for simplicity we first consider the simple mean of the process $E\{X\}$, which for convenience we assume zero. Consider the estimate

$$\overline{X} = (1/n) \sum_{r=1}^{n} X_r.$$

We have $E\{\overline{X}\} = 0$, and further

$$E\{\overline{X}^2\} = (\sigma^2/n) \sum_{s=-n+1}^{n-1} (1 - \frac{|s|}{n}) \rho_s, \quad (1)$$

(this being a special case of formula (14) of the last section with $p = 0$). Thus the variance of our estimate is approximately σ^2/n, provided that the sum in (1) converges to μ. Sufficient conditions for its convergence are obviously that $\Sigma|\rho_s|$ and $\Sigma\,|s|\,|\rho_s|$ converge separately, but these are not necessary.

For example, if $\rho_s = \mathrm{Cos}\,\lambda s$, $\Sigma\,|\rho_s|$ does not converge, but we obtain from formula (16) of the last section that in this case

$$E\{\overline{X}^2\} = \frac{\sigma^2}{n^2}\left[\frac{\mathrm{Sin}^2\,\frac{1}{2}n\lambda}{\mathrm{Sin}^2\,\frac{1}{2}\lambda}\right], \tag{2}$$

so that for $\lambda \neq 0$, the variance of $\overline{X}$ still tends to zero as n increases; indeed, it tends to zero more rapidly than in the case of a completely random series.

Consider next the estimate of a covariance. If we know $E\{X\} = 0$, we may use the estimate

$$C_s = \frac{1}{n-s}\sum_{r=1}^{n-s} X_r X_{r+s}\,. \tag{3}$$

The variance of this quantity is easily evaluated by straightforward algebra. For example, for s = 1, and a completely random series, we find

$$E\{C_s{}^2\} = \frac{1}{(n-1)^2}\,E\sum_{r=1}^{n-1} X_r{}^2 X_{r+1}{}^2 + 2\sum_{r\neq s} X_r X_{r+1} X_s X_{s+1}$$

$$= \sigma^4/(n-1)^2. \tag{4}$$

It follows that the variance of the sample correlation coefficient

$$\frac{\Sigma\, X_r X_{r+1}}{(\Sigma\, X_r{}^2\,\Sigma\, X_{r+1}{}^2)^{\frac{1}{2}}} \sim \frac{\Sigma\, X_r X_{r+1}}{\Sigma\, X_r{}^2}$$

is $1/n{-}1 + O(1/n^2)$. We may also note that if X is measured from its sample mean value $\overline{X}$, we obtain for this 'corrected' covariance C_s, say, a negative bias of amount $-\sigma^2/n + O(1/n^2)$, which is worth allowing for in any approximate test of significance. The exact distribution of the sample correlation coefficient in the case of the 'cyclic definition'

$$\frac{X_1X_2 + X_2X_3 +, \ldots, + X_nX_1 - n\overline{X}^2}{X_1{}^2 + X_2{}^2 +, \ldots, + X_n{}^2 - n\overline{X}^2}$$

and on the assumption of normality is also known (R. L. Anderson, Ann. Math. Stat., 13(1942) 1)*, but we shall content ourselves with approximate but more general results. As above, we shall continue to evaluate only the dominant term in the variance and covariance formulae; and we may therefore for simplicity continue to assume that the mean is known, so that formula (3) is used.

For the covariance of C_s and C_{s+t} we obtain in general

$$E\left\{\frac{1}{(n-s)(n-s-t)}\sum_{r=1}^{n-s} X_r X_{r+s}\sum_{u=1}^{n-s-t} X_u X_{u+s+t}\right\} - \rho_s\,\rho_{s+t}\sigma^4,$$

or using the formula

$$\mu_{1111} = \mu_{1100}\mu_{0011} + \mu_{1010}\mu_{0101} + \mu_{1001}\mu_{0110} + \kappa_{1111}$$

connecting moments and semivariants of the fourth order, we have the expression

$$\frac{1}{(n-s)(n-s-t)}\sum_{r=1}^{n-s}\sum_{u=1}^{n-s-t}\phi(r-u)$$

where

$$\phi(r-u) = \rho_{r-u}\,\rho_{r-u-t} + \rho_{r-u-s-t}\,\rho_{r+s-u} + \kappa_{r-u,s,t},$$

say, or finally we have

$$\frac{1}{n-s}\sum_{v=-(n-s-t)-1}^{(n-s)-1} 1 - \frac{\eta(v)}{n-s-t}\,\phi(v) \qquad (5)$$

where
$$\eta(v) = \begin{cases} |v| & \text{for negative } v \\ 0 & \text{for positive } v \leq t \\ v-t & \text{for } t < v \leq n-s-1. \end{cases}$$

Using the approximate formula

$$\mathrm{Cov}\left(\frac{U}{W}, \frac{V}{W}\right) \sim \frac{1}{w^2}\,\mathrm{var}(U) - \frac{u}{w^2}\,\mathrm{cov}(V,W) - \frac{v}{w^3}\,\mathrm{cov}(U,W) + \frac{uv}{w^4}\,\mathrm{var}(W)$$

when the variation of U, V, and W about u, v, and w, respectively becomes small, and defining

$$R_s = \frac{\frac{1}{n-s}\sum X_r X_{r+s}}{\frac{1}{n}\sum X_r^2}, \qquad (6)$$

---

* See also P. L. Hsu, ibid, 17(1946), 350.

we obtain

$$\operatorname{cov}(R_s,\ R_{s+t}) \sim \frac{1}{n-s}\sum_{v=-\infty}^{\infty}(\rho_v\,\rho_{v-t} + \rho_{v-s-t}\,\rho_{v+s}$$

$$+ 2\rho_s\,\rho_{s+t}\,\rho_v^2 - 2\rho_s\rho_v\,\rho_{v-s-t} - 2\rho_{s+t}\,\rho_v\,\rho_{v-s}), \quad (7)$$

where we have assumed that $\rho_s \to 0$ as s increases rapidly enough for the transition from (5) to the simpler sum $\sum_{v=-\infty}^{\infty}\phi(v)$ to be valid as n increases, and where we have also omitted the component depending on the semivariant $\kappa_{v,s,t}$. This latter component is, of course, zero for normal processes, but it should also be noticed that for any linear process we had

$$\kappa_{v,s,t} = \kappa_4(Y)\sum_{u=-\infty}^{\infty} f(u)\ f(u+s)\ f(u+v)\ f(u+v+s+t;$$

hence

$$\sum_{v=-\infty}^{\infty}\kappa_{v,s,t} = \kappa_4(Y)\sum_{v=-\infty}^{\infty}\ \sum_{u=-\infty}^{\infty} f(u)\ f(u+s)\ f(u+v)\ f(u+v+s+t)$$

$$= \kappa_4(Y)\sum_{u=-\infty}^{\infty} f(u)\ f(u+s)\sum_{w=-\infty}^{\infty} f(w)\ f(w+s+t)$$

$$= \gamma(Y)\ \operatorname{Cov}(X_r, X_{r+s})\ \operatorname{cov}(X_r, X_{r+s+t}) \quad (8)$$

where $\gamma(Y) = \kappa_4(Y)/\sigma^4(Y)$. It will be found that in the expression (7) above for $\operatorname{cov}(R_s, R_{s+t})$ the $\gamma(Y)$ term vanishes.

