A STUDY OF A RANDOM-MATING POPULATION OF FIXED SIZE

by

Samuel Roscoe Knox, A.B., A.M.

Thesis submitted to the Graduate Faculty of the
Virginia Polytechnic Institute
in candidacy for the degree of

DOCTOR OF PHILOSOPHY
in
Statistics

May, 1962

Blacksburg, Virginia

TABLE OF CONTENTS

CHAPTER PAGE

I. INTRODUCTION 5

PART I: ABSORPTION IN A GENETIC POPULATION MODEL

II. INTRODUCTORY SUMMARY 8
- 2.1. Evolution 8
- 2.2. The population model 9
- 2.3. Review and summary 11

III. THE EIGENVECTOR APPROACH 14
- 3.1. Distribution theory 14
- 3.2. Blocking transformation 18
- 3.3. Triangular transformation 34
- 3.4. Some numerical examples and general results 37

IV. THE MOMENT APPROACH 66
- 4.1. Alternate method 66
- 4.2. Moment theory 66
- 4.3. Numerical results 69

V. THE DIFFUSION PROCESS 75
- 5.1. Probability of absorption 75
- 5.2. Moments 77
- 5.3. Numerical results and comparisons . 79

VI. INTERPRETATIONS 84
- 6.1. From moments and graphs 84
- 6.2. From diffusion approximations . . . 84

PART II: A METHOD OF MATRIX INVERSION

VII. DESCRIPTION OF THE METHOD 90
- 7.1. Introduction 90
- 7.2. Determination of the adjoint and inverse 92

TABLE OF CONTENTS

VIII. APPLICATION OF THE METHOD 97
  8.1. Some examples 97
  8.2. Some comparisons 103

IX. SUMMARY 106

X. ACKNOWLEDGEMENTS 109

XI. BIBLIOGRAPHY 110

XII. VITA 112

TABLES AND FIGURES

Figure 1. The probability function $S_1^{(t)}$, M=2 46

Figure 2. The probability function $S_1^{(t)} = S_2^{(t)}$, M=3 . . 47

Figure 3. The probability function $S_1^{(t)} = S_3^{(t)}$, M=4 . . 48

Figure 4. The probability function $S_2^{(t)}$, M=4 49

Figure 5. The probability function $S_1^{(t)} = S_4^{(t)}$, M=5 . . 50

Figure 6. The probability function $S_2^{(t)} = S_3^{(t)}$, M=5 . . 51

Figure 7. The probability function $S_1^{(t)} = S_5^{(t)}$, M=6 . . 52

Figure 8. The probability function $S_2^{(t)} = S_4^{(t)}$, M=6 . . 53

Figure 9. The probability function $S_3^{(t)}$, M=6 54

Figure 10. The probability function $S_1^{(t)} = S_6^{(t)}$, M=7 . . 55

Figure 11. The probability function $S_2^{(t)} = S_5^{(t)}$, M=7 . . 56

Figure 12. The probability function $S_3^{(t)} = S_4^{(t)}$, M=7 . . 57

Figure 13. The probability function $S_1^{(t)} = S_7^{(t)}$, M=8 . . 58

Figure 14. The probability function $S_2^{(t)} = S_6^{(t)}$, M=8 . . 59

Figure 15. The probability function $S_3^{(t)} = S_5^{(t)}$, M=8 . . 60

TABLE OF CONTENTS

Figure 16. The probability function $S_4^{(t)}$, M=8 61

Figure 17. The probability function $S_1^{(t)} = S_8^{(t)}$, M=9 . . 62

Figure 18. The probability function $S_2^{(t)} = S_7^{(t)}$, M=9 . . 63

Figure 19. The probability function $S_3^{(t)} = S_6^{(t)}$, M=9 .. 64

Figure 20. The probability function $S_4^{(t)} = S_5^{(t)}$, M=9 . . 65

Table 4.1. Means and variances of time to absorption . 72

Table 5.1. Diffusion approximations and percent error . 79

Table 7.1. Schematic of the pivotal condensation method for evaluating a fourth-order determinant 91

Table 7.2. Schematic for determining the adjoint and inverse of a fifth-order square matrix . . . 94

Table 8.1. Illustration for a third-order matrix . . . 98

Table 8.2. Illustration for a fourth-order matrix . . . 99

Table 8.3. Illustration for a fifth-order matrix . . . 100

Table 8.4. Illustration for a symmetric matrix 102

Table 8.5. Comparison of adjoint methods 103

Table 8.6. Comparison of inverse methods 104

I. INTRODUCTION

This thesis consists of two parts. The first part is devoted to a study of absorption in a genetic population model and the second part to a new method of matrix inversion.

Since the introduction of the basic concepts of Markov chains in 1907 by A. A. Markov the theory has been developed by a number of leading statisticians and mathematicians. It is only in comparatively recent times, however, that the importance of Markov chain theory to the biological sciences has become recognized. The genetic population considered in Part I is one which can be formulated in terms of a Markov chain with two absorbing states. In such a finite chain with discrete time and discrete states the probability is unity that one or the other of the absorbing states will ultimately be reached. The average time required for the chain to first reach an absorbing state is called the mean first passage time and it is with the first passage time problem that we are primarily concerned. Three approaches are used in an effort to resolve this problem.

One of the important mathematical tools applied in the study of first passage time problems is the inversion of a matrix. Numerous computational methods for determining the inverse of a matrix have been developed but there does not seem to be an overall best method, the advantages and disadvantages of each method being determined by the order of and the restrictions on the matrix involved. In the second part of this study a new procedure, especially applicable to the desk calculator, for determining the adjugate or adjoint of small order matrices is developed. This method is then applied to some problems encountered in the first part.

PART I

ABSORPTION IN A GENETIC POPULATION MODEL

II. INTRODUCTORY SUMMARY

2.1. Evolution. Organic evolution may be defined as a phenomenon described by the theory that plants and animals now living are the modified descendants of somewhat different plants and animals which lived in the past. In turn, these ancestors are thought of as being descendants of predecessors which differed from them, and so on, back to a beginning whose details are unknown at the present.

Living things possess within themselves the mechanisms of their own changes, those units of heredity which are called genes. The hereditary units determine, in large part, the characteristics of an organism, and they form the principal hereditary link between one generation and the next. To a very large extent the characteristics of an offspring are determined by the genes which it receives from its parents.

The genetic properties of a population are influenced in the process of transmission of genes from one generation to the next by a number of agencies, mainly migration, mutation, selection, population size, and random change or genetic drift. Evolution is a result of changes of the gene frequencies in a population due to some or all of these factors. If, in the absence of migration and mutation, the population reaches the state of fixation or absorption, that is, if the population consists of only one gene type, then evolution will come to a stop. The aim of this study is to find the stochastic behavior of the time to a state of absorption for a specific model.

2.2. <u>The population model</u>. Various models of birth and death or branching processes have been studied or are now being considered. These include the cases in which the individuals act independently and hence the population size is the random variable, and those in which the population size is constant and the gene frequency the random variable. Evolutionary agencies may be introduced in the model, but in many of the cases the mathematical treatment is much too complicated and one is forced to introduce certain simplifications.

The population to be considered in this study is one which is kept constant in size by the selection of M individuals in each successive generation. In addition to the restriction of constant size, the model considered is further simplified by assuming that it consists of two types of individuals, representing genes, which will be called A- and a-genes. Thus we only need study the univariate population of A-genes and are not led to the difficulties involved in the study of a bivariate branching process. This is not too serious a restriction, however, for consider the following. Assume a diploid model with k, M-k-h, and h individuals of genotypes aa, Aa, and AA, respectively. The proportion of a-genes is $\frac{1}{2}\,[1 + \frac{k-h}{M}\,] = p$ and the proportion of A-genes is

$\frac{1}{2}[1 - \frac{k-h}{M}] = q$. The probabilities of the birth of one individual under random choice of two parents being of genotypes aa, Aa, or AA are p^2, $2pq$, and q^2, respectively, and thus the probability that the next generation consists of k_1 individuals of genotype aa and h_1 individuals of genotype AA is given by

$$(2.2.1) \qquad \Pr\{k_1, h_1\} = \frac{M!}{k_1!(M-k_1-h_1)!h_1!}(p^2)^{k_1} (2pq)^{M-k_1-h_1}(q^2)^{h_1}.$$

The joint moment-generating function is then

$$E[e^{\theta k_1 + \phi h_1}] = (p^2e^\theta + 2pq + q^2e^\phi)^M.$$

The proportion of a-genes in the new generation is $\frac{1}{2}[1 + \frac{k_1-h_1}{M}]$ and the number of such genes is $M+k_1-h_1$ with moment-generating function

$$E[e^{\theta(M+k_1-h_1)}] = e^{\theta M}(p^2e^\theta + 2pq + q^2e^{-\theta})^M$$
$$= (pe^\theta + q)^{2M}$$

and therefore $M+k_1-h_1$ has the binomial distribution. Thus we conclude that the haploid model using binomial sampling on genes a and A results in the same distribution for gene numbers as the diploid model. If the haploid model is solved, the distribution for the diploid case can be deduced by means of (2.2.1).

If in the t-th generation, $t = 0,1,2,\ldots,$ there are i A-genes (and, of course, M-i a-genes) then we say that the population is at time t in state i ($0 \leq i \leq M$). Assuming random mating, the composition of the following generation is determined by M Bernoulli trials in which the A-gene has survival probability equal to $\frac{i}{M}$. We have, therefore, an absorbing Markov chain with transition probabilities

$$(2.2.2) \qquad p_{ij} = \binom{M}{j} \left(\frac{i}{M}\right)^{j} \left(1 - \frac{i}{M}\right)^{M-j}, \quad i,j = 0,1,2,\ldots,M.$$

2.3. Review and summary. Formulation of the problem in terms of Markov chains is due to Gustave Malécot [14] who also obtained an approximation to the largest non-trivial eigenvalue of the matrix $P = [p_{ij}]$. All stochastic matrices have unity as a trivial eigenvalue, in the present case this is a root of multiplicity two, and all others are of modulus less than or equal to one. The largest non-trivial root is used to judge the speed with which fixation or absorption is reached, that is, if p_t is the probability of no fixation within t generations, then p_{t+1}/p_t is approximately equal to this eigenvalue. The approach of Malécot made use of the following relation which may be found in Recherches Modernes sur le Calcul des Probabilitiés, II, by Fréchet. If a square matrix is such that all elements are positive and if the matrix has column sums all equal to the same constant, say k, then k is an eigenvalue of the matrix and all other eigenvalues are of modulus less than or equal to k. Consider then the sub-matrix Q of the transition matrix P defined as follows,

$$(2.3.1) \qquad Q = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1,M-1} \\ p_{21} & p_{22} & \cdots & p_{2,M-1} \\ \vdots & \vdots & & \vdots \\ p_{M-1,1} & p_{M-1,2} & \cdots & p_{M-1,M-1} \end{bmatrix}$$

If M is quite large, then

$$\begin{aligned} \sum_{i=1}^{M-1} \binom{M}{j} \left(\frac{i}{M}\right)^j \left(1-\frac{i}{M}\right)^{M-j} &= \binom{M}{j} \cdot M \sum_{i=1}^{M} \left(\frac{i}{M}\right)^j \\ &\qquad \cdot \left(1-\frac{i}{M}\right)^{M-j} \cdot \frac{1}{M} \\ &\doteq \binom{M}{j} \cdot M \int_0^1 x^j (1-x)^{M-j} dx \\ &= \binom{M}{j} \cdot M \cdot B(j+1,\ M-j+1) \\ &= \frac{M}{M+1} \, . \end{aligned}$$

That is, the column sums of Q are essentially the constant $\frac{M}{M+1}$. Hence the largest non-trivial eigenvalue of P is approximately this value. This is a remarkably good approximation to the exact value $\frac{M-1}{M}$ and agrees with the results of R. A. Fisher [9] and Sewall Wright [19] who earlier had obtained estimates of the ratio of decay without the use of Markov chain theory.

Extensive work on this problem has been done by William

Feller. In [6] he determined that if the A-genes and a-genes were in the proportion i:M-i, that is, if the population were in state i, then the probability of absorption in state 0 would be $1-\frac{i}{M}$, and in later work, [7], obtained the exact values of all eigenvalues of P. Further attention will be given the work of Feller in this area in the following chapters.

In Chapters III and IV of this study two alternative approaches are given using well known techniques for the theoretical determination of absorption time properties. Although the exact distribution and the moments of time for absorption have not yet been obtained for the general case, results which should prove of value in the advancement of the problem have been determined and exact results for specific values of M have been obtained by the use of both methods.

