

2

PROBABILITY DISTRIBUTIONS ASSOCIATED WITH
FINITE MARKOV CHAINS

Thesis

for the Degree of Doctor of Philosophy

Submitted by

JOHN LEONARD MOTT, M.A.

University of Edinburgh

October, 1956.



Probability Distributions associated with
Finite Markov Chains.

I should like to thank Professor A.C. Aitken both for the interest aroused in this field of mathematics by his course of research lectures 'Linear Operations in Probability', delivered in the University of Edinburgh in the session 1950-51, and for his constant encouragement and guidance throughout the time of preparation of this thesis.

PROBABILITY DISTRIBUTIONS ASSOCIATED WITH
FINITE MARKOV CHAINS.

| | Page |
|---|--------|
| Chapter 1. The Behaviour of Infinite Products of Finite Stochastic Matrices | 1 |
| 1.1 Introduction: stochastic matrices | 1 |
| 1.2 The existence of $\lim_{n \rightarrow \infty} P^n$ | 2 |
| 1.3 Illustration of this result for a binomial matrix | 6 |
| 1.4 Stable matrices | 8 |
| 1.5 Discussion of $\lim_{n \rightarrow \infty} P_1 P_2 \dots P_n$ | 10 |
| 1.6 Range. Conditions for ultimate stability | 12 |
| 1.7 Comparison of the variable and constant cases | 14 |
| 1.8 Range for binomial matrices | 16 |
| 1.9 A more general condition for ultimate stability | 17 |
| 1.10 Examples | 18 |
| 1.11 A second inequality for the range of a product of matrices | 21 |
| 1.12 A proof that the root $\lambda = 1$ of P is simple if $P > 0$ | 22 |
| 1.13 A second proof. | 23 |
| 1.14 An inequality for the subdominant root of P | 24 |
| Chapter 2. Conditions for the Existence of $\lim_{n \rightarrow \infty} P_1 P_2 \dots P_n$ | 28 |
| 2.1 An example in which $S = \lim_{n \rightarrow \infty} P_1 P_2 \dots P_n$ exists | 28 |
| 2.2 Necessary conditions for the existence of S | 30 |
| 2.3 Sufficient conditions for the existence of S | 32 |
| 2.4 Examples | 35 |

| | Page |
|------------|---|
| 2.5 | The conditions for uniformly positive chains 39 |
| 2.6 | A remark on the product of regular matrices 40 |
| 2.7 | Sufficient conditions for the existence of S if the P_i are regular. 41 |
| 2.8 | Existence of S if $P_i \rightarrow P > 0$ as $i \rightarrow \infty$ 44 |
| 2.9 | A generalisation of the conditions of 2.3 47 |
| 2.10 | Discussion of the case in which $P_i \rightarrow I$ as $i \rightarrow \infty$ 49 |
| 2.11 | A remark on the Cesaro summability of the products $P_1 P_2 \dots P_n$ 53 |
| | |
| Chapter 3. | The Frequency of the Occurrence of a State in the Binomial Case 59 |
| 3.1 | Introduction 59 |
| 3.2 | Numerical examples. Symmetric case, $\delta > 0$ 60 |
| 3.3 | Asymmetric case, $\delta > 0$ 64 |
| 3.4 | Symmetric case, $\delta < 0$ 65 |
| 3.5 | The mean score in the constant case 67 |
| 3.6 | The normal law 68 |
| 3.7 | A remark on the variance 73 |
| | |
| Chapter 4. | The Variance of the Frequency Distribution 76 |
| 4.1 | The mean score 76 |
| 4.2 | The c.f. of the score distribution. 79 |
| 4.3 | A matrix transformation 81 |
| 4.4 | A numerical illustration 82 |
| 4.5 | The increments to μ_2 on passing from stage to stage, constant case 84 |

| | Page |
|--|------|
| 4.6 μ_2 in the symmetric case | 88 |
| 4.7 Limiting μ_2 in the symmetric case | 89 |
| 4.8 Extension to variable chains | 90 |
| 4.9 Evaluation of μ_2 in some particular cases | 92 |
| 4.10 μ_2 when $P_i \rightarrow P$ as $i \rightarrow \infty$, symmetric case. | 95 |
| 4.11 The sign of the increments, constant case | 97 |
| 4.12 Proof that $\mu_2^{\omega} = O(n)$ in the variable uniformly positive case | .101 |
| 4.13 Extension to the general variable chain. | .104 |
| 4.14 μ_2 when $P_i \rightarrow P$ as $i \rightarrow \infty$, general case | .108 |
| | |
| Chapter 5. The Proof by the Method of Increments of the Normal Law for a Constant Binomial Chain | .110 |
| 5.1 Introduction | .110 |
| 5.2 Proof that $\mu_3 = O(n)$ | .110 |
| 5.3 Consideration of μ_4 | .115 |
| 5.4 Consideration of μ_5 | .117 |
| 5.5 Consideration of μ_6 | .119 |
| 5.6 Proof by increments of the normal law | .121 |
| | |
| Chapter 6. Proof of the Normal Law for a Variable Binomial Chain | .128 |
| 6.1 Introduction | .128 |
| 6.2 Consideration of μ_4 | .128 |
| 6.3 Consideration of μ_{2r} | .132 |
| 6.4 Generalisation: score increments variable from trial to trial | .134 |

| | Page |
|---|-------|
| Chapter 7. Proof of the Normal Law for a Trinomial Chain and for a Chain of any Finite Order . | . 139 |
| 7.1 Proof in the trinomial case of independent trials | . 139 |
| 7.2 Discussion of the order of the variance in this case | . 141 |
| 7.3 Proof by the method of increments in the independent case | . 145 |
| 7.4 First part of the proof in the case of a variable trinomial chain. | . 147 |
| 7.5 Discussion of the order of $\mu_2^{(n)}$ | . 152 |
| 7.6 Particular recurrence relations | . 155 |
| 7.7 Completion of the proof | . 156 |
| 7.8 Extension of the proof to a chain with k states | . 163 |
| References | . 167 |

CHAPTER I.

1.1. A stochastic matrix P of order k is a square matrix with k rows for which p_{ij} , the element of the i^{th} row and j^{th} column, is such that

- (i) $0 \leq p_{ij} \leq 1$ for all i, j ;
- (ii) $\sum_{j=1}^k p_{ij} = 1$ for all i .

If $p_{ij} > \epsilon > 0$ for all i, j we write $P > \epsilon$ and say that P is positive.

A stochastic matrix can be considered more naturally as a 'matrix of transition probabilities'. Suppose that at successive times an event E occurs and that in every case E is in fact one of the events E_1, E_2, \dots, E_k . Suppose further that the occurrences of these events are not independent but are related in the following way, that if the event E_i occurs at one time, or stage, then the probability of transition at the next stage from this state E_i to the state E_j is p_{ij} . Then clearly a stochastic matrix can be considered as a matrix $[p_{ij}]$ of transition probabilities, and conversely. To emphasise this it would be natural to call such a matrix a 'transition matrix', but this term has been used elsewhere* for a matrix of a different kind, and the term 'stochastic matrix' is in common use.

Dependent processes of this kind were first studied by Markov** and are known as Markov chains, constant if the matrix of transition probabilities is constant from stage to stage, otherwise variable.

* Ledermann. (The names of authors that occur in footnotes relate to their papers listed on p. 167)

** See, for example, Markov, 12, 13, 14, 15.

Consideration of the frequencies of occurrence of the various states leads to the discussion of probability distributions, but our main purpose in this and the succeeding chapter is the study of the related problem of the behaviour, as $n \rightarrow \infty$, of a product $P^{(n)}$ of variable stochastic matrices each of order k :

$$P^{(n)} = P_1 P_2 \dots P_n.$$

As an introduction, and for comparison later with the more general case, we consider first the particular, constant, case in which $P_i = P$ for all i . This leads to a considerable simplification, both of results and of proofs, for it allows us to discuss the behaviour of the matrix power P by the use of classical matrix theory.

1.2. Let $P \equiv [p_{ij}]$ be a stochastic matrix of order k . Then^{*}:

(i) $\lambda = 1$ is a latent root of P .

For $P\underline{x} = \underline{x}$, where \underline{x} is the column vector $\{1, 1, \dots, 1\}$.

(ii) $|\lambda| \leq 1$ for every root λ of P .

For let λ be a root of P and $\underline{x} = \{x_1, \dots, x_k\}$ be an associated non-trivial column vector, so that

$$P\underline{x} = \lambda\underline{x}$$

* Fréchet, pp. 105-107.

Let x_α be an element of x not exceeded in modulus by any other element. Then, abstracting the α^{th} row from the matrix equation above,

$$\lambda x_\alpha = p_{\alpha 1} x_1 + \dots + p_{\alpha k} x_k$$

so that

$$\begin{aligned} |\lambda| |x_\alpha| &\leq p_{\alpha 1} |x_1| + \dots + p_{\alpha k} |x_k| \\ &\leq (p_{\alpha 1} + \dots + p_{\alpha k}) |x_\alpha| \end{aligned}$$

whence

$$|\lambda| \leq 1$$

since $|x_\alpha| \neq 0$ and $p_{\alpha 1} + \dots + p_{\alpha k} = 1$.

(iii) If the diagonal elements of P are positive there is no complex root such that $|\lambda| = 1$.

For, with the notation and method of (ii),

$$|\lambda - p_{\alpha\alpha}| |x_\alpha| \leq (1 - p_{\alpha\alpha}) |x_\alpha|$$

from which, with $p_{\alpha\alpha} > 0$, the result follows.

The root $\lambda=1$ can be multiple, as it is for example in the trivial case of $P = I$, and there can be complex roots such that $|\lambda|=1$, as there are for example in the case of the binomial permutation matrix, but we can show that :

Theorem

If the root $\lambda=1$ is simple and dominates the other roots in modulus, then $\lim_{n \rightarrow \infty} P^n$ exists, and is of rank 1; and conversely*.

* A slightly different proof of this theorem is given by Aitken.

Proof.

Suppose that the roots of P are $1, \lambda_2, \dots, \lambda_k$ (not necessarily distinct), where $|\lambda_i| < 1$ for $2 \leq i \leq k$.

(i) If all the roots of P are simple, a non-singular H exists such that

$$HPH^{-1} = \text{diag} \{ 1, \lambda_2, \dots, \lambda_k \}$$

so that

$$\begin{aligned} HP^n H^{-1} &= \text{diag} \{ 1, \lambda_2^n, \dots, \lambda_k^n \} \\ &\rightarrow \text{diag} \{ 1, 0, \dots, 0 \} \end{aligned}$$

as $n \rightarrow \infty$ since $|\lambda_i| < 1$.

Thus in this case $P^n \rightarrow U$, where U is a stochastic matrix in which all rows are identical.

(ii) If, however, one or more of the roots $\lambda_2, \dots, \lambda_k$ is multiple the same result holds, and the proof is similar. In this case a non-singular H exists such that HPH^{-1} is in the classical canonical form

$$HPH^{-1} = \text{diag} \{ 1, K_2, \dots, K_m \}$$

where each K_i is either one of the roots $\lambda_2, \dots, \lambda_k$ or a block matrix of the form

$$K_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}$$

$$= \lambda_i I + W, \text{ say,}$$

so that W^2, W^3, \dots occupy successively higher superdiagonals and where, if λ_i is an r -fold root, $W^r = 0$.

Thus

$$K_i^n = \lambda^n + \binom{n}{1} \lambda^{n-1} W + \binom{n}{2} \lambda^{n-2} W^2 + \dots + \binom{n}{r-1} \lambda^{n-r+1} W^{r-1}$$

$$\rightarrow \text{diag} \{0, 0, \dots, 0\}$$

as $n \rightarrow \infty$, since

$$\binom{n}{\theta} \lambda^{n-\theta} < \lambda^{-\theta} \frac{n^\theta \lambda^n}{\theta!} \rightarrow 0$$

if $|\lambda| < 1$

Thus $H P^n H^{-1} \rightarrow \text{diag} \{1, 0, \dots, 0\}$, and the result follows as before.

We see at once that the converse of the result is true also. For if the root $\lambda=1$ of P is not simple, or if there is a second root λ_i such that $|\lambda_i|=1$, then at least two of the diagonal elements of the triangular matrix $H P^n H^{-1}$ are not only non-zero but are of unit modulus for all n so that $H P^n H^{-1}$ clearly cannot tend to a limiting matrix of rank less than 2. Then, since rank is invariant under a non-singular transformation, it follows that P^n cannot tend to a limit matrix of rank 1.

Thus if the root $\lambda=1$ of P is simple and dominant

$$U = \lim_{n \rightarrow \infty} P^n \quad \text{exists.} \quad \text{This is the case, for instance,}$$

if $P > 0$, as we shall prove in 1.12. We can, however, relate U more directly to P . For the columns of H above are column vectors associated with the roots $1, \lambda_2, \dots, \lambda_k$ of P , and the rows of H are the corresponding row vectors. Thus each row of U , a matrix of rank 1, is the row vector u ,

normalised so that the sum of its elements is unity, associated with the dominant root $\lambda=1$ of P .

1.3. We illustrate these results by an example, that in which P is the binomial matrix

$$P = \begin{bmatrix} \Sigma & P \\ \Sigma' & P' \end{bmatrix}.$$

Then $\lambda_1=1$ and

$$\begin{aligned} \lambda_2 &= \Sigma P - 1 \\ &= \Sigma + P' - 1 \\ &= P' - P. \end{aligned}$$

Thus $|\lambda_2| < 1$ if $P' - P < 1$, so in the binomial case (but not more generally) the condition that P be positive is equivalent to the condition that $\lambda=1$ exceeds all other roots in modulus.

The row vector associated with the root $\lambda=1$ is, when it is normalised,

$$\frac{1}{P + \Sigma'} [\Sigma', P]$$

so that

$$P^n \rightarrow \frac{1}{P + \Sigma'} \begin{bmatrix} \Sigma' & P \\ \Sigma' & P \end{bmatrix}.$$

Alternatively we can proceed as follows and use a certain matrix transformation. Let

$$H = \begin{bmatrix} 1 & ; \\ -1 & ; \end{bmatrix} \quad \text{so that} \quad H^{-1} = \begin{bmatrix} 1 & ; \\ 1 & ; \end{bmatrix}.$$

Then

$$HPH^{-1} = \begin{bmatrix} 1 & P \\ \cdot & \delta \end{bmatrix}, \quad \text{where} \quad \delta = p' - p,$$

so that

$$HP^2H^{-1} = (HPH^{-1})^2 = \begin{bmatrix} 1 & P + P\delta \\ \cdot & \delta^2 \end{bmatrix},$$

$$HP^3H^{-1} = \begin{bmatrix} 1 & P + P\delta + P\delta^2 \\ \cdot & \delta^3 \end{bmatrix}$$

and, generally,

$$HP^nH^{-1} = \begin{bmatrix} 1 & P(1 + \delta + \delta^2 + \dots + \delta^{n-1}) \\ \cdot & \delta^n \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{P}{1-\delta} \\ \cdot & \cdot \end{bmatrix}$$

whence

$$P^n \rightarrow \frac{1}{1-\delta} \begin{bmatrix} \xi' & P \\ \xi' & P \end{bmatrix},$$

as before. We have also, incidentally, expressed P^n simply in terms of n and the elements of $P^{\#}$.

Thus in the binomial case we can express U in terms of the elements of P ; and this allows us to give, in this particular case, a proof of the existence of the limit U without the use of matrix theory. We write

$$P = U + \alpha_2 Z$$

* The same result is found by the equivalent method of difference equations by Uspensky, pp. 75-76.

where

$$Z = \frac{1}{p+\xi'} \begin{bmatrix} p & -p \\ -\xi' & \xi' \end{bmatrix},$$

a matrix with row sums zero. Then

$$UP = U, \quad ZU = 0 \quad \text{and} \quad Z^2 = Z$$

so that

$$P^2 = U + \lambda_2^2 Z$$

and, generally,

$$P^n = U + \lambda_2^n Z.$$

Since $|\lambda_2| < 1$ it follows that $P^n \rightarrow U$ as $n \rightarrow \infty$.

1.4. Thus, under certain conditions, P^n tends to a limiting matrix of a certain kind that we call stable: a stochastic matrix of rank 1 in which all rows are the same.

Let T be such a stable stochastic matrix. Then T is idempotent: $T^2 = T$; and, more generally, $PT = T$ for any stochastic matrix P . Similarly $TP = S$, a stable matrix; but $S = T$ if and only if $T = U$, where each row of U is a latent row vector associated with the root $\lambda=1$ of P (which root need not be simple here). The proofs of these results are immediate.

Thus the product of a stochastic matrix and a stable matrix

being of rank 1, is stable; and, in particular, the product of two stable matrices is stable. The converse, that if the product of two stochastic matrices is stable then at least one of them is itself stable is true for binomial matrices but is not true in general. The proofs are as follows.

(i) Let P_1, P_2 be binomial stochastic matrices with roots $1, \lambda_1$ and $1, \lambda_2$ respectively. Then since for any square matrices A, B

$$(a) \quad \det (AB) = \det A \cdot \det B$$

and

$$(b) \quad \det A = \text{product of roots of } A,$$

it follows, if the roots of $P_1 P_2$ are $1, \lambda$, that $\lambda = \lambda_1 \lambda_2$. If $\lambda = 0$ then $P_1 P_2$ is stable, and conversely, by the results of 1.1. And if $\lambda = 0$ then either $\lambda_1 = 0$ or $\lambda_2 = 0$, and hence either P_1 or P_2 is stable.

(ii) Now suppose that P_1, P_2 are stochastic matrices of order $k \geq 3$. Then if $P_1 P_2$ is stable and \underline{c} is a column of P_2 , then

$$P_1 \underline{c} = \begin{bmatrix} \alpha \\ \alpha \\ \vdots \\ \alpha \end{bmatrix} = \underline{\alpha}, \text{ say}$$

for some constant α , and if P_1 is non-singular,

$$\underline{c} = P_1^{-1} \underline{\alpha}.$$

Thus the elements of \underline{c} are equal if and only if P_1^{-1} has constant row sums. This is not generally true. As an example consider

$$P_1 = \frac{1}{6} \begin{bmatrix} 2 & 1 & 3 \\ 4 & 1 & 1 \\ 1 & 3 & 2 \end{bmatrix} \quad \text{for which} \quad P_1^{-1} = \frac{1}{4} \begin{bmatrix} -1 & 7 & -2 \\ -7 & 1 & 10 \\ 11 & -5 & -2 \end{bmatrix} .$$

1.5. We now turn to the more general problem mentioned in 1.1, that of discussing the behaviour for large n of

$$P^{(n)} = P_1 P_2 \dots P_n .$$

We have seen that under certain conditions P^n tends to a stable limiting matrix. This behaviour we can interpret in another way and relate the abstract problem of the behaviour of matrix powers and products to the concrete problem of the nature of a probability distribution associated with a Markov chain.

Consider a chain of k states in which the initial probability distribution of the occurrence of the respective states is specified by the row vector \underline{w} , so that the probability distributions after 1, 2, ..., n , stages are $\underline{w}P_1, \underline{w}P_1 P_2, \dots, \underline{w}P^{(n)}, \dots$. Then, in the constant case of 1.2, if P^n tends to a stable matrix, then $\underline{w}P^n$

(i) tends to independence of \underline{w} ;

(ii) tends to a definite limit.

(i) is a consequence of the tendency to stability, and (ii) is a consequence of the existence of a limit for P^n .

Now consider $P^{(\omega)}$. It is evident already from the remarks on stable matrices in 1.4 that $P^{(\omega)}$ does not tend to a limit in general, for consider a product of general stable matrices; and it is only under restricted conditions that a definite limit exists. We discuss these conditions in Chapter 2. The general non-existence of a limit for $P^{(\omega)}$ constitutes the essential difference of the variable from the constant case.

But for a wide class of stochastic matrices $P^{(\omega)}$ does tend to stability. It is easy to verify the following

Lemma

$\underline{w}P^{(\omega)}$ tends to independence of \underline{w} if and only if $P^{(\omega)}$ tends to a stable matrix.

Proof.

(a) The condition is sufficient. For suppose that no two elements of the same column of $P^{(\omega)}$ differ by more than ε . Let $\underline{u}^{(\omega)}$ be any row of $P^{(\omega)}$. Then

$$|\underline{w}P^{(\omega)} - \underline{u}^{(\omega)}| \leq \varepsilon,$$

since the elements of \underline{w} are positive, or zero, and sum to unity. Here and elsewhere $|P| < \varepsilon$ means: each element of P is less than ε in modulus.

(b) The condition is necessary. For suppose, for example, that the first two elements of a given column of $P^{(\omega)}$ are c_1 and c_2 , and consider the two distributions $[1-\alpha, \alpha, 0, \dots, 0]$ and $[1-\beta, \beta, 0, \dots, 0]$. The elements in the corresponding column of

$\underline{w}P^{(n)}$ are $c_1(1-\alpha) + \alpha c_2$ and $c_1(1-\beta) + \beta c_2$ respectively, and these differ by less than ε if and only if $|(\alpha-\beta)(c_1-c_2)| < \varepsilon$; and hence, if $\alpha \neq \beta$, $c_1-c_2 \rightarrow 0$ as $n \rightarrow \infty$. The result follows.

1.6. Thus an initial problem is to find necessary and sufficient conditions under which $P^{(n)}$ tends to a stable matrix. In fact it appears to be not possible to set down general broad conditions of any usefulness; we can only give sufficient conditions, but these are conditions satisfied by most chains.

Matrix theory is not of use now and we have to turn to other methods. It is not difficult to give a sufficient condition; it is that the P_i shall be uniformly positive, that is, that there shall exist an $\varepsilon > 0$ such that $P_i \geq \varepsilon$ for all i . The proof is as follows.

Let ρ_i be the difference between the greatest and the least elements of a column \underline{c} of P , and let $\rho = \max \rho_i$. Call ρ the range of P . Thus P is stable if, and only if, ρ is zero.

If $P \geq \varepsilon$ then $\rho \leq 1 - k\varepsilon$, where k is the order of P , as before. For, since the sum of the elements of any row is unity and each element is not less than ε , the greatest element of any row (and hence of any column) cannot exceed

$1 - (k-1)\varepsilon$. Thus the difference between the greatest and least elements of any column cannot exceed $1 - k\varepsilon$.

We shall show first that if $\rho_1, \rho_2, \rho_{12}$ are the ranges of $P_1, P_2, P_1 P_2$ respectively, then

$$\rho_{12} \leq (1 - ke)\rho_2.$$

Let the elements of a given column of P be $\{c_1, \dots, c_k\}$. For our present purpose we can suppose without loss of generality that

$$\underline{c} = \{c_1, c_2, \dots, c_k\},$$

where $\rho_c = c_1 \geq c_2 \geq \dots \geq c_k = \phi$, and consider the range of P, \underline{c} .

Let $[a_1, \dots, a_k]$ and $[a'_1, \dots, a'_k]$ be two rows of P_1 .

The difference between the corresponding two elements of P, \underline{c} is

$$d = (a_1 - a'_1)c_1 + (a_2 - a'_2)c_2 + \dots + (a_{k-1} - a'_{k-1})c_{k-1},$$

where a_i, a'_i are subject to

$$\sum_1^k a_i = \sum_1^k a'_i = 1 \quad \text{and} \quad |a_i - a'_i| \leq \rho_1 < 1. \quad (1)$$

Since any element a_{ij} of P_1 is such that

$$e \leq a_{ij} \leq 1 - (k-1)e,$$

the greatest possible value of d occurs when

$$a_1 = 1 - (k-1)e, \quad a_2 = a_3 = \dots = a_{k-1} = e$$

and

$$a'_1 = a'_2 = \dots = a'_{k-1} = e,$$

these values of a_i, a'_i being consistent with (1) above. Then

$$d = (1 - ke)c_1$$

so that

$$\rho_{12} \leq (1 - ke)\rho_2.$$

Since

$$\rho_2 \leq 1 - ke$$

we have that

$$\rho_{12} \leq (1 - ke)^2$$

and $\rho^{(n)}$, the range of $P^{(n)}$, is such that

$$\rho^{(n)} \leq (1 - ke)^n.$$

Thus if the P_i are uniformly positive, $P^{(n)}$ tends to a stable matrix.

1.7. The case of a constant chain in which P is regular, i.e. such that $\lim_{n \rightarrow \infty} P^n$ exists and is stable, is exceptional in that the convergence to stability is of a particular kind. Thus if we denote by $M_{(n)}$ and $m_{(n)}$ a greatest element and a least element respectively of a given column of $P^{(n)}$ then, in the constant case, it follows from the fact that $P^n \cdot P = P \cdot P^n$ that

$$M_n \geq M_{n+1} \quad \text{and} \quad m_n \leq m_{n+1},$$

that is, the sequence $\{M_n\}$ is monotonic decreasing and the sequence $\{m_n\}$ is monotonic increasing.

For a variable chain this is not always the case, as the following simple example shows. Let

$$P_1 = \begin{bmatrix} \cdot & 1 \\ \cdot & 1 \end{bmatrix} \quad \text{and} \quad P_2 = \begin{bmatrix} 1 & \cdot \\ 1 & \cdot \end{bmatrix}$$

so that

$$P_1 P_2 = \begin{bmatrix} 1 & \cdot \\ 1 & \cdot \end{bmatrix}.$$

Then, for the first column $M_{(1)} = 0$ but $M_{(2)} = 1$, and for the second column $m_{(1)} = 1$ but $m_{(2)} = 0$. We can, of course, give similar examples in which zero elements do not appear and in which P_1 and P_2 are not stable. However, the significant fact, vital to a discussion of the tendency to stability, is that, whatever the behaviour of the sequences $\{M_n\}$ and $\{m_n\}$, the sequence $\{M_n - m_n\}$, i.e. the sequence of ranges of corresponding columns of $P^{(n)}$, is monotonic decreasing, in the wide sense.

A second difference of the constant from the variable case is that P^n tends to stability if $P^r > 0$ for some finite r . Thus suppose that $P^r > \epsilon > 0$. Then, from the inequality of 1.6 for ranges of products of positive stochastic matrices we have, on writing $P^n = P^{t+ar}$, that

$$\rho(P^n) \leq \rho(P^{ar}) \leq \rho\{P^r \cdot P^{(a-1)r}\} \leq \eta^a,$$

where $\eta = 1 - ke < 1$, whence the result follows on letting $n \rightarrow \infty$. But in the variable case, if $P^{(r)} > 0$, we cannot draw the same conclusion, for suppose that $P_i = I$ for $i > r$. This remark, trivial though it is, does underline this fact, that we cannot infer the stability of a variable chain, in general, from a consideration of a finite number of products. An exception to this is, of course, the case when we know that stability has been achieved for some finite n .

1.8. For a binomial chain we can give a more precise result than the inequality of 1.6. We have already noted in 1.4 that, with the notation of that section,

$$\lambda = \lambda_1 \lambda_2$$

for a binomial chain. If

$$P = \begin{bmatrix} z & p \\ z' & p' \end{bmatrix}$$

then

$$\lambda = p' - p$$

so that

$$\rho = |\lambda|.$$

It follows at once that

$$\rho_{12} = \rho_1 \rho_2.$$

Thus a condition both necessary and sufficient for a binomial chain $P^{(n)}$ to tend to stability is that

$$\rho^{(n)} = \prod_{i=1}^n \rho_i \rightarrow 0$$

as $n \rightarrow \infty$. We shall show later in 1.11 that, for a trinomial chain,

$$\rho_{12} \leq \rho_1 \rho_2$$

so that

$$\rho^{(n)} \leq \prod_{i=1}^n \rho_i.$$

Thus the condition that $\prod_{i=1}^n \rho_i \rightarrow 0$ is, in this case, sufficient

to ensure that $\rho^{(n)} \rightarrow 0$, but it is not necessary. For, as we have remarked in 1.4, the product of two trinomial matrices can be stable even if neither of its factors is stable. If we take P_1 and P_2 as in the example of 1.4, and $P = I$ for $i > 2$ we have an example in which the condition $\prod_{i=1}^{\infty} \rho_i = 0$ is not satisfied although $P^{(n)}$ tends to stability.

1.9. We can generalise a little the result of 1.6. If we suppose that $P_i \geq e_i$ we have by the argument of that section that

$$\rho^{(n)} \leq (1 - ke_1)(1 - ke_2) \dots (1 - ke_n),$$

so that a sufficient condition for $P^{(n)}$ to tend to a stable matrix* is that

$$\sum_1^n e_i \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

Thus there is no need to require of the P_i uniform positivity, but in the variable case positivity in itself is not enough. An example below in 1.10 illustrates this.

The sufficient condition above clearly fails if every P_i of the chain contains a zero element, but a slight adaptation of the proof shows that $\rho^{(n)}$ tends to zero if each P_i has a positive

* This condition is equivalent to that of Hajnal, p. 72, that $\lim_{n \rightarrow \infty} \prod_{i=1}^n (1 - e_i) = 0$, but it indicates the more rapid convergence to stability. Both conditions fail if $e_i = 0$ for all i .

column (not necessarily the same for each P_i) and these columns are uniformly positive, or satisfy the slightly more general condition corresponding to the condition immediately above. Thus suppose for definiteness that c_1 , the first column of P_1 , is positive and that $c_1 > e$. Then the greatest possible value of $a_1 - a'_1$ is $1 - e$, and the greatest value of d consistent with this assumption occurs when

$$a_1 = 1, \quad a'_1 = e, \quad a'_2 = \dots = a'_{k-1} = 0.$$

Then

$$d = (1 - e) b_1$$

so that

$$d \leq (1 - e) \rho_2,$$

and likewise for any positive column of P_i . Thus, if the P_i all have a positive column,

$$\rho^{(n)} \leq (1 - e)^n,$$

and, in the more general case of positivity without uniformity,

$$\rho^{(n)} \leq (1 - e_1)(1 - e_2) \dots (1 - e_n),$$

where e_i is defined in the obvious way.

1.10. We now illustrate some of these results by examples.

Example 1.

Let

$$P_i = \begin{bmatrix} 1 - \frac{1}{i} & \frac{1}{i} \\ \frac{1}{i} & 1 - \frac{1}{i} \end{bmatrix}.$$

The P_i are not uniformly positive, but $\ell_i = \frac{1}{i}$ and $\sum_{i=1}^n \ell_i \rightarrow \infty$ as $n \rightarrow \infty$ so that $\rho^{(n)} \rightarrow 0$. Thus $P^{(n)}$ tends to stability.

But more than this: $P^{(n)}$ tends to a definite limit. For P_i is symmetric and so $P^{(n)}$ is symmetric so that $P^{(n)}$ tends to the only symmetric stable matrix of order 2,

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Example 2.

Let

$$P_i = \begin{bmatrix} 1 - \frac{1}{i^2} & \frac{1}{i^2} \\ \frac{1}{i^2} & 1 - \frac{1}{i^2} \end{bmatrix}$$

for $i \geq 2$. Then $\ell_i = \frac{1}{i^2}$ and the P_i are not uniformly positive. Moreover $\sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$ so that the sufficiency test for stability fails. In general we cannot conclude from this that $P^{(n)}$ does not tend to stability, but in the present binomial case we can show that this is so. For since, from 1.8, $\rho_{12} = \rho_1 \rho_2$ in this case,

$$\rho^{(n)} = \prod_{i=1}^n \left(1 - \frac{2}{i^2}\right)$$

$$\rightarrow \ell \neq 0, \text{ say,}$$

as $n \rightarrow \infty$. Since P_i , and hence $P^{(n)}$, is symmetric it follows that

$$\rho^{(n)} \rightarrow \frac{1}{2} \begin{bmatrix} 1+\ell & 1-\ell \\ 1-\ell & 1+\ell \end{bmatrix}.$$

Here we have the occurrence of a possibility that we have not mentioned explicitly before, that of $P^{(n)}$ tending to a definite limit but not to a stable limit.

Example 3.

Let

$$P_i = \begin{bmatrix} 1 - \frac{1}{i} & \frac{1}{i} \\ \frac{2}{i} & 1 - \frac{2}{i} \end{bmatrix} \quad (i \geq 4)^*$$

Then $\epsilon_i = \frac{1}{i}$ so that $\rho^{(n)} \rightarrow 0$, as in Ex. 1. In fact, again as in Ex. 1, $P^{(n)}$ tends to a definite stable limit. To show this we can use again the particular matrix transformation of 1.3.