As special cases of (7), we have for s large enough, so that ρ_s is small,

$$\operatorname{var}(R_s) \sim \frac{1}{n-s}\sum_{v=-\infty}^{\infty}\rho_v^2,$$

$$\operatorname{cov}(R_s,\ R_{s+t}) \sim \frac{1}{n-s}\sum_{v=-\infty}^{\infty}\rho_v\,\rho_{v+t}, \quad (9)$$

$$\rho(R_s,\ R_{s+t}) \sim \sum_{v=-\infty}^{\infty}\rho_v\,\rho_{v+t}\Big/\sum_{v=-\infty}^{\infty}\rho_v^2$$

The formulae (9) are important in indicating the magnitude of sampling fluctuations in the observed correlogram (graph of the observed autocorrelation coefficients) when the true autocorrelation coefficient ρ_s has dropped to zero.

As an illustration consider the observed autocorrelation coefficient R_s obtained from the autoregressive scheme

$$X_{t+2} = 1.1X_{t+1} - 0.5X_t + Y_{t+2}, \qquad (10)$$

using an actual series of 65 observations obtained by Kendall (E 1) as an artificial series using a rectangular distribution for Y.

The corresponding true correlations ρ_s are easily obtained from the equation

$$\rho_{s+2} = 1.1\,\rho_{s+1} - 0.5\,\rho_s,$$

with $\rho_o = 1$, $\rho_1 = 1.1/1.5$, and the 'correlations' σ_s of the correlations ρ_s as defined in (9) above may then also be calculated. We find also from the table that

$$\sum_{v=-\infty}^{\infty} \rho_v^2 = 2.44, \qquad \sum_{v=-\infty}^{\infty} \sigma_v^2 = 3.42.$$

s	R_s	ρ_s	σ_s	s	R_s	ρ_s	σ_s
1	+.70	+.733	+.832	15	-.30	-.006	-.024
2	+.29	+.307	+.434	16	-.18	-.002	-.012
3	+.01	-.029	+.002	17	+.12	+.001	-.010
4	-.17	-.186	-.286	18	+.29	+.002	+.005
5	-.27	-.190	-.364	19	+.33	+.002	-
6	-.25	-.116	-.276	20	+.22	+.001	-
7	-.13	-.033	-.118	21	+.05	-	-
8	+.07	+.022	+.022	22	-.12	-	-
9	+.12	+.041	+.096	23	-.28	-	-
10	+.05	+.034	+.102	24	-.43	-	-
11	-.05	+.017	+.071	25	-.57	-	-
12	-.17	+.002	+.019	26	-.56	-	-
13	-.27	-.007	-.015	27	+.26	-	-
14	-.31	-.008	-.027	28	+.02	-	-
				29	+.17	-	-
				30	+.27	-	-

Let us consider values of s from 11 to 30, for which ρ_s has become small. We have

$$\text{var}\,(R_s) \sim \frac{2.44}{65 - s},$$

or for the range of s considered,

$$\text{var}(R_s) \sim 0.053.$$

The observed variance of R_s from the table for this range of s is computed to be 0.083. This appears somewhat high, but in attempting to appraise its significance we must not allocate 20 d.f. to the observed value, since we also know adjacent values of R_s are correlated. In fact, treating the values of R_s as a new series for which the variance and autocorrelations are specified by the σ_s quantities, the effective degrees of freedom will be of the order 20/3.42 or only about 6. A variance ratio 0.083/0.053 with 6 d.f. would not reach the 5 per cent significance level, so that while this adaptation of standard tests is rather rough, it indicates that the observed values of R_s are fluctuating somewhat more than would be expected, but not significantly so.

In practice, of course, we do not know the true values ρ_s and hence do not immediately know the sampling errors in R_s. We might, however, for suitable series attempt to fit a stochastic model depending on a few parameters, and then calculate the theoretical correlogram corresponding to this model for comparison with that observed. This raises the question of efficient estimation of such parameters. Let us consider, for example, the case of the process

$$X_{t+2} + aX_{t+1} + bX_t = Y_{t+2}.$$

We obtain the least-squares solution of a and b by making the sum of squares of Y a minimum, i.e.,

$$\sum_{t=1}^{n-2} (X_{t+2} + aX_{t+1} + bX_t)^2 \qquad \text{a minimum.}$$

This gives approximately

$$\begin{cases} a_e + b_e R_1 + R_1 = 0 \\ a_e R_1 + b_e + R_2 = 0 \end{cases} \qquad (11)$$

as our estimates a_e and b_e of a and b respectively.

We cannot quote the usual standard error formulae for these estimates without further justification, since the X's play simultaneously the roles of dependent and independent variables. Mann and Wald (E 5) have, however, shown that these standard error formulae are still asymptotically valid for large samples. A simple (though less rigorous) demonstration is the present case given below.

Write $\Sigma X_{t+1} Y_{t+2} = A, \qquad \Sigma X_t Y_{t+2} = B.$ (12)

Subtracting (11) from (10), and denoting $a_e - a$ by δa, $b_e - b$ by δb, we obtain

$$\begin{cases} \delta A\ \Sigma X_{t+1}^2 + \delta b\ \Sigma X_t X_{t+1} = A \\ \delta A\ \Sigma X_t X_{t+1} + \delta b\ \Sigma X_t^2 \quad = B. \end{cases}$$

Using such results as $E\{Y_{t+1} X_t Y_{u+1} X_u\} = 0$, unless $u = t$ (e.g. if $u > t$, $E\{Y_{t+1} X_t Y_{u+1} X_u\} = E\{Y_{u+1}\}\ E\{X_{t+1} X_t X_u\} = 0$), we have

$$E\{A^2\} \sim E\{B^2\} \sim n\sigma^2(Y)\sigma^2(X),$$

$$E\{AB\} \sim n\ \rho_1 \sigma^2(Y)\sigma^2(X),$$

whence

$$\sigma^2(a_e) \sim \sigma^2(b_e) \sim \frac{\sigma^2(Y)}{n(1 - \rho_1^2)\sigma^2(X)} \sim \frac{1 - b^2}{n}. \qquad (13)$$

It may happen in practice that a random superposed error depresses the observed correlations. Thus if

$$X_t' = X_t + Z_t$$

where Z_t is a completely random series independent of X_t, we have

$$\sigma^2(X') = \sigma^2(X) + \sigma^2(Z)$$

but

$$\operatorname{cov}(X_r'\ X_{r+s}') = \operatorname{cov}(X_r\ X_{r+s}),$$

so that if $\sigma^2(Z)/\sigma^2(X) = \lambda$, we obtain a depression of the correlations by a factor $1/(1 + \lambda)$. If this is suspected, an estimate of the quantity λ will also be necessary.