Relatively small populations require discrete mathematical models, but large populations allow application of a continuous approximation. This leads to diffusion type processes. In Chapter V, the Fokker-Planck diffusion equation is used to obtain approximate results and these are compared with the numerical results obtained in Chapters III and IV in order to ascertain the accuracy of the diffusion approximation.

III. THE EIGENVECTOR APPROACH

3.1. Distribution theory. One approach to the study of the model as given in Section 2.2 uses a method as described in a paper by G. A. Watterson [17]. The method as given by Watterson may be adapted to the particular model under consideration in the following manner.

Let P be the matrix of elements p_{ij} as defined by (2.2.2). The elements of P^t, say $p_{ij}^{(t)}$, are then the t-step transition probabilities. From (2.2.2) it is seen that 0 and M are absorbing states, and hence we have,

$$\begin{aligned} &p_{00} = p_{MM} = 1, \\ (3.1.1)\qquad &p_{01} = p_{02} = \cdots = p_{0M} = p_{M0} = \cdots = p_{M,M-1} = 0. \end{aligned}$$

Let T_i be the time taken for the chain to first reach one or the other absorbing state given i as the initial state, and write $s_i^{(t)}$ for the probability that $T_i = t$. Thus, it follows that

$$(3.1.2)\qquad s_i^{(t)} = p_{i0}^{(t)} + p_{iM}^{(t)} - p_{i0}^{(t-1)} - p_{iM}^{(t-1)}, \quad t= 1,2,\ldots,$$

but in particular,

$$(3.1.3)\qquad s_0^{(0)} = s_M^{(0)} = 1, \quad s_0^{(t)} = s_M^{(t)} = 0, \ 1 \leq t$$

and for $i \neq 0$ or M,

$$(3.1.4)\qquad s_i^{(0)} = 0, \quad s_i^{(1)} = p_{i0} + p_{iM}.$$

Using the convention of underscoring a letter to denote a column vector, let us write

$$\underline{s}^{(t)} = \begin{bmatrix} s_0^{(t)} \\ s_1^{(t)} \\ \vdots \\ s_M^{(t)} \end{bmatrix}.$$

From (3.1.3) and (3.1.4) we have

$$(3.1.5) \qquad \underline{s}^{(1)} = \begin{bmatrix} 0 \\ p_{10} + p_{1M} \\ p_{20} + p_{2M} \\ \vdots \\ p_{M-1,0} + p_{M-1,M} \\ 0 \end{bmatrix},$$

and then (3.1.2) gives

$$(3.1.6) \qquad \underline{s}^{(t)} = (P^t - P^{t-1}) \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$= P^{t-1}\underline{s}^{(1)}, \ t = 1,2,3,\ldots .$$

The calculation of absorption probabilities by (3.1.6) can be simplified if the eigenvalues and eigenvectors of P are known. Knowing these, one can proceed as follows. Let λ_j be the j-th eigenvalue of P and $\underline{K}_j$ the corresponding post-eigenvector. Then

$$P\underline{K}_j = \underline{K}_j \lambda_j, \quad j = 0,1,\ldots,M,$$

that is,

$$PK = KD_\lambda$$

where

$$K = (\underline{K}_0, \underline{K}_1, \ldots, \underline{K}_M), \; D_\lambda = (\lambda_j \delta_{ij}).$$

The columns of K are the post-eigenvectors and the rows of K^{-1} are the pre-eigenvectors, thus we have

$$P = KD_\lambda K^{-1}$$

or, more generally

$$P^{t-1} = KD_\lambda^{t-1}K^{-1}, \; t = 1,2,\ldots, \tag{3.1.7}$$

where D_λ^{t-1} is a diagonal matrix with elements λ_j^{t-1} ($j = 0,1,\ldots,M$) in the diagonal. Upon substituting (3.1.7) into (3.1.6) one obtains

$$\underline{s}^{(t)} = KD_\lambda^{t-1}K^{-1}\underline{s}^{(1)}, \; t = 1,2,\ldots, \tag{3.1.8}$$

and, at least in theory, (3.1.8) gives the distributions we seek.

The above approach is suggested by results which were obtained by Feller [7]. In this article, Feller obtains the exact values of all eigenvalues of P by the following method. Let $x_{j,r}$ denote the j-th element in the post-eigenvector of P corresponding to λ_r. We wish to show that the system of linear equations

$$\sum_{j=0}^{M} p_{ij} x_{j,r} = \lambda_r x_{i,r}, \quad i = 0,1,\dots,M \tag{3.1.9}$$

admits a non-trivial solution for all i. Let us assume that $x_{j,r}$ can always be written as a polynomial in j of degree at most r, that is, using factorial power notation with $j^{(r)} = j(j-1) \dots (j-r+1)$,

$$\begin{aligned} x_{j,r} &= a_0 j^{(r)} + a_1 j^{(r-1)} + a_2 j^{(r-2)} \\ &\quad \dots + a_{r-1} j^{(1)} + a_r \\ &= \sum_{h=0}^{r} a_h j^{(r-h)}. \end{aligned} \tag{3.1.10}$$

Our problem is then equivalent to showing that the system

$$\sum_{j=0}^{M} p_{ij} \sum_{h=0}^{r} a_h j^{(r-h)} = \lambda_r \sum_{h=0}^{r} a_h i^{(r-h)} \tag{3.1.11}$$

possesses a non-trivial solution for the a's. Now

$$\begin{aligned} \sum_{j=0}^{M} p_{ij} \sum_{h=0}^{r} a_h j^{(r-h)} &= \sum_{h=0}^{r} a_h \sum_{j=0}^{M} \binom{M}{j} \left(\frac{i}{M}\right)^j \\ &\qquad \left(1-\frac{i}{M}\right)^{M-j} j^{(r-h)} \\ &= \sum_{h=0}^{r} a_h M^{(r-h)} \left(\frac{i}{M}\right)^{r-h}, \end{aligned}$$

by the well-known formula for factorial moments of a binomial. Thus the system reduces to

$$\sum_{h=0}^{r} a_h M^{(r-h)} \left(\frac{i}{M}\right)^{r-h} = \lambda_r \sum_{h=0}^{r} a_h i^{(r-h)}. \tag{3.1.12}$$

By assigning an arbitrary non-zero value to a_0 and equating coefficients of i, the remaining r-1 coefficients can be determined in terms of a_0, and thus we have proved the existence of a non-trivial solution for the system. An immediate consequence of the above is that, by comparing the coefficients of i^r on both sides of (3.1.12), we have

$$\lambda_r = \frac{M^{(r)}}{M^r} . \tag{3.1.13}$$

Thus Feller obtains the eigenvalues, and, at least in theory, obtains the corresponding eigenvectors.

A simplified derivation of the eigenvalues occurs in a later section of this chapter. Also, as a consequence of some theorems which follow, more specific conclusions can be made concerning the general form of the eigenvectors and their calculation.

3.2. <u>Blocking transformation</u>. Consider the orthogonal matrix C of order M+1 with elements defined as follows:

$$c_{ij} = \begin{cases} 1/\sqrt{2} & \text{for } i = j \neq M/2 \\ 1 & \text{for } i = j = M/2 \\ 1/\sqrt{2} & \text{for } i + j = M, i<j \\ -1/\sqrt{2} & \text{for } i + j = M, j<i \\ 0 & \text{otherwise} \end{cases} , \tag{3.2.1}$$

that is, if M is odd,

$$C = \frac{1}{\sqrt{2}}\begin{bmatrix} I & J \\ -J & I \end{bmatrix},$$

and, if M is even,

$$C = \frac{1}{\sqrt{2}} \begin{bmatrix} I & \underline{0} & J \\ \underline{0}' & \sqrt{2} & \underline{0}' \\ -J & \underline{0} & I \end{bmatrix},$$

where I is the identity matrix and J is the matrix with ones in the secondary diagonal and zeros elsewhere. Define $R = C^{-1}PC$ and let us determine the general element r_{ij}. Since C is orthogonal, we have

$$\begin{aligned} r_{ij} &= \sum_{h=0}^{M} c^{-1}_{ih} \sum_{k=0}^{M} p_{hk} c_{kj} \\ &= \sum_{h=0}^{M} \sum_{k=0}^{M} c_{hi} p_{hk} c_{kj}. \end{aligned} \tag{3.2.2}$$

First we consider those cases where neither i nor j is equal to $\frac{M}{2}$ if M is even, noting that $c_{hi}c_{kj} \neq 0$ only when h=i or M-i and k=j or M-j. Thus (3.2.2) becomes

$$\begin{aligned} r_{ij} &= c_{ii}c_{jj}p_{ij} + c_{ii}c_{M-j,j}p_{i,M-i} \\ &\quad + c_{M-i,i}c_{jj}p_{M-i,j} + c_{M-i,i}c_{M-j,j}p_{M-i,M-j} \\ &= (c_{ii}c_{jj} + c_{M-i,i}c_{M-j,j})p_{ij} + (c_{ii}c_{M-j,j} \\ &\qquad + c_{M-i,i}c_{jj})p_{i,M-j}, \end{aligned} \tag{3.2.3}$$

since

$$\begin{aligned} p_{ij} &= \binom{M}{j}\left(\frac{i}{M}\right)^{j}\left(1-\frac{i}{M}\right)^{M-j} \\ &= \binom{M}{M-j}\left(\frac{M-i}{M}\right)^{M-j}\left(1-\frac{M-i}{M}\right)^{M-(M-j)} \\ &= p_{M-i,M-j}, \end{aligned}$$

and in particular, if M is even, $p_{\frac{M}{2},j} = p_{\frac{M}{2},M-j}$ and $p_{i,\frac{M}{2}} = p_{M-i,\frac{M}{2}}$. Then, using the values for the c_{ij} given by (3.2.1) in (3.2.3), we have, for $i < \frac{M}{2}$, $j < \frac{M}{2}$,

$$\begin{aligned} r_{ij} &= [(\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}}) + (-\tfrac{1}{\sqrt{2}})(-\tfrac{1}{\sqrt{2}})]p_{ij} \\ &\quad + [(\tfrac{1}{\sqrt{2}})(-\tfrac{1}{\sqrt{2}}) + (-\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}})]p_{i,M-j} \\ &= p_{ij} - p_{i,M-j} \\ &= \binom{M}{j}[(\tfrac{i}{M})^j(1-\tfrac{i}{M})^{M-j} - (\tfrac{i}{M})^{M-j}(1-\tfrac{i}{M})^j]. \end{aligned}$$

Similarly, for $i < \frac{M}{2}$, $\frac{M}{2} < j$, we have

$$\begin{aligned} r_{ij} &= [(\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}}) + (-\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}})]p_{ij} \\ &\quad + [(\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}}) + (-\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}})]p_{i,M-j} \\ &= 0 \end{aligned}$$

and, for $\frac{M}{2} < i$, $j < \frac{M}{2}$, we have

$$\begin{aligned} r_{ij} &= [(\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}}) + (\tfrac{1}{\sqrt{2}})(-\tfrac{1}{\sqrt{2}})]p_{ij} \\ &\quad + [(\tfrac{1}{\sqrt{2}})(-\tfrac{1}{\sqrt{2}}) + (\tfrac{1}{\sqrt{2}})(\tfrac{1}{\sqrt{2}})]p_{i,M-j} \\ &= 0, \end{aligned}$$

and, for $\frac{M}{2} < i$, $\frac{M}{2} < j$, we have

$$r_{ij} = [(\frac{1}{\sqrt{2}})(\frac{1}{\sqrt{2}}) + (\frac{1}{\sqrt{2}})(\frac{1}{\sqrt{2}})]p_{ij}$$

$$+ [(\frac{1}{\sqrt{2}})(\frac{1}{\sqrt{2}}) + (\frac{1}{\sqrt{2}})(\frac{1}{\sqrt{2}})]p_{i,M-j}$$

$$= p_{ij} + p_{i,M-j}$$

$$= \binom{M}{j} [(\frac{1}{M})^j(1- \frac{1}{M})^{M-j} + (\frac{1}{M})^{M-j}(1- \frac{1}{M})^j].$$

Now, let us consider the remaining cases in which M is even and either i or j or both are equal to $\frac{M}{2}$. If $i = \frac{M}{2}$, then since $c_{hi}c_{kj} \neq 0$ only when $h=i=\frac{M}{2}$ and $k=j$ or $M-j$, (3.2.2) may be written

$$r_{ij} = c_{\frac{M}{2},\frac{M}{2}}c_{jj}p_{\frac{M}{2},j} + c_{\frac{M}{2},\frac{M}{2}}c_{M-j,j}p_{\frac{M}{2},M-j}$$

$$= c_{\frac{M}{2},\frac{M}{2}}(c_{jj} + c_{M-j,j})p_{\frac{M}{2},j}$$

by using the relation, $p_{\frac{M}{2},j} = p_{\frac{M}{2},M-j}$, as stated earlier.