Let $H = \begin{bmatrix} 1 & \cdot \\ -1 & i \end{bmatrix}$ so that $H^{-1} = \begin{bmatrix} 1 & \cdot \\ 1 & i \end{bmatrix}$.

Then

$$HP_i H^{-1} = \begin{bmatrix} 1 & \frac{1}{i} \\ \cdot & 1 - \frac{3}{i} \end{bmatrix}$$

so that

$$HP^{(n)} H^{-1} = \begin{bmatrix} 1 & \phi_n \\ \cdot & \theta_n \end{bmatrix}$$

where

$$\theta_n = \prod_{i=1}^n \left(1 - \frac{3}{i}\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

and

$$\phi_{n+1} = \frac{1}{n+1} + \left(1 - \frac{3}{n+1}\right) \phi_n.$$

* P_i is stable for $i=3$. We take $P_1 = P_2 = P_3 = I$.

Clearly ϕ_n is monotonic increasing for $n \geq 4$ and so tends to a positive limit, l say, which may be infinite so far as we can say at present. Then

$$HP^{(n)}H^{-1} \rightarrow \begin{bmatrix} 1 & l \\ \cdot & \cdot \end{bmatrix}$$

so that

$$P^{(n)} \rightarrow \begin{bmatrix} 1-l & l \\ 1-l & l \end{bmatrix}.$$

But, since $P^{(n)}$ is a stochastic matrix, $0 < l \leq 1$ so that $P^{(n)}$ tends to a definite stable matrix.

1.11. The convergence to stability is often more rapid than is indicated by the bound of the inequality of 1.6, and we can give a second inequality which is stronger in some cases and is particularly relevant to trinomial matrices. We have already proved, in 1.8, an exact result for binomial chains and mentioned the corresponding result for trinomial chains. The latter is a particular case of the present more general result that:

for chains of order k , $\rho_{12} \leq \left[\frac{1}{2} k \right] \rho_1 \rho_2.$

Proof.

With the notation of 1.6 we have that

$$\sum_{i=1}^{k-1} (a_i - a'_i) = a'_k - a_k$$

so that the greatest possible value of d occurs when $|a'_k - a_k| = \rho,$

the first $\left[\frac{1}{2}k\right]$ coefficients $a_i - a_i' = \rho$ and the last $\left[\frac{1}{2}k-1\right]$ coefficients $a_i - a_i' = -\rho$, and with the coefficient remaining if k is odd equal to zero. Thus

$$\rho_{12} \leq \left[\frac{1}{2}k\right] \rho_1 \rho_2.$$

If $k > 4$, ρ_1 must be sufficiently small for this inequality to have any significance; and it is only of use in cases where the general inequality of 1.6 is very obviously too rough or, sometimes, if the P_i contain zero elements. But the present inequality is always significant in the trinomial case, and it is stronger in this case than the sufficient conditions of 1.6 and 1.9, for it can apply to cases where the P_i are not positive and generally gives a better, and never a worse, bound for ρ_{12} .

1.12. After this discussion of the range of $P^{(n)}$ we can prove the result mentioned at the end of 1.2, that if $P > 0$ then the root $\lambda=1$ of P is simple.

Proof.

For all stochastic matrices there is a stable U associated with a root $\lambda=1$ of P such that $UP = U$; the particular consequence of positivity is that, by the inequality of 1.6, the range of P tends to zero.

Write

$$P = U + E,$$

so that E is a matrix with row sums zero, whence

$$\begin{aligned} P^2 &= UP + EP \\ &= U + EP \end{aligned}$$

and, generally,

$$P^{n+1} = U + EP^n.$$

We introduce here the notation $\rho^{[n]}$ for $\rho(P^n)$. Then if $\rho^{[n]} \leq \varepsilon_n$, each element of a given column of P can be written as $u + \alpha_n$ where $|\alpha_n| \leq \varepsilon_n$. It follows that each element of EP^n does not exceed $k\varepsilon_n$ in modulus. Since $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$, it follows that $EP^n \rightarrow 0$ so that $P^n \rightarrow U$. Thus U is unique and hence $\lambda = 1$ is a simple root if $\rho^{[n]} \rightarrow 0$. Thus $\lambda = 1$ is a simple root, in particular, if $P > 0$.

1.13. We can prove this result in another way, by the use in the k -dimensional case of the transformation we have used in 1.3 and 1.10. Let $P = [P_{ij}]$ and

$$H = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ -1 & & 1 & & \\ \vdots & & & \ddots & \\ -1 & & & & 1 \end{bmatrix} \quad \text{so that} \quad H^{-1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}.$$

Then

$$HPH^{-1} = \begin{bmatrix} 1 & p_{12} & \dots & p_{1k} \\ \vdots & & & \\ \vdots & & Q & \\ \vdots & & & \end{bmatrix}$$

where $Q = [q_{ij}]$ and $q_{ij} = p_{ij} - p_{ji}$ for $i, j = 2, \dots, k$.
 Let γ be a root of Q and $\{y_2, \dots, y_k\}$ be an associated column vector. Then

$$q_{\alpha 2} y_2 + \dots + q_{\alpha k} y_k = \gamma y_\alpha,$$

so that if y_α is the greatest of the y_i in modulus

$$\begin{aligned} |\gamma| &\leq |q_{\alpha 2}| + \dots + |q_{\alpha k}| \\ &\leq (k-1)\rho. \end{aligned}$$

Now consider P^n and suppose that $\rho^{(n)} \rightarrow 0$ as $n \rightarrow \infty$ so that $\rho^{(n)} < \frac{1}{k}$ for $n > n_0$, say. Then since the roots of HPH^{-1} are precisely those of P , and if the roots of P are $1, \lambda_2, \dots, \lambda_k$ so that those of P^n are $1, \lambda_2^n, \dots, \lambda_k^n$, it follows that $|\lambda_i|^n < 1$ for $n > n_0$, and hence that the root $\lambda=1$ of P is simple.

1.14. We have shown incidentally above that if γ is a subdominant root of P , i.e., a root of P greatest in modulus apart from $\lambda=1$, then

$$|\gamma| \leq (k-1)\rho.$$

We can improve this inequality to

$$|\gamma| \leq \left[\frac{1}{2} k \right] \rho.$$

As in 1.11, the result is of significance, apart from the trinomial case, only for sufficiently small ρ , for we already know that $|\gamma| \leq 1$.

We first note that if $\underline{u} = [u_1, \dots, u_k]$ is a row vector associated with a root $\lambda \neq 1$ of P , then $u_1 + \dots + u_k = 0$. For

$$\underline{u} P = \lambda \underline{u}$$

so that

$$u_1 p_{1i} + u_2 p_{2i} + \dots + u_k p_{ki} = \lambda u_i \quad (i = 1, \dots, k) \quad (1)$$

whence, on addition,

$$u_1 + \dots + u_k = \lambda (u_1 + \dots + u_k)$$

so that, if $\lambda \neq 1$,

$$u_1 + \dots + u_k = 0. \quad (2)$$

Suppose for definiteness that $|u_1| \geq |u_2|, \dots, |u_k|$, and so consider (1) in the case of $i = 1$. Using (2) we can write (1) as

$$u_1 (p_{11} - p_{\alpha 1}) + u_2 (p_{21} - p_{\alpha 1}) + \dots + u_k (p_{k1} - p_{\alpha 1}) = \lambda u_1$$

or

$$d_1 u_1 + d_2 u_2 + \dots + d_k u_k = \lambda u_1$$

where

$$d_j = p_{j1} - p_{\alpha 1}.$$

Thus

$$|\lambda| \leq |d_1| + \dots + |d_k|.$$

The choice of α from $1, \dots, k$ is at our disposal, and the problem is to choose it so as to minimise the right-hand side of this inequality. It is equivalent to this: we have k points distributed in some way over a range ρ and we have to find an upper bound to the least of the sums of the distances of the k points from a chosen one of them. It is intuitive (and we could give a formal proof without difficulty) that the upper bound corresponds to the greatest spread of the points, that is, half the points at one extreme of the range and half at the other; if the number of points is odd the position of the remaining point is arbitrary. Then the upper bound sought is $\left[\frac{1}{2} k \rho \right]$, and the result follows. In particular, $|\gamma| \leq \rho$ for trinomial matrices.

The coefficient $\left[\frac{1}{2} k \right]$ in this inequality for the upper bound of the subdominant root γ of P certainly cannot be improved for trinomial matrices, that is, there are trinomial P for which $|\gamma| = \rho$. As an example consider

$$P = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.2 & 0.5 & 0.3 \\ 0.7 & 0.2 & 0.1 \end{bmatrix},$$

for which $\rho = 0.4$. To find the roots $\lambda_1, \lambda_2, \lambda_3$ of P

without further calculation we can use the result* that the roots of

$$Q = \begin{bmatrix} 0.3 & . & . \\ 0.2 & 0.3 & -0.2 \\ 0.7 & . & -0.4 \end{bmatrix},$$

which is derived from P by reducing the elements of the second and third columns of P by 0.2 and 0.5, respectively, are $1 - (0.2 + 0.5) = 0.3$ and λ_2, λ_3 . The roots of Q are clearly 0.3, 0.3, -0.4, whence it follows that λ_2, λ_3 are 0.3, -0.4, so that $|\gamma| = \rho$ for this P.

* Brauer, p. 88.

CHAPTER 2.

2.1. We now consider the special case of stability in which $P^{(n)}$ tends not only to stability but to a definite stable matrix independent of n .^{*} We have noticed already two examples of such chains in 1.10, but these were rather special examples in that, in both cases, $P_n \rightarrow I$ as $n \rightarrow \infty$. The following example is of a slightly more general kind.

Let

$$P_n = T + \frac{1}{n} A$$

where

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} a & b \\ a & b \end{bmatrix}.$$

Thus $P_n \rightarrow T$, a stable matrix, as $n \rightarrow \infty$. We shall show that $P^{(n)} \rightarrow T$ also.

In the particular case in which $a = b$, so that $T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, the matrix P is doubly stochastic, i.e., not only each row but also each column sums to unity. In this case we can give a simple proof. For $\rho_n = \frac{2}{n}$ so that $\rho^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. Thus $P^{(n)}$ tends to stability and, since $P^{(n)}$ also is doubly stochastic, $P^{(n)}$ tends to the limit T , the only stable doubly stochastic matrix of order 2.

But this proof is not available in the more general case of $a \neq b$, and we then proceed as follows. We have that

$$A^2 = 2A, \quad TA = (a-b)A, \quad AT = 0$$

so that

* Sufficient conditions for this have been given by Hajnal.

$$P_1 P_2 = T + \frac{1}{2} TA + \frac{1}{2} A^2,$$

$$P_1 P_2 P_3 = T + \frac{1}{3} TA + \frac{1}{2 \cdot 3} TA^2 + \frac{1}{2 \cdot 3} A^3,$$

and, generally,

$$\begin{aligned} P^{(n)} &= T + (a-b) A \left\{ \frac{1}{n} + \frac{2}{n(n-1)} + \frac{2^2}{n(n-1)(n-2)} + \dots + \frac{2^{n-2}}{n!} \right\} + \frac{2^{n-1} A}{n!} \\ &= T + (a-b) \tau_n A + \frac{2^{n-1} A}{n!} \end{aligned}$$

where

$$\tau_n = \frac{1}{n} + \frac{2}{n(n-1)} + \frac{2^2}{n(n-1)(n-2)} + \dots + \frac{2^{n-2}}{n(n-1)\dots 2}.$$

Thus

$$\tau_{n+1} = \frac{1}{n+1} + \frac{2}{n+1} \tau_n \quad (1)$$

so that

$$\begin{aligned} \tau_n - \tau_{n+1} &= \frac{n-1}{n+1} \tau_n - \frac{1}{n+1} \\ &> \frac{1}{n+1} \left\{ (n-1) \left[\frac{1}{n} + \frac{2}{(n-1)n} \right] - 1 \right\} \quad (n \geq 3) \\ &= \frac{1}{n+1} \left\{ \frac{n+1}{n} - 1 \right\} \\ &= \frac{1}{n(n+1)} \\ &> 0. \end{aligned}$$

Thus the sequence $\{\tau_n\}$ is monotonic decreasing for $n \geq 3$, and since $\tau_n > 0$ for all n , $\tau_n \rightarrow l \geq 0$; and it follows from (1) that in fact $l = 0$. Since $2^{n-1}/n! \rightarrow 0$ as $n \rightarrow \infty$ we have the result that $P^{(n)} \rightarrow T$.

This last example can be regarded as a particular instance of a more general situation, that in which $P_n \rightarrow P > 0$ as $n \rightarrow \infty$. We should expect in such a case that $P^{(n)} \rightarrow U$, where U is the stable limiting matrix associated with P , although possibly only provided that the convergence of P_n to P is not too slow. More generally, we should expect that if U_i is the stable limiting matrix associated with P_i , and if $U_i \rightarrow U$, then $P^{(n)} \rightarrow U$, again perhaps with some restriction on the slowness of convergence. We shall discuss these conjectures in 2.7; but first we consider in more general terms conditions that are necessary and conditions that are sufficient for $P^{(n)}$ to tend to a stable limit.

2.2. Theorem.

The following two conditions are necessary in order that $P^{(n)} \rightarrow S$, a stable matrix independent of n :

- (1) $\rho^{(n)} \rightarrow 0$
- (2) $SP_n - S \rightarrow 0$.

Proof.

- (1) We have merely to note that if

$$P^{(n)} = S + E_n$$

then, since S is stable,

$$\rho^{(n)} = \rho(E_n)$$

and, since $E_n \rightarrow 0$, therefore $\rho(E_n) \rightarrow 0$.

(2) Write

$$P^{(n)} = S + E_n$$

for all n , so that $|E_n| < \epsilon$ for $n > N$, say.

Then

$$P^{(n+1)} = S P_{n+1} + E_n P_{n+1}.$$

But

$$P^{(n+1)} = S + E_{n+1}$$

so that

$$\begin{aligned} |S P_{n+1} - S| &= |E_{n+1} - E_n P_{n+1}| \\ &< \epsilon + k\epsilon. \end{aligned}$$

Hence the result.

These necessary conditions for the existence of a stable limit are not, however, sufficient, as the following example shows.

Let

$$P_i = \begin{bmatrix} 1 - \frac{1}{i} & \frac{1}{i} \\ \frac{1}{i} & 1 - \frac{1}{i} \end{bmatrix}$$

as in Ex. 1 of 1.10. Then, as we have shown,

$$P^{(n)} \rightarrow \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

and $P^{(n)} \rightarrow 0$. Also if

$$T = \begin{bmatrix} a & b \\ a & b \end{bmatrix}$$

then

$$T P_i - T = \frac{1}{i} (b-a) \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$$

so that $T P_i - T \rightarrow 0$ as $i \rightarrow \infty$. But $P^{(n)} \rightarrow T$ only if $a = b = \frac{1}{2}$.

The extreme nature of this example is evident; not only are the P_i symmetric but, much more significant, $P_i \rightarrow I$ as $i \rightarrow \infty$, i.e., P_i tends to a decomposable matrix, a matrix for which the root $\lambda=1$ is not simple. We should expect in a less particular case, say of $P_i \rightarrow P > 0$ or, more generally, $P_i \rightarrow P$ which is regular, that $SP_i - S \rightarrow 0$ implies $P^{(\omega)} \rightarrow S$; and this we shall show later, in 2.7, to be the case.

2.3. Let $\rho_{(i)}^{(\omega)}$ denote the range of P_i, P_{i+1}, \dots, P_n . Then:
Theorem.

Each of the following sets of conditions is sufficient in order that $P^{(\omega)} \rightarrow S$, a fixed stable matrix:

(1) (i)
$$\sum_{n=1}^{\infty} |SP_n - S| = K < \infty$$

(ii) $\rho_{(i)}^{(\omega)} \rightarrow 0$ as $n \rightarrow \infty$, for all i ;

or

(2) (i) $SP_n - S \rightarrow 0$ as $n \rightarrow \infty$

(ii) $\lim_{n \rightarrow \infty} \sum_{k=1}^n \rho_{(i)}^{(k)} = K < \infty$, for all i .

Proof.

We prove (a) that each set of conditions implies that $SP^{(n)} \rightarrow S$, and (b) that this in turn implies $P^{(\omega)} \rightarrow S$.

(a) Write

$$SP_i - S = F_i.$$

Then

$$SP^{(1)} = S + F_1 ;$$

$$\begin{aligned} SP^{(2)} &= SP_2 + F_1 P_2 \\ &= S + F_2 + F_1 P_2 ; \end{aligned}$$

$$\begin{aligned} SP^{(3)} &= SP_3 + F_2 P_3 + F_1 P_2 P_3 \\ &= S + F_3 + F_2 P_3 + F_1 P_2 P_3 ; \end{aligned}$$

and, generally,

$$SP^{(n)} - S = F_1 P_2 P_3 \dots P_n + F_2 P_3 \dots P_n + \dots + F_{n-1} P_n + F_n .$$

Let $|F_i| \leq f_i \leq 1$, and let P be any stochastic matrix of order k and range ρ . Then since each element of any given column of P can be written in the form $\alpha + \frac{1}{2}\theta$, where $|\theta| \leq \rho$, and the row sums of F_i are zero,

$$|F_i P| \leq \frac{1}{2} k \rho f_i .$$

Thus

$$\begin{aligned} |SP_n - S| &\leq \frac{1}{2} k \left\{ f_1 \rho_{(2)}^{(n)} + f_2 \rho_{(3)}^{(n)} + \dots + f_{n-1} \rho_{(n)}^{(n)} \right\} + f_n \\ &= \frac{1}{2} k r_n \quad , \text{ say.} \end{aligned}$$

We now prove that the conditions (1) imply that $r_n \rightarrow 0$.

Since $\rho_{(i)}^{(n)} \rightarrow 0$ as $n \rightarrow \infty$ we can choose n_1 so large that $\rho_{(i)}^{(n)} < \varepsilon$ for $n \geq n_1$; and since $\sum_{i=1}^{\infty} f_i < \infty$ we can choose n_2 so that $\sum_{i=n}^{\infty} f_i < \varepsilon$ for $n \geq n_2$. Then, since $\rho_{(i)}^{(n)} < 1$, $r_n \leq (k+1)\varepsilon$ for $n \geq \max \{ n_1, n_2 \}$. Hence the result in this case; and the proof for (2) is identical except

that the rôles of f and ρ are interchanged.

(b) Let

$$f = [p_1, \dots, p_k]$$

where p_i is the mean of the extreme elements of the i^{th} column of $P^{(n)}$, and P_0 be the stable matrix each of whose rows is \underline{p} . Then

$$|SP^{(n)} - P_0| \leq \frac{1}{2} k \rho^{(n)}.$$

But, in the case of (1) or of (2), $\rho^{(n)} \rightarrow 0$ as $n \rightarrow \infty$, whence

$$|SP^{(n)} - P_0| < \varepsilon$$

for $n \geq n_1$, say. Also

$$|P^{(n)} - P_0| \leq \frac{1}{2} k \rho^{(n)}$$

so that

$$|P^{(n)} - P_0| < \varepsilon$$

for $n \geq n_2$, say. But, from (a),

$$|SP^{(n)} - S| < \varepsilon$$

for $n \geq n_3$, say. Thus, since

$$\begin{aligned} |P^{(n)} - S| &= |(P^{(n)} - P_0) + (P_0 - SP^{(n)}) + (SP^{(n)} - S)| \\ &\leq |P^{(n)} - P_0| + |SP^{(n)} - P_0| + |SP^{(n)} - S| \\ &\leq 3\varepsilon \end{aligned}$$

for $n \geq \max \{n_1, n_2, n_3\}$, it follows that $P^{(n)} \rightarrow S$. Hence the result^H.

2.4. To illustrate these results, and to show that the two sets of sufficient conditions are independent, we consider two examples.

Example 1.

Let

$$P_i = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{2} \frac{i}{i+1} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} .$$

Then $\rho_i = \frac{i}{i+1}$ and, since for binomial stochastic matrices

$\rho_{12} = \rho_1 \rho_2$, it follows that

$$\rho_{(n)}^{(n)} = \frac{i}{n+1} .$$

Thus

$$\begin{aligned} \sum_{i=1}^n \rho_{(i)}^{(n)} &= \frac{1}{n+1} \sum_{i=1}^n i \\ &= \frac{1}{2} n \end{aligned}$$

whence it follows that condition (2)(ii) does not hold. But (1)(ii) does hold; and so does (1)(i), for $SP_i - S = 0$ with

$$S = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} .$$

Example 2.

Let

$$P_i = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{i} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} , \quad i \geq 2 .$$

Then P_i is stable for all i , and so conditions (1)(i) and (2)(i) both hold. Also, for any stable S ,

$$SP_i = P_i$$

so that

$$SP_i - S = P_i - S .$$

Also, with

$$S = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} ,$$

$P_i - S \rightarrow 0$, so condition (2)(i) holds. But $|P_i - S| = \frac{1}{i}$, so that condition (1)(i) does not hold.

Thus there exist chains which satisfy either one of the two sets of conditions but not the other, so that neither set of conditions implies the other. That is, neither of the sets is necessary in itself. But even taken jointly the occurrence of one of other of the sets of sufficient conditions is not necessary. A simple example shows this: take $P_1 = S$ and $P_i = I$, for $i \geq 2$. In this case the P_i (apart from P_1) are not regular. But even if all the P_i are regular the conditions are not necessary, as the following example shows.

Example 3.

Let

$$P_i = \begin{bmatrix} 1 - \frac{1}{i^2} & \frac{1}{i^2} \\ \frac{1}{i^2} & 1 - \frac{1}{i^2} \end{bmatrix}$$

for $i \geq 2$, as in Ex. 2 of 1.6. Then, as in 1.6, and with

the same notation,

$$\prod_{i=2}^{\infty} P_i = \frac{1}{2} \begin{bmatrix} 1-l & 1+l \\ 1+l & 1-l \end{bmatrix},$$

which is not stable so that $\rho_{(2)}^{(n)} \not\rightarrow 0$ as $n \rightarrow \infty$ and condition (ii) of (1) and (2) fails. But if $P_i = T$, any stable matrix, then $P \xrightarrow{(w)} S$, a stable matrix, where

$$S = TP = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{2}(a-b)l \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$$

and

$$T = \begin{bmatrix} a & b \\ a & b \end{bmatrix}.$$

This last example illustrates one way in which the sufficient conditions show themselves to be not necessary; the following example illustrates a second, and perhaps more natural, way.

Example 4.

Let

$$P_i = \begin{bmatrix} 1 - \frac{1}{i} & \frac{1}{i} \\ \frac{2}{i} & 1 - \frac{2}{i} \end{bmatrix}$$

for $i \geq 4$ as in Ex. 3 of 1.10. Then, with the notation of that section,

$$\rho^{(w)} \rightarrow \begin{bmatrix} 1-l & l \\ 1-l & l \end{bmatrix} = S.$$

Thus condition (1)(ii) holds; but (1)(i) does not, for

$$|SP_i - S| = \frac{1}{i}(3l-1)$$

and $l \neq \frac{1}{3}$ (as we can easily verify on noting that the leading term of $P^{(n)}$ is the greater of the two in the first column, and that it decreases monotonically as n increases), so that

$\sum_{i=1}^{\infty} |SP_i - S|$ is divergent. On the other hand, $|SP_i - S| \rightarrow 0$ as $i \rightarrow \infty$ so that (2)(i) holds; but, as we shall now show, (2)(ii) does not.

As $\rho_{12} = \rho_1 \rho_2$ for a binomial chain we have to consider the behaviour, as $n \rightarrow \infty$, of

$$E_n = \left(1 - \frac{3}{4}\right)\left(1 - \frac{3}{5}\right) \dots \left(1 - \frac{3}{n}\right) + \left(1 - \frac{3}{5}\right) \dots \left(1 - \frac{3}{n}\right) + \dots + \left(1 - \frac{3}{n}\right).$$

Consider the last r terms of E . They are

$$\begin{aligned} & \left(1 - \frac{3}{n}\right) + \left(1 - \frac{3}{n}\right)\left(1 - \frac{3}{n-1}\right) + \dots + \left(1 - \frac{3}{n}\right) \dots \left(1 - \frac{3}{n-r+1}\right) \\ & > \left(1 - \frac{3}{n}\right) + \left(1 - \frac{3}{n} - \frac{3}{n-1}\right) + \dots + \left(1 - \frac{3}{n} - \dots - \frac{3}{n-r+1}\right)^* \\ & > \left(1 - \frac{3}{n}\right) + \left(1 - \frac{2 \cdot 3}{n-1}\right) + \dots + \left(1 - \frac{r \cdot 3}{n-r+1}\right) \\ & > r - \frac{3}{n-r+1} \{1+2+\dots+r\} \\ & = r - \frac{3}{2} \frac{r(r+1)}{n-r+1} \\ & > \frac{1}{2}r \end{aligned}$$

for n sufficiently large. Since r can be as great as we please, the result follows.

* Bromwich, p. 95.

2.5. These remarks illustrate once again the difficulty of setting down useful sets of conditions that are both necessary and sufficient. It is important to remark, that the set of conditions (2)(i) above is satisfied by chains that are ^{uniformly} positive, or uniformly column positive; such chains, as we saw in 1.6 and 1.9, tend to stability, and this condition, taken in conjunction with the necessary condition (2)(i), is sufficient in these cases to ensure the existence of a stable limit. In fact, in these cases, the condition (2)(i) is not essentially different from the more simply expressed condition $\rho^{(n)} \rightarrow 0$, as we now show.

Consider first the case of a binomial chain. Then, since

$$\rho_i \leq \rho < 1 \text{ for all } i \text{ and } \rho_{12} = \rho_1 \rho_2, \text{ we have that}$$

$$\sum_{m=1}^{n-1} \rho_{(m)}^{(n)} \leq \rho + \rho^2 + \dots + \rho^n$$

$$< \frac{\rho}{1-\rho}$$

for all n , and the equivalence of the two conditions follows.

Similarly, if the P_i are uniformly positive, or uniformly column positive, and of any order, we have that

$$\sum_{m=1}^{n-1} \rho_{(m)}^{(n)} \leq \eta + \eta^2 + \dots + \eta^n$$

$$< \frac{\eta}{1-\eta},$$

where η denotes $1-ke$ in the first case and $1-e$ in the second, in the notation of 1.6 and 1.9.

We can therefore state the following.

Corollary.

If the P_i are uniformly column positive and if there exists a fixed stable S such that $SP_n - S \rightarrow 0$ as $n \rightarrow \infty$, then $P^{(n)} \rightarrow S$ as $n \rightarrow \infty$.

Since for trinomial stochastic matrices $\rho_{12} \leq \rho_1 \rho_2$, a slight adaptation of the argument shows that if a trinomial chain is not column positive but is such that the ρ_i satisfy $\rho_i < \eta < 1$, then the same result holds.

2.6. We now, and for the two following sections, restrict attention to the case in which all the P_i are regular, i.e., we suppose that $U_i = \lim_{n \rightarrow \infty} P_i^n$ exists and is stable.

So far the various examples have been restricted, almost without exception, to the binomial case, for the obvious reason of the simplicity of calculation that results from the multiplicative property of range peculiar to this case. This restriction has prevented illustration of the fact that, although in the binomial case the product of regular matrices is regular (this result was proved in 1.4), this may not be so in other cases. As an example suppose that

$$P_1 = \begin{bmatrix} \cdot & \frac{1}{2} & \frac{1}{2} & \cdot \\ \cdot & \frac{1}{2} & \frac{1}{2} & \cdot \\ \frac{1}{2} & \cdot & \cdot & \frac{1}{2} \\ \frac{1}{2} & \cdot & \cdot & \frac{1}{2} \end{bmatrix} \quad \text{and} \quad P_2 = \begin{bmatrix} \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \end{bmatrix} .$$

Then

$$P_1^2 = P_2^2 = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = U,$$

so that P_1, P_2 are regular, each with limit U . But

$$P_1 P_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \\ \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

which is not regular.

2.7. With this restriction of the P_i to regular matrices we can now prove the following theorem (which we shall afterwards restate in a simpler form).

Theorem.

If (i) $SP_i - S \rightarrow 0$ as $i \rightarrow \infty$;

(ii) $\sum_{k=1}^{\infty} \rho(P_i^k) < \infty$ for all i ,

then $U_i \rightarrow S$.

Proof.

Write

$$SP_i = S + F_i$$

as in 2.3, so that

$$SP_i^n - S = F_i P_i^{n-1} + F_i P_i^{n-2} + \dots + F_i P_i + F_i.$$

Then

$$\begin{aligned} |SP_i^n - S| &< \{ I + P_i + P_i^2 + \dots + P_i^{n-1} \} \\ &\leq \frac{1}{2} K \{ 1 + \rho_i + \rho(P_i^2) + \dots + \rho(P_i^{n-1}) \} \\ &\leq K \delta_i, \quad \text{by (ii), for all } n. \end{aligned}$$

Thus, by (i),

$$|SP_i^n - S| < \varepsilon$$

for $i > i_1$, say.

But

$$P_i^n = U_i + K_i^{(n)}$$

where $K_i^{(n)} \rightarrow 0$ as $n \rightarrow \infty$, since the P_i are regular, so that

$$\begin{aligned} SP_i^n &= SU_i + SK_i^{(n)} \\ &= U_i + SK_i^{(n)}. \end{aligned}$$

Since $|K_i^{(n)}| < \varepsilon$ for $n > n_1$, say, it follows that

$$|SP_i^n - U_i| < \varepsilon$$

for $n > n_1$. Thus

$$\begin{aligned} |U_i - S| &= |(U_i - SP_i^n) + (SP_i^n - S)| \\ &\leq |SP_i^n - U_i| + |SP_i^n - S| \\ &\leq 2\varepsilon \quad \text{for } i > i_1, n > n_1. \end{aligned}$$

It follows that $U_i \rightarrow S$.

We now consider condition (ii) of the theorem. It has been given in the form natural to the proof of the theorem but, just as in our discussion of the corresponding condition in 2.5, we can simplify it. In fact, we can remove it altogether, for it is a consequence of the supposition that the P_i are regular.

Suppose first that $U_i > 0$. Then $P_i^r > 0$ for all r sufficiently large, $r > r_0$, say, so that

$$\rho(P_i^r) \leq \eta = 1 - \epsilon,$$

where $\epsilon > 0$ is the least element of P_i^r . Then, since

$$\begin{aligned} \rho_{12} &\leq \rho_2, \\ \sum_{m=1}^{\infty} \rho(P_i^m) &\leq r_0 + r_0\eta + r_0\eta^2 + r_0\eta^3 + \dots \\ &\leq \frac{r_0}{1-\eta}, \end{aligned}$$

and the result follows. If, however, U_i is not positive, at least one column of P_i^r is positive for $r > r_0$, say. The result now follows as before on the use of 1.9, except that now $\eta = 1 - \epsilon$.

We can therefore restate the theorem in the following form, and extend it slightly.

Theorem.

If all P_i are regular and $SP_i - S \rightarrow 0$, then $U_i \rightarrow S$. Conversely, if $U_i \rightarrow S$, then $SP_i - S \rightarrow 0$.

Proof

Write

$$S = U_i + M_i$$

where $|M_i| < \varepsilon$ for $i > n$, say. Then

$$\begin{aligned} SP_i &= (U_i + M_i)P_i \\ &= U_i + M_iP_i \\ &= S - M_i + M_iP_i \end{aligned}$$

so that

$$\begin{aligned} |SP_i - S| &= |M_iP_i - M_i| \\ &\leq (k_H) \varepsilon \quad \text{for } i > n. \end{aligned}$$

This proves the converse. Hence the result.

2.8. The conditions of 2.3, apart from being only sufficient conditions for the existence of a limit to $P^{(n)}$, suffer from this weakness, that they do not define the limit (if it exists) in terms of the P_i . It is true that the condition $SP_i - S \rightarrow 0$ suggests, as we have remarked already in 2.1, that the existence of S is related to the existence of a limit of the sequence $\{U_i\}$ in the case of regular P_i and this we have now proved in the case of P_i which satisfy the condition (2)(ii) of 2.3. Thus, taking the latter theorem with that of 2.3. we have, for sequences of regular P_i that satisfy the condition, and so in particular for uniformly positive or uniformly column positive stochastic matrices, the following

Theorem.

If the P_i are regular and satisfy the condition (2)(ii) of 2.3 then a necessary and sufficient condition for the existence of $\lim_{n \rightarrow \infty} P^{(n)}$ is the existence of $\lim_{i \rightarrow \infty} U_i$; and the two limits are then equal.