5. Stationary Processes Specified for Continuous Time.

Since many time-series obviously exist for continuous time, we need to extend the theory of the previous sections to cover such series. We still consider the case of stationary processes. The autocorrelation ρ_T between X_t and X_{t+T} can now be defined for all T, and will be written as a function of T, $\rho(T)$.

We shall find that the theory is very similar to that for discrete time, except for certain further points of rigour associated with the continuous range of t, which raises questions about the 'continuity' or 'integrability' of the random or stochastic function $X(t)$. We shall, following Slutsky, define stochastic 'continuity' by the condition

$$\lim_{h \to 0} E\left\{[X(t+h) - X(t)]^2\right\} = 0; \qquad (1)$$

we may sometimes call this 'continuity in the mean', owing to its correspondence with the notion of 'limit in the mean' in analysis, and use the equivalent notation*

$$\underset{h \to 0}{\text{l.i.m.}} \; X(t+h) = X(t) \qquad (2)$$

It is easily shown that continuity of $X(t)$ in the above sense is closely related to classical continuity of the autocorrelation function. For we have the identity

$$E\left\{[X(t+h) - X(t)]^2\right\} = 2\sigma^2(1 - \rho(h))$$

*There is some danger of confusion in the use of the phrase 'in the mean', and the expression 'in mean square' used by J. E. Moyal (unpublished work), (cf. the French 'en moyenne quadratique', e.g., in F 1) may be preferred.

for stationary processes, so that (1) or (2) is true if $\rho(h)$ is continuous at $h = 0$, and conversely (we assume σ^2 is finite).

When the above condition is satisfied, we may define the stochastic integral

$$Z = \int_a^b \phi(t)\, X(t)dt, \tag{3}$$

where it is assumed that $\phi(t)$ is continuous in the ordinary sense. The existence of such an integral may be established in the standard way as the limit of the Riemann sum

$$Z_n = \sum_{\nu=1}^{n} \phi(t_\nu^{(n)})\; X(t_\nu^{(n)})(t_\nu^{(n)} - t_{\nu-1}^{(n)})$$

where $a = t_o^{(n)} < t_1^{(n)} \;.\;.\;. < t_{n-1}^{(n)} < t_n^{(n)} = b$,

whence it may be shown (see Cramer D 1) that

$$Z = \underset{n\to\infty}{\text{l.i.m.}}\; Z_n.$$

Similarly it may be shown that

$$E\{ZW\} = \int_a^b\int_a^b \phi(t)\psi(u)\; E\{X(t)\; Y(u)\}dt\; du \tag{4}$$

where W is similarly defined to Z and is

$$\int_a^b \psi(t)\; Y(t)dt.$$

Returning to consideration of the autocorrelation function $\rho(T)$, we shall now quote Khintchine's theorem on the consistency conditions for $\rho(T)$. The proof is similar to that indicated for the case of discrete t, and will be omitted. The necessary and sufficient condition that $\rho(T)$ is the autocorrelation of some continuous stationary process is that

$$\rho(T) = \int_o^\infty \text{Cos } Tw\; dF(w) \tag{5}$$

where F(w) is a distribution function defined between o and ; (by 'continuous' we refer to continuity in the sense just defined).

As an example, consider the autocorrelation function

$$\rho(T) = e^{-\mu|T|}. \tag{6}$$

The general inversion formula corresponding to (5) is

$$F(w) = \frac{2}{\pi}\int_0^\infty \left(\frac{\text{Sin } Tw}{T}\right) \rho(T)\, dT, \tag{7}$$

or if f(w) exists,

$$f(w) = \frac{2}{\pi}\int_0^\infty \text{Cos } Tw \;\rho(T)\, dT. \tag{8}$$

From this last formula, we obtain for (6)

$$f(w) = \frac{2\mu}{\pi(\mu^2 + w^2)}, \qquad (w > 0), \tag{9}$$

which is a 'half-Cauchy' distribution.

It is sometimes convenient to define F(w) from $-\infty$ to ∞ instead of from o to ∞; dF(w) (for real X) is symmetric, and $\rho(T)$ is formally equivalent to its characteristic function, i.e.,

$$\rho(T) = \int_{-\infty}^{\infty} e^{iTw} dF(w). \tag{10}$$

The distribution f(w) in (9) would then become the Cauchy distribution defined for all w.

The functions F(w) and f(w) maintain the same fundamental relation to the harmonic analysis of X(t) as we found in the case of discrete t. For example, if X(t) is a strictly harmonic series of the form Cos λt, then $\rho(T)$ is also Cos λT, and the function F(w) in (5) consists simply of one step at $w = \lambda$. In general, F(w) may be a mixture of a step-function whose steps correspond to oscillating components of this strictly harmonic type, and a differentiable function whose derivative f(w) corresponds to a continuous range of 'frequencies' w. F(w) is called the integrated 'spectrum' of the process X(t), its steps corresponding to a 'discrete spectrum' and f(w) to a 'continuous spectrum'.*

6. The Continuous Linear Process.

We shall now consider a particular but fairly wide class of process analogous to the linear process we considered for discrete t. As a specific example, suppose we attempt to formulate Yule's pendulum problem. The general representation of an oscillating pendulum subject to given forces will be

$$\ddot{x}(t) + \alpha\dot{x}(t) + \beta x(t) = y(t) \tag{1}$$

with particular solution

$$x(t) = (D^2 + \alpha D + \beta)^{-1}\, y(t)$$

$$= \frac{1}{\lambda_1 - \lambda_2}\left\{\frac{1}{D-\lambda_1} - \frac{1}{D-\lambda_2}\right\} y(t)$$

$$= \int_{-\infty}^{t}\left[\frac{e^{\lambda_1(t-v)} - e^{\lambda_2(t-v)}}{\lambda_1 - \lambda_2}\right] y(v)\, dv \tag{2}$$

where λ_1 and λ_2 are the roots of $z^2 + \alpha z + \beta = 0$. If the forces are of a stochastic character, we merely denote this by writing capital letters in (2) to obtain

$$X(t) = \int_{-\infty}^{t}\frac{e^{\lambda_1(t-v)} - e^{\lambda_2(t-v)}}{\lambda_1 - \lambda_2}\, Y(v)\, dv. \tag{3}$$

Equation (3) represents the 'steady' or stationary state, the extra component due to initial conditions having damped down to zero. We are particularly interested in the case when the forces Y(t) become impulsive in character, but shall not yet consider this case, which requires a little care, (though not more so than does the handling of impulsive forces in equation (1)).