Thus, for $j< \frac{M}{2}$,

$$r_{ij} = (\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}})p_{\frac{M}{2},j}$$

$$= 0,$$

and for $\frac{M}{2} < j$,

$$r_{ij} = (\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}})p_{\frac{M}{2},j}$$

$$= \sqrt{2}\ p_{\frac{M}{2},j}$$

$$= \sqrt{2}\binom{M}{j}(\frac{1}{2})^M.$$

Now, if $j = \frac{M}{2}$, $c_{hi}c_{kj} \neq 0$ only when $k=j=\frac{M}{2}$ and $h=i$ or $M-i$, and hence (3.2.2) may be written

$$r_{ij} = c_{ii}c_{\frac{M}{2},\frac{M}{2}}p_{i,\frac{M}{2}} + c_{i,M-i}c_{\frac{M}{2},\frac{M}{2}}p_{M-i,\frac{M}{2}} .$$

However, since $p_{i,\frac{M}{2}} = p_{M-i,\frac{M}{2}}$, this may be simplified to give

$$r_{ij} = c_{\frac{M}{2},\frac{M}{2}}(c_{ii} + c_{i,M-i})p_{i,\frac{M}{2}},$$

which, when $i < \frac{M}{2}$, reduces to

$$r_{ij} = \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\right)p_{i,\frac{M}{2}}$$

$$= 0,$$

and, when $\frac{M}{2} < i$, reduces to

$$r_{ij} = \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\right)p_{i,\frac{M}{2}}$$

$$= \sqrt{2}\, p_{i,\frac{M}{2}}$$

$$= \sqrt{2}\binom{M}{M/2}\left[\left(\frac{1}{M}\right)\left(1-\frac{1}{M}\right)\right]^{\frac{M}{2}} .$$

Finally, when both i and j are equal to $\frac{M}{2}$, we note that the only non-zero term in (3.2.2) occurs when $h=i=k=j=\frac{M}{2}$, and thus we have

$$r_{ij} = c_{\frac{M}{2},\frac{M}{2}}c_{\frac{M}{2},\frac{M}{2}}p_{\frac{M}{2},\frac{M}{2}}$$

$$= (1)(1)p_{\frac{M}{2},\frac{M}{2}}$$

or

$$r_{ij} = \binom{M}{M/2}\left(\tfrac{1}{2}\right)^{M} .$$

As a consequence of the above results we may state the following:

THEOREM I. There exists an orthogonal matrix C such that

$$R = C^{-1}PC = \left[\begin{array}{c:c} D & 0 \\ \hdashline 0 & S \end{array}\right]$$

where D and S are square matrices of order $\frac{M+1}{2}$ if M is odd and of orders $\frac{M}{2}$ and $\frac{M}{2}+1$, respectively, if M is even.[1]

Now, R and P are similar matrices and hence have the same eigenvalues. Furthermore, letting I_R, I_D, and I_S denote unit matrices of orders equal to the orders of R, D, and S, respectively, we see that

$$\begin{aligned} |R-\lambda I_R| &= \left| \left[\begin{array}{c:c} D & 0 \\ \hdashline 0 & S \end{array}\right] - \lambda I_R \right| \\ &= \left| \left[\begin{array}{c:c} D & 0 \\ \hdashline 0 & S \end{array}\right] - \lambda \left[\begin{array}{c:c} I_D & 0 \\ \hdashline 0 & I_S \end{array}\right] \right| \\ &= \left| \left[\begin{array}{c:c} D-\lambda I_D & 0 \\ \hdashline 0 & S-\lambda I_S \end{array}\right] \right| \\ &= |D-\lambda I_D| \cdot |S-\lambda I_S| . \end{aligned}$$

---

[1] The notation D and S is here intended to indicate that the elements are "Differences" and "Sums", respectively, of elements in P, and care should be taken not to confuse these matrices with the column vector $\underline{S}^{(t)}$

That is, the matrices D and S separate the eigenvalues of P into two groups, those which are eigenvalues of D and those which are eigenvalues of S.

Theorem I leads us to an important property of the eigenvectors of P which is contained in the following:

COROLLARY. The pre- and post-eigenvectors of P can be written as either symmetric or skew-symmetric vectors.

Proof. Suppose that λ_r is an eigenvalue of D, the matrix of "Differences". Then we may write the corresponding post-eigenvector in the form

$$(x_0, x_1, x_2, \ldots, x_{\left[\frac{M-1}{2}\right]})' = \underline{V}'_{rD},$$

where $\left[\frac{M-1}{2}\right]$ indicates the greatest integer contained in $\frac{M-1}{2}$. Thus, the corresponding post-eigenvector of R can be written in the form

$$(x_0, x_1, \ldots, x_{\left[\frac{M-1}{2}\right]}, 0, 0, \ldots, 0)' = \underline{V}'_{rR},$$

or simply

$$\underline{V}_{rR} = \left[\begin{array}{c} \underline{V}_{rD} \\ \hline \underline{0} \end{array}\right].$$

Therefore, the post-eigenvector of P related to λ_r may now be obtained by pre-multiplying $\underline{V}_{rR}$ by C which gives

$$\underline{V}_{rP} = C\underline{V}_{rR} = \frac{1}{\sqrt{2}} \left[\begin{array}{c|c|c} I & \underline{0} & J \\ \hline \underline{0}' & \sqrt{2} & \underline{0}' \\ \hline -J & \underline{0} & I \end{array}\right] \left[\begin{array}{c} \underline{V}_{rD} \\ \hline 0 \\ \hline \underline{0} \end{array}\right] = \left[\begin{array}{c} \underline{V}_{rD} \\ \hline 0 \\ \hline -J\underline{V}_{rD} \end{array}\right],$$

or

$$\underline{V}'_{rP} = (x_0, x_1, \ldots, x_{\left[\frac{M-1}{2}\right]}, \vdots 0 \vdots, -x_{\left[\frac{M-1}{2}\right]}, \ldots, -x_1, -x_0)'.$$

where the scalars and row and column vectors within dots apply when M is even and are omitted when M is odd. We see then that any post-eigenvector of D is skew-symmetric. Similarly, it can be shown that any eigenvector of P corresponding to an eigenvalue of S is of the form $(x_0, x_1, \ldots, x_{\left[\frac{M-1}{2}\right]}, \vdots 0 \vdots, x_{\left[\frac{M-1}{2}\right]}, \ldots, x_1, x_0)'$. It follows that the pre-eigenvectors can also be written as symmetric or skew-symmetric vectors and the corollary is established.

The separation of the eigenvectors of P into symmetric and skew-symmetric classifications is a very useful result as will be shown in the discussion which follows. Our aim is to determine the distribution of time taken for the chain to first reach one or the other of the absorbing states given that initially the chain is in state i. It has previously been shown, (3.1.8), that $\underline{S}^{(t)} = KD_\lambda^{t-1}K^{-1}\underline{S}^{(1)}$ where $PK = KD_\lambda$, and where

$$\underline{S}^{(1)} = \begin{bmatrix} 0 \\ p_{10} + p_{1M} \\ p_{20} + p_{2M} \\ \vdots \\ p_{M-1,0} + p_{M-1,M} \\ 0 \end{bmatrix}.$$

In the particular problem under consideration, this may be written

$$\underline{S}^{(1)} = \begin{bmatrix} 0 \\ (1-\frac{1}{M})^M + (\frac{1}{M})^M \\ (1-\frac{2}{M})^M + (\frac{2}{M})^M \\ \vdots \\ (\frac{2}{M})^M + (1-\frac{2}{M})^M \\ (\frac{1}{M})^M + (1-\frac{1}{M})^M \\ 0 \end{bmatrix},$$

from which we see that $\underline{S}^{(1)}$ is symmetric.

First, let us consider the form of $\underline{S}^{(t)}$ when M is odd. Theorem I assures us that the orthogonal matrix C, (3.2.1), is such that we may write

$$R = C^{-1}PC = \begin{bmatrix} D & 0 \\ 0 & S \end{bmatrix}.$$

The existence of the eigenvectors of P and the similarity of R and P assures us of the existence of the eigenvector matrix of R, say K_R, such that

$$K_R^{-1}RK_R = \begin{bmatrix} K_{RD}^{-1} & 0 \\ 0 & K_{RS}^{-1} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} K_{RD} & 0 \\ 0 & K_{RS} \end{bmatrix}$$

$$= \begin{bmatrix} D_{\lambda(R)} & 0 \\ 0 & D_{\lambda(S)} \end{bmatrix},$$

where $D_{\lambda(R)}$ and $D_{\lambda(S)}$ denote diagonal matrices with nonzero elements equal to the eigenvalues of R and S, respectively. Then

$$K_R^{-1}C^{-1}PCK_R = \begin{bmatrix} D_{\lambda(R)} & 0 \\ 0 & D_{\lambda(S)} \end{bmatrix},$$

or

$$P = CK_R \begin{bmatrix} D_{\lambda(R)} & 0 \\ 0 & D_{\lambda(S)} \end{bmatrix} K_R^{-1}C^{-1}$$

and from (3.1.6),

$$\underline{s}^{(t)} = CK_R \begin{bmatrix} D^{t-1}_{\lambda(R)} & 0 \\ 0 & D^{t-1}_{\lambda(S)} \end{bmatrix} K_R^{-1}C^{-1}\underline{s}^{(1)},$$

or

$$\underline{s}^{(t)} = \frac{1}{2} \begin{bmatrix} I & J \\ -J & I \end{bmatrix} \begin{bmatrix} K_{RD} & 0 \\ 0 & K_{RS} \end{bmatrix} \begin{bmatrix} D^{t-1}_{\lambda(R)} & 0 \\ 0 & D^{t-1}_{\lambda(S)} \end{bmatrix}$$

$$\cdot \begin{bmatrix} K^{-1}_{RD} & 0 \\ 0 & K^{-1} \end{bmatrix} \begin{bmatrix} I & -J \\ J & I \end{bmatrix} \begin{bmatrix} 0 \\ s_1^{(1)} \\ \vdots \\ s^{(1)}_{\frac{M-1}{2}} \\ s^{(1)}_{\frac{M-1}{2}} \\ \vdots \\ s_1^{(1)} \\ 0 \end{bmatrix}.$$

Multiplying $\underline{S}^{(1)}$ by

$$\begin{bmatrix} I & -J \\ J & I \end{bmatrix}$$

we obtain

$$\begin{bmatrix} \underline{0} \\ \cdots\cdots \\ 2S^{(1)}_{\frac{M-1}{2}} \\ \vdots \\ 2S^{(1)}_{2} \\ 2S^{(1)}_{1} \\ 0 \end{bmatrix},$$

which will be denoted by

$$\begin{bmatrix} \underline{0} \\ \cdots\cdots \\ \underline{S}^{(1)}_{\Delta} \end{bmatrix},$$

where $\underline{S}^{(1)}_{\Delta}$ is used to indicate the column vector $2\underline{S}^{(1)}$ with some elements deleted. Continuing the multiplication, we have

$$\begin{bmatrix} K^{-1}_{RD} & 0 \\ 0 & K^{-1}_{RS} \end{bmatrix} \begin{bmatrix} \underline{0} \\ \underline{S}^{(1)}_{\Delta} \end{bmatrix} = \begin{bmatrix} \underline{0} \\ K^{-1}_{RS} \cdot \underline{S}^{(1)}_{\Delta} \end{bmatrix},$$

which when multiplied by

$$\begin{bmatrix} D_{\lambda(R)} & 0 \\ 0 & D_{\lambda(S)} \end{bmatrix}^{t-1}$$

will give the form

$$\begin{bmatrix} \underline{0} \\ D^{t-1}_{\lambda(S)} K^{-1}_{RS}\, \underline{s}^{(1)}_{\Delta} \end{bmatrix}.$$

Further multiplication by K_R will then yield

$$\begin{bmatrix} \underline{0} \\ K_{RS} D^{t-1}_{\lambda(S)} K^{-1}_{RS}\, \underline{s}^{(1)}_{\Delta} \end{bmatrix}.$$

We note that this column vector does not involve K_{RD} and pre-multiplication by

$$\frac{1}{2}\begin{bmatrix} I & J \\ -J & I \end{bmatrix}$$

will not alter this result. Thus, only the eigenvectors corresponding to S, which have been shown to be the symmetric ones, are needed if M is odd.