We note that the example of 2.6 illustrates the fact that even the condition $U_i = U$ is not sufficient in itself to ensure that $P^{(n)} \rightarrow U$. For we have only to consider $(P_1 P_2)^n$.

As a corollary to the above theorem we have the

Theorem:

If $P_i \rightarrow P > 0$, then $P^{(n)} \rightarrow U$.

Proof.

Since the P_i are uniformly positive, and so satisfy the condition (2)(ii) for i sufficiently large, we have merely to verify that $U_i \rightarrow U$ in this case. For a binomial chain this is immediate; for if

$$P_i = \begin{bmatrix} z_i & p_i \\ z'_i & p'_i \end{bmatrix}$$

then

$$U_i = \frac{1}{p_i + z'_i} \begin{bmatrix} z'_i & p_i \\ z'_i & p_i \end{bmatrix}$$

$$\rightarrow \frac{1}{p + z'} \begin{bmatrix} z' & p \\ z' & p \end{bmatrix}$$

as

$$\begin{bmatrix} q_i & p_i \\ q'_i & p'_i \end{bmatrix} \rightarrow \begin{bmatrix} q & p \\ q' & p' \end{bmatrix}.$$

But in general we cannot express U_i so simply in terms of the elements of P_i ; \underline{u}_i is the solution of the set of consistent equations

$$\underline{u} P_i - \underline{u} = 0$$

as given by Cramer's rule in terms of the cofactors of $P_i - I$ and normalised so that $\sum_1^k u_\alpha = 1$. Now consider the variation in the value of a cofactor when one of its elements is changed by ε . All elements of the cofactor lie between -1 and 1 ; so the greatest value of a cofactor of it (which is a determinant of order $k - 2$) is certainly bounded by $(k - 2)! = K$, say, at very most. It follows that the variation in value of the cofactor cannot exceed $K|\varepsilon|$ numerically. Since, whatever the variation, all elements remain between -1 and 1 it follows that if each of the $(k - 1)^2$ elements of the cofactor vary by not more than ε , then the resultant variation in the value of the cofactor is not more than $(k - 1)^2 K|\varepsilon|$. Now interpret ε throughout as the respective elements of $P_i - P$. Then since by supposition $\varepsilon \rightarrow 0$ as $i \rightarrow \infty$, it follows that $U_i \rightarrow U$. Hence the result.

We notice that, since $P > 0$, there is certainly no restriction on the speed of convergence. The question whether or not there is a restriction for regular P that are not positive

is at present open, and this we now consider. Suppose, therefore, that P is regular. Then, just as above, $U_i \rightarrow U$; we have to verify that, in addition, (2)(ii) is satisfied. Let

$|P_i - P| < \epsilon; < \epsilon$, and consider the possible difference between P_1, P_2 and P^2 . We see at once that this difference is bounded numerically by $3k\epsilon$, and likewise that the difference $|P^{(n)} - P^n|$ is bounded by $3^{n-1} k^n \epsilon$. Take n so large that at least one column of P^n is positive, greater than ϵ say; this choice is certainly possible since P is regular. Now choose i so large, $i > i_0$ say, that $|P_i - P| < \frac{1}{2} \frac{\epsilon}{3^{n-1} k^{n-1}}$. Then the corresponding column of $P_{i_0+1} \dots P_{i_0+n}$ is also positive, and it follows by arguments similar to those used previously that (2)(ii) is satisfied. We have, therefore, as a corollary to the theorem above the following

Theorem.

If P is regular and $P_i \rightarrow P$, then $P^{(\omega)} \rightarrow U$.

2.9. Our statement of sufficient conditions, in 2.3, for the existence of $S = \lim_{n \rightarrow \infty} P^{(n)}$ follows naturally from the known necessary condition that $SP_n - S \rightarrow 0$ as $n \rightarrow \infty$. We can, however, state the conditions in a form that appears to be slightly more general. We replace condition 1(i) by 1(i') below and have the following

* These bounds are very rough, but sufficient.

Theorem.

The two conditions below are sufficient in order that $P^{(n)} \rightarrow S$, a fixed stable matrix;

- (i') There exists a sequence $\{S_i\}$ of stable matrices such that $S_i \rightarrow S$ as $i \rightarrow \infty$ and $\sum_{i=1}^{\infty} |S_{i-1} P_i - S_i| < \infty$;
- (ii) $\rho_{(i)}^{(n)} \rightarrow 0$ as $n \rightarrow \infty$, for all i .

Proof.

The proof is almost identical with that of 2.3. We now define $S_{i-1} P_i - S_i = F_i$ and $S_0 = I$. Then, just as before, we have that

$$P^{(n)} - S_n = F_n + F_{n-1} P_n + F_{n-2} P_{n-1} P_n + \dots + F_1 P_2 P_3 \dots P_n.$$

The convergence to zero of the series on the right follows just as before. Finally, writing

$$P^{(n)} - S_n = (P^{(n)} - S) + (S - S_n)$$

it follows that

$$P^{(n)} \rightarrow S.$$

This theorem is a slightly more general form of a known theorem^{*} in that the condition $\rho_{(i)}^{(n)} \rightarrow 0$ as $n \rightarrow \infty$ is implied by $\lim_{n \rightarrow \infty} \prod_{i=1}^n (1 - \epsilon_i) = 0$ but not conversely, except in the binomial case.

The condition (i') above is apparently more general than the corresponding condition of 2.3 only if $\sum_{i=1}^{\infty} |S_i - S| = \infty$. For, since

$$S_{i-1} P_i - S_i = (S P_i - S) + (S - S_i) + (S_{i-1} - S) P_i, \quad (*)$$

* Hajnal, pp. 69-71.

we have that, if $\sum_{i=1}^{\infty} |S_i - S| < \infty$ then $\sum_{i=1}^{\infty} |S_{i-1} P_i - S_i| < \infty$
if and only if $\sum_{i=1}^{\infty} |S P_i - S| < \infty$.

The condition (2)(i) of 2.3 is unchanged by the present replacement of S by S_i . This follows at once from (*) above.

2.10. We now return to the consideration of the existence of $\lim_{n \rightarrow \infty} P^{(n)}$ and consider the case in which $P_n \rightarrow I$ as $n \rightarrow \infty$, i.e., a case in which P_n tends to a (particular) decomposable limit. In the light of the examples of 1.10 it is perhaps reasonable to conjecture that if $P_n \rightarrow I$ then $P^{(n)}$ tends to a limit, stable or otherwise. This result, however, is not generally true, even in the binomial case.

We first prove that the result is true for one particular class of binomial stochastic matrices, that of symmetric matrices. Examples 1 and 2 of 1.10 are instances of this case.

Suppose that

$$P_i = \begin{bmatrix} 1 - a_i & a_i \\ a_i & 1 - a_i \end{bmatrix} .$$

Then each diagonal element of $P_i P_j$ is

$$(1-a_i)(1-a_j) + a_i a_j$$

$$= 1 - (a_i + a_j) + 2a_i a_j,$$

which is not greater than $1 - a_i$ if $a_j(1-2a_i) \geq 0$. If $P_n \rightarrow I$ then $a_i \leq \frac{1}{2}$ for sufficiently large i whence, in particular, the diagonal elements of $P_i P_{i+1}$ do not exceed those of P_i . By a repetition of this argument we see that the sequence of diagonal elements of the successive $P_i, P_i P_{i+1}, P_i P_{i+1} P_{i+2}, \dots$ is monotonic decreasing and, since it is bounded, tends to a limit. It follows that the off-diagonal elements of the matrix sequence also tend to a limit, and the result follows.

But if we drop the condition that the matrices be symmetric the result may fail. This can be shown by an example. Suppose that

$$X_\alpha = \begin{bmatrix} 1 & \cdot \\ \theta_\alpha & 1-\theta_\alpha \end{bmatrix} \quad \text{and} \quad Y_\beta = \begin{bmatrix} 1-\phi_\beta & \phi_\beta \\ \cdot & 1 \end{bmatrix}.$$

Then, as $n \rightarrow \infty$,

$$X_\alpha^n \rightarrow \begin{bmatrix} 1 & \cdot \\ 1 & \cdot \end{bmatrix} = A \quad \text{and} \quad Y_\beta^n \rightarrow \begin{bmatrix} \cdot & 1 \\ \cdot & 1 \end{bmatrix} = B$$

however small θ_α or ϕ_β may be, provided that $\theta_\alpha, \phi_\beta > 0$. We

notice that $AB = B$ and $BA = A$. A counter-example can now be constructed roughly as follows. Each P_i is either an X_α or a Y_β for some choice of $\theta_\alpha, \phi_\beta$. For the first n_1 P_i choose a sequence of X_1 so that $P^{(n_1)}$ approximates to A ; for the next n_2 P_i choose a sequence of Y_2 so that $P^{(n_1+n_2)}$ approximates to B ; for the next n_3 P_i choose a sequence of X_3 so that $P^{(n_1+n_2+n_3)}$ approximates to A ; and so on, the value of $P^{(n)}$ oscillating roughly between the bounds A and B , so that $P^{(\omega)}$ does not tend to a limit.

To make this argument precise, suppose that, given $\varepsilon > 0$, n_1 is chosen sufficiently large to ensure that

$$P^{(n_1)} = \begin{bmatrix} 1-a & a \\ 1-a' & a' \end{bmatrix} = A^\circ$$

where $a, a' < \varepsilon$, and n_2 is chosen sufficiently large to ensure that

$$P_{n_1+1} \dots P_{n_1+n_2} = \begin{bmatrix} b & 1-b \\ b' & 1-b' \end{bmatrix} = B^\circ$$

where $b, b' < \varepsilon$. Then we find that $A^\circ B^\circ$ differs from B° by not more than ε ; and similarly for a product $B^\circ A^\circ$. Each n_i of the sequence n_1, n_2, n_3, \dots is finite, and the sequence itself can be continued indefinitely. It remains only to choose, as we can, the sequences $\theta_1, \theta_3, \theta_5, \dots$ and $\phi_2, \phi_4, \phi_6, \dots$ such that θ_{2r+1} and ϕ_{2r} tend to zero with r , and the example is complete.

Conversely, suppose that $P^{(\omega)}$ tends to a non-stable limit.



Then, certainly for a chain of order $k \geq 3$ it does not necessarily follow that $P_n \rightarrow I$. For example, if

$$P_n = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & T \end{array} \right] = P,$$

where T is stable, then $P^{(n)} = P$ for all n , and P is not stable. However, in the binomial case the existence of a non-stable limit does imply that $P_n \rightarrow I$. For in this case

$$\rho^{(n)} = \rho_1 \rho_2 \dots \rho_n$$

so that if

$$\rho^{(n)} \rightarrow \rho \neq 0$$

it follows that

$$\rho_n \rightarrow 1.$$

Thus

$$P_n \rightarrow \begin{bmatrix} 1 & \\ & i \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} & \\ i & 1 \end{bmatrix}.$$

We can exclude the second possibility by a direct argument.

Suppose that the first row of $P^{(n)}$ is $[\alpha, 1-\alpha]$ and the first column of P_{n+1} is $\{\theta, 1-\phi\}$, where θ, ϕ can be as small as we please if n is sufficiently large. The element in $P^{(n+1)}$ that corresponds to α differs from α by

$$\begin{aligned} & |\alpha(\theta-1) + (1-\alpha)(1-\phi)| \\ &= |1-2\alpha + \alpha(\theta+\phi) - \phi|. \end{aligned}$$

Unless $1-2\alpha = 0$ this difference cannot be made arbitrarily small, and it follows that $P^{(w)}$ does not tend to a limit unless $\alpha = \frac{1}{2}$; and likewise for the other diagonal element. Thus $P^{(w)}$ does not tend to a limit unless this limit is

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

which is stable. Hence if $P^{(w)}$ tends to a non-stable limit

$$P_n \rightarrow \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}.$$

2.11. We have seen that only under special conditions does $U_n = P^{(w)}$ tend to a limit as $n \rightarrow \infty$; we have cited some sufficient conditions (in 2.3 and 2.9) and, in some cases, conditions both sufficient and necessary (in 2.8). We now consider whether or not the limit exists if we widen our definition of convergence and consider the behaviour as $n \rightarrow \infty$ of the more general

$$V_n = \frac{1}{n} \{ U_1 + U_2 + \dots + U_n \}. \quad (1)$$

We know, by the consistency theorem of Cesàro summability, that if $U = \lim_{n \rightarrow \infty} U_n$ exists then so does $V = \lim_{n \rightarrow \infty} V_n$. We also know that in the constant case of $P_i = P$ for all i , then V

exists for all P . A proof of this is as follows*. Firstly, V_n is itself a stochastic matrix for all n so that its elements are positive (or zero) and bounded above by 1. Consider a sequence of corresponding elements of the sequence $\{V_n\}$. Since this sequence is bounded it has at least one point of accumulation, by the Bolzano-Weierstrass theorem, and likewise, by the extension of this theorem to space of k^2 dimensions, we have the corresponding result that the sequence $\{V_n\}$ has at least one 'point' of accumulation, V say. To establish our result we have to show that $\{V_n\}$ has no other point of accumulation. To prove this we notice that $\{V_n\}$ includes a subsequence $\{V_{n_i}\}$ having V as its limit. Then, for each n_i ,

$$PV_{n_i} = V_{n_i}P = V_{n_i} + \frac{1}{n} (P^{n+1} - P)$$

so that, on taking the limit as $n \rightarrow \infty$,

$$PV = VP = V.$$

It follows that

$$P^r V = V P^r = V \tag{2}$$

for all integral r . Now suppose that there exists $\{W_n\}$, a subsequence of $\{V_n\}$ having a limit W distinct from V . Then, for each W_n , we have from (1) and (2) that

$$V W_n = W_n V = V$$

* See Doob, p. 175. Cf. Fréchet, p. 109.

given in 1.2 we see that the first superdiagonal of K_c^h contains terms of modulus n , that is, contains terms of unbounded modulus. But since the elements of H and H^{-1} are of fixed finite modulus and the elements of the stochastic matrix P^h do not exceed unity in modulus, this leads to a contradiction, and the result is established.

The Cesàro summability now follows at once. Let

$$HPH^{-1} = \text{diag} \left\{ 1, \dots, 1, \lambda_1, \dots, \lambda_r, K_1, \dots, K_s \right\}$$

where $|\lambda_i| = 1$, $\lambda_i \neq 1$, for $1 \leq i \leq r$ and the block matrices K_j are associated with roots λ_j for which $|\lambda_j| < 1$. Since

$$\lambda_i + \dots + \lambda_i^n \rightarrow \frac{\lambda_i}{1-\lambda_i} \text{ as } n \rightarrow \infty \text{ it follows that}$$

$\lim_{h \rightarrow \infty} H(P + P^2 + \dots + P^h)H^{-1}$ exists, and the result follows*.

We now consider the question whether or not the corresponding result is true in the variable case, i.e., we consider the existence of

$$\lim_{n \rightarrow \infty} V_n = \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ P^{(1)} + P^{(2)} + \dots + P^{(n)} \right\}.$$

The following example shows that this limit does not always exist. Consider the two binomial stochastic matrices

$$I = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \quad \text{and} \quad J = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}.$$

Then if each P_c is either I or J , so is each $P^{(n)}$; for

$$I^2 = I, \quad J^2 = JI = J, \quad J^2 = I.$$

* A somewhat similar proof, due to D.G. Kendall, is given by Bartlett, p. 32.

It follows, further, that by a suitable choice of P_n we can ensure that $P^{(n)}$ is either I or J as desired. The necessary and sufficient condition that $\lim_{n \rightarrow \infty} V_n$ shall exist is that $\lim_{n \rightarrow \infty} n/n$ shall exist, where n , is the number of choices of I among $P_1^{(1)}, \dots, P_n^{(n)}$; and clearly this limit may not exist. For example, suppose that $P^{(1)}$ is I, $P^{(2)}$ and $P^{(3)}$ are J, so that n/n at this stage is $1/3$, $P^{(4)}$, $P^{(5)}$ and $P^{(6)}$ are I, so that n/n is $2/3$, $P^{(7)}, \dots, P^{(12)}$ are J, so that n/n is $1/3$, $P^{(13)}, \dots, P^{(24)}$ are I, so that n/n is $2/3$, and so on so that the successive values of n/n ($n \geq 2$) oscillate between the bounds $1/3$ and $2/3$.

The Cesàro limit may not exist even if $\rho^{(n)} \rightarrow 0$. We can show this by a slight modification of the preceding example. We replace I and J by two distinct stable matrices A and B. Then

$$A^2 = A, \quad AB = B, \quad BA = A \quad \text{and} \quad B^2 = B,$$

and it follows that by a suitable choice of P_n , either as A or as B, we can ensure that $P^{(n)}$ is either A or B as desired. The rest of the proof now follows as before on the replacement of I and J by A and B respectively.

CHAPTER 3.

3.1. We now leave the consideration of $P^{(n)}$ and turn to the study of the frequency of occurrence of a given state. We shall show that the relative frequency of the occurrence of any given state in n stages tends to be normally distributed as $n \rightarrow \infty$.

We consider first the binomial and constant case in which there are two states, E_0 and E_1 , and such that the occurrence of E_0 ('failure') or E_1 ('success') results in the addition of 0 or 1 respectively to the score. In this case of a constant chain, with initial probability distribution $[q_0, p_0]$ say, the probability generating function (p.g.f.) of the distribution after n trials is

$$[q_0, p_0]^t \left\{ P(t) \right\}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where

$$P(t) = \begin{bmatrix} q & pt \\ q' & p't \end{bmatrix}.$$

We wish to study the ultimate distribution of x_n/n , where x_n is the number of successes, or score, after n trials.

3.2. It is interesting to have at hand some numerical examples of such probability distributions and to observe in them the tendency of the (discrete) probability distribution to approximate more and more closely to a (continuous) normal distribution.

We shall give some numerical examples of such distributions to show the way in which the distribution, and the rate of convergence to a limiting distribution, depend upon the parameters n and $\delta = p' - p$, and also upon the initial distribution.

Suppose first that

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

so that $\delta = 0.6$, and that the initial distribution is $[0.5, 0.5]$. This is a case of complete symmetry; at every stage and always the probability of m successes is equal to the probability of m failures. Thus the probability distribution is symmetric, and skewness does not appear in this case. This simplification makes the example the more useful for the study of (i) the forms of the distributions of x_n for successive values of n , and (ii) the limiting form of the distribution about the mean of x_n/n .

Evidently a sufficient condition for the existence of a

limiting non-zero variance of the distribution of (ii) is that the increments to the variance of the distributions of (i) for successive values of n shall tend to a non-zero limit.

The successive probability distributions and variance increments are listed below, in Tables 1 and 2 at the end of this chapter. They were found by direct computation of the p.g.f.'s for successive n , using a desk hand-operated calculating machine and rounding off the entries to 5 places of decimals at alternate stages. The probability distributions as given may, therefore, be in error by a unit in the final decimal place and the variance increments in the third decimal place; but these errors, if they occur, do not affect the general validity of the conclusions.

At first the distribution is U-shaped, but soon a low crest rises from the centre of the trough. As n increases the crest becomes more pronounced; steadily the peaks at the extremities of the range fall relatively to the crest until, as the process develops, they lie below it. At this stage the curve approximating to the distribution still has two hollow troughs; but these, too, disappear as the extremities fall still farther. The shape of the curve is then that of an inverted U or, more exactly, of a bell, which is the shape characteristic of an approximation to the normal curve.

We have to find the means in calculating the variances,

and we find that the successive increments to the mean on passing from one stage to the next are constant, and equal to 0.5. This is a result of the symmetry and what we should expect, for the score increments consequent on success and failure are 1 and 0 respectively,

The apparent convergence of the sequence of $\Delta\mu_2$, the successive increments to the variance, is evident. If, as a first approximation, we judge from the last three increments that the successive increments decrease in geometrical progression with common ratio $\frac{2}{3}$ we estimate the limit to be 1.001. In fact it follows from the theory of 4.6 that the limit is 1.000.

We now consider the effect of a change of the initial distribution from $[0.5, 0.5]$ to $[0.7, 0.3]$. This change destroys the symmetry of the successive probability distributions; these are shown below in Table 3. They were calculated in the same way as before, but the calculations were made more lengthy by the loss of symmetry.

Apart from this lack of symmetry the curves approximating to the successive probability distributions show the same trend as before. A curve of distorted U shape gives place to a curve with a central crest lying between two troughs. The extremities fall relatively to the crest and in consequence one of the troughs disappears. At the stage to which we have taken the calculations the second trough is still present but

is about to disappear; the calculation of a few probabilities of the distribution at the next stage shows that the trough does then disappear. Apart from skewness the approximating probability curve is now of the same type as that of the preceding example.

Finally we consider the effect of the further, and most extreme, change in the initial distribution, to $[1, \cdot]$. This change increases the asymmetry, but once again we see, from Table 4, that the asymmetry becomes less with increasing n . At first the approximating probability curve is J-shaped but with increasing n the steepness of the central part of the curve lessens and the curve comes to have one crest and one trough. This is the general form of a curve at one stage of the preceding example and we are led to assume that the general development of the successive curves now proceeds in essentially the same way as there.

To sum up, this sequence of examples leads us to believe that, whatever the initial distribution, the curve approximating to the discrete probability distribution tends, as n increases, to be unimodal and bell-shaped, that asymmetry (if present) tends to disappear, and that the variance of the distribution of x_n/n about the mean tends to a finite limit.

3.3. In the examples above P is symmetric, but the introduction of asymmetry through the use of a more general P produces an effect not essentially different from that produced above by a change of initial probability distribution. The following example, in which

$$P = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$$

and the initial probability distribution is $[1, \cdot]$ illustrates this. Also, in this case $\delta = 0.3$ as compared with the previous 0.6, and we observe the effect of this difference on the convergence to a limit of the successive increments to the variance. Further, we calculate the third moment μ_3 and hence $\mu_3/\mu_2^{3/2}$, a measure of the skewness, of the distributions for successive values of n .

The probability distributions are given in Table 5 below. The approximating frequency curves are bell-shaped, though distorted, even for small values of n . A comparison of the nature of these curves with those of the third of those of the preceding set in which the initial distribution was $[1, \cdot]$ as here -- or, in fact, with the curves in the other cases -- shows clearly the steadying effect of the smaller δ .

The effect shows again in a comparison of the sequences

of variance increments, given in Table 6. Despite the presence of asymmetry these increments are now essentially constant to 3 decimal places even for $n = 6$.

In the calculation of successive μ_3 it was necessary to retain 6 decimal places in the probability distributions. We see in Table 6 that the skewness, at first positive, decreases to zero and then becomes negative and is most negative at the point where we have left the calculations. But consideration of the sequence, or a graph, indicates a tendency to return towards zero with increasing n . The result of 3.6 of this chapter shows that, in fact, the skewness does converge to zero.

3.4. In the preceding section the value of δ for the P in question was positive. Certain difficulties appear when we study an example in which δ is negative.

We consider a symmetric case corresponding to the first example of the preceding section. We take

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{bmatrix}$$

so that $\delta = -0.6$, and initial probability distribution

$[0.5, 0.5]$. In this case success and failure tend to alternate, so we expect to find that the probability distribution of the score is bunched closely about a central mean. The probability distributions for successive values of n , found as before, are shown below in Table 7.

The probability curves are symmetric and bell-shaped throughout, and closely bunched about the mean. But the successive increments to the variance given in Table 8, do not tend monotonically to a limit; they oscillate. If we take alternative increments we have the two subsequences

0.130 ; 0.087 ; 0.071 ; 0.066 ;

and

-0.050 ; 0.022 ; 0.048 ; 0.057 ; 0.066 ;

each of which shows a tendency to convergence to some limit, and possibly to the same limit. In fact later theory will show that this is the case, and that the common limit is 0.0625. We notice also that the rate of convergence appears to be slower than in the previous corresponding case.

We should expect that asymmetry, if introduced by the choice of an initial probability distribution other than $[0.5, 0.5]$, would tend to disappear in the same manner as in the preceding section.

We notice further that, in the second sequence of increments above, the first increment is negative. Thus there is a decrease of the variance with increase in n . This event could not occur in the case of a sum of independent variables where the variance of the sum is the sum of the variances of the terms of the sum, and in 4.11 we shall comment further on the possibility of its occurrence for a sum of dependent variables.

3.5. It is not difficult to find directly the mean score in the constant binomial case. The mean after n trials is

$$\bar{x}_n = [q_0, p_0] \{P + P^2 + \dots + P^n\} \begin{bmatrix} \cdot \\ 1 \end{bmatrix}$$

if the initial distribution is $[q_0, p_0]$; for the mean of a sum of random variables, independent or not, is the sum of the means.

If P is regular, $P^n \rightarrow U$, so that

$$P + P^2 + \dots + P^n \rightarrow nU.$$

Hence

$$\bar{x}_n \rightarrow n [q_0, p_0] U \begin{bmatrix} \cdot \\ 1 \end{bmatrix}.$$

If we take as before

$$P = \begin{bmatrix} q & p \\ q' & p' \end{bmatrix}$$

then

$$U = \frac{1}{p+q'} \begin{bmatrix} q' & p \\ q' & p \end{bmatrix}$$

so that

$$\begin{aligned} \bar{x}_n &\rightarrow \frac{1}{p+q'} \{ q_0 p + p_0 p \} \\ &= \frac{p}{1-\delta}, \end{aligned}$$

which is independent of the initial probability distribution $[q_0, p_0]$.

If the initial distribution is the probability vector \underline{w} associated with the root $\lambda=1$ of P , i.e. if

$$\underline{w} = \frac{1}{p+q'} [q', p], \text{ then the mean after } n \text{ trials is } \frac{np}{1-\delta}$$

for all n . If, further, P is symmetric, so that

$\underline{w} = [0.5, 0.5]$ and $p = \frac{1}{2}(1-\delta)$ then $\bar{x}_n = \frac{1}{2}n$ for all n . This was the case in the symmetric examples of the preceding sections.

3.6. In the case of a constant chain we are able, as in 1.1, to make use of canonical matrix theory, and this we now do^{*}; but we shall have to replace this powerful, but limited, method by another when we come to extend the results to the

* Cf. Uspensky, pp. 297-301.

variable case.

We consider the distribution of x_n/n , but first 'change the origin' so that 'success' results in the addition of $z/(z'+p)$ to the score and 'failure' results in the addition of $-p/(z'+p)$. The asymptotic value of the mean, given for score increments 1 and 0 in the preceding section, is then zero.

If the initial probability distribution is $[z_0, p_0]$, the characteristic function (c.f.) of the score distribution is

$$G_n(t) = \left[q_0 e^{-\frac{p}{z'+p} \frac{it}{n}}, p_0 e^{\frac{z'}{z'+p} \frac{it}{n}} \right] \{P(t)\}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where

$$P(t) = \begin{bmatrix} q e^{-\frac{p}{z'+p} \frac{it}{n}} & p e^{\frac{z'}{z'+p} \frac{it}{n}} \\ z' e^{-\frac{p}{z'+p} \frac{it}{n}} & p' e^{\frac{z'}{z'+p} \frac{it}{n}} \end{bmatrix}.$$

Consider first

$$\lim_{n \rightarrow \infty} \{P(t)\}^n.$$

The roots of $P(t)$ are given by

$$\begin{vmatrix} q e^{-\frac{p}{z'+p} \frac{it}{n}} - \lambda & p e^{\frac{z'}{z'+p} \frac{it}{n}} \\ z' e^{-\frac{p}{z'+p} \frac{it}{n}} & p' e^{\frac{z'}{z'+p} \frac{it}{n}} - \lambda \end{vmatrix} = 0,$$

$$\text{i.e. } \lambda^2 - \lambda \left\{ \rho e^{-\frac{p}{\sigma'+p}} \frac{\alpha}{n} + \rho' e^{\frac{\sigma'}{\sigma'+p}} \frac{\alpha}{n} \right\} + (\rho' - \rho \sigma') e^{\frac{\sigma'-p}{\sigma'+p}} \frac{\alpha}{n} = 0. \quad (1)$$

The roots are expressible as power series in $\frac{\alpha}{n}$.

To a first approximation they are 1 and $\rho' - \rho \sigma'$,
i.e., 1 and $\rho' - \rho = \delta$.

Since $|\delta| < 1$, $\lim_{n \rightarrow \infty} \delta^n = 0$, and we shall need a closer approximation only to the root approximating to unity.

To find a second approximation to it put $\lambda = 1 + \frac{\theta}{n}$ in (1) and retain only terms $O(n^{-1})$. We get

$$\begin{aligned} \left(1 + \frac{\theta}{n}\right)^2 - \left(1 + \frac{\theta}{n}\right) \left\{ \rho + \rho' + \frac{\alpha}{n} \left(-\frac{\rho p}{\sigma'+p} + \frac{\rho' \sigma'}{\sigma'+p} \right) \right\} \\ + (\rho' - \rho) \left\{ 1 + \frac{\sigma'-p}{\sigma'+p} \frac{\alpha}{n} \right\} = 0, \end{aligned}$$

$$\frac{2\theta}{n} - \frac{\theta}{n} (\rho + \rho') + \frac{\alpha}{n} \left[\frac{\rho \sigma' - \rho' \sigma'}{\sigma'+p} + \frac{(\rho' - \rho)(\sigma' - p)}{\sigma'+p} \right] = 0.$$

The coefficient of $\frac{\alpha}{n}$ is zero, because

$$\begin{aligned} \rho \sigma' - \rho' \sigma' + (\rho' - \rho)(\sigma' - p) &= \rho \sigma' - \rho' \sigma' + \rho' \sigma' - \rho \sigma' - \rho \rho' + \rho^2 \\ &= \rho(1 - \rho) - \rho(1 - \rho') - \rho \rho' + \rho^2 \\ &= 0 \end{aligned}$$

Hence

$$\theta = 0.$$

This merely confirms the value found previously for the mean.

To get a third approximation, put $\lambda = 1 + \frac{\phi}{n^2}$ in (1).

We get

$$\left(1 + \frac{\phi}{n^2}\right)^2 - \left(1 + \frac{\phi}{n^2}\right) \left\{ \varrho + p' + \frac{1}{2} \frac{(\phi)^2}{n^2} \left[\varrho \left(\frac{p}{\varrho+p}\right)^2 + p' \left(\frac{\varrho'}{\varrho'+p}\right)^2 \right] \right\} \\ + (p'-p) \left[1 + \frac{1}{2} \frac{(\phi)^2}{n^2} \left(\frac{\varrho'-p}{\varrho'+p}\right)^2 \right] + O(n^{-3}) = 0,$$

since the terms $O(n^{-1})$ cancel.

Thus

$$\frac{2\phi}{n^2} - \frac{\phi}{n^2} (\varrho+p') + \frac{1}{2} \frac{(\phi)^2}{n^2} \left\{ - \frac{\varrho p^2 + p' \varrho'^2}{(\varrho'+p)^2} + (p'-p) \left(\frac{\varrho'-p}{\varrho'+p}\right)^2 \right\} + O(n^{-3}) = 0,$$

$$\phi \{ \varrho+p' \} = \frac{1}{2} \frac{(\phi)^2}{(\varrho'+p)^2} \left\{ (p'-p) (\varrho'-p)^2 - \varrho p^2 - p' \varrho'^2 \right\},$$

$$\phi \{ \varrho+p' \}^3 = - \frac{1}{2} (\phi)^2 p \varrho' (\varrho+p').$$

It will follow that the variance of the distribution of t/\sqrt{n} is given by

$$\sigma^2 = \frac{p \varrho' (\varrho+p')}{(\varrho'+p)^3},$$

which is non-zero since $p > 0$.

Column vectors associated with the roots approximating to

$1, \delta$ are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -p \\ \varrho' \end{bmatrix}$ to the first order (which is

all we need here) so that, on replacing t/n of the preceding work by $t/\sigma\sqrt{n}$ we have that $\phi_n(t) \rightarrow \phi(t)$ as $n \rightarrow \infty$, where $\phi_n(t)$ is the c.f. of the distribution about its mean of $x_n/\sigma\sqrt{n}$ and

$$\begin{aligned} \phi(t) &= [z_0, p_0] \begin{bmatrix} 1 & -p \\ 1 & z' \end{bmatrix} \begin{bmatrix} e^{-\frac{1}{2}t^2} & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} z' & p \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{p+z'} \\ &= [z_0, p_0] e^{-\frac{1}{2}t^2} \begin{bmatrix} \frac{z'}{z'+p} & \frac{p}{z'+p} \\ \frac{z'}{z'+p} & \frac{p}{z'+p} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= e^{-\frac{1}{2}t^2}, \end{aligned}$$

which is independent of the initial distribution. Thus the distribution of $x_n/\sigma\sqrt{n}$ tends, for all initial distributions, to a normal distribution, and the proof is complete*.