---

*For further details see Wiener (F 12); cf. also F 2 and F 10.

It is evident that any linear operator on $x(t)$ of the type illustrated in (1) would lead to a linear operator solution of the type (3); we shall in general consider the case

$$X(t) = \int_{-\infty}^{t} g(t - v)\, Y(v)\, dv, \tag{4}$$

or as before, if we define $g(v) = 0$ for $v < 0$,

$$X(t) = \int_{-\infty}^{\infty} g(t - v)\, Y(v)\, dv, \tag{5}$$

where $g(v)$ is assumed to be continuous (except for the artificial discontinuity at $v = 0$).

Under these conditions, if we assume $Y(t)$ is continuous 'in the mean', we have

$$E\{X(t_1)X(t_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t_1-u)\, g(t_2-v)\, E\{Y(u)Y(v)\}\, du dv$$

or

$$\sigma^2(X)\, \rho_X(t_1-t_2) = \sigma^2(Y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t_1-u)\, g(t_2-v)\rho_Y(v-u)\, du dv \tag{6}$$

This equation may be written

$$\sigma^2(X)\rho_X(T) = \sigma^2(Y)\int_{-\infty}^{\infty} g(T-z)\left[\int_{-\infty}^{\infty} g(w-z)\rho_Y(w)\, dw\right] dz,$$

in which form it is evident that if we write

$$f_Y(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iTw}\, \rho_Y(T)\, dT$$

and

$$G(w) = \int_{-\infty}^{\infty} e^{-iTw}\, g(T)\, dT,$$

(assuming that these functions exist), then we have the rather simpler formula for these transforms

$$\sigma^2(X)\, f_X(w) = \sigma^2(Y)\, G(w)\, \overline{G}(w)\, f_Y(w). \tag{7}$$

In the case of equation (3), where

$$g(T) = \frac{e^{\lambda_1 T} - e^{\lambda_2 T}}{\lambda_1 - \lambda_2}, \qquad (T > 0),$$

we have

$$G(w) = \frac{1}{\lambda_1 - \lambda_2}\; \frac{1}{iw-\lambda_1} - \frac{1}{iw-\lambda_2} = 1/\Phi(iw), \tag{8}$$

where $(z) = z^2 + \alpha z + \beta$ (this result being obviously generalizable to any differential operator $\phi(D)$), or

$$\sigma^2(X)\, f_X(w) = \frac{\sigma^2(Y)\, f_Y(w)}{\phi(iw)\phi(-iw)} . \tag{9}$$

Thus if $\rho_Y(T) = e^{-\mu|T|}$,

$$f_X(w) = \frac{\sigma^2(Y)}{\sigma^2(X)} \frac{1}{(\beta - w^2)^2 + w^2\alpha^2} \quad \frac{\mu}{\pi(\mu^2 + w^2)} . \tag{10}$$

We may represent the limiting case of impulsive forces by letting μ become large; we then obtain

$$f_X(w) \rightarrow \frac{C}{(\beta - w^2)^2 + w^2\alpha^2} \tag{11}$$

where by integration from $-\infty$ to ∞ we have $C = \alpha\beta/\pi$. Comparing this result with (10), we note that we must suppose at the same time that $\sigma^2(Y)/\mu$ remains finite if $\sigma^2(X)$ is to be finite, in fact $\sigma^2(Y)/[\mu\sigma^2(X)] \rightarrow \alpha\beta$.

The exact expression for $\rho_X(T)$ corresponding to (10) is obtainable by the usual inversion. We have

$$\rho_X(T) = \psi(T)/\psi(0),$$

where

$$\psi(T) = \int_{-\infty}^{\infty} \frac{\mu e^{iTw}\, dw}{\pi\{(\beta - w^2)^2 + w^2\alpha^2\}\ (\mu^2 + w^2)},$$

or (for $T > 0$) evaluating the residues at the relevant poles $+ i\mu$, $- i\lambda_1$, $- i\lambda_2$, we have

$$\psi(T) = \frac{e^{-\mu T}}{(\beta+\mu^2)^2 - \mu^2\alpha^2} + \frac{\mu e^{-\lambda_1 T}}{(\mu^2-\lambda_1^2)\lambda_1(\lambda_1^2-\lambda_2^2)} + \frac{\mu e^{-\lambda_2 T}}{(\mu^2-\lambda_2^2)\lambda_2(\lambda_2^2-\lambda_1^2)} .$$

As μ becomes large, this gives

$$\rho_X(T) \rightarrow \frac{\lambda_2 e^{\lambda_1 T} - \lambda_1 e^{\lambda_2 T}}{\lambda_2 - \lambda_1}, \tag{12}$$

or in terms of α and β,

$$\frac{e^{-\frac{1}{2}\alpha T} \operatorname{Cos}\left[\sqrt{(\beta - \frac{1}{4}\alpha^2)}T - \theta\right]}{\operatorname{Cos}\theta}, \quad \left(\tan\theta = \frac{\frac{1}{2}\alpha}{\sqrt{(\beta - \frac{1}{4}\alpha^2)}}\right). \tag{13}$$

The above method is rather roundabout if we are only interested in the limiting case of 'random impulses', and we shall now briefly indicate the direct analogue of the theory of linear processes for discrete time to this case.