Now let us consider the case when M is even. The matrix C in this case has the form

$$\frac{1}{\sqrt{2}}\begin{bmatrix} I & \underline{0} & J \\ \underline{0}' & \sqrt{2} & \underline{0}' \\ -J & \underline{0} & I \end{bmatrix},$$

where I and J are of order M/2. Again performing computations similar to the preceding, we find in this case also that $\underline{s}^{(t)}$ does not involve K_{RD}. Therefore, in order to determine the distribution only one-half of the eigenvectors are needed, these being the ones which are symmetric, and thus, only the matrix S need be considered.

Let us now examine more closely the eigenvalues of D and of S. We shall now show that the eigenvalues of D are the odd ones of P, that is, $\lambda_1, \lambda_3, \lambda_5, \ldots$, and that the eigenvalues of S are the even ones of P, that is $\lambda_0, \lambda_2, \lambda_4, \ldots$. Note, since $\lambda_0 = \lambda_1 = 1$, the previous statement is ambiguous. However, for convenience, the eigenvalues λ_0 and λ_1 will be assigned to S and D, respectively.

The statements made in the preceding paragraph are consequences of the following:

THEOREM II. The post-eigenvectors corresponding to the odd eigenvalues are skew-symmetric. Thus, we need to prove that for $r = 0, 1, \ldots, \left[\frac{M-1}{2}\right]$, the system of linear equations

$$\sum_{j=0}^{M} p_{ij} x_{j,2r+1} = \lambda_{2r+1} x_{i,2r+1}, \quad i = 0, 1, \ldots, M,$$

admits a non-trivial solution $(x_0, x_1, x_2, \ldots, -x_2, -x_1, -x_0)$.

Proof. Let

$$x_{j,2r+1} = \sum_{h=0}^{r} a_{2r-2h+1} \left(j - \frac{M-2r+2h}{2}\right)^{(2r-2h+1)}$$

and note that $x_{j,2r+1} = -x_{M-j,2r+1}$. That the x's are so related follows by considering the expressions

$$\left(j - \frac{M-2r+2h}{2}\right)^{(2r-2h+1)} = \left[\left(j - \frac{M}{2}\right) + (r-h)\right]^{(2r-2h+1)}$$

and

$$\left(M - j - \frac{M-2r+2h}{2}\right)^{(2r-2h+1)} = \left[\left(\frac{M}{2} - j\right) + (r-h)\right]^{(2r-2h+1)}.$$

The i-th factor of $\left[\left(j - \frac{M}{2}\right) + (r-h)\right]^{(2r-2h+1)}$ is

$$\left[\left(j - \frac{M}{2}\right) + (r-h) - (i-1)\right]$$

and the (2r-2h-i+2)-th factor of $\left[\left(\frac{M}{2} - j\right) + (r-h)\right]^{(2r-2h+1)}$ is

$$\left[\left(\frac{M}{2}-j\right) + (r-h) - (2r-2h-i+1)\right] = \left[\left(\frac{M}{2}-j\right)-r+h+i-1\right]$$
$$= -\left[\left(j-\frac{M}{j}\right) + (r-h) - (i-1)\right].$$

Since there is an odd number of factors, it follows that

$$\left(j-\frac{M-2r+2h}{2}\right)^{(2r-2h+1)} = -\left(M-j-\frac{M-2r+2h}{2}\right)^{(2r-2h+1)}$$

and thus $x_{j,2r+1} = -x_{M-j,2r+1}$. We must then show that

$$\begin{aligned}&\sum_{j=0}^{M} p_{ij} \sum_{h=0}^{r} a_{2r-2h+1}\left(j-\frac{M-2r+2h}{2}\right)^{(2r-2h+1)} \\ &= \frac{M^{(2r+1)}}{M^{2r+1}} \sum_{h=0}^{r} a_{2r-2h+1}\left(i-\frac{M-2r+2h}{2}\right)^{(2r-2h+1)}\end{aligned} \tag{3.2.4}$$

admits a solution for all i.

Consider the left-hand side of the above equation. It may be written as follows.

$$\begin{aligned}&\sum_{j=0}^{M} p_{ij} \sum_{h=0}^{r} a_{2r-2h+1}\left(j-\frac{M-2r+2h}{2}\right)^{(2r-2h+1)} \\ &= \sum_{h=0}^{r} a_{2r-2h+1} \sum_{j=0}^{M} p_{ij}\left(j+\frac{2r-2h-M}{2}\right)^{(2r-2h+1)} \\ &= \sum_{h=0}^{r} a_{2r-2h+1} \sum_{j=0}^{M} p_{ij} \sum_{k=0}^{2r-2h+1} \binom{2r-2h+1}{k} \\ &\qquad \cdot \left(\frac{2r-2h-M}{2}\right)^{(k)} j^{(2r-2h+1-k)} \\ &= \sum_{h=0}^{r} a_{2r-2h+1} \sum_{k=0}^{2r-2h+1} \binom{2r-2h+1}{k}\left(\frac{2r-2h-M}{2}\right)^{(k)} \\ &\qquad \cdot \sum_{j=0}^{M} p_{ij}\, j^{(2r-2h+1-k)}\end{aligned}$$

$$= \sum_{h=0}^{r} a_{2r-2h+1} \sum_{k=0}^{2r-2h+1} \binom{2r-2h+1}{k} \left(\frac{2r-2h-M}{2}\right)^{(k)}$$

$$\cdot M^{(2r-2h+1-k)} \left(\frac{i}{M}\right)^{2r-2h+1-k}$$

$$= \sum_{h=0}^{r} \sum_{k=0}^{2r-2h+1} a_{2r-2h+1} \binom{2r-2h+1}{k} \left(\frac{2r-2h-M}{2}\right)^{(k)}$$

$$\cdot \frac{M^{(2r-2h+1-k)}}{M^{2r-2h+1-k}} \; i^{2r-2h+1-k},$$

which has been obtained by means of the relation

$$(x+y)^{(n)} = \sum_{i=0}^{n} \binom{n}{i} x^{(i)} y^{(n-i)}$$

as may be found in Cramer [3]. Let us denote this last form of the left-hand side by (1). Using the above factorial relation in the right-hand member of (3.2.4) we may write

$$\frac{M^{(2r+1)}}{M^{2r+1}} \sum_{h=0}^{r} a_{2r-2h+1} \left(i - \frac{M-2r+2h}{2}\right)^{(2r-2h+1)}$$

$$= \frac{M^{(2r+1)}}{M^{2r+1}} \sum_{h=0}^{r} a_{2r-2h+1} \sum_{k=0}^{2r-2h+1} \binom{2r-2h+1}{k}$$

$$\cdot \left(\frac{2r-2h-M}{2}\right)^{(k)} \; i^{(2r-2h+1-k)}$$

$$= \frac{M^{(2r+1)}}{M^{2r+1}} \sum_{h=0}^{r} \sum_{k=0}^{2r-2h+1} a_{2r-2h+1} \binom{2r-2h+1}{k}$$

$$\cdot \left(\frac{2r-2h-M}{2}\right)^{(k)} \; i^{(2r-2h+1-k)}$$

$$= \frac{M^{(2r+1)}}{M^{2r+1}} \sum_{h=0}^{r} \sum_{k=0}^{2r-2h+1} a_{2r-2h+1} \binom{2r-2h+1}{k}$$

$$\left(\frac{2r-2h-M}{2}\right)^{(k)}$$

$$\sum_{m=0}^{2r-2h+1-k} S_m^{2r-2h+1-m}\, i^m$$

$$= \frac{M^{(2r+1)}}{M^{2r+1}} \sum_{h=0}^{r} \sum_{k=0}^{2r-2h+1} \sum_{m=0}^{2r-2h+1-k} a_{2r-2h+1} \binom{2r-2h+1}{k} \left(\frac{2r-2h-M}{2}\right)^{(k)} S_m^{2r-2h+1-k}\, i^m$$

where S_m^n denotes Stirling numbers of the first kind, that is $S_m^{n+1} = -nS_m^n + S_{m-1}^n$, $S_1^1 = 1$, and $S_m^n = 0$ if $n<m$, as defined in Richardson [15]. Label this form of the right-hand side by (2).

Now (1) must equal (2) for $i = 0,1,\ldots,M$. We have then reduced the problem to that of showing that the equation (1) = (2) has a solution for all i.

The coefficient of i^t in relation (1) is

$$\sum_{h,k \ni 2r-2h+1-k=t} \sum a_{2r-2h+1} \binom{2r-2h+1}{k} \left(\frac{2r-2h-M}{2}\right)^{(k)} \cdot \frac{M^{(2r-2h+1-k)}}{M^{2r-2h+1-k}}$$

$$= \frac{M^{(t)}}{M^t} \sum_{k=0}^{2r+1-t} a_{k+t} \binom{k+t}{t} \left(\frac{k+t-M-1}{2}\right)^{(k)}$$

and the coefficient of i^t in relation (2) is

$$\frac{M^{(2r+1)}}{M^{2r+1}} \sum_{h,k \ni 2r-2h+1-k \geq t} \sum a_{2r-2h+1} \binom{2r-2h+1}{k} \cdot \left(\frac{2r-2h-M}{2}\right)^{(k)} S_m^{2r-2h+1-k}$$

$$= \frac{M^{(2r+1)}}{M^{2r-1}} \sum_{h=t}^{2r+1} \sum_{k=0}^{2r+1-h} a_{h+k} \binom{h+k}{h} \left(\frac{h+k-1-M}{2}\right)^{(k)} S_t^h .$$

Equating coefficients of i^t we have

$$\frac{M^{(t)}}{M^t} \sum_{k=0}^{2r+1-t} a_{k+t} \binom{k+t}{t} \left(\frac{k+t-M-1}{2}\right)^{(k)}$$

$$= \frac{M^{(2r+1)}}{M^{2r+1}} \sum_{k=t}^{2r+1} \sum_{k=0}^{2r+1-h} a_{h+k} \binom{h+k}{h} \cdot \left(\frac{h+k-1-M}{2}\right)^{(k)} S_t^h$$

for $t=2r+1, 2r, \ldots, 2, 1, 0$. Taking successive values of $t=2r+1$, $2r, \ldots,$ gives us a method for determining a_n in terms of $a_{n+1}, a_{n+2}, \ldots, a_{2r+1}$. The coefficient of a_t is

$$\frac{M^{(t)}}{M^t} - \frac{M^{(2r+1)}}{M^{2r+1}}$$

which is not zero except for $t=2r+1$, and hence a non-trivial solution exists, a_{2r+1} being arbitrary. Thus it follows that the post-eigenvectors corresponding to the even eigenvalues are the ones which are symmetric.

Therefore, we may summarize the work in this section by stating that only the eigenvectors corresponding to the even eigenvalues are needed in order to determine the first passage time distributions.

3.3. Triangular transformation. Before applying the theory developed in Section 3.2 to numerical examples, let us consider a transformation of matrix P which will reduce the necessary computations for determining the eigenvectors.

Let B denote the matrix which has elements defined by

$$(B)_{ij} = \binom{i}{j}, \quad i,j = 0,1,\ldots,M.$$

where the subscripts denote the row and column of the elements. The elements of B^{-1} are found to be $(B^{-1})_{ij} = (-1)^{i+j} \cdot \binom{i}{j}$, for consider

$$(B^{-1}B)_{ij} = \sum_{k=0}^{M} (-1)^{i+k}\binom{i}{k}\binom{k}{j}$$

$$= \sum_{k=j}^{i} (-1)^{i+k}\binom{i}{k}\binom{k}{j}.$$

Since both B and B^{-1} are lower triangular and have all diagonal elements equal to unity, it is obvious that $(B^{-1}B)_{ij}$ is equal to zero when $i<j$ and that $(B^{-1}B)_{ii} = 1$. If $j< i$, we have,

$$(B^{-1}B)_{ij} = \frac{(-1)^i i!}{j!} \sum_{k=j}^{i} \frac{(-1)^k}{(i-k)!(k-j)!}$$

which, upon replacing the index of summation, k, by k+j gives

$$(B^{-1}B)_{ij} = \frac{(-1)^i i!}{j!} \sum_{k=0}^{i-j} \frac{(-1)^{k+j}}{(i-j-k)!k!}$$

$$= \frac{i!}{j!(i-j)!} \sum_{k=0}^{i-j} \binom{i-j}{k} 1^k (-1)^{i-j-k}$$

and the summation is simply $\binom{i}{j}(1-1)^{i-j} = 0$.

We may now prove the following:

THEOREM III. Pre- and post-multiplication of the matrix P by the non-singular matrices B^{-1} and B, respectively, gives an upper triangular matrix, $T=B^{-1}PB$.