* The theorem is proved similarly by the use of matrix theory for a chain with k states in Aitken; but in this general case no simple expression for the variance can be given, and a separate discussion of the positivity of the variance is needed. See 7.5 and the references there.

3.7. The variance of the limiting distribution is

$\frac{p\xi'(\xi+p')}{(\xi'+p)^3}$, and this allows us to calculate the limiting values of the increments to the variance that we have already quoted in 3.1. We notice again, as before in the examples of 3.1, the effect of dependence: although $P^n \rightarrow U$, the limiting distribution of x/\sqrt{n} differs from that of a sequence of independent trials with probabilities of success and failure proportional to q', p , for the variance of the independent case is now multiplied by a factor

$$\frac{\xi^2+p'}{\xi'+p} = \frac{1-p+p'}{1-p'+p} = \frac{1+\delta}{1-\delta}$$

which is unity only if $\delta = 0$, i.e. only if successive trials are, in fact, independent. We notice that this factor can take any value between 0 and ∞ , these bounds being given by $\delta = -1$ and $\delta = 1$ respectively, and so correspond to the (non-regular) matrices

$$\begin{bmatrix} & 1 \\ i & \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & \\ & i \end{bmatrix}.$$

The first of these matrices corresponds to the case in which successes and failures occur alternately, the first trial being assigned arbitrarily either as success or failure, so that the score after n trials is either $\lfloor \frac{1}{2}n \rfloor$ or $\lfloor \frac{1}{2}n \rfloor + 1$, from which it is evident that $\sigma^2 \rightarrow 0$ whatever the outcome

of the initial trial.

The second of the matrices corresponds to the case in which the result of n trials consists entirely of n successes or entirely of n failures; and which of the two possibilities in fact occurs depends on the initial distribution $[q_0, p_0]$. If $p_0 q_0 \neq 0$ it is evident that

$$\sigma^2 \rightarrow \infty.$$

If we make the restriction as in 3.6 above that P be positive, $P \geq e$ say, then the least possible, and greatest possible, values of the variance occur when $P = P_1$ and $P = P_2$ respectively, where

$$P_1 = \begin{bmatrix} e & 1-e \\ 1-e & e \end{bmatrix} \quad \text{and} \quad P_2 = \begin{bmatrix} 1-e & e \\ e & 1-e \end{bmatrix}.$$

This is easy to see intuitively and not difficult to prove.

For we can write

$$\sigma^2 = \frac{1+\delta}{(1-\delta)^3} (p' - \delta)(1 - p'),$$

where $\delta = p' - p$, so that the least value of $(p' - \delta)(1 - p')$, subject to the restriction mentioned, occurs when $p' = 1 - e$ or $\delta + e$. Then

$$\sigma^2 = e \frac{1+\delta}{(1-\delta)^3} (1 - e - \delta),$$

and the least value occurs when δ , which is now necessarily

non-positive, is as negative as possible, i.e. $-\delta = 1 - 2e$,
so that P is P_1 above. The corresponding result for the
greatest value of the variance follows likewise.

TABLE 1.

| x_n n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1. | 0.4 | 0.2 | 0.4 | | | | | | | | | |
| 2. | 0.32 | 0.18 | 0.18 | 0.32 | | | | | | | | |
| 3. | 0.256 | 0.160 | 0.168 | 0.160 | 0.256 | | | | | | | |
| 4. | 0.2048 | 0.1408 | 0.1544 | 0.1544 | 0.1408 | 0.2048 | | | | | | |
| 5. | 0.16384 | 0.12288 | 0.14016 | 0.14624 | 0.14016 | 0.12288 | 0.16384 | | | | | |
| 6. | 0.13107 | 0.10650 | 0.12595 | 0.13648 | 0.13648 | 0.12595 | 0.10650 | 0.13107 | | | | |
| 7. | 0.10486 | 0.09175 | 0.11223 | 0.12585 | 0.13062 | 0.12585 | 0.11223 | 0.09175 | 0.10486 | | | |
| 8. | 0.08389 | 0.07865 | 0.09928 | 0.11489 | 0.12329 | 0.12329 | 0.11489 | 0.09928 | 0.07865 | 0.08389 | | |
| 9. | 0.06711 | 0.06711 | 0.08730 | 0.10400 | 0.11503 | 0.11890 | 0.11503 | 0.10400 | 0.08730 | 0.06711 | 0.06711 | |
| 10. | 0.05369 | 0.05704 | 0.07634 | 0.09347 | 0.10629 | 0.11317 | 0.11317 | 0.10629 | 0.09347 | 0.07634 | 0.05704 | 0.05369 |

TABLE 2.

| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------------------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| $\Delta\mu_2^{(n)}$ | 0.73 | 0.838 | 0.903 | 0.942 | 0.965 | 0.979 | 0.987 | 0.992 | 0.995 |

TABLE 3

| x_n n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1. | 0.56 | 0.20 | 0.24 | | | | | | | | | |
| 2. | 0.488 | 0.188 | 0.172 | 0.192 | | | | | | | | |
| 3. | 0.3584 | 0.1728 | 0.1680 | 0.1472 | 0.1536 | | | | | | | |
| 4. | 0.28672 | 0.15616 | 0.15934 | 0.14896 | 0.12544 | 0.12288 | | | | | | |
| 5. | 0.22938 | 0.13926 | 0.14912 | 0.14624 | 0.13120 | 0.10650 | 0.09830 | | | | | |
| 6. | 0.18350 | 0.12238 | 0.13701 | 0.14039 | 0.13258 | 0.11489 | 0.09011 | 0.07864 | | | | |
| 7. | 0.14680 | 0.10748 | 0.12442 | 0.13248 | 0.13068 | 0.11920 | 0.10008 | 0.07598 | 0.06288 | | | |
| 8. | 0.11744 | 0.09332 | 0.11178 | 0.12326 | 0.12629 | 0.12035 | 0.10654 | 0.08679 | 0.06393 | 0.05030 | | |
| 9. | 0.09392 | 0.08056 | 0.09966 | 0.11336 | 0.12010 | 0.11894 | 0.11000 | 0.09458 | 0.07502 | 0.05362 | 0.04024 | |
| 10. | 0.07514 | 0.06914 | 0.08814 | 0.10334 | 0.11285 | 0.11545 | 0.11091 | 0.09976 | 0.08360 | 0.06457 | 0.04491 | 0.03219 |
| 11. | 0.06008 | 0.05910 | 0.07748 | 0.09336 | 0.10490 | 0.11058 | 0.10972 | 0.10256 | 0.08996 | 0.07356 | 0.05538 | 0.03756 |

TABLE 4.

| x_n n | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|--------------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 0.8 | 0.2 | | | | | |
| 2 | 0.64 | 0.20 | 0.16 | | | | |
| 3 | 0.512 | 0.192 | 0.168 | 0.128 | | | |
| 4 | 0.4096 | 0.1792 | 0.1680 | 0.1408 | 0.1024 | | |
| 5 | 0.32768 | 0.16384 | 0.16256 | 0.14624 | 0.11776 | 0.08192 | |
| 6 | 0.26214 | 0.14746 | 0.15360 | 0.14624 | 0.12672 | 0.09830 | 0.06554 |

TABLE 5.

| $n \backslash x_n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 0.6 | 0.4 | | | | | | | | | |
| 2 | 0.36 | 0.36 | 0.28 | | | | | | | | |
| 3 | 0.216 | 0.288 | 0.300 | 0.196 | | | | | | | |
| 4 | 0.1296 | 0.2160 | 0.2736 | 0.2436 | 0.1372 | | | | | | |
| 5 | 0.07776 | 0.15552 | 0.22896 | 0.24768 | 0.19404 | 0.09604 | | | | | |
| 6 | 0.04666 | 0.10886 | 0.18144 | 0.22680 | 0.21672 | 0.15229 | 0.06723 | | | | |
| 7 | 0.02799 | 0.07465 | 0.13841 | 0.19440 | 0.21449 | 0.18487 | 0.11813 | 0.04706 | | | |
| 8 | 0.01679 | 0.05039 | 0.10265 | 0.15910 | 0.19673 | 0.19604 | 0.15460 | 0.09076 | 0.03294 | | |
| 9 | 0.01008 | 0.03359 | 0.07446 | 0.12579 | 0.17109 | 0.19099 | 0.17453 | 0.12723 | 0.06918 | 0.02306 | |
| 10 | 0.00605 | 0.02217 | 0.05307 | 0.09680 | 0.14298 | 0.17534 | 0.17960 | 0.15213 | 0.10334 | 0.05238 | 0.01614 |

TABLE 6.

| | | | | | | | | | | |
|-----------------------|-------|-------|-------|--------|-------|--------|-------|--------|-------|--------|
| $\Delta \mu_2^{(n)}$ | | 0.394 | 0.439 | 0.452 | 0.453 | 0.456 | 0.455 | 0.455 | 0.455 | 0.455 |
| $\mu_3 / \mu_2^{3/2}$ | 0.407 | 0.144 | 0.010 | -0.057 | -- | -0.108 | -- | -0.120 | -- | -0.121 |

TABLE 7.

| x_n n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1. | 0.1 | 0.8 | 0.1 | | | | | | | | |
| 2. | 0.02 | 0.48 | 0.48 | 0.02 | | | | | | | |
| 3. | 0.004 | 0.160 | 0.672 | 0.160 | 0.004 | | | | | | |
| 4. | 0.0008 | 0.0448 | 0.4544 | 0.4544 | 0.0448 | 0.0008 | | | | | |
| 5. | 0.00016 | 0.01152 | 0.19584 | 0.58496 | 0.19584 | 0.01152 | 0.00016 | | | | |
| 6. | 0.00003 | 0.00282 | 0.06835 | 0.42880 | 0.42880 | 0.06835 | 0.00282 | 0.00003 | | | |
| 7. | 0.00001 | 0.00066 | 0.02115 | 0.21693 | 0.52250 | 0.21693 | 0.02115 | 0.00066 | 0.00001 | | |
| 8. | 0.00000 | 0.00015 | 0.00605 | 0.08863 | 0.40517 | 0.40517 | 0.08863 | 0.00605 | 0.00015 | 0.00000 | |
| 9. | 0.00000 | 0.00004 | 0.00163 | 0.03163 | 0.22891 | 0.47558 | 0.22891 | 0.03163 | 0.00163 | 0.00004 | 0.00000 |

TABLE 8.

| n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $\Delta\mu_2^{(n)}$ | -0.05 | 0.130 | 0.022 | 0.087 | 0.048 | 0.071 | 0.057 | 0.066 | 0.061 |

CHAPTER 4.

4.1. We now extend the result of 3.6 to the case of a variable chain. To do this we need to consider first the mean. We have already found the asymptotic value of the mean in the constant case; we now find the value of the mean for any value of n for either a constant or a variable chain.

Just as in 3.5 we know that, if the initial distribution is $[1, \cdot]$,

$$\bar{x}_n = [1, \cdot] \{ P^{(1)} + P^{(2)} + \dots + P^{(n)} \} \begin{bmatrix} \cdot \\ 1 \end{bmatrix}. \quad (1)$$

But now \bar{x}_n/n does not, in general, tend to a limit as $n \rightarrow \infty$; except in a few particular cases $P^{(n)}$ tends to a stable matrix but one dependent on n . We discuss the value of \bar{x}_n in this simple case by a method that we shall use later to discuss the higher moments of the distribution: we find the increments to \bar{x}_i on transition from the i^{th} stage to the $(i+1)^{\text{th}}$ and then sum these increments to find \bar{x}_n . Consider, then, the increment to \bar{x}_i . The term

$$[1, \cdot] P^{(i+1)} \begin{bmatrix} \cdot \\ 1 \end{bmatrix}$$

is added to the right-hand side of (1), and this term is the element in the (1, 2) position of $P^{(i+1)}$. To find this element consider first

$$P_1 P_2 = \begin{bmatrix} z_1 & p_1 \\ z_1' & p_1' \end{bmatrix} \begin{bmatrix} z_2 & p_2 \\ z_2' & p_2' \end{bmatrix}.$$

The element sought is

$$\begin{aligned} & z_1 p_2 + p_1 p_2' \\ &= (1-p_1)p_2 + p_1(p_2 + \delta_2) \\ &= p_2 + p_1 \delta_2, \end{aligned}$$

and it follows that the corresponding term in $P^{(i)}$ is

$$p_i + p_{i-1} \delta_i + p_{i-2} \delta_{i-1} \delta_i + \dots + p_1 \delta_2 \delta_3 \dots \delta_i.$$

Thus, on summation,

$$\begin{aligned} \bar{x}_n &= p_n + p_{n-1}(1+\delta_n) + p_{n-2}(1+\delta_{n-1} + \delta_{n-1}\delta_n) + \dots \\ &\dots + p_1(1+\delta_2 + \delta_2\delta_3 + \dots + \delta_2\delta_3\dots\delta_n). \end{aligned}$$

In the particular case of a constant chain the increment above becomes

$$\begin{aligned} & p + \delta p + \delta^2 p + \dots + \delta^{n-1} p \\ &= p \frac{1-\delta^n}{1-\delta}, \end{aligned}$$

whence

$$\begin{aligned} \bar{x}_n &= \frac{p}{1-\delta} \left\{ (1-\delta) + (1-\delta^2) + \dots + (1-\delta^{n-1}) \right\} \\ &\rightarrow \frac{p}{1-\delta} \left\{ (n-1) - \frac{\delta}{1-\delta} \right\} \\ &\rightarrow \frac{np}{1-\delta} \end{aligned}$$

as before.

We can prove the result also by use of the HPH^{-1} transformation used previously in 1.3. We have that

$$[1 \cdot \cdot] P^{(n)} \begin{bmatrix} \cdot \\ \cdot \\ 1 \end{bmatrix} = [1 \cdot \cdot] \left\{ \prod_{i=1}^n \begin{bmatrix} 1 & p_i \\ \cdot & \delta_i \end{bmatrix} \right\} \begin{bmatrix} \cdot \\ \cdot \\ 1 \end{bmatrix},$$

since $[1 \cdot \cdot]$ is unchanged by post-multiplication by H^{-1} and $\begin{bmatrix} \cdot \\ \cdot \\ 1 \end{bmatrix}$ is unchanged on premultiplication by H . The result now follows at once.

If we take the initial score distribution as $[\cdot \cdot]$ instead of $[1 \cdot \cdot]$ we have for the increment to \bar{x}_i the element in the (2, 2) position of $P^{(i+1)}$. Since $\delta_{12} = \delta_1 \delta_2$ for binomial matrices it follows that this increment is

$$p_i + p_{i-1} \delta_i + p_{i-2} \delta_{i-1} \delta_i + \dots + p_1 \delta_2 \dots \delta_i + \delta_1 \delta_2 \dots \delta_i.$$

The effect of this is to add to the value of \bar{x}_n already found for initial probability distribution $[1 \cdot \cdot]$ a term

$$\delta_1 + \delta_1 \delta_2 + \delta_1 \delta_2 \delta_3 + \dots + \delta_1 \delta_2 \dots \delta_n$$

which, in the constant case, becomes

$$\frac{\delta(1-\delta^n)}{1-\delta},$$

so that $\bar{x}_n \rightarrow \frac{nP}{1-\delta}$ as before.

By combining these two results we see that the increment to \bar{x}_n if the initial distribution is $[\xi_0, p_0]$ is

$$p_i + p_{i-1} \delta_i + \dots + p_1 \delta_2 \dots \delta_i + p_0 \delta_1 \delta_2 \dots \delta_i.$$

It follows that the increment to the mean, and the mean, tend to independence of the initial distribution.

In the particular case of complete symmetry, in which each P_i is symmetric so that $p_i = \frac{1}{2}(1 - \delta_i)$ and the initial distribution is $[0.5, 0.5]$ we can verify that the formula above for the increment gives a constant increment of 0.5 for all i . This result is evident otherwise.

4.2. The matrix proof of the normal law in the constant case was based on the use of the c.f.; and we use likewise the c.f. of the distribution in the extension of the result to the variable case. If a typical matrix of the chain is

$$P_j = \begin{bmatrix} q_j & p_j \\ q'_j & p'_j \end{bmatrix},$$

then $G_n(t)$, the c.f. of the score distribution after n trials, is

$$G_n(t) \equiv [q_0, p_0 e^{it}] \left\{ \prod_{j=1}^n P_j(e^{it}) \right\} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where

$$P_j(e^{it}) = \begin{bmatrix} q_j & p_j e^{it} \\ q'_j & p'_j e^{it} \end{bmatrix}.$$

We wish to study the distribution of the score about the mean; and here we encounter the first difficulty resultant on the generalisation to a variable chain: the mean varies from stage to stage and, moreover, does not generally tend to a limit. We have no alternative, therefore, to the making of a separate change of origin at each stage.

We illustrate the process of formation of successive $\phi_n(t)$, where $\phi_n(t)$ is the c.f. of the distribution of score after n trials about the mean, by reference to a particular numerical constant case, that of 3.3. We have

$$\begin{aligned} G_1(t) &= [1 \ .] \begin{bmatrix} 0.6 & 0.4e^{it} \\ 0.3 & 0.7e^{it} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= 0.6 + 0.4e^{it} \end{aligned}$$

$\phi_1(t)$ is got from this by multiplying $G_1(t)$ by $e^{-0.4it}$, since the mean at this stage is 0.4, whence

$$\phi_1(t) = 0.6e^{-0.4it} + 0.4e^{0.6it}$$

The mean at the next stage is 0.92, so that

$$\phi_2(t) = e^{-0.92it} [1 \ .] \begin{bmatrix} 0.6 & 0.4e^{it} \\ 0.3 & 0.7e^{it} \end{bmatrix}^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

and, generally,

$$\phi_n(t) = e^{-\bar{x}_n it} [1 \ .] \begin{bmatrix} 0.6 & 0.4e^{it} \\ 0.3 & 0.7e^{it} \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

4.3. We have already found a formula for \bar{x}_n in terms of the elements of the matrices P_j , and hence the result above allows us to calculate $\phi_n(t)$ for any particular n . In the variable case this amounts to the evaluation of the successive products

$$\prod_{j=1}^m \left\{ P_j(e^{it}) \right\}$$

for $m = 2, 3, \dots, n$. It is no more difficult to evaluate the successive $\phi_m(t)$ themselves, and this we do by noting the recurrence relation between $\phi_m(t)$ and its successor $\phi_{m+1}(t)$.

To this end it is convenient to transform $P_j(e^{it})$. We replace $P_j(e^{it})$ by $H P_j(e^{it}) H^{-1}$, where

$$H = \begin{bmatrix} 1 & \cdot \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad H^{-1} = \begin{bmatrix} 1 & \cdot \\ 1 & 1 \end{bmatrix}.$$

Then

$$\begin{aligned} \phi_n(t) &= e^{-\bar{x}_n it} [1 \cdot] H^{-1} \left\{ \prod_{j=1}^n Q_j(e^{it}) \right\} H \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= e^{-\bar{x}_n it} [1 \cdot] \left\{ \prod_{j=1}^n Q_j(e^{it}) \right\} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \end{aligned}$$

where

$$Q_j(e^{it}) = \begin{bmatrix} \alpha_j + \beta_j e^{it} & \beta_j e^{it} \\ -\delta_j + \delta_j e^{it} & \delta_j e^{it} \end{bmatrix}.$$

Then, writing

$$\phi_n(t) = [\alpha_n(t), \beta_n(t)] \begin{bmatrix} 1 \\ \cdot \end{bmatrix},$$

we have that

$$\phi_n(t) = \alpha_n(t).$$

To study the successive $\phi_j(t)$, therefore, is to study the sequence of vectors

$$\underline{\alpha}_j = [\alpha_j(t), \beta_j(t)];$$

and for this sequence we have the recurrence relation

$$\underline{\alpha}_{j+1} = e^{-\Delta x_j} \underline{\alpha}_j Q_{j+1}(e^{it}).$$

We have already a formula for the increment to the mean Δx_j , and so we can calculate the $\underline{\alpha}_j$ successively.

4.4. We illustrate the process in our present particular case.

$$(i) \quad \alpha_1 = e^{-0.4it} [1 \cdot] \begin{bmatrix} 0.6 + 0.4e^{it} & 0.4e^{it} \\ -0.3 + 0.3e^{it} & 0.3e^{it} \end{bmatrix}$$

$$= e^{-0.4it} [0.6 + 0.4e^{it}, 0.4e^{it}]$$

$$= \left[1 + 0.240 \frac{(it)^2}{2!} + 0.0480 \frac{(it)^3}{3!} + \dots \right],$$

$$0.4 + 0.24(it) + 0.144 \frac{(it)^2}{2!} + 0.0864 \frac{(it)^3}{3!} + \dots]$$

Thus $\mu_2 = 0.240$ and $\mu_3 = 0.0480$.

(ii) At the second stage the increment to the mean is 0.52, so that

$$\underline{\alpha}_2 = e^{-0.52it} \underline{\alpha}_1 \begin{bmatrix} 0.6 + 0.4e^{it} & 0.4e^{it} \\ -0.3 + 0.3e^{it} & 0.3e^{it} \end{bmatrix},$$

which gives

$$\underline{\alpha}_2 = \left[1 + 0.6336 \frac{(it)^2}{2!} + 0.072576 \frac{(it)^3}{3!} + \dots, \right. \\ \left. 0.52 + 0.3216 (it) + 0.3161648 \frac{(it)^2}{2!} + \dots \right].$$

Thus $\mu_2 = 0.6336$ and $\mu_3 = 0.072576$, which values check with those calculated directly from the probability distribution given in 3.5.

(iii) At the third stage the increment to the mean is 0.556, so

$$\underline{\alpha}_3 = e^{-0.556it} \underline{\alpha}_2 \begin{bmatrix} 1 + 0.4(it) + 0.4 \frac{(it)^2}{2!} + \dots, & 0.4 + 0.4(it) + 0.4 \frac{(it)^2}{2!} + \dots \\ 0.3(it) + 0.3 \frac{(it)^2}{2!} + \dots, & 0.3 + 0.3(it) + 0.3 \frac{(it)^2}{2!} + \dots \end{bmatrix}.$$

Thus

$$\begin{aligned} \phi_3(t) &= \left\{ 1 - 0.556(it) + 0.309136 \frac{(it)^2}{2!} + \dots \right\} \\ &\quad \left\{ 1 + (it) [0.4 + (0.3)(0.52)] + \frac{(it)^2}{2!} [0.4 + 0.6336 + (0.52)(0.3) + 2(0.3126)(0.3)] + \dots \right\} \\ &= (1 - 0.556(it) + 0.309136 \frac{(it)^2}{2!} + \dots) (1 + 0.556(it) + 1.38256 \frac{(it)^2}{2!} + \dots) \\ &= 1 + \frac{(it)^2}{2!} [1.38256 - 2(0.556)^2 + 0.309136] + \dots \end{aligned}$$

$$\begin{aligned}
 &= 1 + \frac{(it)^2}{2!} [1.38256 - (0.556)^2] + \dots \\
 &= 1 + 1.073424 \frac{(it)^2}{2!} + \dots
 \end{aligned}$$

Thus $\mu_2 = 1.073424$, as already found by direct calculation from the probability distribution.

4.5. We have set out above the working involved in the calculation of the successive μ_2 in some detail because it serves as a numerical illustration of our present consideration of the increments to μ_2 at each stage in the general case. We shall consider first a constant binomial chain and use a method that is immediately extensible to the corresponding variable case. Notice that, as we are dealing with dependent variables, we have no direct method of calculating $\mu_2^{(n)}$, the variance of the score distribution after n trials, corresponding to that used for calculating \bar{x}_n in 4.1. An expression for $\mu_2^{(n)}$ is found later.

The elements $\alpha_j(t)$, $\beta_j(t)$ of the vector $\underline{\alpha}_j$ above are power series in (it) ; and, since $\alpha_j(t)$ is the c.f. of a distribution with its mean as origin, $\alpha_j(t)$ has no term in (it) . Let us suppose that

$$\underline{\alpha}_j = \left[1 + \mu_2^{(j)} \frac{(it)^2}{2!} + \mu_3^{(j)} \frac{(it)^3}{3!} + \dots, \nu_0^{(j)} + \nu_1^{(j)} (it) + \nu_2^{(j)} \frac{(it)^2}{2!} + \dots \right]$$

and that

$$P = \begin{bmatrix} q & p \\ q' & p' \end{bmatrix}.$$

Then, if for convenience we write Δ for $\Delta \bar{x}_j$, μ, ν for $\mu^{(j)}, \nu^{(j)}$ respectively, we have

$$\alpha_{j+1} = e^{-\Delta it} \alpha_j \begin{bmatrix} q + pe^{it} & pe^{it} \\ -\delta + \delta e^{it} & \delta e^{it} \end{bmatrix},$$

where $\delta = p' - p$ just as in 4.4, and so, expanding the terms in powers of (it) , we have that

$$\phi_{j+1}(t) = \alpha_{j+1}$$

$$= (1 - \Delta(it) + \Delta^2 \frac{(it)^2}{2!} - \dots) \left[1 + \mu_2 \frac{(it)^2}{2!} + \dots, \nu_0 + \nu_1(it) + \nu_2 \frac{(it)^2}{2!} + \dots \right] p_i$$

where

$$p_i = \left\{ 1 + p(it) + p \frac{(it)^2}{2!} + \dots, \delta(it) + \delta \frac{(it)^2}{2!} + \dots \right\}.$$

Thus

$$\begin{aligned} \phi_{j+1}(t) &= \left\{ 1 - \Delta(it) + \Delta^2 \frac{(it)^2}{2!} - \dots \right\} \\ &\quad \left\{ 1 + (p + \nu_0 \delta)(it) + \frac{(it)^2}{2!} (p + \mu_2 + \nu_0 \delta + 2\nu_1 \delta) + \dots \right\} \\ &= 1 + (it)(p + \nu_0 \delta - \Delta) \\ &\quad + \frac{(it)^2}{2!} \left\{ p + \mu_2 + \nu_0 \delta + 2\nu_1 \delta - 2\Delta(p + \nu_0 \delta) + \Delta^2 \right\} + \dots \end{aligned}$$

Since $\phi_{j+1}(t)$ is the c.f. of a distribution with zero mean the coefficient of (it) is zero so that

$$p + \nu_0^{(j)} \delta - \Delta_j = 0. \tag{2}$$

Again, from the coefficient of $(t)^2/2!$ in $\phi_{j+1}(t)$, we have that $\Delta\mu_2^{(j)}$, the increment to μ_2 on passing from the j^{th} to the $(j+1)^{\text{th}}$ stage, is given by

$$\begin{aligned}\Delta\mu_2^{(j)} &= p + v_0\delta + 2v_1\delta - 2\Delta(p + v_0\delta) + \Delta^2 \\ &= \Delta_j - \Delta_j^2 + 2v_1^{(j)}\delta\end{aligned}\quad (3)$$

on using (1) above.

We need now to express $v_1^{(j)}$ in terms of p and δ .

From (1),

$$\beta_{j+1} = (1 - \Delta(t) + \Delta^2 \frac{(t)^2}{2!} - \dots) \alpha_j p_2$$

where

$$p_2 = \left\{ p + p(t) + p \frac{(t)^2}{2!} + \dots, \delta + \delta(t) + \delta \frac{(t)^2}{2!} + \dots \right\}$$

so that

$$\beta_{j+1} = (p + v_0\delta) + (t) [p + v_0\delta + v_1\delta - \Delta(p + v_0\delta)] + \dots$$

Thus

$$v_0^{(j+1)} = p + v_0^{(j)}\delta \quad (4)$$

whence, on comparison with (2),

$$v_0^{(j+1)} = \Delta_j.$$

Also,

$$\begin{aligned}v_1^{(j+1)} &= p + v_0^{(j)}\delta + v_1^{(j)}\delta - \Delta(p + v_0^{(j)}\delta) \\ &= \Delta_j - \Delta_j^2 + v_1^{(j)}\delta.\end{aligned}\quad (5)$$

This recurrence formula enables us to express $v_1^{(j)}$ in terms of p and δ . For we have, from (5),

$$\begin{aligned}
 v_1^{(n)} &= \Delta_{n-1} - \Delta_{n-1}^2 + \delta v_1^{(n-1)}, \\
 \delta v_1^{(n-1)} &= \delta (\Delta_{n-2} - \Delta_{n-2}^2) + \delta^2 v_1^{(n-2)}, \\
 \delta^2 v_1^{(n-2)} &= \delta^2 (\Delta_{n-3} - \Delta_{n-3}^2) + \delta^3 v_1^{(n-3)}, \\
 &\dots \\
 \delta^{n-1} v_1^{(1)} &= \delta^{n-1} (\Delta_0 - \Delta_0^2) + \delta^n v_1^{(n)}.
 \end{aligned}$$

If the initial probability distribution is $[q_0, p_0]$,

$$v_0^{(1)} = p_0(1-p_0)$$

so we have, on addition, that

$$v_1^{(n)} = \sum_{j=0}^{n-1} (\Delta_j - \Delta_j^2) \delta^{n-j-1} + \delta^n p_0(1-p_0). \quad (6)$$

In the symmetric case, and with $p_0 = 0.5$, we have that

$\Delta_j = \frac{1}{2}$ and so

$$v_1^{(n)} = \frac{1}{4} \frac{1-\delta^{n+1}}{1-\delta};$$

but in all cases for which $|\delta| < 1$,

$$\Delta_j \rightarrow \frac{f}{1-\delta} = \Delta,$$

say, and

$$v_1^{(n)} \rightarrow \frac{\Delta(1-\Delta)}{1-\delta}$$

as $n \rightarrow \infty$. For if we write

$$\Delta_j(1-\Delta_j) = f_j \rightarrow f = \Delta(1-\Delta)$$

as $j \rightarrow \infty$, then

$$v_1^{(n)} = f_0 \delta^{n-1} + f_1 \delta^{n-2} + \dots + f_{n-1} + p_0(1-p_0) \delta^n$$

so that

$$v_1^{(n)} - \frac{f}{1-\delta} = v_1^{(n)} - f \{1 + \delta + \delta^2 + \dots + \delta^{n-1}\} - \frac{f \cdot \delta^n}{1-\delta} + p_0(1-p_0) \delta^n$$

$$= (f_0 - f) \delta^{n-1} + (f_1 - f) \delta^{n-2} + \dots + (f_{n-1} - f) + p_0 (1 - p_0) \delta^n - \frac{f \cdot \delta^n}{1 - \delta}.$$

The result now follows if

$$g_0 \delta^{n-1} + g_1 \delta^{n-2} + \dots + g_{n-1} \rightarrow 0$$

as $n \rightarrow \infty$, where

$$g_j = f_j - f \rightarrow 0$$

as $j \rightarrow \infty$. But the sum of the first $[\frac{1}{2}n]$ terms does not exceed $K \delta^{[\frac{1}{2}n]} / (1 - \delta)$, where K is an upper bound of the convergent sequence $\{g_j\}$; and the sum of the remaining terms does not exceed $g_{[\frac{1}{2}n+1]} / (1 - \delta)$. By taking n sufficiently large each of these sums can be made arbitrarily small. The result now follows. An incidental consequence of this result is that $v_1^{(w)} = o(1)$.

By the use of (3) and (6) we can express $\Delta \mu_2^{(j)}$, and hence $\mu_2^{(w)}$, in terms of p and δ . This we do later, in 4.8, but we first note the following particular case.

4.6. In the case of symmetry $\Delta \mu_2$ can be expressed in a concise form. For in that case

$$v_1^{(w)} = \frac{1}{4} \frac{1 - \delta^{n+1}}{1 - \delta}$$

so that

$$\begin{aligned} \Delta \mu_2^{(j)} &= \frac{1}{4} + 2 \cdot \frac{1}{4} \delta \left(\frac{1 - \delta^{j+1}}{1 - \delta} \right) \\ &= \frac{1}{4} \left\{ \frac{1 + \delta}{1 - \delta} - \frac{2 \cdot \delta^{j+2}}{1 - \delta} \right\}. \end{aligned}$$

Hence

$$\mu_2^{(n)} = \frac{1}{4} n \frac{1+\delta}{1-\delta} - \frac{1}{4} \frac{2\delta^2}{1-\delta} \left(\frac{1-\delta^{n+1}}{1-\delta} \right).$$

Thus $\mu_2^{(n)}$ tends to differ from $\frac{1}{4} n \frac{1+\delta}{1-\delta}$ by a constant;
and

$$\mu_2/n \rightarrow \frac{1}{4} \frac{1+\delta}{1-\delta}$$

as $n \rightarrow \infty$.