We replace the 'integrated force' $\int_0^t Y(v)\,dv$ by a Stieltjes integral* corresponding to 'integrated independent impulses'

$$\int_0^t dI(v),$$

the cumulant function for which satisfied the equation

$$K(\theta,\ t = t_1 + t_2) = K(\theta,\ t_1) + K(\theta,\ t_2). \tag{14}$$

Equation (14) is equivalent (cf. Cramer, A 1, Chapter VIII) to the equation

$$K(\theta, t) = tK_I(\theta), \tag{15}$$

where $K_I(\theta)$ represents the rate of change of $K(\theta,t)$. If we now consider

$$X(t) = \int_{-\infty}^{t} g(t-v)\,dI(v)$$

for a bounded continuous function g, the argument in the case of discrete time may evidently be extended to the above integral (by the use of approximating sums) to give the corresponding formulae

$$\begin{aligned} K_X(\theta) &= \int_{-\infty}^{\infty} K(g(v)\theta)dv, \\ K_{X_1,X_2}(\theta,\phi) &= \int_{-\infty}^{\infty} K_I(g(v)\theta + g(v+T)\phi)dv, \end{aligned} \tag{16}$$

*Cf. Doob, F 4.

etc, where it is assumed that all cumulants of K_I exist, with $\kappa_1 = 0$, and that $\int_{-\infty}^{\infty} g^2(v)dv$ exists, (we have also $g(v) = 0$ for $v < 0$, and $T = t_2 - t_1$). In particular, we have the formulae

$$\sigma^2(X) = \sigma_I^{\,2} \int_{-\infty}^{\infty} g^2(v)dv,$$

$$\rho_X(T) = \frac{\int_{-\infty}^{\infty} g(v)\ g(v+T)dv}{\int_{-\infty}^{\infty} g^2(v)dv}. \tag{17}$$

Applying these formulae to the case

$$g(v) = \frac{e^{\lambda_1 v} - e^{\lambda_2 v}}{\lambda_1 - \lambda_2}, \qquad (v > 0)$$

we readily re-obtain the autocorrelation function of equation (13), and also the relation

$$\sigma^2(X) = \sigma_I^{\,2}/(2\alpha\beta). \tag{18}$$

7. Standard Errors in the Case of Continuous Time.

There is no particular difficulty in the adaptation of the sampling error formulae of section 4 to the continuous case; for example, the approximate sampling variances and covariances of the correlation coefficient

$$R(T) = \frac{\int_0^T X_t X_{t+T}\ dt}{\int_0^T X_t^{\,2}\ dt} \tag{1}$$

are exactly as in the case of discrete sums, with integrals replacing sums and T replacing n (cf. Bartlett, E 6).

To illustrate the estimation of the coefficients in continuous linear processes, let us consider the estimation of α and β in equation (1) of the last section when the disturbing forces consist of independent random impulses of the type referred to in the last two sections. First of all, we note that the solution (2) (section 6) has differential coefficients $\dot{x}(t)$ and $\ddot{x}(t)$ satisfying (1). We have

$$\dot{x}(t) = \int_{-\infty}^{t} \left[\frac{\lambda_1 e^{\lambda_1(t-v)} - \lambda_2 e^{\lambda_2(t-v)}}{\lambda_1 - \lambda_2} \right] y(v)\, dv, \tag{2}$$

$$\ddot{x}(t) = y(t) + \int_{-\infty}^{t} \left[\frac{\lambda_1^2 e^{\lambda_1(t-v)} - \lambda_2^2 e^{\lambda_2(t-v)}}{\lambda_1 - \lambda_2} \right] y(v)\, dv. \tag{3}$$

The first of these coefficients, $\dot{x}(t)$, still exists if $y(v)dv$ is replaced by $d\,I(v)$, but the second only does so in a formal sense, since the first component becomes $d\,I(v)/dv$, which may be infinite. However, by regarding our equation as equivalent to the 'limit' of the equation integrated over an interval Δt when Δt becomes small, we have no difficulty in interpreting our formulae. The estimation of α and β by least squares, i.e., minimizing the sum of squares

$$\Sigma_{\Delta t} \left[I(t + \Delta t) - I(t) \right]^2,$$

gives as $\Delta t \rightarrow o$, the formal equations

$$\left. \begin{aligned} \int_o^T (x_t \ddot{x}_t + \alpha_e x_t \dot{x}_t + \beta_e x_t^2) dt = 0 \\ \int_o^T (\dot{x}_t \ddot{x}_t + \alpha_e \dot{x}_t^2 + \beta_e \dot{x}_t x_t) dt = 0 \end{aligned} \right\} \tag{4}$$

where x_t is the realized series between 0 and T, and α_e and β_e are our estimates of α and β. In these equations we have

$$\int_o^T x_t \dot{x}_t \, dt = \left[x_t^2 \right]_o^T, \tag{5}$$

or a quantity small compared with quantities like $\int_o^T x_t^2 dt$, when T becomes large. We must be careful about interpreting integrals involving $\ddot{x}_t$, which, however, we can replace by $(\dot{x}_{t+\Delta t} - \dot{x}_t)/\Delta t$. Thus, from the approximate orthogonality caused by (5), we have

$$\left. \begin{aligned} \alpha_e \sim - \operatorname*{Lt}_{\Delta t \rightarrow o} \frac{\int_o^T \dot{x}_t \left[(\dot{x}_{t+\Delta t} - \dot{x}_t)/\Delta t \right] dt}{\int_o^T \dot{x}_t^2 \, dt} \\ \beta_e \sim - \operatorname*{Lt}_{\Delta t \rightarrow o} \frac{\int_o^T x_t \left[(\dot{x}_{t+\Delta t} - \dot{x}_t)/\Delta t \right] dt}{\int_o^T x_t^2 \, dt}, \end{aligned} \right\} \tag{6}$$

or integrating the numerator in this last equation by parts, so that

$$\underset{\Delta t \to 0}{\mathrm{Lt}} \int_0^T x_t \left[(x_{t+\Delta t} - x_t)/\Delta t \right] dt$$

$$= \underset{\Delta t \to 0}{\mathrm{Lt}} \left\{ \left[x_t (x_{t+\Delta t} - x_t)/\Delta t \right]_0^T - \int_0^T \dot{x}_t \left[(x_{t+\Delta t} - x_t)/\Delta t \right] dt \right\}$$

$$\sim \quad - \int_0^T \dot{x}_t^{\,2}\, dt,$$

$$\beta_e \sim \frac{\int_0^T \dot{x}_t^{\,2}\, dt}{\int_0^T x_t^2\, dt}, \tag{7}$$

The limiting variances and covariance of α_e and β_e in repeated sampling are asymptotically

$$\mathrm{var}(\alpha_e) \sim \frac{\sigma_I^{\,2}}{\int_0^T \dot{x}_t^{\,2}\, dt} \sim \frac{2\alpha}{T}; \qquad \mathrm{var}(\beta_e) \sim \frac{\sigma_I^{\,2}}{\int_0^T x_t^2\, dt} \sim \frac{2\alpha\beta}{T};$$

$$\mathrm{cov}\,(\alpha_e, \beta_e) \sim 0. \tag{8}$$

If the data only exist as discrete observations, for other reasons, it may not be possible to use formulae (6) and (7) directly, but they will still indicate the maximum asymptotic accuracy obtainable with which the accuracy obtained by any consistent method of estimation, e.g., from the autocorrelation coefficients, may be compared.