Proof. Post-multiplication of P by B will give

$$
\begin{aligned}
(PB)_{ij} &= \sum_{k=0}^{M} \binom{M}{k}\left(\frac{i}{M}\right)^{k}\left(1-\frac{i}{M}\right)^{M-k}\binom{k}{j} \\
&= \sum_{k=j}^{M} \binom{M}{k}\binom{k}{j}\left(\frac{i}{M}\right)^{k}\left(1-\frac{i}{M}\right)^{M-k} \\
&= \frac{M!}{j!}\sum_{k=j}^{M} \frac{1}{(M-k)!(k-j)!}\left(\frac{i}{M}\right)^{k}\left(1-\frac{i}{M}\right)^{M-k}
\end{aligned}
$$

and replacing the index, k, by $k+j$, we have

$$
\begin{aligned}
(PB)_{ij} &= \frac{M!}{j!}\sum_{k=0}^{M-j} \frac{1}{(M-j-k)!k!}\left(\frac{i}{M}\right)^{k+j}\left(1-\frac{i}{M}\right)^{M-j-k} \\
&= \frac{M!}{j!(M-j)!}\left(\frac{i}{M}\right)\sum_{k=0}^{M-j} \binom{M-j}{k}\left(\frac{i}{M}\right)^{k}\left(1-\frac{i}{M}\right)^{M-j-k} \\
&= \binom{M}{j}\left(\frac{i}{M}\right)^{j}.
\end{aligned}
$$

Then pre-multiplication of PB by B^{-1} will give

$$
\begin{aligned}
(T)_{ij} &= (B^{-1}PB)_{ij} \\
&= \sum_{k=0}^{M} (-1)^{i+k}\binom{i}{k}\binom{M}{j}\left(\frac{k}{M}\right)^{j} \\
&= \frac{M^{(j)}}{j!M^{j}}\sum_{k=0}^{i} (-1)^{i+k}\binom{i}{k}\, k^{j} \qquad (3.3.1) \\
&= \frac{i!M^{(j)}}{j!M^{j}}\sum_{k=0}^{i} (-1)^{i+k}\frac{k^{j}}{k!(i-k)!} \\
&= \frac{i!M^{(j)}}{j!M^{j}}\mathscr{S}_{j}^{i}
\end{aligned}
$$

where the $\mathscr{S}_{i}^{j}$ are Stirling numbers of the second kind,

$\mathscr{S}_i^{j+1} = i\mathscr{S}_i^j + \mathscr{S}_{i-1}^j$, $\mathscr{S}_i^j = 0$ if $j< i$, $\mathscr{S}_i^i = 1$, as defined by Richardson [15]. Since $\mathscr{S}_i^j = 0$ for $j< i$, we see that $T = B^{-1}PB$ is upper triangular. Also, since $\mathscr{S}_i^i = 1$, we have $(B^{-1}PB)_{ii} = \frac{M^{(i)}}{M^i} = \lambda_i$, and, P and $B^{-1}PB$ being similar matrices implies that the eigenvalues of P are these λ_i. This follows since the elements of the principal diagonal of a triangular matrix are the eigenvalues of the matrix. These eigenvalues agree with the results of Feller [7] and the derivation is quite simple.

3.4. Some numerical examples and general results.

The theory of the previous section suggests an easy computational technique for determining the distributions for specific numerical values of M. If we represent the matrix of post-eigenvectors of T by K_T, then we may determine the matrix of post-eigenvectors, K, of P by the relation $K=BK_T$. Application of formula (3.1.8) will then give us the distributions we seek.

As an example, let us consider the case of M = 5. The transition matrix P is given by

$$(3.4.1) \qquad P = \frac{1}{3125}\begin{bmatrix} 3125 & 0 & 0 & 0 & 0 & 0 \\ 1024 & 1280 & 640 & 160 & 20 & 1 \\ 243 & 810 & 1080 & 720 & 240 & 32 \\ 32 & 240 & 720 & 1080 & 810 & 243 \\ 1 & 20 & 160 & 640 & 1280 & 1025 \\ 0 & 0 & 0 & 0 & 0 & 3125 \end{bmatrix}.$$

Post-multiplication of (3.4.1) by

$$(3.4.2) \qquad B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 4 & 6 & 4 & 1 & 0 \\ 1 & 5 & 10 & 10 & 5 & 1 \end{bmatrix}$$

and then pre-multiplication of the resulting product by

$$B^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ -1 & 3 & -3 & 1 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 \\ -1 & 5 & -10 & 10 & -5 & 1 \end{bmatrix}$$

will, by (3.3.1), give

$$T = \frac{1}{3125} \begin{bmatrix} 3125 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3125 & 1250 & 250 & 25 & 1 \\ 0 & 0 & 2500 & 1500 & 350 & 30 \\ 0 & 0 & 0 & 1500 & 900 & 150 \\ 0 & 0 & 0 & 0 & 600 & 240 \\ 0 & 0 & 0 & 0 & 0 & 120 \end{bmatrix}.$$

The matrix of post-eigenvectors, K_T, may now be determined by solving the systems $TK_T = K_T D_\lambda$ which is quite straightforward since each of the systems is triangular. However, according to the results of Section 3.2, a complete solution is not needed and we may simply solve the system

$$TK_T^{even} = K_T^{even} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2500/3125 & 0 \\ 0 & 0 & 600/3125 \end{bmatrix}.$$

Solving this system, we obtain as one set of solutions,

$$K_T^{even} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & -\frac{4}{19} \\ 0 & 1 & \frac{23}{38} \\ 0 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and pre-multiplication by (3.4.2) gives column vectors which are proportional to

$$(3.4.3) \qquad K^{even} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 8 \\ 1 & 3 & -7 \\ 1 & 3 & -7 \\ 1 & 2 & 8 \\ 1 & 0 & 0 \end{bmatrix}.$$

Now, since the pre-eigenvectors are also symmetric, $(K^{even})^{-1}$ can be obtained by solving

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_3 & x_2 & x_1 \\ y_1 & y_2 & y_3 & y_3 & y_2 & y_1 \\ z_1 & z_2 & z_3 & z_3 & z_2 & z_1 \end{bmatrix} . K^{even} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which gives

$$(3.4.4) \qquad (K^{even})^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} \\ -\frac{15}{76} & \frac{7}{76} & \frac{8}{76} & \frac{8}{76} & \frac{7}{76} & -\frac{15}{76} \\ -\frac{1}{76} & \frac{3}{76} & -\frac{2}{76} & -\frac{2}{76} & \frac{3}{76} & -\frac{1}{76} \end{bmatrix}.$$

Using the eigenvector matrices (3.4.3) and (3.4.4), we may now use relation (3.1.8) to obtain

$$\underline{S}^{(t)} = K^{even}(D_{\lambda}^{even})^{t-1}(K^{even})^{-1}\underline{S}^{(1)}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 8 \\ 1 & 3 & -7 \\ 1 & 3 & -7 \\ 1 & 2 & 8 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & (\frac{4}{5})^{t-1} & 0 \\ 0 & 0 & (\frac{24}{125})^{t-1} \end{bmatrix}$$

$$\cdot \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} \\ -\frac{15}{76} & \frac{7}{76} & \frac{8}{76} & \frac{8}{76} & \frac{7}{76} & -\frac{15}{76} \\ -\frac{1}{76} & \frac{3}{76} & -\frac{2}{76} & -\frac{2}{76} & \frac{3}{76} & -\frac{1}{76} \end{bmatrix}$$

$$\cdot \begin{bmatrix} 0 \\ \frac{41}{125} \\ \frac{11}{125} \\ \frac{11}{125} \\ \frac{41}{125} \\ 0 \end{bmatrix}$$

$$= \frac{1}{4750}\begin{bmatrix} 0 \\ 750(\frac{4}{5})^{t-1} + 808(\frac{24}{125})^{t-1} \\ 1125(\frac{4}{5})^{t-1} - 707(\frac{24}{125})^{t-1} \\ 1125(\frac{4}{5})^{t-1} - 707(\frac{24}{125})^{t-1} \\ 750(\frac{4}{5})^{t-1} + 808(\frac{24}{125})^{t-1} \\ 0 \end{bmatrix}, \text{ for } 0< t,$$

and for t=0,

$$\underline{s}^{(0)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

which are the distributions we seek. That is, if the population is of size five and consists of one a-gene and four A-genes or four a-genes and one A-gene, then the probability of absorption in state 0 or 5 in t generations is given by

$$s_1^{(t)} = s_4^{(t)} = \frac{1}{4750}[750(\tfrac{4}{5})^{t-1} + 808(\tfrac{24}{125})^{t-1}],\ 0<t.$$

If however, the population consists of two genes of one type and three of the other, then the probability of absorption is given by

$$s_2^{(t)} = s_3^{(t)} = \frac{1}{4750}[1125(\tfrac{4}{5})^{t-1} - 707(\tfrac{24}{125})^{t-1}],\ 0<t.$$

The means and variances of these and the following distributions, as well as the graphic presentation, are given in

Figures 1-20 at the end of this section. Some interpretations of these results will be found in Chapter VI.

Using the preceding approach, additional distributions have been obtained and the results are as follows, the domain of t being t = 1,2,... .

M=2: $s_1^{(t)} = (\frac{1}{2})^t$

M=3: $s_1^{(t)} = s_2^{(t)} = \frac{1}{3}(\frac{2}{3})^{t-1}$

M=4: $s_1^{(t)} = s_3^{(t)} = \frac{1}{2688}[600(\frac{3}{4})^{t-1} + 261(\frac{3}{32})^{t-1}]$

$s_2^{(t)} = \frac{1}{168}[50(\frac{3}{4})^{t-1} - 29(\frac{3}{32})^{t-1}]$

M=6: $s_1^{(t)} = s_5^{(t)} = .117095(\frac{5}{6})^{t-1}$

$+ .206518(\frac{5}{18})^{t-1} + .011370(\frac{5}{324})^{t-1}$

$s_2^{(t)} = s_4^{(t)} = .187352(\frac{5}{6})^{t-1}$

$- .066086(\frac{5}{18})^{t-1} - .032103(\frac{5}{324})^{t-1}$

$s_3^{(t)} = .210770(\frac{5}{6})^{t-1}$

$- .223039(\frac{5}{18})^{t-1} + .043519(\frac{5}{324})^{t-1}$

M=7: $s_1^{(t)} = s_6^{(t)} = .090145(\frac{6}{7})^{t-1}$

$+ .218975(\frac{120}{343})^{t-1} + .030798(\frac{720}{16807})^{t-1}$

$s_2^{(t)} = s_5^{(t)} = .150242(\frac{6}{7})^{t-1}$

$+ .012165(\frac{120}{343})^{t-1} - .067387(\frac{720}{16807})^{t-1}$

$$s_3^{(t)} = s_4^{(t)} = .180290(\tfrac{6}{7})^{t-1}$$

$$- .197077(\tfrac{120}{343})^{t-1} + .039337(\tfrac{720}{16807})^{t-1}$$

M=8: $$s_1^{(t)} = s_7^{(t)} = .071483(\tfrac{7}{8})^{t-1}$$

$$+ .218029(\tfrac{105}{256})^{t-1} + .052792(\tfrac{315}{4096})^{t-1}$$

$$+ .001305(\tfrac{315}{131072})^{t-1}$$

$$s_2^{(t)} = s_6^{(t)} = .122542(\tfrac{7}{8})^{t-1}$$

$$+ .071193(\tfrac{105}{256})^{t-1} - .089023(\tfrac{315}{4096})^{t-1}$$

$$- .004584(\tfrac{315}{131072})^{t-1}$$

$$s_3^{(t)} = s_5^{(t)} = .153178(\tfrac{7}{8})^{t-1}$$

$$- .137937(\tfrac{105}{256})^{t-1} - .000258(\tfrac{315}{4096})^{t-1}$$

$$+ .008691(\tfrac{315}{131072})^{t-1}$$

$$s_4^{(t)} = .163389(\tfrac{7}{8})^{t-1} - .227818(\tfrac{105}{256})^{t-1}$$

$$+ .082872(\tfrac{315}{4096})^{t-1} - .010631(\tfrac{315}{131072})^{t-1}$$

M=9: $$s_1^{(t)} = s_8^{(t)} = .058050(\tfrac{8}{9})^{t-1} + .210149(\tfrac{112}{243})^{t-1}$$

$$+ .073606(\tfrac{560}{6561})^{t-1} + .004635(\tfrac{4480}{531441})^{t-1}$$

$$s_2^{(t)} = s_7^{(t)} = .101587(\tfrac{8}{9})^{t-1} + .111642(\tfrac{112}{243})^{t-1}$$

$$- .094919(\tfrac{560}{6561})^{t-1} - .014149(\tfrac{4480}{531441})^{t-1}$$

$$s_3^{(t)} = s_6^{(t)} = .130612(\tfrac{8}{9})^{t-1} - .075991(\tfrac{112}{243})^{t-1}$$

$$- .047946(\tfrac{560}{6561})^{t-1} + .019388(\tfrac{4480}{531441})^{t-1}$$

$$s_4^{(t)} = s_5^{(t)} = .145125(\tfrac{8}{9})^{t-1} - .206396(\tfrac{112}{243})^{t-1}$$

$$+ .076508(\tfrac{560}{6561})^{t-1} - .009519(\tfrac{4480}{531441})^{t-1}$$

In order to obtain the exact distributions for general M using the eigenvector approach, an attempt was made to obtain algebraic expressions for the eigenvectors. However, except for the properties of symmetry and skew-symmetry, no pattern was discerned and this approach was abandoned. Expressions which were obtained for the post-eigenvectors corresponding to the first seven eigenvalues are as follows, where i denotes the row of the eigenvector.