The expression for $\Delta\mu_2$ above shows that the two occurrences we noted in the symmetrical examples of 3.2 and 3.3 are typical of the symmetric case:

- (i) the increments $\Delta\mu_2$ tend to a limit with the rapidity of a geometrical progression;
- (ii) the convergence of the $\Delta\mu_2$ to their limit is monotonic if $\delta > 0$ and oscillatory if $\delta < 0$.

In the example of 3.2 we estimated roughly the common ratio of the G.P. as $2/3$; we now see that it is, in fact, 0.6 and that the limit of the sequence of $\Delta\mu_2$ is 1.0 precisely.

4.7. The limit of $\mu_2^{(n)}/n$ found above is but a particular case of the general result. For, since

$$v_1^{(n)} \rightarrow \frac{\Delta(1-\Delta)}{1-\delta}$$

as $n \rightarrow \infty$, it follows that

$$\Delta\mu_2 \rightarrow \Delta(1-\Delta) + 2\delta \cdot \frac{\Delta(1-\Delta)}{1-\delta}$$

$$\begin{aligned}
 &= \Delta(1-\Delta) \left\{ .1 + \frac{2\delta}{1-\delta} \right\} \\
 &= \Delta(1-\Delta) \frac{1+\delta}{1-\delta},
 \end{aligned}$$

whence μ_2/n tends to the same limit. We have thus confirmed by the method of increments the limiting value of the variance of the distribution of the score that we have already found by a matrix method in 3.6.

4.8. We now extend the results of the previous sections to variable chains. We use the fundamental recurrence formula

$$\underline{\alpha}_{j+1} = e^{-\Delta \bar{x}_j u} \underline{\alpha}_j Q_{j+1}(e^{iu})$$

of 4.3 and find from it, as in 4.5, that

$$p_{j+1} + v_0^{(j)} \delta_{j+1} - \Delta_j = 0. \quad (1)$$

The only difference is the replacement of p and δ , which were constant from stage to stage, by the variable p_{j+1} and δ_{j+1} .

Likewise we find that

$$\Delta \mu_2^{(j)} = \Delta_j - \Delta_j^2 + 2v_1^{(j)} \delta_{j+1} \quad (2)$$

and

$$v_0^{(j+1)} = p_{j+1} + v_0^{(j)} \Delta_j \quad (3)$$

so that, by comparison of (1) and (3),

$$v_0^{(j+1)} = \Delta_j,$$

as before.

Also

$$v_1^{(j+1)} = \Delta_j - \Delta_j^2 + v_1^{(j)} \delta_{j+1} \quad (4)$$

so that we can express the $v_1^{(j)}$ successively in terms of the various p_j and δ_j of the matrices of the chain.

The general formula for $\mu_2^{(n)}$ simplifies in the particular symmetric case, which we consider first. In that case, as we have noted in 4.1, $\Delta_j = \frac{1}{2}$ for all j , so that

$$v_1^{(j+1)} = \frac{1}{4} + v_1^{(j)} \delta_{j+1}.$$

It follows that

$$v_1^{(n)} = \frac{1}{4} \{ 1 + \delta_n + \delta_n \delta_{n-1} + \dots + \delta_n \delta_{n-1} \dots \delta_1 \}$$

and that

$$\Delta \mu_2^{(n)} = \frac{1}{4} \{ 1 + 2\delta_{n+1} + 2\delta_{n+1} \delta_n + \dots + 2\delta_{n+1} \delta_n \dots \delta_1 \}.$$

Hence

$$\mu_2^{(n)} = \frac{1}{4} + \frac{1}{4} \{ 1 + 2\delta_1 \} + \frac{1}{4} \{ 1 + 2\delta_2 + 2\delta_2 \delta_1 \}$$

$$+ \frac{1}{4} \{ 1 + 2\delta_3 + 2\delta_3 \delta_2 + 2\delta_3 \delta_2 \delta_1 \} + \dots$$

$$+ \frac{1}{4} \{ 1 + 2\delta_n + 2\delta_n \delta_{n-1} + \dots + 2\delta_n \delta_{n-1} \dots \delta_1 \}.$$

4.9. This result allows us to evaluate $\lim_{n \rightarrow \infty} \mu_2^{(n)}$ in particular cases in which this limit exists. Thus, to take a simple example, if $\delta_i = \delta$ for all i ,

$$\begin{aligned} \Delta \mu_2^{(n)} &= \frac{1}{4} \{ 1 + 2\delta + 2\delta^2 + \dots + 2\delta^{n+1} \} \\ &\rightarrow \frac{1}{4} \left\{ 1 + \frac{2\delta}{1-\delta} \right\} \\ &= \frac{1}{4} \frac{1+\delta}{1-\delta} \end{aligned}$$

so that

$$\mu_2^{(n)} \rightarrow \frac{1}{4} n \frac{1+\delta}{1-\delta},$$

a particular case of the result found for a general constant chain in 4.7.

Again, suppose that $\delta_i = (-1)^i \delta$, where $\delta > 0$. Then $\Delta \mu_2^{(n)}$ tends to one limit for n even and a second limit for n odd; for n even

$$\Delta \mu_2^{(n)} \rightarrow 1 - 2\delta - 2\delta^2 + 2\delta^3 + 2\delta^4 - \dots$$

and

$$\Delta \mu_2^{(n+1)} \rightarrow 1 + 2\delta - 2\delta^2 - 2\delta^3 + 2\delta^4 + \dots$$

so that

$$\begin{aligned} \Delta \mu_2^{(n)} + \Delta \mu_2^{(n+1)} &\rightarrow 2 \{ 1 - 2\delta^2 + 2\delta^4 - 2\delta^6 + \dots \} \\ &= 2 \left\{ 1 - \frac{2\delta^2}{1+\delta^2} \right\} \\ &= 2 \frac{1-\delta^2}{1+\delta^2}. \end{aligned}$$

Thus

$$\mu_2/n \rightarrow \frac{1-\delta^2}{1+\delta^2}$$

In the particular case of symmetry we can, of course, prove the result also by a matrix method corresponding to that used already in 3.6. Thus, if we now suppose, in order to preserve the symmetry, that the score increments corresponding to 'failure' and 'success' are -1 and 1 respectively, the c.f. of the distribution after $2n$ trials is

$$\left[\frac{1}{2} e^{-it}, \frac{1}{2} e^{it} \right] \{ P(t) \}^{2n} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where

$$\begin{aligned} P(t) &= \begin{bmatrix} \frac{1}{2}(1+\delta)e^{-it} & \frac{1}{2}(1-\delta)e^{it} \\ \frac{1}{2}(1-\delta)e^{-it} & \frac{1}{2}(1+\delta)e^{it} \end{bmatrix} \begin{bmatrix} \frac{1}{2}(1-\delta)e^{-it} & \frac{1}{2}(1+\delta)e^{it} \\ \frac{1}{2}(1+\delta)e^{-it} & \frac{1}{2}(1-\delta)e^{it} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{4} \{ (1-\delta^2)e^{-2it} + (1-\delta^2) \} & \frac{1}{4} \{ (1+\delta)^2 + (1-\delta)^2 e^{2it} \} \\ \frac{1}{4} \{ (1-\delta)^2 e^{-2it} + (1+\delta)^2 \} & \frac{1}{4} \{ (1-\delta^2) + (1-\delta^2)e^{2it} \} \end{bmatrix} \end{aligned}$$

We find that the characteristic equation of $P(t)$ is

$$\lambda^2 - (1-\delta^2) \{ 1 + (it)^2 \} \lambda - \delta^2 = 0,$$

retaining powers of (it) up to the second only. On replacement of t by t/\sqrt{n} the equation becomes

$$\lambda^2 - (1-\delta^2) \left\{ 1 + \left(\frac{it}{\sqrt{n}} \right)^2 \right\} \lambda - \delta^2 = 0.$$

The roots are $1, -\delta^2$ to a first approximation. To find a better approximation to the dominant root put

$$\lambda = 1 + \phi \left(\frac{it}{\sqrt{n}} \right)^2.$$

This gives

$$\phi = \frac{1 - \delta^2}{1 + \delta^2}$$

and hence the same value for μ_2 as before.

We notice that there is no trouble here in the extreme case of $\delta = 1$; in this case $\mu_2 = 0$. This is otherwise obvious on consideration of the possible outcomes of a sequence of trials, for the outcome is determinate after the first trial.

Now consider the case in which δ_{4r+1} , δ_{4r+2} , δ_{4r+3} , δ_{4r+4} are δ , δ , $-\delta$, $-\delta$ respectively for any positive integer r . Then $4\Delta\mu_2^{(n)}$ tends to one of the following four forms according to the value of n :

- (i) $1 + 2\delta + 2\delta^2 - 2\delta^3 + 2\delta^4 + 2\delta^5 + 2\delta^6 - 2\delta^7 + 2\delta^8 + \dots$;
- (ii) $1 + 2\delta - 2\delta^2 + 2\delta^3 + 2\delta^4 + 2\delta^5 - 2\delta^6 + 2\delta^7 + 2\delta^8 + \dots$;
- (iii) $1 - 2\delta + 2\delta^2 + 2\delta^3 + 2\delta^4 - 2\delta^5 + 2\delta^6 + 2\delta^7 + 2\delta^8 - \dots$;
- (iv) $1 - 2\delta - 2\delta^2 - 2\delta^3 + 2\delta^4 - 2\delta^5 - 2\delta^6 - 2\delta^7 + 2\delta^8 - \dots$;

so that the sum of the four increments tends to

$$\begin{aligned} & 4 + 8\delta^4 + 8\delta^8 + \dots \\ &= 4 \left\{ 1 + \frac{2\delta^4}{1-\delta^4} \right\} \\ &= 4 \frac{1+\delta^4}{1-\delta^4} . \end{aligned}$$

Thus

$$\mu_2/n \rightarrow \frac{1}{4} \frac{1+\delta^4}{1-\delta^4}.$$

In this case, in contrast to the preceding one, μ_2 is not $O(n)$ for $\delta = 1$.

These two examples are particular instances of the general situation in which the stochastic matrices P_i occur in regular cycles; μ_2/n tends to a limit in all such cases. We can calculate this limit in the same way for all such cyclic cases although the details will generally be more complicated. Thus, for example, even in the simple case of δ_{3r+1} , δ_{3r+2} , $\delta_{3r+3} = -\delta, -\delta, \delta$ respectively we find that

$$4. \frac{\mu_2^{(n)}}{n} \rightarrow \frac{1}{3} \left\{ \frac{3 - 2\delta - 2\delta^2 + 3\delta^3}{1-\delta^3} \right\}.$$

4.10. We showed in 2.8 that if $P_i \rightarrow P > 0$ as $i \rightarrow \infty$ then $P^{(n)} \rightarrow U$, where U is the stable limit of P^n . We can now prove the corresponding result for the score distribution associated with the convergent sequence $\{P_i\}$ in the present symmetric binomial case, that, with the previous notation,

$$\Delta \mu_2^{(n)} \rightarrow \frac{1}{4} \frac{1+\delta}{1-\delta}$$

i.e., that $\Delta\mu_2^{(w)}$ for the variable chain tends to $\Delta\mu_2^{(w)}$ for the corresponding constant chain. We need the following

Lemma.

Let $|\delta_i| \leq d < 1$ for all i ; $\delta_i = \delta + \varepsilon_i$ and $|\varepsilon_i| \leq \varepsilon$ for all i .

Then

$$|\delta_1 \delta_2 \dots \delta_n - \delta^n| \leq n d^{n-1} \varepsilon.$$

Proof.

The result is clearly true in the case of $n = 1$. Now assume that the result is true for $n = m$. Then

$$\begin{aligned} |\delta_1 \dots \delta_{m+1} - \delta^{m+1}| &= |\delta_1 \dots \delta_m (\delta + \varepsilon_{m+1}) - \delta^{m+1}| \\ &= |\delta \{ \delta_1 \dots \delta_m - \delta^m \} + \varepsilon_{m+1} \delta_1 \dots \delta_m| \\ &\leq \delta \cdot d^m + \varepsilon d^m, \end{aligned}$$

by the assumption above. Thus

$$|\delta_1 \dots \delta_{m+1} - \delta^{m+1}| \leq (m+1) d^m \varepsilon,$$

and the truth of the result for all values of n follows by induction.

Now suppose, given $\varepsilon > 0$, that $|\varepsilon_i| \leq \varepsilon$ for $i \geq n_0$, and take $n \geq n_0$.

Then

$$\begin{aligned} \Delta\mu_2^{(n)} &= \frac{1}{4} \{ 1 + 2\delta_{n+1} + \dots + 2\delta_{n+1} \dots \delta_{n_0} \} + \frac{1}{4} \{ 2\delta_{n+1} \dots \delta_{n_0-1} + \dots + 2\delta_{n+1} \dots \delta_2 \delta_1 \} \\ &= A + B, \end{aligned}$$

say. Since $\delta_i \leq d < 1$ for all i ,

$$\begin{aligned}
 B &\leq \frac{1}{2} \left\{ d^{n-n_0+3} + d^{n-n_0+4} + \dots + d^{n+1} \right\} \\
 &< \frac{1}{2} \frac{d^{n-n_0+3}}{1-d} \\
 &\leq \varepsilon
 \end{aligned}$$

for n sufficiently large. Also

$$A = \frac{1}{4} \left\{ 1 + 2\delta + 2\delta^2 + \dots + 2\delta^{n-n_0+2} \right\} + R_n$$

where, using the lemma,

$$\begin{aligned}
 |R_n| &\leq \frac{1}{2} \left\{ \varepsilon + 2d\varepsilon + 3d^2\varepsilon + \dots + (n+1)d^n\varepsilon \right\} \\
 &\leq \frac{1}{2} \varepsilon \frac{1}{(1-d)^2}.
 \end{aligned}$$

Since

$$1 + 2\delta + 2\delta^2 + \dots + 2\delta^{n-n_0+2} = \frac{1+\delta}{1-\delta} + C$$

where, by the same argument as above, $C < \varepsilon$ for n sufficiently large, the result follows.

4.11. We know from the formula for $\mu_2^{(n)}$ in 4.6 that in the present symmetrical binomial case with constant δ and $|\delta| < 1$, the variance $\mu_2 = O(n)$ and that a value of δ close to 1 results in a large value of μ_2/n and a value

of δ close to -1 results in a small value of μ_2/n . The extreme cases themselves, of $\delta = 1$ and $\delta = -1$, correspond to the limit of μ_2/n being infinite and zero respectively, and in both these cases the score distribution does not tend to a normal distribution, as we have remarked in 3.7.

A sufficient, although a priori not a necessary, condition for the normal law to hold is that μ_2/n be uniformly bounded from infinity and zero, i.e. $\mu_2 = o(n)$ precisely, and we shall have to find conditions under which this is so in the variable case. We shall certainly expect to find that $\mu_2 = o(n)$ provided that the P_i are uniformly positive so that

$$|\delta_i| \leq \delta < 1$$

for all i . This is the case we investigate first.

It is not difficult to show that $\mu_2 = o(n)$ at most. For we have that

$$\begin{aligned} \Delta \mu_2^{(n)} &\leq \frac{1}{4} \left\{ 1 + 2\delta + 2\delta^2 + \dots + 2\delta^{n+1} \right\} \\ &= \frac{1}{4} \left\{ 1 + \frac{2\delta(1-\delta^{n+1})}{1-\delta} \right\} \\ &\leq \frac{1}{4} \left\{ 1 + \frac{2\delta}{1-\delta} \right\} \\ &= \frac{1}{4} \frac{1+\delta}{1-\delta}, \end{aligned}$$

and the result that $\mu_2 = o(n)$ at most follows.

To complete the proof we have to show that $\mu_2 = o(n)$ at least. If $\delta_i \geq 0$ for all i this further result is immediate, for then

$$\Delta\mu_2^{(n)} \geq \frac{1}{4}$$

for all n . But it is not so easy to establish this second result in the more general case in which some or all of the

δ_i are negative. The difficulty of discussing the result by consideration of the successive increments lies in the fact that in this case an increment $\Delta\mu_2$ may be negative. We have already noticed an instance of this in 4.4, and it is perhaps useful to illustrate the difficulty by first discussing briefly the constant case.

We then have, on putting $\delta_i = \delta$ for all i , that

$$\begin{aligned}\Delta\mu_2^{(n)} &= \frac{1}{4} \{ 1 + 2\delta + 2\delta^2 + \dots + 2\delta^{n+1} \} \\ &= \frac{1}{4(1-\delta)} \{ 1 + \delta - 2\delta^{n+2} \}.\end{aligned}$$

If $\delta \neq 1$, so that $1 - \delta > 0$, then $\Delta\mu_2^{(n)} < 0$ if and only if

$$1 + \delta - 2\delta^{n+2} < 0.$$

This cannot be so if $\delta > 0$ (as we already know) for then

$$1 + \delta - 2\delta^{n+2} > 1 + \delta - 2\delta = 1 - \delta > 0;$$

but it can be for $\delta < 0$. We can illustrate this by reference

to the example of 4.4. If we take $n = 0$, $\delta = -0.6$, we have that

$$\begin{aligned}\Delta\mu_2 &= \frac{1}{4} \frac{1}{1.6} \{ 0.4 - 2(0.36) \} \\ &= -0.05,\end{aligned}$$

as we found. However, if $|\delta| < 1$, $\delta^n \rightarrow 0$ as $n \rightarrow \infty$ so that

$$2|\delta|^{n+2} < 1 + \delta$$

for all sufficiently large n so that the increments are ultimately positive; the closer δ is to -1 , the longer is this positivity deferred. We shall see that such an argument cannot be used in the variable case.

Negative increments $\Delta\mu_2$ can occur likewise in the variable case. For example, suppose that $\delta_i = \delta$ for $i = 1, 2, \dots, n$ but that $\delta_{n+1} = -\delta$, where $0 < \delta < 1$. Then

$$\begin{aligned}\Delta\mu_2^{(n)} &= \frac{1}{4} \{ 1 - 2\delta [1 + \delta + \dots + \delta^n] \} \\ &= \frac{1}{4} \left\{ 1 - 2\delta \left[\frac{1}{1-\delta} - \frac{\delta^{n+1}}{1-\delta} \right] \right\} \\ &= \frac{1}{4} \left\{ \frac{1-3\delta}{1-\delta} + \frac{2\delta^{n+2}}{1-\delta} \right\}.\end{aligned}$$

If $\delta < \frac{1}{3}$, then $\Delta\mu_2^{(n)} > 0$ for all n . But if $\delta > \frac{1}{3}$, then $\Delta\mu_2^{(n)} < 0$ for sufficiently large n . This conclusion merely confirms what we should expect. For the effect of the first n transitions is to create a score distribution which is widely scattered about the mean whereas the $(n + 1)^{\text{th}}$ transition tends to bunch the distribution more closely.

We notice, however, that we now have a proof that $\mu_2^{(n)} = o(n)$ at least in a particular case where some or all of the δ_i are negative. This is the case in which the δ_i , whether positive or negative, are such that $|\delta_i| \leq \delta < \frac{1}{3}$; this follows at once from the expression above. However, we wish to prove the same result under the much less restrictive assumption that $|\delta_i| < \delta < 1$, and this we now do.

4.12. It is evident in the particular example given above to show how a negative increment can arise from a variable chain that, although $\Delta\mu_2^{(n)}$ is negative, the previous increments are all positive. And in the earlier case of negative increments $\Delta\mu_2^{(i)}$ in a constant chain, the increments (after the first) are negative for all i less than a certain fixed value of n (depending on δ) but are positive thereafter. In fact, to prove the result that $\mu_2^{(n)} = o(n)$ we have to consider, not a particular set, but the totality of increments. We proceed as

follows[‡]. We can express $\Delta\mu_2^{(n)}$ in the following form:

$$\frac{1}{4} \Delta\mu_2^{(n)} = v_{n+1}^2 - \delta_{n+1}^2 v_n^2,$$

where for convenience we have written v_n for the previous $v_1^{(n)}$. This result follows from (2) and (4) of 4.7. For in the present particular case of $\Delta_i = \frac{1}{2}$ we have that

$$\begin{aligned} \Delta\mu_2^{(n)} &= \frac{1}{4} + 2 \delta_{n+1} v_n \\ &= \left(\frac{1}{2} + 2 \delta_{n+1} v_n\right)^2 - 4 \delta_{n+1}^2 v_n^2 \\ &= 4 \left(\frac{1}{4} + \delta_{n+1} v_n\right)^2 - 4 \delta_{n+1}^2 v_n^2 \\ &= 4 (v_{n+1}^2 - \delta_{n+1}^2 v_n^2), \end{aligned}$$

from (5) of 4.5. We now have on summation that

$$\begin{aligned} \mu_2^{(n)} &= \frac{1}{4} + 4(v_1^2 - \delta_1^2 v_0^2) + 4(v_2^2 - \delta_2^2 v_1^2) + \dots \\ &\quad \dots + 4(v_n^2 - \delta_n^2 v_{n-1}^2) \end{aligned}$$

which we can write in the form

$$\begin{aligned} \frac{1}{4} \mu_2^{(n)} &= \frac{1}{16} (1 - \delta_1^2) + (1 - \delta_2^2) v_1^2 + (1 - \delta_3^2) v_2^2 + \dots \\ &\quad \dots + (1 - \delta_n^2) v_{n-1}^2 + v_n^2, \end{aligned}$$

since $v_0 = v_1^{(0)} = \frac{1}{4}$. Then, since $|\delta_i| \leq \delta$,

$$\frac{1}{4} \mu_2^{(n)} \geq (1 - \delta^2) \{ v_0^2 + v_1^2 + \dots + v_n^2 \}.$$

‡ The method is that of Markov, 14, pp. 32-33. It is reproduced in Bernstein, pp. 30-31.

Also, since

$$v_{nn} = \frac{1}{4} + \delta_{nn} v_n,$$

it follows that

$$v_{nn}^2 = \frac{1}{16} + \frac{1}{2} \delta_{nn} v_n + \delta_{nn}^2 v_n^2,$$

so that

$$\begin{aligned} v_n^2 + v_{nn}^2 &= \frac{1}{16} + \frac{1}{2} \delta_{nn} v_n + (1 + \delta_{nn}^2) v_n^2 \\ &= (1 + \delta_{nn}^2) \left\{ \left(v_n + \frac{1}{4} \frac{\delta_{nn}}{1 + \delta_{nn}^2} \right)^2 + \frac{1}{16} \left(\frac{1}{1 + \delta_{nn}^2} - \frac{\delta_{nn}^2}{(1 + \delta_{nn}^2)^2} \right) \right\} \\ &\geq \frac{1}{16} \frac{1}{1 + \delta_{nn}^2}, \end{aligned}$$

this value being given by

$$v_n = -\frac{1}{4} \frac{\delta_{nn}}{1 + \delta_{nn}^2}.$$

Thus

$$v_n^2 + v_{nn}^2 \geq \frac{1}{16} \frac{1}{1 + \delta^2}$$

and so

$$\begin{aligned} \frac{1}{4} \mu_2^{(n)} &\geq \frac{1}{2} (1 - \delta^2) \left\{ v_0^2 + n \cdot \frac{1}{16} \frac{1}{1 + \delta^2} + v_n^2 \right\}, \\ \mu_2^{(n)} &\geq \frac{n}{8} \frac{1 - \delta^2}{1 + \delta^2}. \end{aligned}$$

It follows that $\mu_2^{(n)} = O(n)$ at least, as we wished to prove.

This result is sufficient for our needs but it is

evident from the method of proof, and also otherwise, that

the bound for $\mu_2^{(n)}$ above is by no means a 'best possible' bound. For example, if $\delta_i = -\delta$ and $\delta > 0$,

$$\mu_2^{(n)} \sim \frac{1}{4} n \frac{1-\delta}{1+\delta}$$

and

$$\frac{1-\delta}{1+\delta} - \frac{1}{2} \frac{1-\delta^2}{1+\delta^2} = \frac{(1-\delta)^3}{(1+\delta)(1+\delta^2)} > 0$$

unless $\delta = 1$. It is natural to conjecture (from the known nature of the distribution in the constant case rather than this present result) that the best possible bound corresponds, in fact, to the case of $\delta_i = -\delta$ and to suppose that, if

$|\delta_i| \leq \delta < 1$, then

$$\mu_2/n \geq \frac{1}{4} \frac{1-\delta}{1+\delta} - \varepsilon$$

where $\varepsilon > 0$, for sufficiently large n and fixed arbitrary ε . This conjecture we have been able to prove only for low values of n .

4.13. We now extend these results to the case of a general binomial chain and so make no assumptions of symmetry. We have from 4.5 that

$$\Delta \mu_2^{(n)} = \Delta_i - \Delta_i^2 + 2v_1^{(i)} \delta_{i+1}$$

and

$$v_1^{(i)} = \Delta_i - \Delta_i^2 + v_1^{(i)} \delta_{i+1}$$

so that, on writing

$$\Delta i - \Delta_i^2 = m_{i+1}$$

we have

$$\Delta \mu_2^{(i)} = m_{i+1} + 2\delta_{i+1} m_i + 2\delta_{i+1} \delta_i m_{i-1} + \dots + 2\delta_{i+1} \dots \delta_1 m_0.$$

From 4.1, $\Delta i = o(1)$ if $|\delta_i| \leq \delta < 1$ so that, as before,

$\Delta \mu_2 = o(1)$, and $\mu_2 = O(n)$, at most. To show that $\mu_2 = O(n)$ precisely we use a transformation similar to that of the preceding section. Firstly, we have from (1) above, on addition of the increments, that

$$\begin{aligned} \mu_2^{(n)} &= m_{n+1} + m_n (1 + 2\delta_{n+1}) + m_{n-1} (1 + 2\delta_n \delta_{n+1}) \\ &\quad + m_{n-2} (1 + 2\delta_{n-1} + 2\delta_{n-1} \delta_n + 2\delta_{n-1} \delta_n \delta_{n+1}) + \dots \\ &\quad \dots + m_0 (1 + 2\delta_1 + 2\delta_1 \delta_2 + \dots + 2\delta_1 \delta_2 \dots \delta_{n+1}). \end{aligned}$$

Now write

$$t_i = 1 + \delta_i + \delta_i \delta_{i+1} + \dots + \delta_i \delta_{i+1} \dots \delta_{n+1}, \quad t_{n+2} = 1.$$

Then

$$t_i = 1 + \delta_i t_{i+1}$$

and

$$t_i^2 = 1 + 2\delta_i t_{i+1} + \delta_i^2 t_{i+1}^2$$

so that

$$\begin{aligned} \mu_2^{(n)} &= m_{n+1} + m_n (t_{n+1}^2 - \delta_{n+1}^2 t_{n+2}^2) + m_{n-1} (t_n^2 - \delta_n^2 t_{n+1}^2) + \dots \\ &\quad \dots + m_0 (t_1^2 - \delta_1^2 t_2^2) \\ &= t_{n+2}^2 (m_{n+1} - \delta_{n+1}^2 m_n) + t_{n+1}^2 (m_n - \delta_n^2 m_{n-1}) + \dots \\ &\quad \dots + t_2^2 (m_1 - \delta_1^2 m_0) + m_0 t_1^2. \end{aligned}$$

We now consider the sign of a typical coefficient

$$m_{i+1} - \delta_{i+1}^2 m_i = c_i$$

say. Then

$$\begin{aligned} c_i &= \Delta_i (1 - \Delta_i) - \delta_{i+1}^2 \Delta_{i-1} (1 - \Delta_{i-1}) \\ &= (p_{i+1} + \delta_{i+1} \Delta_{i-1}) (\xi_{i+1} - \delta_{i+1} \Delta_{i-1}) - \delta_{i+1}^2 \Delta_{i-1} (1 - \Delta_{i-1}), \end{aligned}$$

from 4.1, so that

$$\begin{aligned} c_i &= p_{i+1} \xi_{i+1} + \xi_{i+1} \delta_{i+1} \Delta_{i-1} - p_{i+1} \delta_{i+1} \Delta_{i-1} - \delta_{i+1}^2 \Delta_{i-1}^2 \\ &\quad - \delta_{i+1}^2 \Delta_{i-1} + \delta_{i+1}^2 \Delta_{i-1}^2 \\ &= p_{i+1} \xi_{i+1} + \Delta_{i-1} \delta_{i+1} \{ 1 - 2p_{i+1} - \delta_{i+1} \}. \end{aligned}$$

Suppose that p_{i+1} is fixed. Then the least value of

$$\delta_{i+1} \{ 1 - 2p_{i+1} - \delta_{i+1} \} \tag{1}$$

occurs when

$$\delta_{i+1} = -p_{i+1} \quad \text{or} \quad \delta_{i+1} = 1 - p_{i+1}$$

in general. But in the uniformly positive case, where we suppose that each element of P_{i+1} is not less than ϵ , the corresponding values are

$$\delta_i = -p_{i+1} + \epsilon \quad \text{or} \quad \delta_{i+1} = 1 - p_{i+1} - \epsilon,$$

for the previous values occur when an element of P_{i+1} is 0 or 1 respectively. Thus in the uniformly positive case the least value of (1) is

$$(-p_{i+1} + \epsilon)(1 - p_{i+1} - \epsilon)$$

which is negative or zero. Since $0 \leq \Delta_{i+1} \leq 1$ it follows that

$$\begin{aligned} c_i &\geq p_{i+1}(1 - p_{i+1}) + (-p_{i+1} + \epsilon)(1 - p_{i+1} - \epsilon) \\ &= p_{i+1}(1 - p_{i+1}) - p_{i+1}(1 - p_{i+1} - \epsilon) + \epsilon(1 - p_{i+1} - \epsilon) \\ &= \epsilon(1 - \epsilon). \end{aligned}$$

Thus $c_i \geq c = \epsilon(1 - \epsilon)$ for all i , so that

$$\mu_2^{(n)} \geq c \left\{ t_1^2 + t_2^2 + \dots + t_{n+1}^2 + t_{n+2}^2 \right\}.$$

We now proceed much as in the preceding section. We have that

$$t_i = 1 + \delta_i t_{i+1}$$

so that

$$\begin{aligned}
t_i^2 + t_{i+1}^2 &= (1 + \delta_i^2) t_{i+1}^2 + 2\delta_i t_{i+1} + 1 \\
&\geq \frac{1}{1 + \delta_i^2} \\
&\geq \frac{1}{1 + \delta^2}
\end{aligned}$$

since $|\delta_i| \leq \delta$. It now follows that

$$\mu_2^{(n)} \geq \frac{1}{2} c n \frac{1}{1 + \delta^2},$$

so that $\mu_2^{(n)} = O(n)$ as in the previous particular case.

Once again this lower bound for $\mu_2^{(n)}/n$ is clearly not the best possible and we conjecture that the best possible lower bound corresponds ultimately to the same constant, and in fact symmetric, chain as before.

4.14. We can extend the result proved in 4.10 for the symmetric binomial chain to the general binomial chain, i.e. we show that if $P_i \rightarrow P$, then the values of $\mu_2^{(n)}$, for any n , of the score distributions associated with the chains $\{P_i\}$ and $\{P\}$ tend to equality. We have from 4.13 that

$$\Delta \mu_2^{(n)} = m_{n+1} + 2\delta_{n+1} m_n + 2\delta_{n+1} \delta_n m_{n-1} + \dots + 2\delta_{n+1} \delta_n \dots \delta_1 m_0,$$

where

$$m_i = \Delta_i - \Delta_i^2$$

and

$$\Delta_i = p_{i+1} + \delta_{i+1} p_i + \delta_{i+1} \delta_i p_{i-1} + \dots + \delta_{i+1} \delta_i \delta_{i-1} \dots \delta_2 p_1.$$

We have in the present case that $p_i \rightarrow p$ and $\delta_i \rightarrow \delta$ as $i \rightarrow \infty$, and we show first that, in consequence, $\Delta_i \rightarrow \Delta$, where Δ is the corresponding score increment associated with the chain $\{P\}$. To do this we write

$$\begin{aligned} \Delta_n &= \{ p_{n+1} + \delta_{n+1} p_n + \dots + \delta_{n+1} \delta_n \dots \delta_{n_1+1} p_{n_1} \} \\ &\quad + \{ \delta_{n+1} \dots \delta_{n_1} p_{n_1-1} + \dots + \delta_{n+1} \dots \delta_p p_1 \} \\ &= A_1 + A_2, \end{aligned}$$

say. Consider the differences of the sums A_1 and A_2 from the corresponding sums in the constant case. If these differences are D_1 and D_2 respectively, and if n_1 is chosen so large that $|p_i - p| \leq \varepsilon$ for $n \geq n_1$, then clearly

$$|D_1| \leq \frac{\varepsilon}{1-d} \quad \text{and} \quad |D_2| \leq \frac{d^{n-n_1}}{1-d^2},$$

using the notation and the lemma of 4.10. The result that

$\Delta_i \rightarrow \Delta$ now follows and, as a corollary of this, we have that $m_i \rightarrow m$. Finally we have, by essentially the same argument as above but with m_i in place of p_i , that

$\Delta \mu_2^{(w)}$ for the variable chain tends to the $\Delta \mu_2^{(w)}$ of the corresponding constant chain.