Thus from the first two coefficients R_1 and R_2 obtained from the data formula (13) of the last section may be used to provide estimates of α and β, (with the aid of interpolatory methods). It has been verified (Bartlett, E 6) that the use of the first available coefficients R_1 and R_2 appears most reasonable from a comparison of the asymptotic accuracy actually reached by this procedure (in the case of α small) with the maximum asymptotic accuracy.

Incidentally, it should be noticed that, while a process defined for continuous time permits a series of discrete observations to be obtained from it (the converse does not necessarily hold), the discrete process so specified does not in the case of the process treated in detail above allow an exact discrete specification of the autoregressive type treated in section 2 (for a further discussion of this point, see Bartlett, E 6).

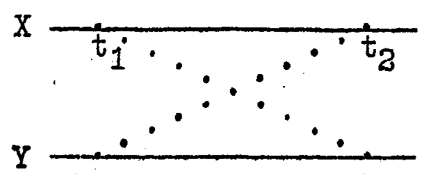
To illustrate the estimation procedure for actual data, Yule's original analysis (E 7) of Wolfer's sunspot numbers was repeated using the present model in place of his original autoregressive scheme. The comparison here is made on the basis of Yule's 'smoothed' data, (though it is pointed out that such smoothing is dangerous in analyzing time-series for periods). The 'period', defined in the present model as $2\pi/\sqrt{(\beta - \frac{1}{4}\alpha^2)}$, can be assigned an approximate standard error from the known approximate variances and covariance of α_e and β_e; thus we obtain from the limiting formulae (8) a coefficient of variation of order $\sqrt{[\alpha(2\beta + \frac{1}{2}\alpha^2)/T]}$. The period estimated was 10.8 ± 1.2 years, in comparison with Yule's estimate of 11.2 years. The corresponding autocorrelation coefficients ρ_s up to $s = 5$ are given in the table below.

Autocorrelations of Wolfer's Sunspot Numbers.

s	Observed (Yule's) smoothed data)	Calculated (Yule, E 7)	Calculated (Bartlett, E 6)
1	0.8407	(0.8407)	(0.8407)
2	0.4714	(0.4714)	(0.4713)
3	0.0470	0.0397	0.0605
4	-0.2641	-0.3181	-0.2562
5	-0.4043	-0.5139	-0.4091

8. Simultaneous Time-Series.

As we often wish to correlate more than one time-series, it is important to see how the preceding theory generalizes to multivariate or simultaneous stationary processes. For definitions we shall consider the case of two variables, $X(t)$ and $Y(t)$ with zero means. Here, in addition to the autocorrelations,



we have a cross-correlation between $X(t_1)$ and $Y(t_2)$ and another cross-correlation, not necessarily the same, between $X(t_2)$ and $Y(t_1)$. The covariance between $X(t_1)$ and $X(t_2)$ thus generalizes to a variance matrix

$$E\left\{\begin{pmatrix} X(t_1) \\ X(t_1) \end{pmatrix}\begin{pmatrix} X(t_2) & Y(t_2) \end{pmatrix}\right\} = E\left\{\begin{pmatrix} X(t_1)X(t_2) & X(t_1)Y(t_2) \\ X(t_2)Y(t_1) & Y(t_1)Y(t_2) \end{pmatrix}\right\}$$

or

$$\underset{\sim}{V}(T) = \begin{pmatrix} v_{11}(T) & v_{12}(T) \\ v_{21}(T) & v_{22}(T) \end{pmatrix}, \qquad (1)$$

say, where $T = t_2 - t_1$. The consistency properties of this matrix may be summarized in a theorem which is an extension of Khintchine's theorem for one series (the theorem, due to Cramer (D 1), is quoted in the next section).

Let us now consider the structure of $X(t)$ and $Y(t)$ for a specific example. We assume two 'coupled' series:

$$\begin{aligned} \dot{X}(t) + a_{11}X(t) + a_{12}Y(t) &= A(t) \\ \dot{Y}(t) + a_{21}X(t) + a_{22}Y(t) &= B(t), \end{aligned} \qquad (2)$$

whence operationally

$$\begin{aligned} X(t) &= \left[\Delta(D)\right]^{-1}\left[(D + a_{22}A(t) - a_{12}B(t)\right] \\ Y(t) &= \left[\Delta(D)\right]^{-1}\left[(D + a_{11}B(t) - a_{21}A(t)\right], \end{aligned} \tag{3}$$

where $\Delta(D) = (D + a_{11})(D + a_{22}) - (a_{11}a_{22} - a_{12}a_{21})$. These equations, (which can still be given a valid interpretation in the limiting case of random 'impulses'), are particular cases of the general linear processes:

$$\left.\begin{aligned} X(t) &= \int_{-\infty}^{\infty}\left[g_{11}(t-v)A(t) + g_{21}(t-v)B(t)\right]\ dt \\ Y(t) &= \int_{-\infty}^{\infty}\left[g_{21}(t-v)A(t) + g_{22}(t-v)B(t)\right]\ dt. \end{aligned}\right\} \tag{4}$$

The possible existence of <u>two</u> series of random disturbances $A(t)$ and $B(t)$, with unknown mutual relation, is a serious complication to the straightforward generalization of the sampling error formulae previously obtained for one series. For example, if we wish to estimate the coefficients a_{ij} in equation (2) from actual data, the least-squares solution in the limiting case of impulsive disturbances will require minimizing a quadratic form in the two sets of disturbances, with unknown ratio of variances, and unknown correlation, which have in consequence also to be estimated. The estimation problem simplifies, however, if $A(t)$ and $B(t)$ are assumed mutually uncorrelated, since the coefficients in each equation may then be estimated separately.