λ_0: M

λ_1: $M-2i$

λ_2: $i(M-i)$

λ_3: $i(M-i)(M-2i)$

λ_4: $i(M-i)[(M^3-6M^2+11M-6) + (-5M+6)(i-1)(M-i-1)]$

λ_5: $i(M-i)(M-2i)[M^3-(7i+1)M^2 + (7i^2+12i)M - 12i^2]$

$$\begin{aligned}\lambda_6:\quad i(M-i)\Big\{&[M^2-(M-4)^{(2)}][M^4-(M-2)^{(4)}](i-1)(M-i-1)^{(2)}\\ &-[M^4-(M-2)^{(4)}](3M-5)(M-4)^{(2)}(i-1)(M-i-1)\\ &+(M-1)(M-2)^{(4)}[2M^2(M-2)+(M-1)(M-4)^{(2)}]\Big\}\end{aligned}$$

Also, as an alternative, an attempt was made to determine the general eigenvectors of P indirectly by means of matrix T, (3.3.1). Again, no generalization was obtained. The eigenvector matrix of T is upper triangular and some of the elements are as follows.

$$x_{i,k} = 0, \quad k<i$$

$$x_{k,k} = 1$$

$$x_{k-1,k} = -\frac{(M-k+1)}{2}$$

$$x_{k-2,k} = \frac{(k-2)(M-k+2)^{(2)}(6M-3k+5)}{24[M^2-(M-k+2)^{(2)}]}$$

$$x_{k-3,k} = -\frac{(k-3)(M-k+3)^{(3)}(2M-k+2)}{48[M^2-(M-k+2)^{(2)}]}$$

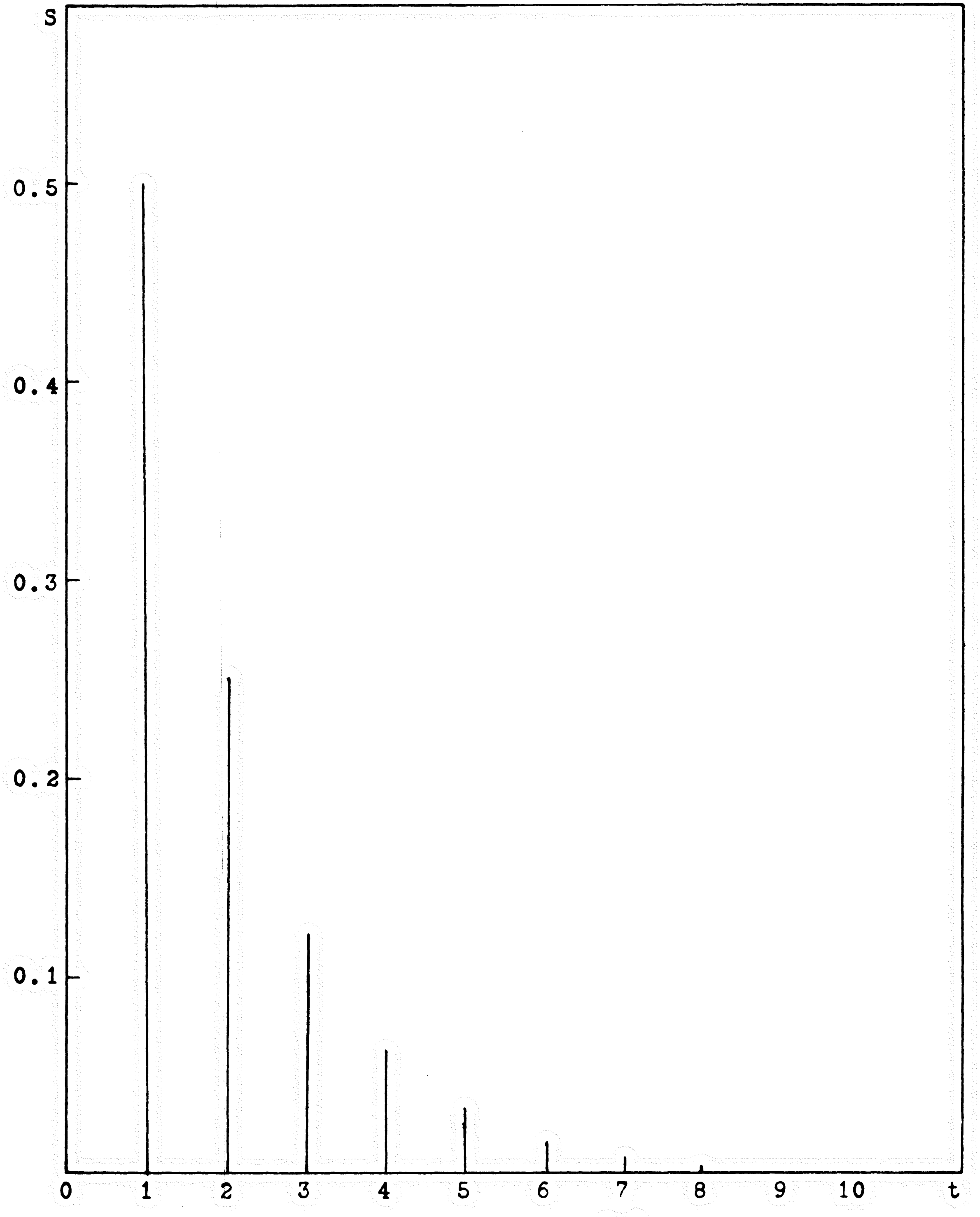

FIGURE 1. The probability function $S_1^{(t)}$, M=2.

$\mu = 2$, $\sigma^2 = 2$

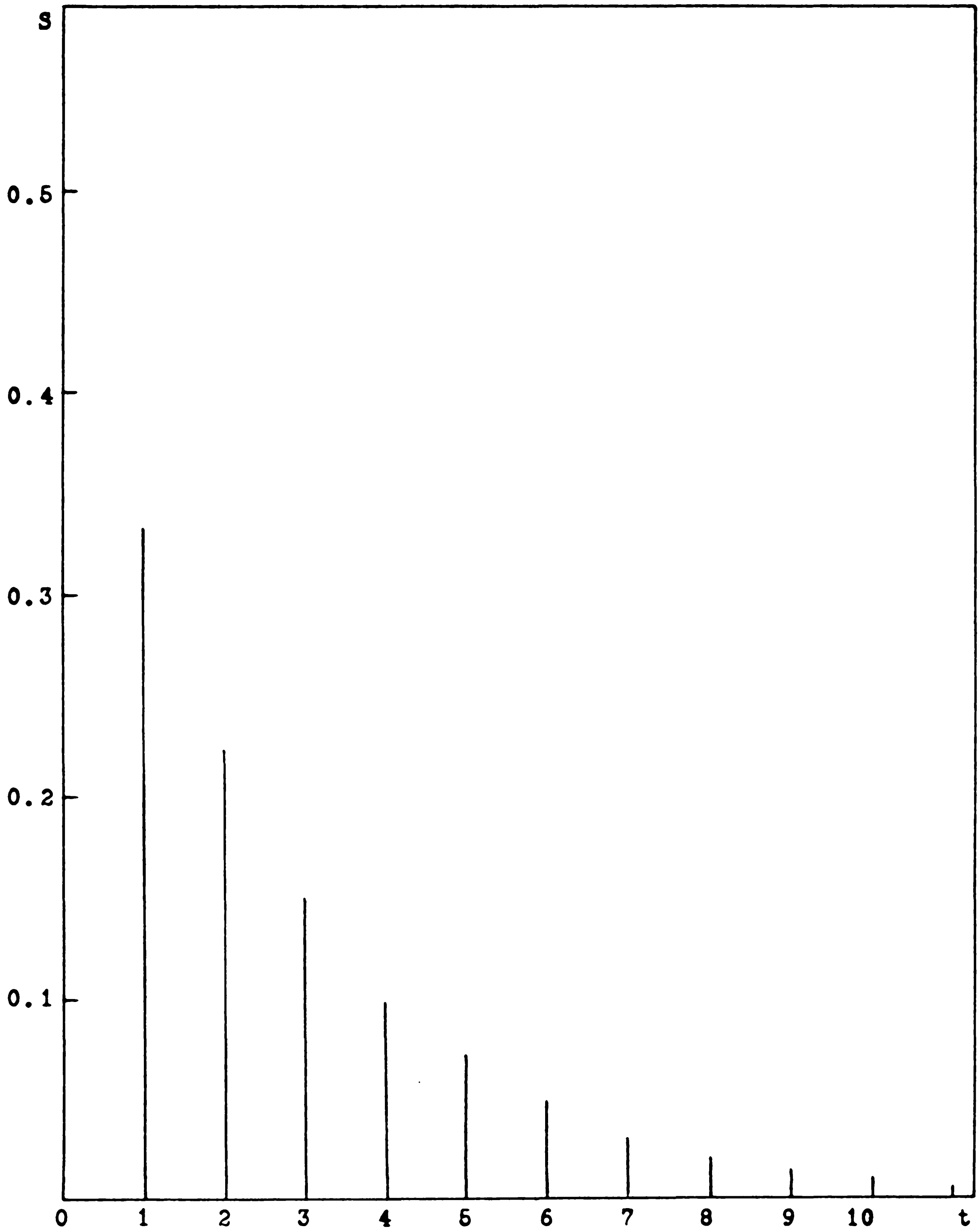

FIGURE 2. The probability function $S_1^{(t)}=S_2^{(t)}$, M=3.
$\mu = 3$, $\sigma^2 = 6$

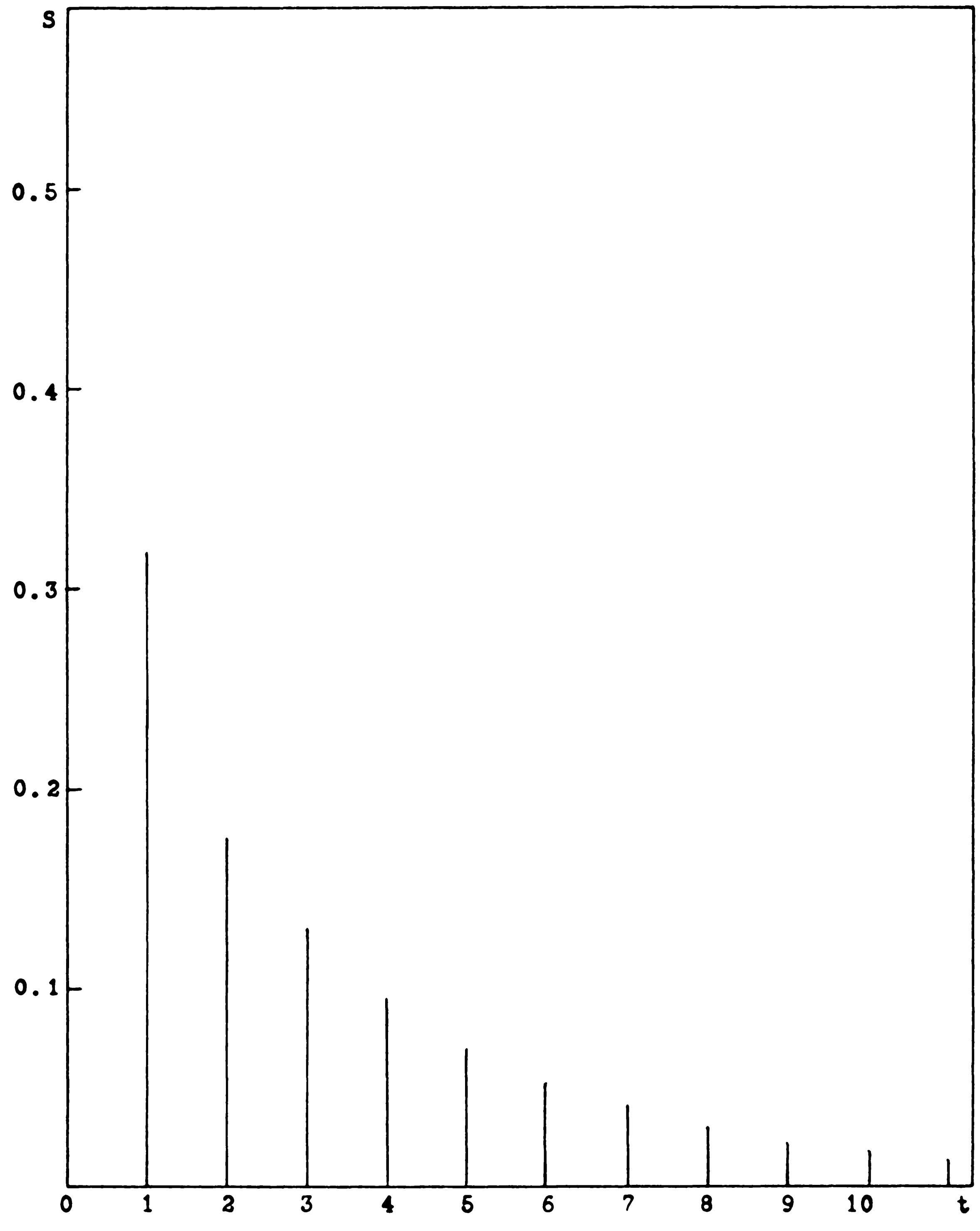

FIGURE 3. The probability function $S_1^{(t)}=S_3^{(t)}$, M=4.