CHAPTER 5.

5.1. We have considered the mean and the variance for a binomial chain and now consider the moments of higher order by a similar method of increments. We shall once again consider first the case of a constant chain, and then extend the results to the variable case in the next chapter. We shall show that, in the uniformly positive case (to which we restrict ourselves)

$$\mu_{2m+1}^{(n)} = O(n^m)$$

for moments of odd order, and

$$\mu_{2m}^{(n)} = (2m-1)(2m-3) \dots 3 \left\{ \mu_2^{(n)} \right\}^m + O(n^{m-1})$$

for moments of even order. It then follows that the distribution of x^n/\sqrt{n} about the mean tends to the normal distribution.

We shall consider first some moments of low order and then, having noted the methods of proof in these particular cases, use these methods and results to prove the general results by induction.

5.2. We show first that $\mu_3 = O(n)$. It then follows, since $\mu_2 = O(n)$, that 'skewness', defined as $\mu_3/\mu_2^{3/2}$, tends to zero as $n \rightarrow \infty$. We know that this is so, of

course, in the constant case from our previous matrix proof of 3.6, and we have noted the fact in the particular numerical example of 3.3.

Consider the increment to μ_3 on passing from the i^{th} to the $(i+1)^{\text{th}}$ stage. Writing Δ for Δ_i ; δ for δ_{i+1} ; $\lambda_\theta, \nu_\theta$ for $\lambda_\theta^{(i)}, \mu_\theta^{(i)}$ for the present, we have that

$$\underline{\alpha}_{i+1} = e^{-\Delta i} \left[1 + \mu_2 \frac{(i)^2}{2!} + \mu_3 \frac{(i)^3}{3!} + \dots, \nu_0 + \nu_1 (i) + \nu_2 \frac{(i)^2}{2!} + \dots \right]$$

$$\left[\begin{array}{cc} 1 + p(i) + p^2 \frac{(i)^2}{2!} + p^3 \frac{(i)^3}{3!} + \dots & , p + p(i) + p \frac{(i)^2}{2!} + p \frac{(i)^3}{3!} + \dots \\ \delta(i) + \delta \frac{(i)^2}{2!} + \delta \frac{(i)^3}{3!} + \dots & , \delta + \delta(i) + \delta \frac{(i)^2}{2!} + \delta \frac{(i)^3}{3!} + \dots \end{array} \right] \quad \text{--- (A)}$$

whence

$$\alpha_{i+1} = \left\{ 1 - \Delta(i) + \Delta^2 \frac{(i)^2}{2!} - \Delta^3 \frac{(i)^3}{3!} + \dots \right\} \left[1 + (p + \nu_0 \delta)(i) + \frac{(i)^2}{2!} (\mu_2 + p + \nu_0 \delta + 2\nu_1 \delta) + \frac{(i)^3}{3!} (\mu_3 + 3p\mu_2 + p + \nu_0 \delta + 3\nu_1 \delta + 3\nu_2 \delta) + \dots \right],$$

so that $\Delta \mu_3^{(i)}$, the increment to μ_3 on passing from the i^{th} to the $(i+1)^{\text{th}}$ stage, is given by

$$\Delta \mu_3 = [3p\mu_2 + p + \nu_0 \delta + 3\nu_1 \delta + 3\nu_2 \delta] - 3\Delta [\mu_2 + p + \nu_0 \delta + 2\nu_1 \delta] + 3\Delta^2 (p + \nu_0 \delta) - \Delta^3.$$

We know from 4.5 that ν_0 and ν_1 are $O(1)$ and, from 4.13,

that μ_2 is $O(n)$. The order of v_2 is unknown at present, but we can say that

$$\begin{aligned}\Delta\mu_3 &= 3(p-\Delta)\mu_2 + 3v_2\delta + O(1) \\ &= 3(v_2 - v_0\mu_2)\delta + O(1)\end{aligned}$$

since, from 4.5,

$$\Delta = p + v_0\delta.$$

Thus $\Delta\mu_3$ is $O(1)$ if and only if $v_2 - v_0\mu_2$ is $O(1)$.

We have now to consider v_2 . We find from the recurrence relation (A) above that

$$\begin{aligned}v_2^{(L_n)} &= \Delta + p\mu_2 + \delta(2v_1 + v_2) - 2\Delta(\Delta + v_1\delta) + \Delta^3 \\ &= \Delta - 2\Delta^2 + \Delta^3 + v_2\delta + p\mu_2 + 2v_1(1-\Delta)\delta.\end{aligned}$$

This result expresses $v_2^{(L_n)}$ in terms of $v_2^{(L)}$ and known functions and so we can express $v_2^{(L)}$ recursively in terms of known functions. However, we are interested at present only in the order of $v_2^{(L)}$, and hence only in the dominant terms. It is easy to see that $v_2^{(L_n)}$ differs from

$$\frac{p\mu_2}{1-\delta} (1-\delta^n)$$

only by terms $O(1)$. For we can write

$$v_2^{(L_n)} = p\mu_2^{(L)} + Ki,$$

where K_i is $O(1)$ and uniformly bounded : $K_i \leq K$ say, for all i , whence $v_2^{(n+i)}$ differs from

$$\frac{p\mu_2}{1-\delta} (1-\delta^n),$$

which is $O(n)$, by less than $K/(1-\delta)$, which is $O(1)$.

Thus

$$v_2^{(n+i)} = \frac{p\mu_2}{1-\delta} (1-\delta^n) + o(1).$$

We have, from 4.5 and 4.1, that

$$v_0^{(i)} = \frac{p(1-\delta^i)}{1-\delta},$$

and so we see at once that $v_2 - v_0 \mu_2 = O(1)$. Thus

$\Delta \mu_3 = O(1)$ and so, on summation, $\mu_3 = O(n)$ at most, as we wished to show.

We can, however, prove this result otherwise without appeal to the explicit formulae for $\mu_2^{(n)}$ and $v_2^{(n)}$. It is this more general method that we shall use in the proof of the similar results that arise from the discussion of moments of higher order. Thus we have as above

$$v_2^{(n+i)} = p\mu_2^{(i)} + \delta v_2^{(i)} + K_i$$

so that, since $\Delta \mu_2^{(i)} = O(1)$,

$$v_2^{(n)} = \rho \mu_2^{(n)} + \delta v_2^{(n)} + K'_i,$$

where $K'_i = O(1)$.

Then, using the result that

$$v_0^{(n)} = \rho + \delta v_0^{(n)},$$

it follows that

$$\begin{aligned} v_2^{(n)} - \mu_2^{(n)} v_0^{(n)} &= \delta (v_2^{(n)} - \mu_2^{(n)} v_0^{(n)}) + O(1) \\ &= \delta (v_2^{(n)} - \mu_2^{(n)} v_0^{(n)}) + O(1). \end{aligned}$$

Thus, writing

$$v_2^{(n)} - \mu_2^{(n)} v_0^{(n)} = f_i,$$

we have the recurrence relation

$$f_{i+1} = \delta f_i + r_{i+1},$$

where $r_i = O(1)$. This is a recurrence relation of a type we have met before, in 4.5. We find, as then, that

$$\begin{aligned} f_i &= r_i + \delta r_{i-1} + \dots + \delta^{i-1} r_1 \\ &\leq \{1 + \delta + \dots + \delta^{i-1}\} r \\ &< \frac{r}{1-\delta}, \end{aligned}$$

where $r = \overline{\text{bd}} r_i$, this bound existing by definition since

$r_i = O(1)$. Thus $v_2^{(n)} - v_0^{(n)} \mu_2^{(n)} = O(1)$ and $\Delta \mu_3 = O(1)$,
as before.

5.3. We now consider μ_4 . We have the general recurrence formula for the c.f. :

$$\underline{\alpha_{i+1}} = e^{-\Delta i t} \underline{\alpha_i} \begin{bmatrix} \rho + p e^{i t} & p e^{i t} \\ -\delta + \delta e^{i t} & \delta e^{i t} \end{bmatrix}$$

where

$$\underline{\alpha_i} = \left[1 + \mu_2 \frac{(i t)^2}{2!} + \mu_3 \frac{(i t)^3}{3!} + \mu_4 \frac{(i t)^4}{4!} + \dots, \nu_0 + \nu_1 (i t) + \nu_2 \frac{(i t)^2}{2!} + \dots \right].$$

From this we find that

$$\begin{aligned} \underline{\alpha_{i+1}} &= e^{-\Delta i t} \left[1 + (p + \nu_0 \delta) (i t) + \frac{(i t)^2}{2!} (\mu_2 + p + 2\nu_1 \delta + \nu_0 \delta) \right. \\ &\quad + \frac{(i t)^3}{3!} (p + 3\mu_2 p + \mu_3 + \nu_0 \delta + 3\nu_1 \delta + 3\nu_2 \delta) \\ &\quad \left. + \frac{(i t)^4}{4!} (\mu_4 + 4p\mu_3 + 6p\mu_2 + p + 4\nu_3 \delta + 6\nu_2 \delta + 4\nu_1 \delta + \nu_0 \delta) + \dots \right], \end{aligned}$$

whence the increment to μ_4 on transition from the i^{th} to the $(i + 1)^{\text{th}}$ stage is given by

$$\begin{aligned} \Delta \mu_4 &= [4p\mu_3 + 6p\mu_2 + p + 4\nu_3 \delta + 6\nu_2 \delta + 4\nu_1 \delta + \nu_0 \delta] - 4\Delta [p + 3\mu_2 p \\ &\quad + \nu_0 \delta + 3\nu_1 \delta + 3\nu_2 \delta] + 6\Delta^2 [\mu_2 + p + 2\nu_1 \delta + \nu_0 \delta] - 4\Delta^3 (p + \nu_0 \delta) + \Delta^4. \end{aligned}$$

In this expression μ_3 , μ_2 and ν_2 are $O(n)$ and all other terms apart from ν_3 , whose order is as yet

unknown, are $O(1)$. Noting these facts and the relation

$$\Delta = p + v_0 \delta,$$

the expression for $\Delta \mu_4$ can be written more concisely in the form

$$\Delta \mu_4 = 4\delta(v_3 - v_0 \mu_3) + 6\mu_2(\Delta - \delta^2) + 6\delta(1 - 2\Delta)(v_2 - v_0 \mu_2) + O(1).$$

We know already, from (5.2) above, that $v_2 - v_0 \mu_2 = O(1)$, and we are therefore led to the study of $v_3 - v_0 \mu_3$.

From the fundamental recurrence relation for the c.f. we find that

$$\begin{aligned} v_3^{(n)} &= p\mu_3 + 3p\mu_2 + p + v_0\delta + 3v_1\delta + 3v_2\delta + v_3\delta \\ &\quad - 3\Delta [p + \mu_2 p + v_0\delta + 2v_1\delta + v_2\delta] \\ &\quad + 3\Delta^2 [p + v_1\delta + v_0\delta] - \Delta^3 (p + v_0\delta) \\ &= \delta v_3^{(i)} + p\mu_3^{(i)} + 3\Delta i(1 - \Delta i)\mu_2^{(i)} + 3\delta(1 - \Delta i)(v_2^{(i)} - v_0^{(i)}\mu_2^{(i)}) + O(1). \end{aligned}$$

It follows that we can write the preceding equation in the form

$$v_3^{(n)} - v_0^{(n)}\mu_3^{(i)} = \delta(v_3^{(i)} - v_0^{(i)}\mu_3^{(i)}) + 3\Delta i(1 - \Delta i)\mu_2^{(i)} + O(1)$$

since $\Delta i = v_0^{(n)}$, whence, since $\Delta \mu_3 = O(1)$ as μ_3 is $O(n)$,

$$v_3^{(n)} - v_0^{(n)}\mu_3^{(i)} = \delta(v_3^{(i)} - v_0^{(i)}\mu_3^{(i)}) + 3\Delta i(1 - \Delta i)\mu_2^{(i)} + O(1).$$

In the present constant case we have that $\Delta i \rightarrow \Delta$ and

hence

$$v_3^{(n)} - v_0^{(n)} \mu_3^{(n)} \rightarrow \frac{3\Delta(1-\Delta)}{1-\delta} \mu_2^{(n)} + o(1).$$

Thus

$$\begin{aligned} \Delta \mu_4^{(i)} &\rightarrow 4\delta \cdot \frac{3\Delta(1-\Delta)}{1-\delta} \mu_2^{(i)} + 6\mu_2^{(i)}(\Delta - \Delta^2) \\ &= 6\mu_2^{(i)} \left\{ \Delta - \Delta^2 + \frac{2\delta(1-\Delta)\Delta}{1-\delta} \right\} \\ &= 6\mu_2^{(i)} \left\{ \Delta - \Delta^2 + 2v_1\delta \right\}. \end{aligned}$$

$$\therefore \Delta \mu_4^{(i)} \rightarrow 6\mu_2^{(i)} \cdot \Delta \mu_2^{(i)}$$

so that

$$\mu_4 \rightarrow 3\mu_2^2.$$

5.4. We have now sufficient material to establish the corresponding results for all moments by induction in the constant case. But the general method of proof is perhaps made clearer if, in this preliminary particular binomial case, we consider first μ_5 and μ_6 , the two moments of next highest order.

From the fundamental recurrence relation for $\underline{\alpha}_i$ we find that

$$\Delta \mu_5 = 5\delta (v_4 - v_0 \mu_4) + O(n)$$

so that $\Delta \mu_5 = O(n)$ if $v_4 - v_0 \mu_4 = O(n)$.

Again consider the recurrence relation, and omit from Δ all terms $O(n)$ or less. We know that μ_2, v_2 and μ_3 are $O(n)$, and so also is v_3 , for we have shown, in 5.3, that $v_3 - v_0 \mu_3 = O(n)$, and also $\mu_3 = O(n)$. Then we have that $v_4^{(n)}$ is the coefficient of $\frac{(ix)^4}{4!}$ in

$$\{1\} \left[\mu_4 \frac{(ix)^4}{4!}, v_4 \frac{(ix)^4}{4!} \right] \begin{bmatrix} p \\ \delta \end{bmatrix}$$

apart from terms $O(n)$. Thus

$$\begin{aligned} v_4^{(n)} &= p \mu_4^{(n)} + \delta v_4^{(n)} \\ &= (\Delta i - v_0^{(n)} \delta i_1) \mu_4^{(n)} + \delta i_1 v_4^{(n)} + O(n). \end{aligned}$$

But $\Delta i = v_0^{(n)}$ so that

$$v_4^{(n)} - v_0^{(n)} \mu_4^{(n)} = \delta i_1 (v_4^{(n)} - v_0^{(n)} \mu_4^{(n)}) + O(n)$$

and thus, since $\mu_4^{(n)} - \mu_4^{(n)} = O(n)$,

$$v_4^{(n)} - v_0^{(n)} \mu_4^{(n)} = \delta i_1 (v_4^{(n)} - v_0^{(n)} \mu_4^{(n)}) + K_i$$

where $K_i = O(n)$, so that $K_i \leq Kn$ for all n . It now follows, just as in the preceding section, that

$$v_4^{(n)} - v_0^{(n)} \mu_4^{(n)} = O(n)$$

and hence that $\Delta \mu_5 = O(n)$ and $\mu_5 = O(n^2)$.

5.5. We now show that $\mu_6 = 5.3. \mu_2^2$. We find

$\Delta \mu_6$, the increment to μ_6 on passing from the i^{th} to the $(i+1)^{\text{th}}$ stage, from the recurrence relation; since we are concerned with the ultimate order of μ_6 we retain only terms $O(n^2)$ so that, to this order, $\mu_6^{(i+1)}$ is the coefficient of $\frac{(i+1)^6}{6!}$ in

$$\left\{ 1 - \Delta(i) + \Delta^2 \frac{(i)^2}{2!} \right\} \left[\mu_4 \frac{(i)^4}{4!} + \mu_5 \frac{(i)^5}{5!} + \mu_6 \frac{(i)^6}{6!}, \nu_4 \frac{(i)^4}{4!} + \nu_5 \frac{(i)^5}{5!} + \nu_6 \frac{(i)^6}{6!} \right]$$

$$\left[\begin{array}{c} 1 + p(i) + p \frac{(i)^2}{2!} \\ \delta(i) + \delta \frac{(i)^2}{2!} \end{array} \right].$$

We notice that the only relevant powers of (i) in $e^{-\Delta i}$ and the transformed characteristic transition matrix are those not greater than the second. This has been so in consideration of μ_2 and μ_4 , and will be so in consideration of μ_{2r} for any integral r . Thus

$$\begin{aligned} \mu_6^{(i+1)} &= (\mu_6 + 6p\mu_5 + 5.3p\mu_4 + 6\delta\mu_5 + 5.3.8\nu_4) \\ &\quad - 6\Delta(5p\mu_4 + 5\delta\nu_4 + \mu_5) + 5.3\mu_4\Delta^2 + O(n). \end{aligned}$$

On substitution of $\Delta = \nu_0\delta$ for p throughout we have that

$$\begin{aligned}
 \Delta \mu_6 &= 6(\Delta - \nu_0 \delta) \mu_5 + 15(\Delta - \nu_0 \delta) \mu_4 + 6\delta \mu_5 + 15\delta \nu_4 \\
 &\quad - 30\delta \nu_4 \Delta - 30\Delta(\Delta - \nu_0 \delta) \mu_4 - 6\mu_5 \Delta + 15\mu_4 \Delta^2 + O(n) \\
 &= -30\delta \Delta (\nu_4 - \nu_0 \mu_4) + 15\delta (\nu_4 - \nu_0 \mu_4) \\
 &\quad - 15\mu_4 \Delta^2 + 15\mu_4 \Delta + 6\delta (\nu_5 - \nu_0 \mu_5) + O(n) \\
 &= 15\Delta(1-\Delta)\mu_4 + 6\delta (\nu_5 - \nu_0 \mu_5) + O(n).
 \end{aligned}$$

We now consider ν_5 . We have from the fundamental recurrence relation that $\nu_5^{(n)}$ is the coefficient of $(it)^5/5!$ in

$$\{1 - \Delta(it)\} \left[(p\mu_4 + \delta\nu_4) \frac{(it)^4}{4!} + \frac{(it)^5}{5!} (5p\mu_4 + p\mu_5 + 5\nu_4\delta + \nu_5\delta) \right]$$

so that

$$\begin{aligned}
 \nu_5^{(n)} &= 5p\mu_4 + p\mu_5 + 5\nu_4\delta + \nu_5\delta - 5\Delta(p\mu_4 + \nu_4\delta) + O(n) \\
 &= 5(\Delta - \nu_0\delta)\mu_4 + (\Delta - \nu_0\delta)\mu_5 + 5\nu_4\delta + \nu_5\delta \\
 &\quad - 5\Delta\nu_4\delta - 5\Delta(\Delta - \nu_0\delta)\mu_4 + O(n)
 \end{aligned}$$

whence

$$\begin{aligned}
 \nu_5^{(n)} - \nu_0^{(n)} \mu_5^{(n)} &= \delta(\nu_5^{(i)} - \nu_0^{(i)} \mu_5^{(i)}) + 5\Delta(1-\Delta)\mu_4 \\
 &\quad + 5\delta(1-\Delta)(\nu_4^{(i)} - \nu_0^{(i)} \mu_4^{(i)}) + O(n),
 \end{aligned}$$

where we have replaced $\mu_5^{(n)}$ on the L.H.S. by $\mu_5^{(n)}$; this we can do since this adds only a term $O(n)$. It now follows, since $v_4 - v_0 \mu_4 = O(n)$, that

$$v_5^{(n)} - v_0^{(n)} \mu_5^{(n)} \rightarrow \frac{5\Delta(1-\Delta)}{1-8} \mu_4 + O(n).$$

Thus

$$\begin{aligned} \Delta \mu_6 &\rightarrow 15 \Delta(1-\Delta) \mu_4 + 68 \frac{5\Delta(1-\Delta)}{1-8} \mu_4 + O(n) \\ &= 15 \mu_4 \Delta(1-\Delta) \frac{1+8}{1-8} + O(n) \\ &= 15 \mu_4 \cdot \Delta \mu_2 + O(n) \\ &= 15 (3\mu_2^2 \cdot \Delta \mu_2) + O(n). \end{aligned}$$

$$\therefore \mu_6 \rightarrow 5 \cdot 3 \cdot \mu_2^3 + O(n^2).$$

5.6. The general pattern of proof in the constant case has now become clear. The main general results, which we have proved above in some particular cases, are:

1. $\mu_{2r+1}^{(n)} = O(n^r)$ for any integer r .

2. $\mu_{2r}^{(n)} = O(n^r)$ or, more precisely,

$$\mu_{2r}^{(n)} = (2r-1)(2r-3) \dots \cdot 3(\mu_2)^r + O(n^{r-1}),$$

To prove 1 we need

$$3. \quad v_{2r}^{(u)} - v_0^{(u)} \mu_{2r}^{(u)} = O(n^{r-1}).$$

To prove 2 we need

$$4. \quad v_{2r-1}^{(u)} - v_0^{(u)} \mu_{2r-1}^{(u)} = (2r-1) \cdot \frac{\Delta(1-\Delta)}{1-\delta} \mu_{2r-2}^{(u)} + O(n^{r-2})$$

and

$$5. \quad \Delta \mu_{2r}^{(u)} = \binom{2r}{2} \Delta(1-\Delta) \mu_{2r-2}^{(u)} + 2r\delta (v_{2r-1}^{(u)} - v_0^{(u)} \mu_{2r-1}^{(u)}) + O(n^{r-2}).$$

We have already proved these results for certain low values of r and we can use inductive arguments to complete the proof.

We verify first that 4 and 5 imply 2. Substituting from 4 in 5,

$$\begin{aligned} \Delta \mu_{2r} &= \binom{2r}{2} \Delta(1-\Delta) \mu_{2r-2} + 2r(2r-1) \frac{\delta \Delta(1-\Delta)}{1-\delta} \mu_{2r-2} + O(n^{r-2}) \\ &= \binom{2r}{2} \mu_{2r-2} \left\{ \Delta(1-\Delta) + 2\delta \frac{\Delta(1-\Delta)}{1-\delta} \right\} + O(n^{r-2}) \\ &= \binom{2r}{2} \mu_{2r-2} \cdot \Delta \mu_2 + O(n^{r-2}) \\ &= (2r-1)(2r-3) \dots 3 \left\{ r \mu_2^{r-1} \Delta \mu_2 \right\} + O(n^{r-2}), \end{aligned}$$

assuming the result proved for all integers up to $r-1$, as in later inductive arguments.

$$\therefore \mu_{2r} = (2r-1)(2r-3) \dots 3 \mu_2^r + O(n^{r-1}).$$

We now prove 1. We have, neglecting terms $O(u^{r-1})$, that $\mu_{2r+1}^{(k)}$ is the coefficient of $(it)^{2r+1} / (2r+1)!$ in

$$\{1 - \Delta(it)\} \left[\mu_{2r} \frac{(it)^{2r}}{(2r)!} + \mu_{2r+1} \frac{(it)^{2r+1}}{(2r+1)!}, \nu_{2r} \frac{(it)^{2r}}{(2r)!} + \nu_{2r+1} \frac{(it)^{2r+1}}{(2r+1)!} \right] \begin{bmatrix} 1+p(it) \\ \delta(it) \end{bmatrix}$$

so that

$$\mu_{2r+1}^{(k)} = \mu_{2r+1} + (2r+1)\rho\mu_{2r} + (2r+1)\delta\nu_{2r} - (2r+1)\mu_{2r}\Delta + O(u^{r-1}),$$

whence

$$\Delta\mu_{2r+1} = (2r+1)(\nu_{2r} - \nu_0\mu_{2r}) + O(u^{r-1}),$$

since $\Delta = \rho + \nu_0\delta$. The result will follow from 3, to which we now turn. The detail in the proof here of the general case is completely analogous to that of the particular case, with $r = 3$, considered above, and it is not necessary to give it again. The only point to notice is that we assume in this proof that 1 and 2 are known to be true for all integers up to $r - 1$.

We now consider the more fundamental results of 4 and 5. We have to consider ν_{2r-1} . We use the recurrence relation for $\frac{\alpha c}{\nu_{2r-1}^{(k)}}$ and omit all terms $O(u^{r-2})$. Then, to this order, $\nu_{2r-1}^{(k)}$ is the coefficient of $(it)^{2r-1} / (2r-1)!$ in

$$\{1 - \Delta(it)\} \left[\mu_{2r-2} \frac{(it)^{2r-2}}{(2r-2)!} + \mu_{2r-1} \frac{(it)^{2r-1}}{(2r-1)!}, \nu_{2r-2} \frac{(it)^{2r-2}}{(2r-2)!} + \nu_{2r-1} \frac{(it)^{2r-1}}{(2r-1)!} \right] \begin{bmatrix} \rho + \rho(it) \\ \delta + \delta(it) \end{bmatrix}$$

i.e. in

$$\left\{ 1 - \Delta(u) \right\} \left\{ \frac{(u)^{2r-2}}{(2r-2)!} (p\mu_{2r-2} + \delta v_{2r-2}) + \frac{(u)^{2r-1}}{(2r-1)!} \left[(p\mu_{2r-1} + v_{2r-1}\delta) + (2r-1)(p\mu_{2r-2} + v_{2r-2}) \right] \right\}$$

We have assumed here that 1 and 2 have already been proved for all integers up to $r - 2$. It follows that

$$v_{2r-1}^{(u)} = p\mu_{2r-1} + v_{2r-1}\delta + (2r-1)(1-\Delta)(p\mu_{2r-2} + v_{2r-2}\delta) + O(u^{r-2}).$$

On replacing p by $\Delta_i - \delta v_0^{(i)}$ and noting that $v_0^{(u)} = \Delta_i$ we have

$$\begin{aligned} v_{2r-1}^{(u)} - v_0^{(u)} \mu_{2r-1}^{(i)} &= \delta(v_{2r-1} - v_0 \mu_{2r-1}) + (2r-1)\Delta(1-\Delta)\mu_{2r-1} \\ &+ (2r-1)(1-\Delta)\delta(v_{2r-2} - v_0 \mu_{2r-2}) + O(u^{r-2}). \end{aligned}$$

We now assume that 3 is proved for all integers up to $r - 1$, and likewise for 1. By use of the latter we can replace $\mu_{2r-1}^{(i)}$ by $\mu_{2r-1}^{(u)}$ on the L.H.S., so that if we write f_i for $v_{2r-1}^{(i)} - v_0^{(i)} \mu_{2r-1}^{(i)}$, we have the recurrence relation

$$f_{i+1} = \delta f_i + (2r-1)\Delta_i(1-\Delta_i)\mu_{2r-1}^{(i)} + O(u^{r-2}).$$

Hence

$$f_n = (2r-1) \sum_{i=1}^n \Delta_i(1-\Delta_i) \delta^{n-i} \mu_{2r-1}^{(i)} + O(u^{r-2}).$$

Thus we have that

$$f_n \rightarrow (2r-1) \frac{\Delta(1-\Delta)}{1-\delta} \mu_{2r-1}^{(i)}.$$

Finally we consider 5. If we neglect all terms $O(n^{r-2})$ we have from the recurrence relation that, to this order,

$\mu_{2r}^{(1)}$ is the coefficient of $(it)^{2r} / (2r)!$ in

$$\left\{ 1 - \Delta(it) + \frac{\Delta^2(it)^2}{2!} \right\} \left[\mu_{2r-2} \frac{(it)^{2r-2}}{(2r-2)!} + \mu_{2r-1} \frac{(it)^{2r-1}}{(2r-1)!} + \mu_{2r} \frac{(it)^{2r}}{(2r)!} \right]$$

$$+ \left[\nu_{2r-2} \frac{(it)^{2r-2}}{(2r-2)!} + \nu_{2r-1} \frac{(it)^{2r-1}}{(2r-1)!} + \nu_{2r} \frac{(it)^{2r}}{(2r)!} \right] \left[1 + p(it) + p \frac{(it)^2}{2!} \right]$$

$$+ \left[\delta(it) + \delta \frac{(it)^2}{2!} \right]$$

and so in

$$\left\{ 1 - \Delta(it) + \frac{\Delta^2(it)^2}{2!} \right\} \left\{ \mu_{2r-2} \frac{(it)^{2r-2}}{(2r-2)!} + \frac{(it)^{2r-1}}{(2r-1)!} \left[\mu_{2r-1} + (2r-1)(p\mu_{2r-2} + \delta\nu_{2r-2}) \right] \right.$$

$$\left. + \frac{(it)^{2r}}{(2r)!} \left[\mu_{2r} + 2r(p\mu_{2r-1} + \nu_{2r-1}\delta) + \binom{2r}{2}(p\mu_{2r-2} + \nu_{2r-2}\delta) \right] \right\}$$

whence

$$\Delta\mu_{2r} = 2r(p\mu_{2r-1} + \nu_{2r-1}\delta) + \binom{2r}{2}(p\mu_{2r-2} + \nu_{2r-2}\delta)$$

$$- 2r\Delta \left[\mu_{2r-1} + (2r-1)(p\mu_{2r-2} + \nu_{2r-2}\delta) \right]$$

$$+ \Delta^2 \binom{2r}{2} \mu_{2r-2} + O(n^{r-2})$$

$$= 2r\delta(\nu_{2r-1} - \nu_0\mu_{2r-1}) + \binom{2r}{2}(\Delta - \Delta^2)\mu_{2r-2} + O(n^{r-2}),$$

on substituting $\Delta i - \delta\nu_0^{(i)}$ for p throughout and assuming that 3 is true for all values up to $r - 1$. This establishes

5, and the truth of 1 - 5 now follows by induction.
The proof by the method of increments of the normal law
in the constant binomial case is thus complete.

CHAPTER 6.

6.1. The proof that the score distribution about its mean also tends to a normal distribution in the case of a variable binomial chain is, in part, essentially the same as the inductive proof by increments in the case of a constant chain, and there is no need to repeat those parts of the proof that are unchanged apart from slight differences of notation. Thus, of the five results of that proof in 5.6, the results 1 and 3 are the same, apart from the replacement of the constant p, δ by the variable p_{i+1}, δ_{i+1} . But differences do appear in the results corresponding to 2, 3 and 4, and to illustrate these differences we shall first prove 2 in the particular case of $m = 2$. This we do by showing, as before, that

$$\Delta \mu_4^{(n)} = 6 \mu_2^{(n)} \Delta \mu_2^{(n)} + o(1).$$

6.2. We have in the variable case, corresponding to a result of 5.3, that

$$\Delta \mu_4^{(i)} = 4 \delta_{i+1} (v_3^{(i)} - v_0^{(i)} \mu_3^{(i)}) + 6 \mu_2^{(i)} (\Delta i - \Delta i^2) + o(1),$$

provided that

$$v_2^{(n)} - v_0^{(n)} \mu_2^{(n)} = o(1).$$

We proved this latter result in the constant case in 5.2, and

a slight adaptation^{at} of that proof shows that the result certainly holds also in the variable case provided that the P_i are uniformly positive. For we have, corresponding to a result of 5.2, that

$$v_2^{(i)} - \mu_2^{(i)} v_0^{(i)} = \delta_{i+1} (v_2^{(i)} - \mu_2^{(i)} v_0^{(i)}) + o(1)$$

so that, on writing

$$v_2^{(i)} - \mu_2^{(i)} v_0^{(i)} = g_i$$

we have the corresponding recurrence relation

$$g_{i+1} = \delta_{i+1} g_i + r_{i+1}$$

where $r_i = o(1)$. Thus

$$\begin{aligned} g_i &= r_i + \delta_i r_{i-1} + \delta_i \delta_{i-1} r_{i-2} + \dots + \delta_i \delta_{i-1} \dots \delta_1 r_0 \\ &\leq \{1 + \delta + \delta^2 + \dots + \delta^i\} r \\ &\leq \frac{r}{1-\delta}, \end{aligned}$$

where $r = \overline{\text{bd}} r_i$ as before. The result now follows.