The approximate sampling errors for the cross-correlations between <u>any</u> two stationary processes can be determined by obvious extension of the previous methods for one series. We note a useful special case when the two series are entirely independent:

$$\left.\begin{aligned} \sigma^2\left\{R_{12}(t)\right\} &\sim \frac{1}{T-t}\int_{-\infty}^{\infty}\rho_{11}(v)\,\rho_{22}(v)\ dv, \\ \operatorname{cov}\left\{R_{12}(t),\ R_{12}(t+s)\right\} &\sim \frac{1}{T-t}\int_{-\infty}^{\infty}\rho_{11}(v)\,\rho_{22}(v+s)\ dv. \end{aligned}\right\} \tag{5}$$

We note also the joint cumulant function for $X(t_1)$ and $Y(t_2)$ in the case of linear processes of type (4), when the disturbances become impulsive in character. If the joint cumulant function for the disturbances integrated over the unit interval is $K_{I,J}(\theta,\phi)$, then

$$K_{X_1,Y_2}(\theta,\ \phi) = \int_{-\infty}^{\infty} K_{I,J}(\alpha,\beta)\ dv, \tag{6}$$

where

$$\begin{cases} \alpha = g_{11}(t_1 - v)\theta + g_{12}(t_2 - v)\phi \\ \beta = g_{21}(t_1 - v)\theta + g_{22}(t_2 - v)\phi\ . \end{cases}$$

*9. Further Notes on Simultaneous Time-Series.

We conclude with some notes on the formal extension of some of the results mentioned in the last section to the case of any number of series.

We denote the series by the column vector $\underset{\sim}{X}(t) \equiv \{X_i(t)\}$; it is also useful to consider the single series, defined for arbitrary λ_i,

$$X(t,\underset{\sim}{\lambda}) = \underset{\sim}{\lambda}'\ \underset{\sim}{X}(t), \tag{1}$$

where $\underset{\sim}{\lambda}$ is the column vector of coefficients λ_i. For this representation to be sufficiently general, however, we require $\underset{\sim}{\lambda}$ to be complex, corresponding to the possibility of separating the distinct terms $v_{ij}(T)$ and $v_{ji}(T)$ when we consider the autocovariance of $X(t,\underset{\sim}{\lambda})$.

We therefore first give the extension of Khintchine's theorem to complex valued processes. We define

$$v(T) = E\left\{X(t)\ \overline{X}(t + T)\right\},$$

and then we must have

$$v(T) = \int_{-\infty}^{\infty} e^{iTw}\ dU(w), \tag{2}$$

where $U(w)$ is a never-decreasing real function such that $U(-\infty) = 0$, $U(+\infty) = \sigma^2(X) = E\left\{X(t)\,\overline{X}(t)\right\}$. With this extension the real and symmetric character of $v(T)$ has disappeared, and $v(T)/\sigma^2(X)$ has the form of a characteristic function in general; the proof of the theorem is, however, essentially unchanged.*

Cramer's extension of (2) to simultaneous series (see D 1) may be written in matrix notation

$$\underset{\sim}{V}(T) = \int_{-\infty}^{\infty} e^{iTw}\, d\underset{\sim}{U}(w), \tag{3}$$

where $\Delta \underset{\sim}{U}(w)$ is a non-negative definite Hermitian matrix. In terms of the series defined by (1), we have from (2)

$$\underset{\sim}{\lambda}'\underset{\sim}{V}(T)\underset{\sim}{\lambda} = \int_{-\infty}^{\infty} e^{iTw}\, dU(w,\underset{\sim}{\lambda}), \tag{4}$$

for arbitrary $\underset{\sim}{\lambda}$, where $\Delta U(w,\underset{\sim}{\lambda})$ is non-negative; and result (3) amounts to the identification in (4) of the Hermitian form $\underset{\sim}{\lambda}'\underset{\sim}{U}(w)\underset{\sim}{\lambda}$ with the increasing quantity $U(w,\underset{\sim}{\lambda})$.

The relation of (3) to the harmonic analysis of $\underset{\sim}{X}(t)$ is also worth noting. It has been shown by Cramer (F 2) that $X(t)$ is equivalent to an integral of the form $\int_{-\infty}^{\infty} e^{itu}\, dI(u)$, where the successive stochastic increments $dI(u)$ are uncorrelated. We correspondingly obtain for the vector process $\underset{\sim}{X}(t)$,

$$\begin{aligned}\underset{\sim}{V}(T) = E\left\{\underset{\sim}{X}(t_1)\overline{\underset{\sim}{X}}'(t_1+T)\right\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(t_1u+t_2v)} E\left\{d\underset{\sim}{I}(u)\,\overline{\underset{\sim}{I}}'(v)\right\} \\ &= \int_{-\infty}^{\infty} e^{iTw}\, d\underset{\sim}{U}(w), \quad (t_1 + T = t_2).\end{aligned}$$

The generalized linear process we define by

$$\underset{\sim}{X}(t) = \int_{-\infty}^{\infty} \underset{\sim}{G}(t-v)\, \underset{\sim}{Y}(v)\, dv, \tag{5}$$

where $\underset{\sim}{G}(v)$ is a matrix function $\left\{g_{ij}(v)\right\}$. We then have

$$\begin{aligned}\underset{\sim}{V}_X(T) &= E\left\{\underset{\sim}{X}(t)\overline{\underset{\sim}{X}}'(t+T)\right\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underset{\sim}{G}(v)\, \underset{\sim}{V}_Y(v-u)\overline{\underset{\sim}{G}}'(u-T)\, du\,dv,\end{aligned} \tag{6}$$

*For an alternative proof, see F 8.

with corresponding transform

$$\underset{\sim}{T}_X(w) = \underset{\sim}{H}(w)\ \underset{\sim}{T}_Y(w)\ \overline{\underset{\sim}{H}}'(w), \tag{7}$$

say, where we assume $\underset{\sim}{T} = d\underset{\sim}{U}/dw$ to exist.

In the limiting case of independent 'impulses', we write in place of (5),

$$\underset{\sim}{X}(t) = \int_{-\infty}^{\infty} \underset{\sim}{G}(t-v)\ d\underset{\sim}{I}(v) \tag{8}$$

where $\int_0^t d\underset{\sim}{I}(v)$ has the cumulant function $K_{\underset{\sim}{I}}(\underset{\sim}{\theta})$. We then have (for real quantities $\underset{\sim}{G}$ and $\underset{\sim}{I}$)

$$K_{\underset{\sim}{X}}(\underset{\sim}{\theta}) = \int_{-\infty}^{\infty} K_{\underset{\sim}{I}}(\underset{\sim}{G}'(v)\underset{\sim}{\theta})dv, \tag{9}$$

$$K_{\underset{\sim}{X}(t_1),\underset{\sim}{X}(t_1+T)}(\underset{\sim}{\theta},\underset{\sim}{\phi}) = \int_{-\infty}^{\infty} K_{\underset{\sim}{I}}(\underset{\sim}{G}'(v)\underset{\sim}{\theta} + \underset{\sim}{G}'(v-T)\underset{\sim}{O})dv$$

etc,. In particular, we have

$$\underset{\sim}{V}_X(T) = \int_{-\infty}^{\infty} \underset{\sim}{G}(v)\ \underset{\sim}{J}\ \underset{\sim}{G}'(v-T)dv \tag{10}$$

where $E\{\underset{\sim}{I}\ \underset{\sim}{I}'\} = \underset{\sim}{J}$, as a special case of (6).