$\mu = 3.68966, \sigma^2 = 11.5291$

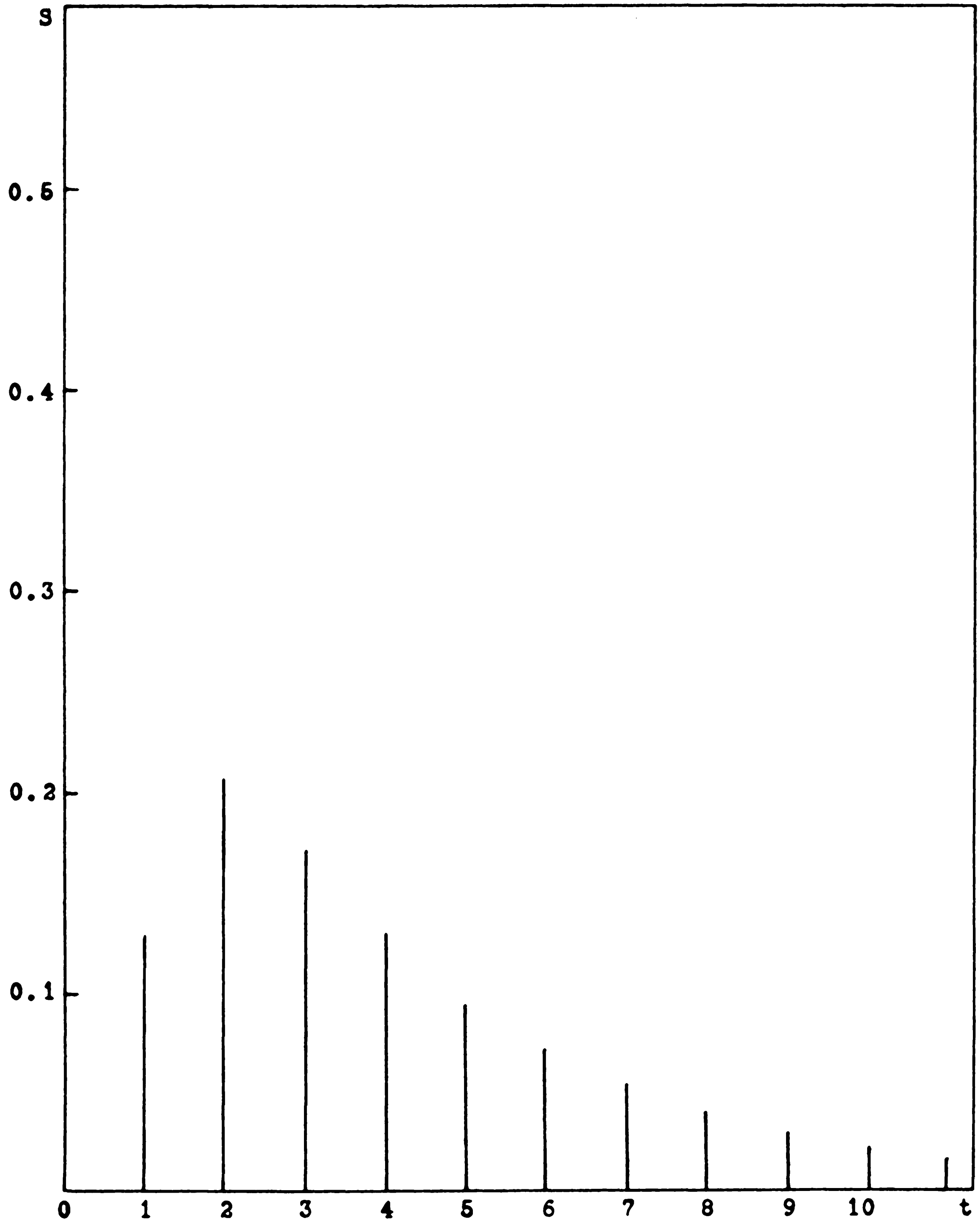

FIGURE 4. The probability function $S_2^{(t)}$, M=4.

$\mu = 4.55172$, $\sigma^2 = 12.3615$

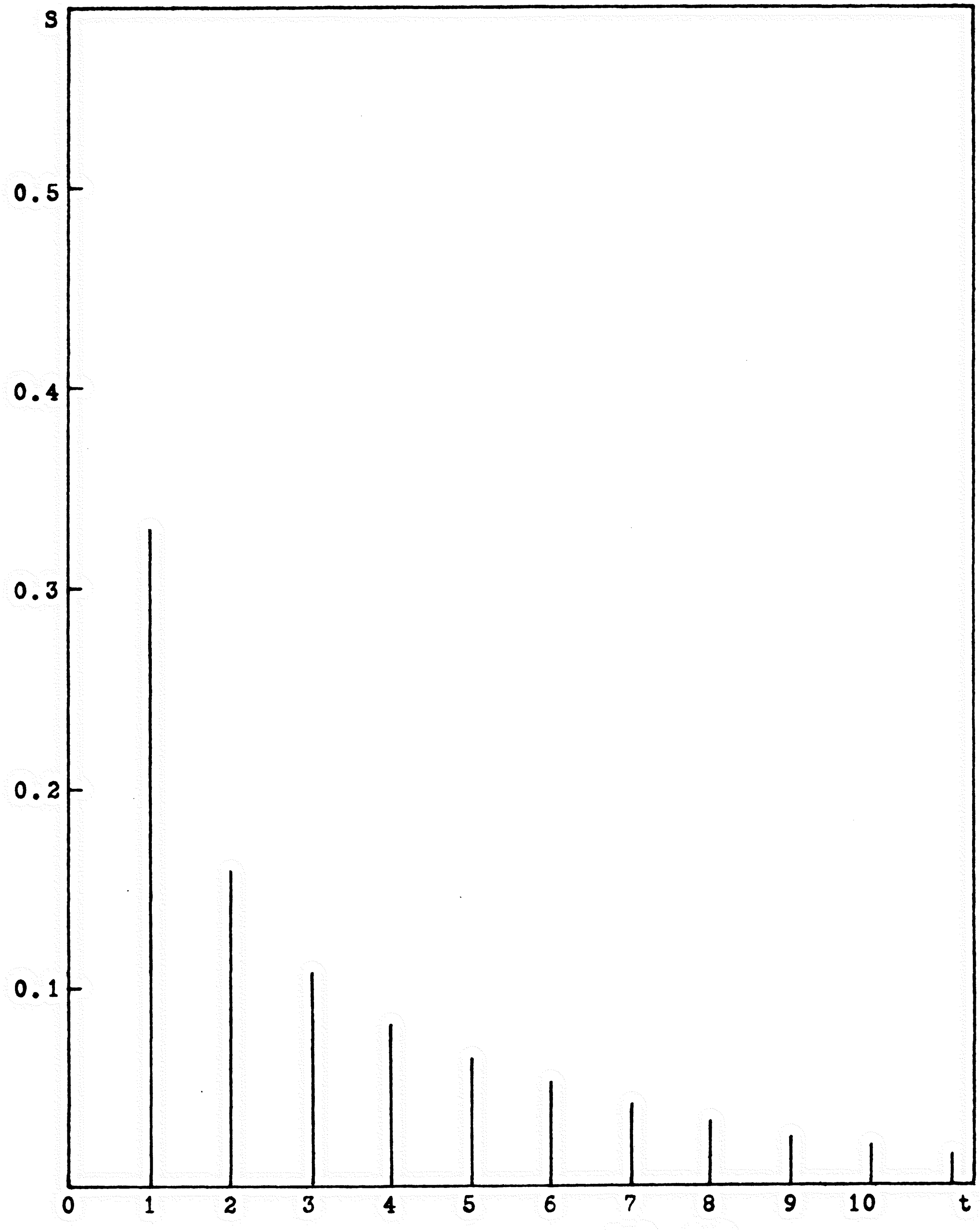

FIGURE 5. The probability function $S_1^{(t)}=S_4^{(t)}$, M=5.

$\mu = 4.20792$, $\sigma^2 = 18.2041$

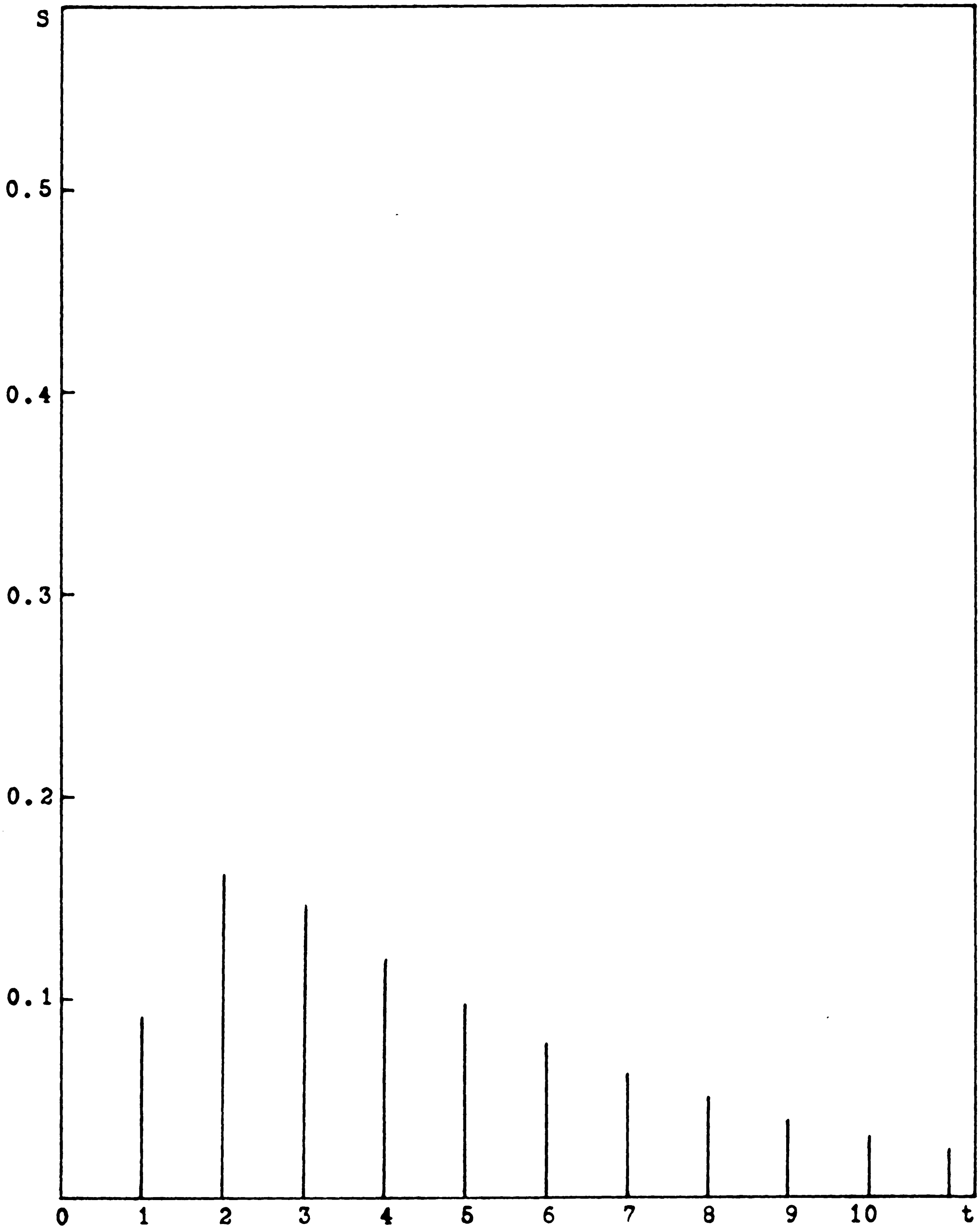

FIGURE 6. The probability function $S_2^{(t)} = S_3^{(t)}$, M=5.