We also have, corresponding to a result of 5.3, that

$$\begin{aligned} v_3^{(i)} &= \delta_{i+1} v_3^{(i)} + p_{i+1} \mu_3^{(i)} + 3\Delta_i (1-\Delta_i) \mu_2^{(i)} \\ &\quad + 3\delta_{i+1} (1-\Delta_i) (v_2^{(i)} - v_0^{(i)} \mu_2^{(i)}) + o(1) \end{aligned}$$

which we can write, using the result just proved and the result that

$$v_0^{(i)} = p_{i+1} + \delta_{i+1} v_0^{(i)},$$

in the form

$$v_3^{(i)} - v_0^{(i)} \mu_3^{(i)} = \delta_{i+1} (v_3^{(i)} - v_0^{(i)} \mu_3^{(i)}) + 3 \Delta_i (1 - \Delta_i) \mu_2^{(i)} + o(1).$$

Thus, if we write

$$v_3^{(i)} - v_0^{(i)} \mu_3^{(i)} = f_i$$

and

$$\Delta_i (1 - \Delta_i) = m_{i+1}$$

we have

$$\Delta \mu_4^{(i)} = 4 \delta_{i+1} f_i + 6 m_{i+1} + o(1), \quad (1)$$

where

$$f_{i+1} = \delta_{i+1} f_i + 3 m_{i+1} + o(1). \quad (2)$$

We have also proved before, in 4.13, that

$$\Delta \mu_2^{(i)} = \Delta_i (1 - \Delta_i) + \delta_{i+1} \cdot 2 v_1^{(i)} \quad (3)$$

and

$$v_1^{(i)} = \Delta_i - \Delta_i^2 + \delta_{i+1} v_1^{(i)}. \quad (4)$$

We now see what these results, taken with our prospective result, that

$$\Delta \mu_4^{(i)} = b \mu_2^{(i)} \Delta \mu_2^{(i)} + o(1), \quad (5)$$

imply. With (1)

$$6\mu_2^{(i)} \Delta\mu_2^{(i)} = 4\delta_{i+1} f_i + 6\Delta_i(1-\Delta_i)\mu_2^{(i)} + o(1),$$

and now, from (3),

$$6\mu_2^{(i)} [\Delta_i(1-\Delta_i) + 2\delta_{i+1} v_1^{(i)}] = 4\delta_{i+1} f_i + 6\Delta_i(1-\Delta_i)\mu_2^{(i)} + o(1)$$

so that

$$f_i = 3\mu_2^{(i)} v_1^{(i)} + o(1). \quad (6)$$

The steps of this argument are reversible, so that if we prove (6) we have in fact proved (5). We have from (2) and (4) that

$$f_{i+1} = \delta_{i+1} f_i + 3\mu_2^{(i)} [v_1^{(i+1)} - \delta_{i+1} v_1^{(i)}] + o(1)$$

whence

$$f_{i+1} - 3\mu_2^{(i)} v_1^{(i+1)} = \delta_{i+1} (f_i - 3\mu_2^{(i)} v_1^{(i)}) + o(1)$$

which, on use of the results that $\mu_2^{(i+1)} - \mu_2^{(i)} = o(1)$ and

$v_1^{(i+1)} = o(1)$, we can write in the form

$$f_{i+1} - 3\mu_2^{(i+1)} v_1^{(i+1)} = \delta_{i+1} (f_i - 3\mu_2^{(i)} v_1^{(i)}) + K_{i+1}$$

where $K_i \leq K$ for all i . It follows from this recurrence relation that

$$f_n - 3\mu_2^{(n)} v_1^{(n)} = R_n$$

where

$$R_n = K_n + \delta_n K_{n-1} + \delta_n \delta_{n-1} K_{n-2} + \dots + \delta_n \delta_{n-1} \dots \delta_1 (f_1 - 3\mu_2^{(1)} v_1^{(1)}),$$

If we choose the K above so that $f_i - 3\mu_2^{(i)}v_1^{(i)} < K$ we have that

$$|R_n| < K [1 + \delta + \delta^2 + \dots + \delta^n]$$

where

$$\delta = \max |\delta_i| < 1$$

so that

$$|R_n| < K/(1-\delta) = o(1).$$

The result now follows; and hence that

$$\Delta\mu_4^{(i)} = 6\mu_2^{(i)}\Delta\mu_2^{(i)} + o(1)$$

and

$$\Delta\mu_4^{(n)} = 3 \left\{ \mu_2^{(n)} \right\}^2 + o(n).$$

6.3. The proof in the general case of $\mu_{2r}^{(i)}$ follows a similar pattern. We have, corresponding to (1) and (2) above, that

$$\Delta\mu_{2r}^{(i)} = 2r\delta_{i+1}f_i + \binom{2r}{2}(\Delta_i - \Delta_i^2)\mu_{2r-2}^{(i)} + o(n^{r-2}), \quad (1)$$

where now

$$f_i = v_{2r-1}^{(i)} - v_0^{(i)}\mu_{2r-1}^{(i)}$$

and

$$f_{i+1} = \delta_{i+1}f_i + (2r-1)\Delta_i(1-\Delta_i)\mu_{2r-2}^{(i)} + o(n^{r-2}), \quad (2)$$

which was proved in the course of the proof of (4) in 5.6. We use (3) and (4) unchanged, and have from (2) and (4) that

$$f_{i+1} - (2r-1)\mu_{2r-2}^{(i)} v_1^{(i+1)} = \delta_{i+1} \left\{ f_i - (2r-1)\mu_{2r-2}^{(i)} v_1^{(i)} \right\} + O(n^{r-2})$$

which, since

$$\mu_{2r-2}^{(i+1)} - \mu_{2r-2}^{(i)} = O(n^{r-2})$$

we can write in the form

$$f_{i+1} - (2r-1)\mu_{2r-2}^{(i+1)} v_1^{(i+1)} = \delta_{i+1} \left\{ f_i - (2r-1)\mu_{2r-2}^{(i)} v_1^{(i)} \right\} + K_{i+1},$$

where $K_i = O(n^{r-2})$ for all i . It follows, by exactly the same argument as in the preceding section, that

$$f_n - (2r-1)\mu_{2r-2}^{(n)} v_1^{(n)} = O(n^{r-2})$$

and hence that

$$\binom{2r}{2} \mu_{2r-2}^{(i)} \left[\Delta_i(1-\Delta_i) + 2\delta_{i+1} v_1^{(i)} \right] = 2r\delta_{i+1} f_i + \binom{2r}{2} \Delta_i(1-\Delta_i) \mu_{2r-2}^{(i)} + O(n^{r-2})$$

and so, from (1) and (3), that

$$\Delta \mu_{2r} = \binom{2r}{2} \mu_{2r-2}^{(i)} \Delta \mu_2^{(i)} + O(n^{r-2}),$$

whence, as in the constant case,

$$\mu_{2r} = (2r-1)(2r-3) \dots \dots \dots 3\mu_2^r + O(n^{r-1}).$$

This completes the proof in the variable case.*

* Cf. the proofs in this binomial case given by Bernstein.

6.4. In all the previous discussions the score increments on each occurrence of the event E_0 or E_1 were 0 or 1 respectively, so that the score after the n^{th} stage is, in fact, the number of occurrences of the event E_1 in these n stages. We can generalise this and suppose that the score increments on the occurrence of E_0 and E_1 are 0 and x respectively; but this clearly amounts to no more than a change of scale, and, with this change, the previous results still hold. Again, we can suppose that the score increments are x_0 and x_1 , but this case is essentially the same as that in which the score increments are 0 and $x_1 - x_0$; the difference is merely that of a change of origin.

An essentially different case arises if we suppose that the score increments vary from stage to stage. By the same argument as above, the case in which the score increments consequent on the occurrence of E_0 and E_1 at the j^{th} stage are $x_0^{(j)}$ and $x_1^{(j)}$ is equivalent to that in which the score increments are 0 and $x^{(j)}$, and this is the case we now consider. The work runs parallel to that of preceding sections and we give briefly the results, using the previous notation.

From the recurrence formula (cf. 4.8)

$$\underline{\alpha}_{j+1} = e^{-\Delta \bar{x}_j} \underline{\alpha}_j Q_{j+1} (e^{\Delta \bar{x}_j})$$

where

$$Q_j(e^{it}) = \left[\begin{array}{l} 1 + p_j x_j(it) + p_j x_j^2 \frac{(it)^2}{2!} + \dots, \quad p_j + p_j x_j(it) + p_j x_j^2 \frac{(it)^2}{2!} + \dots \\ \delta_j x_j(it) + \delta_j x_j^2 \frac{(it)^2}{2!} + \dots, \quad \delta_j + \delta_j x_j(it) + \delta_j x_j^2 \frac{(it)^2}{2!} + \dots \end{array} \right]$$

we find, corresponding to the results of 4.8, that

$$\Delta \mu_2^{(j)} = x_{j+1}^2 v_0^{(j)} - x_{j+1}^2 [v_0^{(j)}]^2 + 2 v_1^{(j)} x_{j+1} \delta_{j+1}$$

and

$$v_1^{(j)} = x_{j+1} \{ v_0^{(j)} - [v_0^{(j)}]^2 \} + v_1^{(j)} \delta_{j+1}$$

Thus, writing

$$v_0^{(j)} (1 - v_0^{(j)}) = a_j,$$

we have that

$$\begin{aligned} \Delta \mu_2^{(j)} &= x_{j+1} \{ a_{j+1} x_{j+1} + 2 a_j x_j \delta_{j+1} + 2 a_{j-1} x_{j-1} \delta_{j+1} \delta_j + \dots \\ &\quad \dots + 2 a_0 x_0 \delta_{j+1} \delta_j \dots \delta_2 \delta_1 \} \\ &= x_{j+1} \{ y_{j+1} + 2 \delta_{j+1} y_j + 2 \delta_{j+1} \delta_j y_{j-1} + \dots + 2 \delta_{j+1} \dots \delta_2 \delta_1 y_0 \}, \end{aligned}$$

where

$$x_j a_j = y_j.$$

We now show that $\mu_2^{(n)} = O(n)$ if the P_i are uniformly positive. The method is analogous to that of 4.13. We write

$$Y_{n+1} = y_{n+1} + \delta_{n+1} y_n + \delta_{n+1} \delta_n y_{n-1} + \dots + \delta_{n+1} \delta_n \dots \delta_2 \delta_1 y_0.$$

Then

$$Y_{n+1} = y_{n+1} + \delta_{n+1} Y_n$$

and so

$$Y_{n+1}^2 - \delta_{n+1}^2 Y_n^2 = y_{n+1} \{ y_{n+1} + 2 \delta_{n+1} Y_n \}.$$

Thus

$$\Delta u_2^{(j)} = \frac{1}{a_{j+1}} \{ Y_{j+1}^2 - \delta_{j+1}^2 Y_j^2 \}.$$

We note that, in the uniformly positive case, $a_j > 0$ for all j . For we see from the recurrence relation that, if

$$[1, p_0] \prod_{i=1}^n \begin{bmatrix} 1 & p_i \\ \cdot & \delta_i \end{bmatrix} = [a_n, b_n],$$

then

$$v_0^{(n)} = b_n.$$

Thus $v_0^{(n)}$ is the second element of

$$[z_0, p_0] P^{(n)}$$

whence

$$e \leq v_0^{(n)} \leq 1 - e$$

for all n if

$$p_i \geq e$$

for all i . It follows that

$$a_j \geq e(1 - e)$$

where we have written ρ , δ for ρ_j , δ_j and v for $v_\delta^{(j-1)}$. The result that the c_j are uniformly positive now follows precisely as in 4.13.

Finally,

$$\begin{aligned} Y_j^2 + Y_{j+1}^2 &= Y_j^2 + (y_{j+1} + \delta_{j+1} Y_j)^2 \\ &= y_{j+1}^2 + 2 y_{j+1} \delta_{j+1} Y_j + (1 + \delta_{j+1}^2) Y_j^2 \\ &> \frac{y_{j+1}^2}{1 + \delta_{j+1}^2} \\ &> \frac{y_{j+1}^2}{1 + \delta^2} \end{aligned}$$

if $|\delta_j| \geq \delta$ for all j , in this uniformly positive case.

The result that $\mu_2^{(n)} = o(n)$ therefore follows provided that

$$\sum_{i=1}^{n-1} y_i^2 = o(n).$$

This is the case if the x_i are uniformly bounded above zero, or more generally, if

$$\sum_{i=1}^{n-1} x_i^2 = o(n).$$

The rest of the argument concerning the order of, and relations between, the higher moments follows in essentially the same way as before. We omit the detail here because it occurs in a more general form in the next chapter in the corresponding trinomial case.

CHAPTER 7.

7.1. As a preliminary we consider first a sequence of n independent trials with the probabilities of success variable from trial to trial. Let the probabilities of the occurrence of the states E_1, E_2, E_3 be $p_1^{(j)}, p_2^{(j)}, p_3^{(j)}$ at the j^{th} trial and, corresponding to 6.4, suppose that the occurrence of these states results in the addition of x_1, x_2, x_3 respectively to the score. The probability distribution of the score after n trials is a multinomial distribution which tends, in general, to a normal distribution as n tends to infinity. A sketch of the usual proof is as follows.*

The c.f. after n trials with the current mean as origin (the increment to the mean varies from trial to trial) is

$$\phi_n(t) \equiv \prod_{j=1}^n \left\{ e^{-\Delta_j i t} \left(p_1^{(j)} e^{i x_1 t} + p_2^{(j)} e^{i x_2 t} + p_3^{(j)} e^{i x_3 t} \right) \right\},$$

where

$$\Delta_j = p_1^{(j)} x_1 + p_2^{(j)} x_2 + p_3^{(j)} x_3.$$

Thus, writing p_α for $p_\alpha^{(j)}$ and Δ for Δ_j ,

$$\phi_n(t) \equiv \prod_{j=1}^n \left\{ \left[1 - \Delta(i t) + \frac{\Delta^2 (i t)^2}{2!} - \dots \right] \left[1 + \Delta(i t) + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) \frac{(i t)^2}{2!} + \dots \right] \right\}.$$

* Cramér, 215.

It follows that

$$\begin{aligned} \Phi_n(t) &\equiv \prod_{j=1}^n \left\{ 1 + \left[(p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) - (p_1 x_1 + p_2 x_2 + p_3 x_3)^2 \right] \frac{(t)^2}{2!} + \dots \right\} \\ &= \prod_{j=1}^n \left\{ 1 + \mu_2^{(j)} \frac{(t)^2}{2!} + \mu_3^{(j)} \frac{(t)^3}{3!} + \dots \right\}, \end{aligned}$$

where $\mu_2^{(j)}$, $\mu_3^{(j)}$, are the second, third, moments of the distribution about the mean of the increment to the score at the j^{th} trial. Now replace t by t/σ_n , where

$$\begin{aligned} \sigma_n^2 &= \sum_{j=1}^n \mu_2^{(j)} \\ &= \sum_{j=1}^n \left\{ p_1^{(j)} x_1^2 + p_2^{(j)} x_2^2 + p_3^{(j)} x_3^2 - (p_1^{(j)} x_1 + p_2^{(j)} x_2 + p_3^{(j)} x_3)^2 \right\}. \end{aligned}$$

Then we can write $\Phi_n\left(\frac{t}{\sigma_n}\right)$ in the form

$$\Phi_n\left(\frac{t}{\sigma_n}\right) = \prod_{j=1}^n \left\{ 1 + \frac{\mu_2^{(j)}}{\sigma_n^2} \frac{(t)^2}{2!} + \lambda_j \frac{\rho_j^3}{\sigma_n^2} \frac{t^3}{3!} \right\},$$

where λ_θ is a constant such that $|\lambda_\theta| \leq 1$ and ρ_j^3 is the third absolute moment of the distribution. Then, under certain conditions

$$\Phi_n\left(\frac{t}{\sigma_n}\right) \rightarrow e^{-t^2}$$

as $n \rightarrow \infty$, so that the distribution tends to a normal distribution. A sufficient condition for this to be so is that

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_n} \left\{ \sum_{j=1}^n \rho_j^3 \right\}^{1/3} = 0.$$

If we suppose in the present case that the score increments x_1, x_2, x_3 are bounded in modulus then clearly

$$\sum_{j=1}^3 \rho_j^3 = O(n)$$

at most so that the condition certainly holds provided that $\sigma_n^2 = O(n)$ precisely. A similar situation arises later in the consideration of a sequence of dependent trials and for that reason we now discuss σ_n^2 in more detail.

7.2. Let us drop the suffix j for the present and suppose that p_1, p_2, p_3 are the probabilities of the occurrence of E_1, E_2, E_3 in a typical trial. We know, in this present sequence of independent trials, that

$$E \equiv p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2 - (p_1 x_1 + p_2 x_2 + p_3 x_3)^2$$

is non-negative. But it can be zero, and this can happen in two obvious cases:

(i) when $x_1 = x_2 = x_3$, so that the score increment is independent of the outcome of the trial;

(ii) when one of p_1, p_2, p_3 is unity.

But E can also be zero if neither (i) nor (ii) is true, e.g. if $p_1 = 0$ and $x_2 = x_3$. It is intuitive, in fact, that the expression is zero if, and only if, there is no

uncertainty in the score increment at that stage. This result we shall now prove. First transform E by writing

$$x_1 = x, \quad x_2 - x_1 = \xi, \quad x_3 - x_1 = \eta.$$

Then

$$E = p_1 x^2 + p_2 (\xi + x)^2 + p_3 (\eta + x)^2 - [p_1 x + p_2 (\xi + x) + p_3 (\eta + x)]^2$$

where $p_1 + p_2 + p_3 = 1$, so that

$$E = (p_2 - p_2^2) \xi^2 - 2p_2 p_3 \xi \eta + (p_3 - p_3^2) \eta^2.$$

If both $p_2 = 0$ and $p_3 = 0$ then $p_1 = 1$ and the outcome of the trial is fixed. Otherwise at least one of p_2 and p_3 is non-zero. Suppose that $p_2 \neq 0$. Then

$$E = \frac{1}{p_2 - p_2^2} \left\{ [(p_2 - p_2^2) \xi - p_2 p_3 \eta]^2 + \eta^2 [(p_2 - p_2^2)(p_3 - p_3^2) - p_2^2 p_3^2] \right\},$$

since $p_2 - p_2^2 \neq 0$ if we exclude the case of $p_2 = 1$, another case of certainty. Thus

$$\begin{aligned} E &= \frac{1}{p_2 - p_2^2} \left\{ [(p_2 - p_2^2) \xi - p_2 p_3 \eta]^2 + p_1 p_2 p_3 \eta^2 \right\} \\ &> \frac{p_1 p_2 p_3}{p_2 - p_2^2} \eta^2. \end{aligned}$$

Now if $p_1 = 0$, then $1 - p_2 = p_3$, so that

$$\begin{aligned} E &= \frac{1}{p_2 p_3} [p_2 p_3 (\xi - \eta)]^2 \\ &= p_2 p_3 (\xi - \eta)^2. \end{aligned}$$

If we exclude cases of certainty neither p_2 nor p_3 can be

zero so that E is zero if and only if $\xi = \eta$, in which case $x_1 = x_2 = x_3$, and this again is a case in which the score increment is fixed. Likewise if $p_3 = 0$. If none of p_1, p_2, p_3 is zero and if $\eta > 0$ then $E > 0$. If $\eta = 0$ then $E = (p_2 - p_2^2) \xi^2$ so that $E > 0$ except if $p_2 = 0$ or 1, each of which possibilities correspond to certainty, or $\eta = 0$, i.e. $x_1 = x_2$, which is also a case of certainty, since $p_2 = 0$. Thus $E > 0$ except in cases of certainty; and in all these cases $E = 0$.

We can set lower bounds to E . Thus we have the following three typical cases:

(a) $p_1 p_2 p_3 \neq 0, \xi \eta \neq 0$. Then

$$E \geq \frac{p_1 p_3}{p_1 + p_3} \eta^2.$$

If we suppose that $p_i \geq e_i$, this gives

$$E \geq \frac{1}{2} e_i \eta^2.$$

(b) $\eta = 0$. Then

$$\begin{aligned} E &= p_2 (1 - p_2) \xi^2 \\ &\geq e_i (1 - e_i) \xi^2. \end{aligned}$$

(c) $p_1 = 0$. Then

$$\begin{aligned} E &= p_2 p_3 (\xi - \eta)^2 \\ &\geq e_i (1 - e_i) (\xi - \eta)^2. \end{aligned}$$

In a sequence of trials one or more of these cases can arise. Suppose, however, that the $p_\alpha^{(\theta)}$ are always positive and that

$|\xi|, |\eta| \geq m$; then clearly

$$E_j \geq \frac{1}{2} e_j m_j^2$$

(in fact equality is not possible); and if we now suppose further that $e_j \geq e$, $m_j \geq m$, we have

$$E_j \geq \frac{1}{2} e m^2$$

for all j and so

$$\sigma_n^2 \geq \frac{1}{2} e m^2 n.$$

It follows that the score distribution tends to a normal distribution in this case, and likewise in any case in which the increment to the variance at any stage, E_j , is such that

$$\sum_{j=1}^n E_j = o(n)$$

precisely. This argument leaves open the nature of the distribution where this is not so. For instance, suppose that

$\xi_j = \eta_j = c$ for all j . Then

$$\begin{aligned} E &= c^2 \{ (p_2 - p_2^2) - 2p_2 p_3 + (p_3 - p_3^2) \} \\ &= c^2 \{ p_2 + p_3 - (p_2 + p_3)^2 \} \\ &= c^2 p_1 (1 - p_1). \end{aligned}$$

Now suppose that $p_i^{(j)} = 1/j$. Then

$$\frac{1}{n} \sum_{j=1}^n E_j \rightarrow 0$$

as $n \rightarrow \infty$ and the condition above does not hold. In fact, since we have essentially here the distribution of the number

of occurrences of one of two possible states in a certain sequence of n trials we know* that the distribution does tend to a normal distribution.

7.3. An essential feature of the proof of 7.1 is the use of the multiplicative property of the c.f.'s of the independent variables of a sum; the result then follows from a fundamental exponential limit theorem. But this multiplicative property does not hold for dependent random variables, and we now give a second proof which, although longer, does generalise to the dependent case. Suppose that after n trials the c.f. referred to the then current origin as mean is

$$\phi_n(t) = \left\{ 1 + \mu_2^{(n)} \frac{(it)^2}{2!} + \mu_3^{(n)} \frac{(it)^3}{3!} + \dots \right\}.$$

After $n + 1$ trials the c.f. is

$$\begin{aligned} \phi_{n+1}(t) &= e^{-\Delta_n it} \left\{ 1 + \mu_2 \frac{(it)^2}{2!} + \mu_3 \frac{(it)^3}{3!} + \dots \right\} \\ &\quad \left\{ 1 + (it)(p_1 x_1 + p_2 x_2 + p_3 x_3) + \frac{(it)^2}{2!} (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) + \dots \right\} \\ &= e^{-\Delta_n it} \left\{ 1 + \Delta_n (it) + \frac{(it)^2}{2!} \left[\mu_2 + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) \right] \right. \\ &\quad \left. + \frac{(it)^3}{3!} \left[\mu_3 + 3\mu_2 \Delta_n + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) \right] \right. \\ &\quad \left. + \frac{(it)^4}{4!} \left[\mu_4 + 4\mu_3 \Delta_n + 6\mu_2 (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) + (p_1 x_1^4 + p_2 x_2^4 + p_3 x_3^4) \right] + \dots \right\} \end{aligned}$$

* Cramér, 218.

so that

$$\begin{aligned} \phi_{n+1}(t) = & 1 + \frac{(it)^2}{2!} \left\{ \mu_2 + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) - 2\Delta_n^2 + \Delta_n^2 \right\} \\ & + \frac{(it)^3}{3!} \left\{ \mu_3 + 3\mu_2 \Delta_n + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) - 3\Delta_n [\mu_2 + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2)] + 3\Delta_n^3 - \Delta_n^3 \right\} \\ & + \frac{(it)^4}{4!} \left\{ \left[\mu_4 + 4\mu_3 \Delta_n + 6\mu_2 (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) + (p_1 x_1^4 + p_2 x_2^4 + p_3 x_3^4) \right] \right. \\ & \quad \left. - 4\Delta_n [\mu_3 + 3\mu_2 \Delta_n + (p_1 x_1^3 + p_2 x_2^3 + p_3 x_3^3)] \right. \\ & \quad \left. + 6\Delta_n^2 [\mu_2 + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2)] - 4\Delta_n^4 + \Delta_n^4 \right\} \end{aligned}$$

We wish to study the behaviour of $\phi_n(t)$ as $n \rightarrow \infty$.

We have that

$$\Delta_n = p_1 x_1 + p_2 x_2 + p_3 x_3$$

so that

$$\Delta \mu_2 = (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) - \Delta_n^2$$

whence

$$\mu_2^{(n)} = \sum_{j=1}^{n-1} \left\{ (p_1^{(j)} x_1^2 + p_2^{(j)} x_2^2 + p_3^{(j)} x_3^2) - \Delta_j^2 \right\},$$

as before. Suppose that the $p_\alpha^{(j)}$ are such that $\mu_2^{(n)} = o(n)$.

Now consider μ_3 . We have that

$$\mu_3^{(n)} = \mu_3^{(j)} + p_1 x_1^3 + p_2 x_2^3 + p_3 x_3^3 - 3\Delta_n (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) + 2\Delta_n^2$$

and we see at once that $\mu_3 = o(n)$, for there is no term in μ_2 . Likewise when we consider μ_{2r+1} we find that there is no term in μ_{2r} , so that μ_{2r+1} is $o(n^r)$. Now consider μ_4 . We have that

$$\Delta \mu_4 = 6\mu_2(p_1x_1^2 + p_2x_2^2 + p_3x_3^2) + (p_1x_1^4 + p_2x_2^4 + p_3x_3^4) - 12\mu_2\Delta_n^2 - 4\Delta_n(p_1x_1^3 + p_2x_2^3 + p_3x_3^3) + 6\mu_2\Delta_n^2 + 6\Delta_n^2(p_1x_1^2 + p_2x_2^2 + p_3x_3^2) - 3\Delta_n^4.$$

We notice that there is no term in μ_3 . We know that

$$\Delta_n = O(1) \text{ and } \mu_2 = O(n), \text{ and we wish to show that}$$

$$\mu_4 = 3\mu_2^2 + O(n).$$

Now

$$\begin{aligned} \Delta \mu_4 &= 6\mu_2 \{ (p_1x_1^2 + p_2x_2^2 + p_3x_3^2) - 2\Delta_n^2 + \Delta_n^2 \} + O(1) \\ &= 6\mu_2 \Delta \mu_2 + O(1) \end{aligned}$$

whence

$$\mu_4 = 3\mu_2^2 + O(n),$$

and similarly for any μ_{2r} , as in 6.3.

This sketch of a proof serves as a model for a proof in the corresponding case of a variable trinomial chain. We notice the dependence of the method of proof on the order of

$$\mu_2^{(n)}.$$

7.4. We shall now generalise the preceding proof to the dependent case, and so consider a variable Markov chain with three states E_1, E_2, E_3 and such that the occurrence of one of these states results in the addition of x_1, x_2, x_3 respectively to the score. The proof follows the same general pattern as that of the proof we have already given in the

binomial case; but the extension to the trinomial case, and the use of the general increments x_1, x_2, x_3 which are necessary here, instead of the previous particular 0, 1, both introduce new features. The final generalisation, however, to a chain of k states with corresponding score increments x_1, \dots, x_k is almost immediate, and the only additional difficulty is one of notation.

A typical trinomial stochastic matrix is

$$P_j = \begin{bmatrix} p_1^{(j)} & p_2^{(j)} & p_3^{(j)} \\ q_1^{(j)} & q_2^{(j)} & q_3^{(j)} \\ r_1^{(j)} & r_2^{(j)} & r_3^{(j)} \end{bmatrix},$$

but for simplicity of writing we shall omit the suffixes j , and likewise when they occur in association with other variables, whenever their inclusion is not made necessary by the context. If the initial probability distribution is $[\alpha, \beta, \gamma]$, the c.f. after n trials is

$$[\alpha e^{itx_1}, \beta e^{itx_2}, \gamma e^{itx_3}] \left\{ \prod_{j=1}^n \begin{bmatrix} p_1^{(j)} e^{itx_1} & p_2^{(j)} e^{itx_2} & p_3^{(j)} e^{itx_3} \\ q_1^{(j)} e^{itx_1} & q_2^{(j)} e^{itx_2} & q_3^{(j)} e^{itx_3} \\ r_1^{(j)} e^{itx_1} & r_2^{(j)} e^{itx_2} & r_3^{(j)} e^{itx_3} \end{bmatrix} \right\} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

But at present the origin of score is not the mean; the effect of referring the distribution to the then current mean score \bar{x}_n as origin is to multiply the preceding c.f. by

$e^{-\pi_n it}$. It is convenient to utilize the particular nature of the matrices, namely, that their row sums are unity, by making the matrix transformation corresponding to that used previously in the binomial case: replace each matrix () by $H()H^{-1}$, where

$$H = \begin{bmatrix} 1 & \cdot & \cdot \\ -1 & 1 & \cdot \\ -1 & \cdot & 1 \end{bmatrix} \quad \text{and} \quad H^{-1} = \begin{bmatrix} 1 & \cdot & \cdot \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \end{bmatrix}.$$

Under this transformation a typical 'characteristic matrix' becomes

$$P(t) \equiv \begin{bmatrix} p_1 e^{itx_1} + p_2 e^{itx_2} + p_3 e^{itx_3} & , & p_2 e^{itx_2} & , & p_3 e^{itx_3} \\ (q_1 - p_1) e^{itx_1} + (q_2 - p_2) e^{itx_2} + (q_3 - p_3) e^{itx_3} & , & (q_2 - p_2) e^{itx_2} & , & (q_3 - p_3) e^{itx_3} \\ (r_1 - p_1) e^{itx_1} + (r_2 - p_2) e^{itx_2} + (r_3 - p_3) e^{itx_3} & , & (r_2 - p_2) e^{itx_2} & , & (r_3 - p_3) e^{itx_3} \end{bmatrix}.$$

The terms in this matrix can be expanded in powers of (it) . We shall find that we need not retain powers of (it) above the second so that, to this order, the matrix $P(t)$ is as shown overleaf. The definition of the various δ_{ab} is as shown there. In the previous binomial case we had, with a corresponding notation, $\delta_{21} = -\delta_{22} = -\delta$. This particularisation, however, tended to obscure the generalisation of the proof to the trinomial, and higher, cases.