In the case of operational equations (with real coefficients) of the type

$$\underset{\sim}{\Phi}(D)\underset{\sim}{X}(t) = \underset{\sim}{Y}(t), \tag{11}$$

we have the solution

$$\underset{\sim}{X}(t) = \left[\underset{\sim}{\Phi}^{-1}(D)\right] Y(t), \tag{12}$$

for which (7) becomes

$$\underset{\sim}{T}_X(w) = \left[\underset{\sim}{\Phi}^{-1}(iw)\right]\ \underset{\sim}{T}_Y(w)\left[\underset{\sim}{\Phi}^{-1}(-iw)\right]'. \tag{13}$$

For example, the generalizations of equation (2) of the last section gives

$$\dot{\underset{\sim}{X}}(t) + \underset{\sim}{A}\underset{\sim}{X}(t) = \underset{\sim}{Y}(t) \tag{14}$$

with solution

$$\underset{\sim}{X}(t) = (D + \underset{\sim}{A})^{-1}\ \underset{\sim}{Y}(t), \tag{15}$$

and

$$\underset{\sim}{T}_X(w) = (iw + \underset{\sim}{A})^{-1}\ \underset{\sim}{T}_Y(w)(-iw + \underset{\sim}{A}')^{-1}. \tag{16}$$

IV. SELECTED BIBLIOGRAPHY

(This list was recommended as further reading matter in connection with the course, and is a very restricted list. Further references will be found at the end of many of the works cited.)

A. Advanced Probability.

(1) Cramér, H., Random Variables and Probability Distributions, Cambridge, 1937.

(2) ________, Mathematical Methods of Statistics, Princeton, 1946.

(3) Fréchet, M., "Recherches théoriques modernes sur la théorie des probabilities," Traité du Calcul des Probabilities, (ed. Borel), Tome 1, Fasc. 3 Paris, 1937-38.

B. Applications in Physics.

(1) Arley, N., "On the Theory of Stochastic Processes and their Application to the Theory of Cosmic Radiation," Copenhagen, 1943.

(2) Chandrasekhar, S., "Stochastic Problems in Physics and Astronomy," Reviews of Modern Physics, 15 (1943), 1.

(3) Rice, S. O., "Mathematical Analysis of Random Noise," Bell System Techn. Journ., 23 (1944), 282; 24 (1945, 46).

C. Population Growth and Industrial Renewal Theory.

(1) Feller, W., "On the Integral Equation of Renewal Theory," Ann. Math. Stat., 12 (1941, 243.

(2) Leslie, P. H., "On the Use of Matrices in Certain Population Mathematics," Biometrika, 34 (1945) 212.

(3) Lotka, A. J., "A Contribution to the Theory of Self-Renewing Aggregates, with special reference to industrial replacement," Ann. Math. Stat., 10 (1939), 1.

D. Theory of Stationary Processes.

(1) Cramér, H., "On the Theory of Stationary Random Processes," Ann. Math., 41 (1940), 215.

(2) Khintchine, A., "Korrelationstheorie der Stationären Stochastischen Prozesse," Math. Annalen., 109 (1933-34), 604.

(3) Wold, H., "Analysis of Stationary Time-Series," Uppsala, 1938.

E. Analysis of Time-Series.

(1) Kendall, M. G., "On Autoregressive Time-Series," Biometrika, 33 (1944), 105.

(2) ________, "On the Analysis of Oscillatory Time-Series," J. Roy. Stat. Soc., 107 (1945).

(3) ________, Researches in Oscillatory Time-Series, Cambridge, 1946.

(4) ________, Advanced Theory of Statistics, London, vol. II, 1946.

(5) Mann, H. B. and Wald, A., "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrica, 11 (1943), 173.

(6) Symposium on Time-Series (Bartlett, M. S.; Foster, G. A. R.; Cunningham, L. B. C.; and Hynd, W.R.B,) J. Roy Stat. Soc., (Suppl.), 7 (1946).

(7) Yule, G. U., "On a Method of Investigating Periodicities in Disturbed Series, with special reference to Wolfer's sunspot numbers," Phil. Trans., A 226 (1927), 267.

F. Further Selected List of References on the General Theory.

(1) Bass, J., "Les fonctions aléatoires et leur interpretation mécanique," La Revue Scientifique, 83 (1945), 3.

(2) Cramér, H., "On Harmonic Analysis in Certain Functional Spaces," Arkiv for Matematik, Astronomi och Fysik, 28 (1942), No. 12.

(3) Doob, J. L., "Stochastic Processes Depending on a Continuous Parameter," Trans. Am. Math. Soc., 42 (1937), 107.

(4) Doob, J. L., "The Elementary Gaussian Processes," <u>Ann. Math. Stat.</u>, 15(1944), 229.

(5) Feller, W., "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein Wahrscheinlichkeits-theoretischer Behandlung," <u>Acta. Biotheoretica</u>, 5(1939), 11.

(6) Kolmogoroff, A., "Uber die analytische Methoden in der Wahrscheinlichkeitsrechnung," <u>Math. Annalen</u>, 104(1931), 415.

(7) ________, "Grundbegriffe der Wahrscheinlichkeits-rechnung," <u>Ergebnisse der Mathematik</u>, 2, No. 3 (Berlin, 1933).

(8) Loeve, M., "Sur les fonctions aleatoires stationnaires de second ordre," <u>La Revue Scientifique</u>, 83 (1945), 297.

(9) Slutsky, E., "Sur les fonctions eventuelles continues, integrables et derivables dans le sens stochastique," <u>Comptes Rendues Ac. Sci.</u>, 187 (1928), 370.

(10) ________, "Sur les fonctions aleatoires presque periodiques et sur la decomposition des fonctions aleatoires stationnaires en composantes," <u>Actualites Scientifiques et Industrielles</u>, 738, (Paris, 1938).

(11) Uspensky, J. V. <u>Introduction to Mathematical Probability</u>, New York (1937), (Chapters 5 and 8).

(12) Wiener, N., "Generalized Harmonic Analysis," <u>Acta. Mathematica</u>, 55 (1930), 117.