We now have that the c.f. after n trials of the distribution referred to the current origin as mean is

$$P(t) \equiv \begin{bmatrix} 1 + (p_1 x_1 + p_2 x_2 + p_3 x_3)(t) + (p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2) \frac{(t)^2}{2!} & p_2 + p_2 x_2(t) + p_2 x_2^2 \frac{(t)^2}{2!} & p_3 + p_3 x_3(t) + p_3 x_3^2 \frac{(t)^2}{2!} \\ (\delta_{21} x_1 + \delta_{22} x_2 + \delta_{23} x_3)(t) + (\delta_{21} x_1^2 + \delta_{22} x_2^2 + \delta_{23} x_3^2) \frac{(t)^2}{2!} & \delta_{22} + \delta_{22} x_2(t) + \delta_{22} x_2^2 \frac{(t)^2}{2!} & \delta_{23} + \delta_{23} x_3(t) + \delta_{23} x_3^2 \frac{(t)^2}{2!} \\ (\delta_{31} x_1 + \delta_{32} x_2 + \delta_{33} x_3)(t) + (\delta_{31} x_1^2 + \delta_{32} x_2^2 + \delta_{33} x_3^2) \frac{(t)^2}{2!} & \delta_{32} + \delta_{32} x_2(t) + \delta_{32} x_2^2 \frac{(t)^2}{2!} & \delta_{33} + \delta_{33} x_3(t) + \delta_{33} x_3^2 \frac{(t)^2}{2!} \end{bmatrix}$$

where the δ 's are defined by

$$\begin{bmatrix} 1 & p_2 & p_3 \\ x_1 - p_1 & x_2 - p_2 & x_3 - p_3 \\ x_1 - p_1 & x_2 - p_2 & x_3 - p_3 \end{bmatrix} \equiv \begin{bmatrix} 1 & p_2 & p_3 \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix}.$$

$$\phi_n(t) = e^{-\bar{x}_n it} \left[\alpha e^{x_1 it} + \beta e^{x_2 it} + \gamma e^{x_3 it}, \quad \beta e^{x_2 it}, \quad \gamma e^{x_3 it} \right] \left\{ \prod_{j=1}^n P_j(t) \right\} \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

$$= \left[1 + \mu_2^{(n)} \frac{(it)^2}{2!} + \mu_3^{(n)} \frac{(it)^3}{3!} + \dots, \quad \nu_0^{(n)} + \nu_1^{(n)} (it) + \nu_2^{(n)} \frac{(it)^2}{2!} + \dots, \quad \omega_0^{(n)} + \omega_1^{(n)} (it) + \omega_2^{(n)} \frac{(it)^2}{2!} + \dots \right] \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

say. If we denote the row vector above by

$$\underline{\alpha}_n = [\alpha_n, \beta_n, \gamma_n]$$

then, because of the effect of the column vector $\{1 \cdot \cdot \cdot\}$ in post-multiplication, we have as a result of the transformation that the first element of the row vector

$$\alpha_n = 1 + \mu_2^{(n)} \frac{(it)^2}{2!} + \mu_3^{(n)} \frac{(it)^3}{3!} + \dots$$

gives the c.f. at this stage. We have to study the change in this c.f. on passing from the j^{th} to the $(j+1)^{\text{th}}$ stage, and to do this we consider the corresponding change in the vector $\underline{\alpha}_j$. We have the fundamental recurrence relation

$$\underline{\alpha}_{j+1} = e^{-\Delta_j it} \underline{\alpha}_j P_{j+1}(t),$$

where Δ_j is the increment to the mean on passing from the j^{th} to the $(j+1)^{\text{th}}$ stage. It will simplify the expression of these relations if we introduce a vector notation. Let

$$\underline{\delta}_2 = [\delta_{21}, \delta_{22}, \delta_{23}], \quad \underline{\delta}_3 = [\delta_{31}, \delta_{32}, \delta_{33}], \quad \underline{P} = [P_1, P_2, P_3]$$

and

$$\underline{x} = [x_1, x_2, x_3], \quad \underline{x}^2 = [x_1^2, x_2^2, x_3^2].$$

Then with ' \cdot ' denoting scalar product in the usual way we have, omitting powers of (it) above the second, that

$$P(t) = \left[\begin{array}{l} 1 + p \cdot x (it) + p \cdot x^2 \frac{(it)^2}{2!}, \\ \delta_2 \cdot x (it) + \delta_2 \cdot x^2 \frac{(it)^2}{2!}, \\ \delta_3 \cdot x (it) + \delta_3 \cdot x^2 \frac{(it)^2}{2!}, \end{array} \right. \quad \text{as before} \left. \right].$$

We now have from the recurrence relation that

$$\begin{aligned} \alpha_{j+1} &= \left\{ 1 - \Delta(it) + \Delta^2 \frac{(it)^2}{2!} - \dots \right\} \left\{ 1 + (it) [p \cdot x + v_0(\delta_2 \cdot x) + w_0(\delta_3 \cdot x)] \right. \\ &\quad \left. + \frac{(it)^2}{2!} [\mu_2 + p \cdot x^2 + 2v_1(\delta_2 \cdot x) + 2w_1(\delta_3 \cdot x) + v_0(\delta_2 \cdot x^2) + w_0(\delta_3 \cdot x^2)] + \dots \right. \\ &= \left. 1 + (it) \left[\mu_2 + \frac{p \cdot x}{1} + \frac{v_0(\delta_2 \cdot x)}{1} + w_0(\delta_3 \cdot x) - \Delta \right] \right. \\ &\quad \left. + \frac{(it)^2}{2!} \left\{ [\mu_2 + p \cdot x^2 + 2v_1(\delta_2 \cdot x) + 2w_1(\delta_3 \cdot x) + v_0(\delta_2 \cdot x^2) + w_0(\delta_3 \cdot x^2)] \right. \right. \\ &\quad \left. \left. - 2\Delta [p \cdot x + v_0(\delta_2 \cdot x) + w_0(\delta_3 \cdot x)] + \Delta^2 \right\} + \dots \right. \end{aligned}$$

Since the distribution of score is referred to the mean as origin, the coefficient of (it) here is zero so that

$$\Delta = p \cdot x + v_0(\delta_2 \cdot x) + w_0(\delta_3 \cdot x).$$

Using this result we have that $\Delta \mu_2^{(j)}$, the increment to μ_2 on passing from the j^{th} to the $(j+1)^{\text{th}}$ stage, is given by

$$\Delta \mu_2^{(j)} = 2v_1(\delta_2 \cdot x) + 2w_1(\delta_3 \cdot x) + v_0(\delta_2 \cdot x^2) + w_0(\delta_3 \cdot x^2) + p \cdot x^2 - \Delta^2.$$

7.5. So far the pattern of proof has been directly analogous to that in the binomial case and, corresponding to that case, the next step is to examine the order of $\mu_2^{(n)}$. We shall need in the completion of the proof the fact that $\mu_2^{(n)} = O(n)$ precisely. This fact is essential to the

method of proof, although for the score distribution to tend to a normal distribution the condition is not necessary.* It is in this discussion of the order of $\mu_2^{(n)}$ that we encounter the essential difficulty of the proof. The discussion of $\mu_2^{(n)}$ in the independent case of 7.2 showed that, in this case, $\mu_2^{(n)} = O(n)$ certainly if the score increments were fixed and distinct (or, more generally, if their differences are uniformly bounded above zero) and also the probabilities of occurrence of each event are uniformly positive. This same condition is also sufficient in the case of a constant chain~~***~~, and it is natural to expect that the corresponding conditions apply likewise in the more general case of a variable chain, certainly provided that (i) the P_j are uniformly positive (ii) the $x_1^{(j)}, x_2^{(j)}, x_3^{(j)}$ are not all equal and such that the smallest difference is uniformly positive. An example~~***~~ shows that if these conditions do not hold then $\mu_2^{(n)}$ can be bounded. For suppose that $x_1 = -1, x_2 = 0, x_3 = 1$ and

$$P_j = \begin{bmatrix} \cdot & 1 & \cdot \\ p_{21} & \cdot & p_{23} \\ p_{31} & \cdot & p_{33} \end{bmatrix}$$

* See, e.g., Bernstein, 40.

~~***~~ Fréchet, p. 84 or Doeblin, p. 86. Romanovsky does not discuss the conditions under which the variance is non-zero.

~~***~~ Doeblin, p. 90.

for all j . Then clearly the score oscillates between the bounds -1 and 1 whatever the initial state. We should expect that, with the conditions (i) and (ii) mentioned above, the least value of $\mu_2^{(n)}$ would occur when there is the greatest certainty of outcome. For instance, if we suppose that the score increments are $1, 0, 0$, corresponding to the distribution of the frequency of occurrence of a particular state, we should expect the result, analogous to the conjecture of 4.12, that the least value of $\mu_2^{(n)}$ occurs when

$$P_j = \begin{bmatrix} e & 1-2e & e \\ e & e & 1-2e \\ 1-2e & e & e \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} e & e & 1-2e \\ 1-2e & e & e \\ e & 1-2e & e \end{bmatrix}$$

for all j so that the states tend strongly to occur in cyclic order. These conjectures we have been unable to prove, and we shall proceed under the assumption that, under the conditions stated above, $\mu_2^{(n)} = O(n)$.

7.6. Again, from the recurrence relation,

$$\begin{aligned} \beta_{r+1} &= \left\{ 1 - \Delta(i\kappa) + \frac{\Delta^2(i\kappa)^2}{2!} - \dots \right\} \left\{ (p_2 + \delta_{22}v_0 + \delta_{32}\omega_0) \right. \\ &\quad \left. + (i\kappa) \left[(p_2 + \delta_{22}v_0 + \delta_{32}\omega_0)x_2 + \delta_{22}v_1 + \delta_{32}\omega_1 \right] \right. \\ &\quad \left. + \frac{(i\kappa)^2}{2!} \left[(p_2 + \delta_{22}v_0 + \delta_{32}\omega_0)x_2^2 + 2(\delta_{22}v_1 + \delta_{32}\omega_1)x_2 + (p_2\mu_2 + \delta_{22}v_2 + \delta_{32}\omega_2) \right] + \dots \right\} \\ &= (p_2 + \delta_{22}v_0 + \delta_{32}\omega_0) + (i\kappa) \left[(p_2 + \delta_{22}v_0 + \delta_{32}\omega_0)(x_2 - \Delta) + (\delta_{22}v_1 + \delta_{32}\omega_1) \right] \\ &\quad + \frac{(i\kappa)^2}{2!} \left[(p_2 + \delta_{22}v_0 + \delta_{32}\omega_0)(x_2 - \Delta)^2 + 2(\delta_{22}v_1 + \delta_{32}\omega_1)(x_2 - \Delta) + (p_2\mu_2 + \delta_{22}v_2 + \delta_{32}\omega_2) \right] + \dots \end{aligned}$$

from which we have the particular recurrence relations

$$v_0^{(r+1)} = p_2 + \delta_{22}v_0 + \delta_{32}\omega_0, \quad (1)$$

$$\begin{aligned} v_1^{(r+1)} &= (p_2 + \delta_{22}v_0 + \delta_{32}\omega_0)(x_2 - \Delta) + (\delta_{22}v_1 + \delta_{32}\omega_1) \\ &= \delta_{22}v_1 + \delta_{32}\omega_1 + (x_2 - \Delta)v_0^{(r+1)}, \end{aligned}$$

on using (1), and

$$\begin{aligned} v_2^{(r+1)} &= (p_2 + \delta_{22}v_0 + \delta_{32}\omega_0)(x_2 - \Delta)^2 + 2(\delta_{22}v_1 + \delta_{32}\omega_1)(x_2 - \Delta) \\ &\quad + p_2\mu_2 + \delta_{22}v_2 + \delta_{32}\omega_2. \end{aligned}$$

The corresponding results for $\omega_0^{(r+1)}$, $\omega_1^{(r+1)}$, $\omega_2^{(r+1)}$ are completely analogous.

7.7. We wish to show that, as $n \rightarrow \infty$, $\mu_{2m}^{(n)}$ is asymptotic to $(2m-1)(2m-3)\dots 3(\mu_2)^m$ and that the order of $\mu_{2m+1}^{(n)}$ is the same as, or no higher than, that of $\mu_{2m}^{(n)}$. We prove the first result as before; we show by induction that

$$\Delta \mu_{2m}^{(n)} = \binom{2m}{2} \mu_{2m-2} \Delta \mu_2^{(n)} + O(n^{m-2}).$$

We notice that, since $\mu_0 = 1$, the result is true for $m = 1$.

From the fundamental recurrence relation we have that

$\mu_{2r}^{(n+1)}$ is the coefficient of $(it)^{2r} / (2r)!$ in

$$\left\{ 1 - \Delta(it) + \Delta^2 \frac{(it)^2}{2!} - \dots \right\} \left\{ \mu_{2m-2} \frac{(it)^{2m-2}}{(2m-2)!} + \frac{(it)^{2m-1}}{(2m-1)!} \left[\mu_{2m-1} + (2m-1)(p \cdot x) \mu_{2m-2} \right. \right. \\ \left. \left. + (2m-1)(\delta_2 \cdot x) v_{2m-2} + (2m-1)(\delta_3 \cdot x) \omega_{2m-2} \right] \right. \\ \left. + \frac{(it)^{2m}}{(2m)!} \left[\mu_{2m} + 2m(p \cdot x) \mu_{2m-1} + 2m(\delta_2 \cdot x) v_{2m-1} + 2m(\delta_3 \cdot x) \omega_{2m-1} \right. \right. \\ \left. \left. + m(2m-1)(p \cdot x^2) \mu_{2m-2} + m(2m-1)(\delta_2 \cdot x^2) v_{2m-2} + m(2m-1)(\delta_3 \cdot x^2) \omega_{2m-2} \right] \right\}.$$

We have used here the result, to be proved later, that the orders of $v_r^{(n)}$ and $\omega_r^{(n)}$ are the same as, or not higher than, that of $\mu_r^{(n)}$; we have used this for values of r up to $2m - 3$. Then, using the result of 7.4 that

$$\Delta = p \cdot x = (\delta_2 \cdot x) v_0 + (\delta_3 \cdot x) \omega_0,$$

we have that

$$\begin{aligned} \Delta \mu_{2m} &= 2m (v_{2m-1} - v_0 \mu_{2m-1}) (\delta_2 \cdot x) + 2m (\omega_{2m-1} - \omega_0 \mu_{2m-1}) (\delta_3 \cdot x) \\ &+ m(2m-1) \left\{ (\rho \cdot x^2) \mu_{2m-2} + (\delta_2 \cdot x^2) v_{2m-2} + (\delta_3 \cdot x^2) \omega_{2m-2} \right\} \\ &- 2m(2m-1) \Delta^2 \mu_{2m-2} + m(2m-1) \Delta^2 \mu_{2m-2} \\ &+ 2m(2m-1) \left\{ (v_{2m-2} - v_0 \mu_{2m-2}) (\delta_2 \cdot x) + (\omega_{2m-2} - \omega_0 \mu_{2m-2}) (\delta_3 \cdot x) \right\} \\ &+ O(n^{m-2}). \end{aligned}$$

We have found above, in 7.4, that

$$\Delta \mu_2 = \rho \cdot x^2 + 2v_1 (\delta_2 \cdot x) + 2\omega_1 (\delta_3 \cdot x) + v_0 (\delta_2 \cdot x^2) + \omega_0 (\delta_3 \cdot x^2) - \Delta^2,$$

so the result that

$$\Delta \mu_{2m}^{(n)} = \binom{2m}{2} \mu_{2m-2} \cdot \Delta \mu_2^{(n)} + O(n^{m-2})$$

will follow if

$$\begin{aligned} &2m (v_{2m-1} - v_0 \mu_{2m-1}) (\delta_2 \cdot x) + 2m (\omega_{2m-1} - \omega_0 \mu_{2m-1}) (\delta_3 \cdot x) \\ &+ m(2m-1) \left\{ (\delta_2 \cdot x^2) v_{2m-2} + (\delta_3 \cdot x^2) \omega_{2m-2} \right\} \\ &+ 2m(2m-1) \left\{ (v_{2m-2} - v_0 \mu_{2m-2}) (\delta_2 \cdot x) + (\omega_{2m-2} - \omega_0 \mu_{2m-2}) (\delta_3 \cdot x) \right\} \\ &= m(2m-1) \left\{ 2v_1 (\delta_2 \cdot x) + 2\omega_1 (\delta_3 \cdot x) + v_0 (\delta_2 \cdot x^2) + \omega_0 (\delta_3 \cdot x^2) \right\} + O(n^{m-2}), \end{aligned}$$

that is, if

$$\begin{aligned} &2m \left\{ [(v_{2m-1} - v_0 \mu_{2m-1}) - (2m-1)v_1 \mu_{2m-2}] (\delta_2 \cdot x) \right. \\ &\quad \left. + [(\omega_{2m-1} - \omega_0 \mu_{2m-1}) - (2m-1)\omega_1 \mu_{2m-2}] (\delta_3 \cdot x) \right\} \\ &+ m(2m-1) \left\{ (v_{2m-2} - v_0 \mu_{2m-2}) (\delta_2 \cdot x) + (\omega_{2m-2} - \omega_0 \mu_{2m-2}) (\delta_3 \cdot x) \right\} = O(n^{m-2}). \end{aligned}$$

This is so if

(1)

$$v_{2r}^{(n)} - v_0^{(n)} \mu_{2r}^{(n)} = O(n^{r-1})$$

and

$$w_{2r}^{(n)} - w_0^{(n)} \mu_{2r}^{(n)} = O(n^{r-1}) ;$$

(2)

$$(v_{2r-1}^{(n)} - v_0^{(n)} \mu_{2r-1}^{(n)}) - (2r-1) v_1^{(n)} \mu_{2r-2}^{(n)} = O(n^{r-2})$$

and

$$(w_{2r-1}^{(n)} - w_0^{(n)} \mu_{2r-1}^{(n)}) - (2r-1) w_1^{(n)} \mu_{2r-2}^{(n)} = O(n^{r-2}) .$$

We need, to complete the proof above, that (1) is already proved for values of r up to $2m - 2$ only, and not up to $2m$. This detail is important in the sequence of the induction. Since, as a particular case of the next result, $v_0^{(n)} = O(1)$ and $v_1^{(n)} = O(1)$ it will follow from (1) and (2) that the orders of $v_r^{(n)}$ and $w_r^{(n)}$ do not exceed those of $\mu_r^{(n)}$. This fact we have used earlier in the proof for values of r up to $2m - 3$.

We consider (1), and hence the recurrence relations for v_{2r} and w_{2r} . We have that $v_{2r}^{(n)}$ is the coefficient of $(it)^r / (2r)!$ in

$$\{1\} \left[\mu_{2r} \frac{(it)^{2r}}{(2r)!}, v_{2r} \frac{(it)^{2r}}{(2r)!}, w_{2r} \frac{(it)^{2r}}{(2r)!} \right] \begin{bmatrix} p_2 \\ \delta_{22} \\ \delta_{32} \end{bmatrix} ,$$

where we have omitted all terms $O(n^{r-1})$. But

$$p_2 = v_0^{(n)} - v_0 \delta_{22} - w_0 \delta_{32}$$

so that

$$v_{2r}^{(jH)} - v_0^{(jH)} \mu_{2r}^{(j)} = \delta_{22} (v_{2r} - v_0 \mu_{2r}) + \delta_{32} (\omega_{2r} - \omega_0 \mu_{2r}) + O(n^{r-1})$$

which, since

$$\mu_{2r}^{(jH)} - \mu_{2r}^{(j)} = O(n^{r-1}),$$

We can write as

$$v_{2r}^{(jH)} - v_0^{(jH)} \mu_{2r}^{(jH)} = \delta_{22} (v_{2r} - v_0 \mu_{2r}) + \delta_{32} (\omega_{2r} - \omega_0 \mu_{2r}) + O(n^{r-1})$$

Similarly

$$\omega_{2r}^{(jH)} - \omega_0^{(jH)} \mu_{2r}^{(jH)} = \delta_{23} (v_{2r} - v_0 \mu_{2r}) + \delta_{33} (\omega_{2r} - \omega_0 \mu_{2r}) + O(n^{r-1}).$$

We wish to deduce from these two last equations that

$$v_{2r}^{(n)} - v_0^{(n)} \mu_{2r}^{(n)} = O(n^{r-1})$$

and

$$\omega_{2r}^{(n)} - \omega_0^{(n)} \mu_{2r}^{(n)} = O(n^{r-1}).$$

If we write

$$v_{2r}^{(j)} - v_0^{(j)} \mu_{2r}^{(j)} = x_j \quad \text{and} \quad \omega_{2r}^{(j)} - \omega_0^{(j)} \mu_{2r}^{(j)} = y_j$$

the equations can be written

$$x_{j+1} = \delta_{22} x_j + \delta_{32} y_j + O(n^{r-1})$$

and

$$y_{j+1} = \delta_{23} x_j + \delta_{33} y_j + O(n^{r-1})$$

or compactly in matrix notation as

$$\underline{x}_{j+1} = \underline{x}_j D_{j+1} + \underline{a}_{j+1}$$

where

$$\underline{x}_j = [x_j, y_j], \quad D_j = \begin{bmatrix} \delta_{22}^{(j)} & \delta_{32}^{(j)} \\ \delta_{23}^{(j)} & \delta_{33}^{(j)} \end{bmatrix}$$

and \underline{a}_j is a row vector whose elements are $O(n^{r-1})$: all elements less than Kn^{r-1} , say, for every j . Consider the $n+1$ such equations for $j = n, n-1, \dots, 2, 1, 0$. Multiply these equations by $D_{n+1}, D_n D_{n+1}, \dots, D_1 D_2 \dots D_{n+1}$ respectively and add. We then have \underline{x}_{n+1} expressed in terms of the D_i and a_i ; the result that $\underline{x}_n = O(n^{r-1})$ follows if

$$D_n + D_{n-1} D_n + \dots + D_1 D_2 \dots D_n = O(1).$$

Consider first the product D_{12} of two typical matrices D_1, D_2 . We can relate this product to P_{12} , the product of the two stochastic matrices P_1 and P_2 from which D_1 and D_2 are derived, for we have

$$P_{12} = P_1 P_2$$

so that

$$H P_{12} H^{-1} = H P_1 H^{-1} H P_2 H^{-1},$$

that is,

$$\begin{bmatrix} 1 & x & x \\ \cdot & & \\ \cdot & D_{12} & \end{bmatrix} = \begin{bmatrix} 1 & x & x \\ \cdot & & \\ \cdot & D_1 & \end{bmatrix} \begin{bmatrix} 1 & x & x \\ \cdot & & \\ \cdot & D_2 & \end{bmatrix}$$

and hence

$$D_{12} = D_1 D_2.$$

We now use the result of 1.6 that if the P_i are such that $P_i \geq e$ for all i then

$$\rho_{12} \leq (1-3e)\rho_2 \leq (1-3e)^2.$$

It follows that the greatest element of $D_{1,2}$ does not exceed $(1-3e)^2$ in modulus. Thus the greatest element of

$$D_n + D_{n-1} D_n + \dots + D_1 D_2 \dots D_n$$

is not greater than

$$\eta + \eta^2 + \dots + \eta^n$$

in modulus, where $\eta = 1 - 3e < 1$, and so does not exceed $\eta/(1-\eta)$, a fixed constant. Thus $\underline{x} = O(n^{r-1})$.

We now prove (2). We consider the recurrence relation for v_{2r-1} .

We have that $v_{2r-1}^{(j+1)}$ is the coefficient of $(it)^{2r-1}/(2r-1)!$ in

$$\{1 - \Delta(it)\} \left\{ (p_2 \mu_{2r-2} + \delta_{22} v_{2r-2} + \delta_{32} \omega_{2r-2}) \frac{(it)^{2r-2}}{(2r-2)!} + \frac{(it)^{2r-1}}{(2r-1)!} \left[(p_2 \mu_{2r-1} + \delta_{22} v_{2r-1} + \delta_{32} \omega_{2r-1}) + (2r-1)x_2 (p_2 \mu_{2r-2} + \delta_{22} \omega_{2r-2} + \delta_{32} \omega_{2r-2}) \right] \right\}$$

whence, using the result that

$$p_2 = v_0^{(j+1)} - v_0 \delta_{22} - \omega_0 \delta_{32},$$

we have that

$$v_{2r-1}^{(j+1)} - v_0^{(j+1)} \mu_{2r-1} = \delta_{22} (v_{2r-1} - v_0 \mu_{2r-1}) + \delta_{32} (\omega_{2r-1} - \omega_0 \mu_{2r-1}) + (2r-1)(x_2 - \Delta) v_0^{(j+1)} \mu_{2r-2} + O(n^{r-2})$$

since

$$v_{2r-2} - v_0 \mu_{2r-2} = O(n^{r-2})$$

and

$$w_{2r-2} - w_0 \mu_{2r-2} = O(n^{r-2})$$

But

$$v_1^{(j+1)} = \delta_{22} v_1 + \delta_{32} w_1 + (x_2 - \Delta) v_0^{(j+1)}$$

so that

$$\begin{aligned} & v_{2r-1}^{(j+1)} - v_0^{(j+1)} \mu_{2r-1} - (2r-1) v_1^{(j+1)} \mu_{2r-2} \\ &= \delta_{22} (v_{2r-1} - v_0 \mu_{2r-1} - (2r-1) v_1 \mu_{2r-2}) + \delta_{32} (w_{2r-1} - w_0 \mu_{2r-1} - (2r-1) w_1 \mu_{2r-2}) \\ & \qquad \qquad \qquad + O(n^{r-2}) \end{aligned}$$

which we can write as

$$\begin{aligned} v_{2r-1}^{(j+1)} - v_0^{(j+1)} \mu_{2r-1} - (2r-1) v_1^{(j+1)} \mu_{2r-2} \\ = \text{as above,} \end{aligned}$$

since

$$\Delta \mu_{2r-2}^{(w)}, \Delta \mu_{2r-1}^{(w)} = O(n^{r-2}).$$

The corresponding result for $w_{2r-1}^{(j+1)} - w_0^{(j+1)} \mu_{2r-1} - (2r-1) w_1^{(j+1)} \mu_{2r-2}$ follows similarly. These two recurrence relations can be combined into a single matrix relation in form identical with the one above. The rest of the proof then follows as before. This completes the proof that

$$\Delta \mu_{2r} = \binom{2r}{2} \mu_{2r-2} \cdot \Delta \mu_2 + O(n^{r-2}).$$

We now prove that $\mu_{2r+1} = O(n^r)$. We have that $\mu_{2r+1}^{(j)}$ is the coefficient of $(it)^{2r+1} / (2r+1)!$ in $\{1 - \Delta(x)\} \left\{ u_{2r} \frac{(it)^{2r}}{(2r)!} + \frac{(it)^{2r+1}}{(2r+1)!} \left[\mu_{2r+1} + (2r+1)\mu_{2r}(p \cdot x) + (2r+1)v_{2r}(\delta_2 \cdot x) + (2r+1)w_{2r}(\delta_3 \cdot x) \right] \right\}$

so that, using

$$\Delta - p \cdot x = v_0(\delta_2 \cdot x) + w_0(\delta_3 \cdot x)$$

we have that

$$\Delta \mu_{2r+1} = (2r+1)(v_{2r} - v_0 \mu_{2r})(\delta_2 \cdot x) + (2r+1)(w_{2r} - w_0 \mu_{2r})(\delta_3 \cdot x).$$

The result now follows since

$$v_{2r} - v_0 \mu_{2r} = O(n^{r-1})$$

and

$$w_{2r} - w_0 \mu_{2r} = O(n^{r-1}).$$

The proof is now complete.

7.8. We now extend the theorem, just proved for a trinomial chain, to a chain with any finite number of states. As we have already remarked the difficulty is only one of notation.

A typical matrix of transition probabilities is

$$P_j = [p_{\theta\phi}^{(j)}]$$

where $\theta, \phi = 1, \dots, k$ and we suppose as before that $P_j \geq \epsilon$ for all j . If the initial probability distribution is $\underline{w} = [w_1, \dots, w_k]$ and the score increments resulting on the occurrence of the states E_1, \dots, E_k are x_1, \dots, x_k then the c.f. after n trials is

$$w(t) \prod_{j=1}^k P_j(t) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

where

$$P_j(t) = \left[p_{\theta\phi}^{(j)} e^{itx_\phi} \right]$$

and

$$w(t) = \left[w_\phi e^{itx_\phi} \right].$$

We use the $H()H^{-1}$ transformation once again, where now

$$H = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ -1 & & \ddots & & \\ \vdots & & & \ddots & \\ -1 & & & & 1 \end{bmatrix},$$

and find as before that the c.f. of the distribution after n trials, and referred to the then current mean as origin, is

$$\underline{\alpha}^{(n)} \{1, 0, \dots, 0\}$$

where

$$\underline{\alpha}^{(n)} = \left[\alpha_1^{(n)}, \dots, \alpha_k^{(n)} \right]$$

and $\underline{\alpha}^{(n)}$ satisfies the recurrence relation

$$\underline{\alpha}^{(j+1)} = e^{-\Delta_j it} \underline{\alpha}^{(j)} P_{j+1}(t).$$

Here Δ_j is the increment to the mean on transition from the j^{th} to the $(j+1)^{\text{th}}$ stage and $P_j(t)$ is the obvious generalisation to a chain of k states of the particular case of 7.4 in which $k = 3$.

The elements of $\underline{\alpha}^{(n)}$ can be expressed as power series in (it) ; we write

$$\alpha_{\phi}^{(n)} = \mu_{\phi,0}^{(n)} + \mu_{\phi,1}^{(n)} (it) + \mu_{\phi,2}^{(n)} \frac{(it)^2}{2!} + \dots$$

Thus the term

$$v_0^{(n)} + v_1^{(n)} (it) + v_2^{(n)} \frac{(it)^2}{2!} + \dots$$

of 7.4 would now be written

$$\mu_{2,0}^{(n)} + \mu_{2,1}^{(n)} (it) + \mu_{2,2}^{(n)} \frac{(it)^2}{2!} + \dots,$$

and the term

$$1 + \mu_2^{(n)} \frac{(it)^2}{2!} + \mu_3^{(n)} \frac{(it)^3}{3!} + \dots$$

would be written

$$\mu_{1,0}^{(n)} + \mu_{1,1}^{(n)} (it) + \mu_{1,2}^{(n)} \frac{(it)^2}{2!} + \dots$$

wherein

$$\mu_{1,0}^{(n)} = 1 \quad \text{and} \quad \mu_{1,1}^{(n)} = 0$$

for all n . These latter results arise likewise in the present general case, for they are consequences of the fact that once again the element $\alpha_1^{(n)}$ of $\underline{\alpha}^{(n)}$ is the c.f. of the distribution referred to the current mean as origin.

The generalisation of the results from the trinomial to the general case is now apparent from the use of a vector notation in 7.4 and later so that, on making the equivalent assumption about the order of $\mu_{1,2}^{(n)}$, the proof proceeds as before. In particular the matrix results of 7.7 generalise immediately; the matrices D_j are now of order $k-1$ instead of 2 and, correspondingly, $\eta = 1 - ke$.

A further generalisation arises from the replacement of the constant score increments x_1, \dots, x_k by the increments $x_1^{(j)}, \dots, x_k^{(j)}$ at the j^{th} stage and variable from stage to stage. The proof is essentially unchanged.

REFERENCES.

1. A.C. Aitken - Lectures on linear operations in probability
(Edinburgh, 1950-51, unpublished.)
2. M.S. Bartlett - An Introduction to Stochastic Processes
(Cambridge University Press, 1955.)
3. S. Bernstein - Sur l'extension du théorème du calcul des
probabilités aux sommes de quantités dépendantes
(Mathematische Annalen, 97 (1927), 1 - 59)
4. A. Brauer - Limits for the characteristic roots of a
matrix. IV: Applications to stochastic matrices.
(Duke Mathematical Journal, 19 (1952), 75-91).
5. T.J.I'a. Bromwich - An Introduction to the Theory of
Infinite Series (Macmillan, 1908).
6. H. Cramér - Mathematical Methods of Statistics
(Princeton University Press, 1946).
7. W. Doeblin - Sur les propriétés asymptotiques de mouve-
ment régis par certains types de chaînes simples
(Bulletin Mathématique de la Société Roumaine des
Sciences, 39 (1) (1937), 57 - 115).
8. J.L. Doob - Stochastic Processes (New York, 1953).
9. M. Fréchet - Recherches Théoriques Modernes sur les
Probabilités, Second Livre (Gauthier-Villars,
Paris, 1938).

REFERENCES (Contd).

10. J. Hajnal - The ergodic properties of non-homogeneous finite Markov chains (Proceedings of the Cambridge Philosophical Society, 52 (1956), 67 - 77).
 11. W. Ledermann - On the asymptotic probability distribution for certain Markoff processes (Proceedings of the Cambridge Philosophical Society, 46 (1950), 581 - 594).
 12. A. Markov - On the extension of the limit theorems of probability to the sum of values in a chain (in Russian) (Mémoires de l'Académie Impériale des Sciences de St.-Petersbourg, VIII, 22, No. 9, 1908).
 13. A. Markov - Recherches sur un cas remarquable d'épreuves dépendantes (Acta Mathematica, 33 (1910), 87 - 104).
 14. A. Markov - Researches on the general case of a chain of trials (in Russian) (Mémoires de l'Académie Impériale des Sciences de St.-Petersbourg VIII, 25, No. 3 (1910)).
 15. A. Markov - Wahrscheinlichkeitsrechnung (Leipzig, 1912).
 16. V. Romanovsky - Recherches sur les Chaînes de Markoff (Acta Mathematica, 66 (1936), 147-251).
 17. J.V. Uspensky - Introduction to Mathematical Probability (McGraw Hill, New York, 1937).
-