$\mu = 5.69307$, $\sigma^2 = 20.5421$

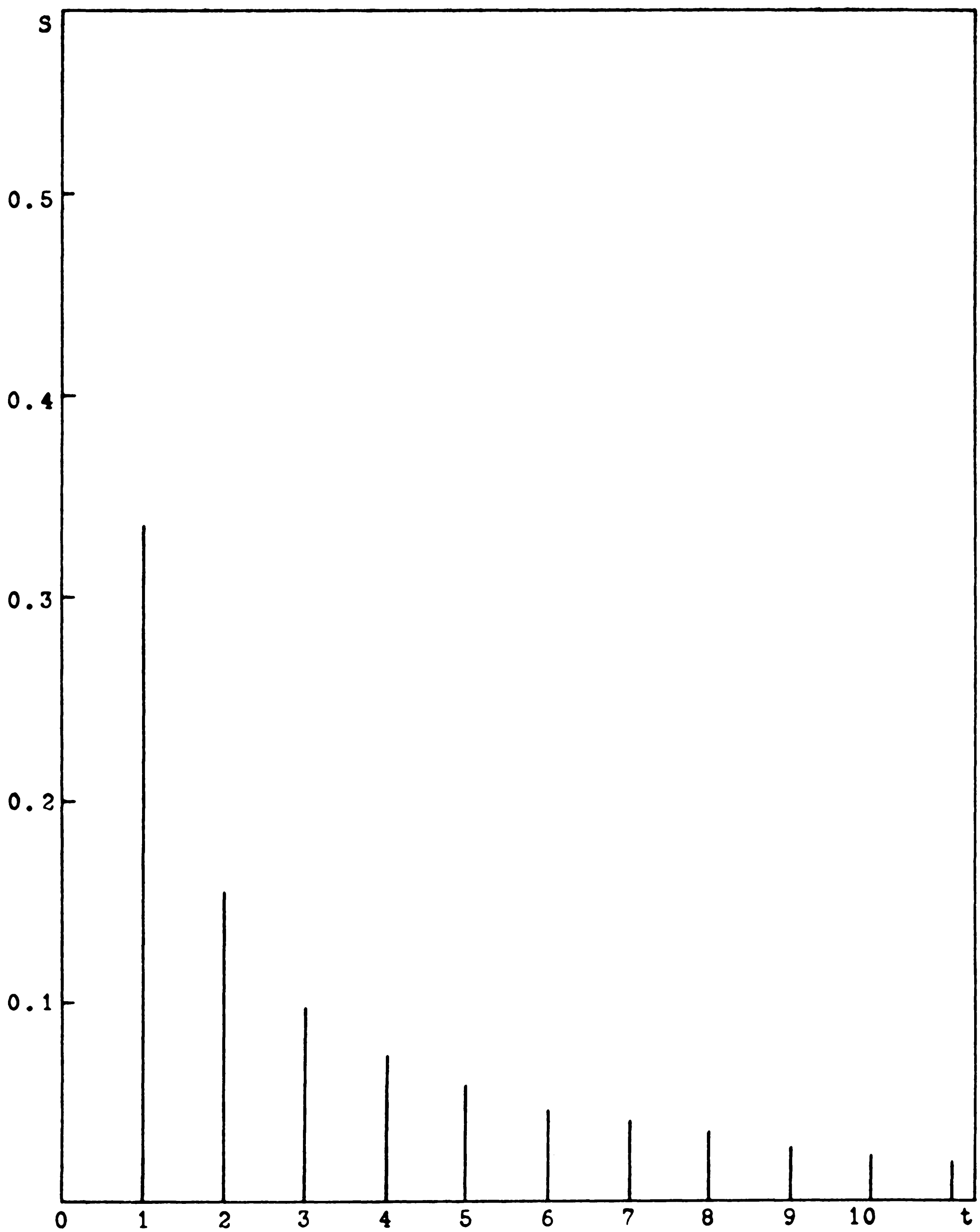

FIGURE 7. The probability function $S_1^{(t)} = S_5^{(t)}$, $M = 6$.

$\mu = 4.62300$, $\sigma^2 = 25.7099$

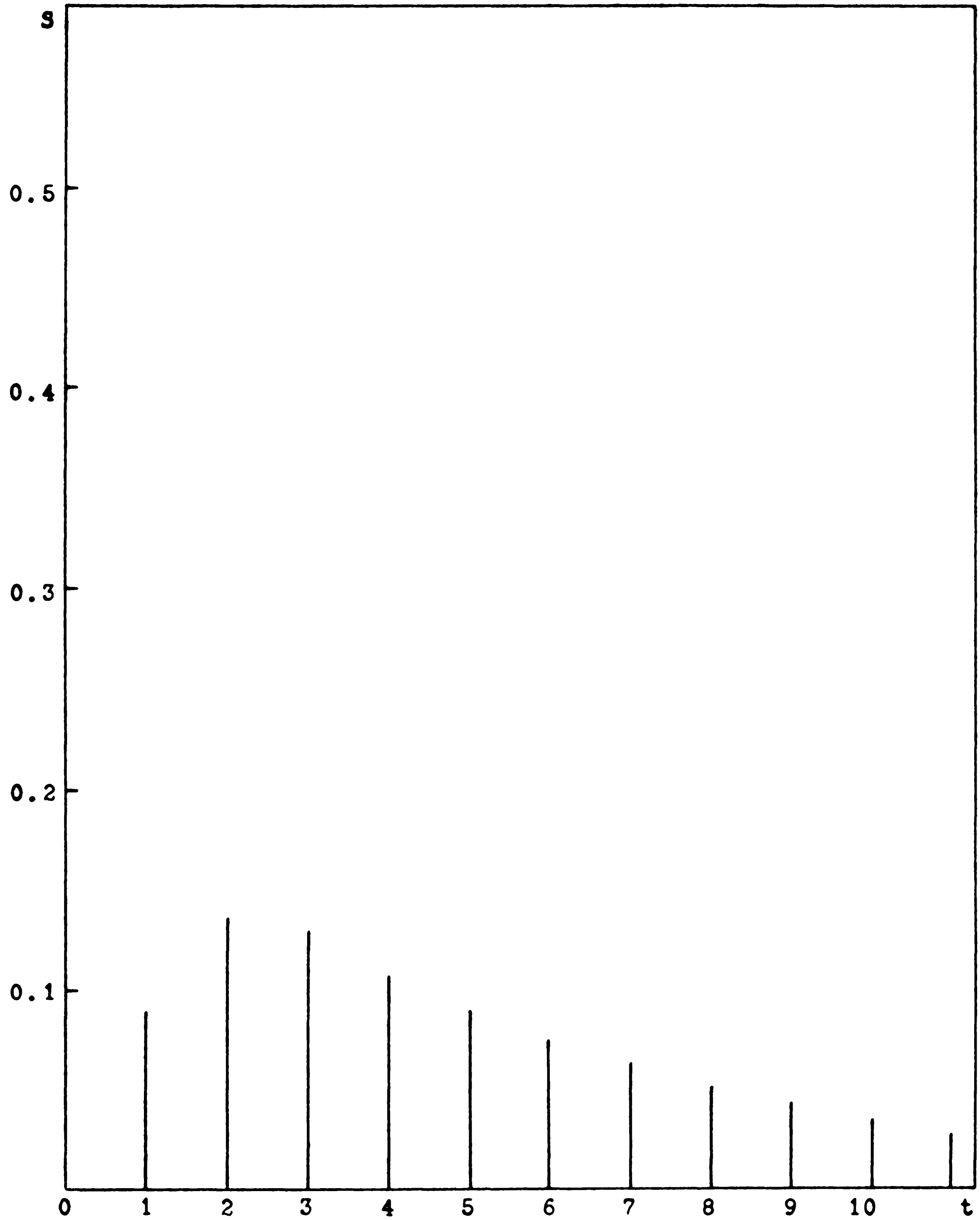

FIGURE 8. The probability function $S_2^{(t)} = S_4^{(t)}$, M=6.

$\mu = 6.58484$, $\sigma^2 = 30.5728$

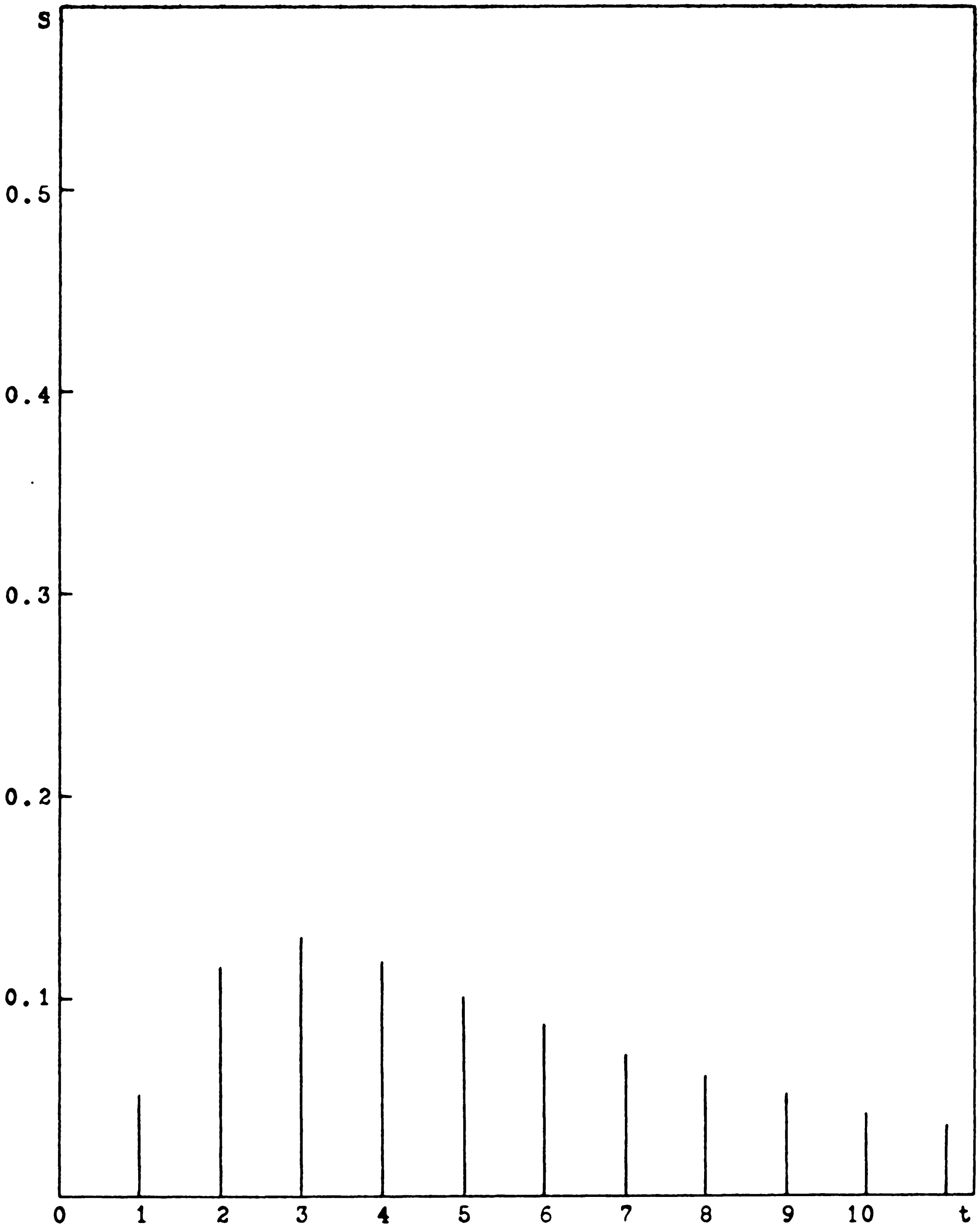

FIGURE 9. The probability function $S_3^{(t)}$, M=6.

$\mu = 7.20503$, $\sigma^2 = 30.8424$

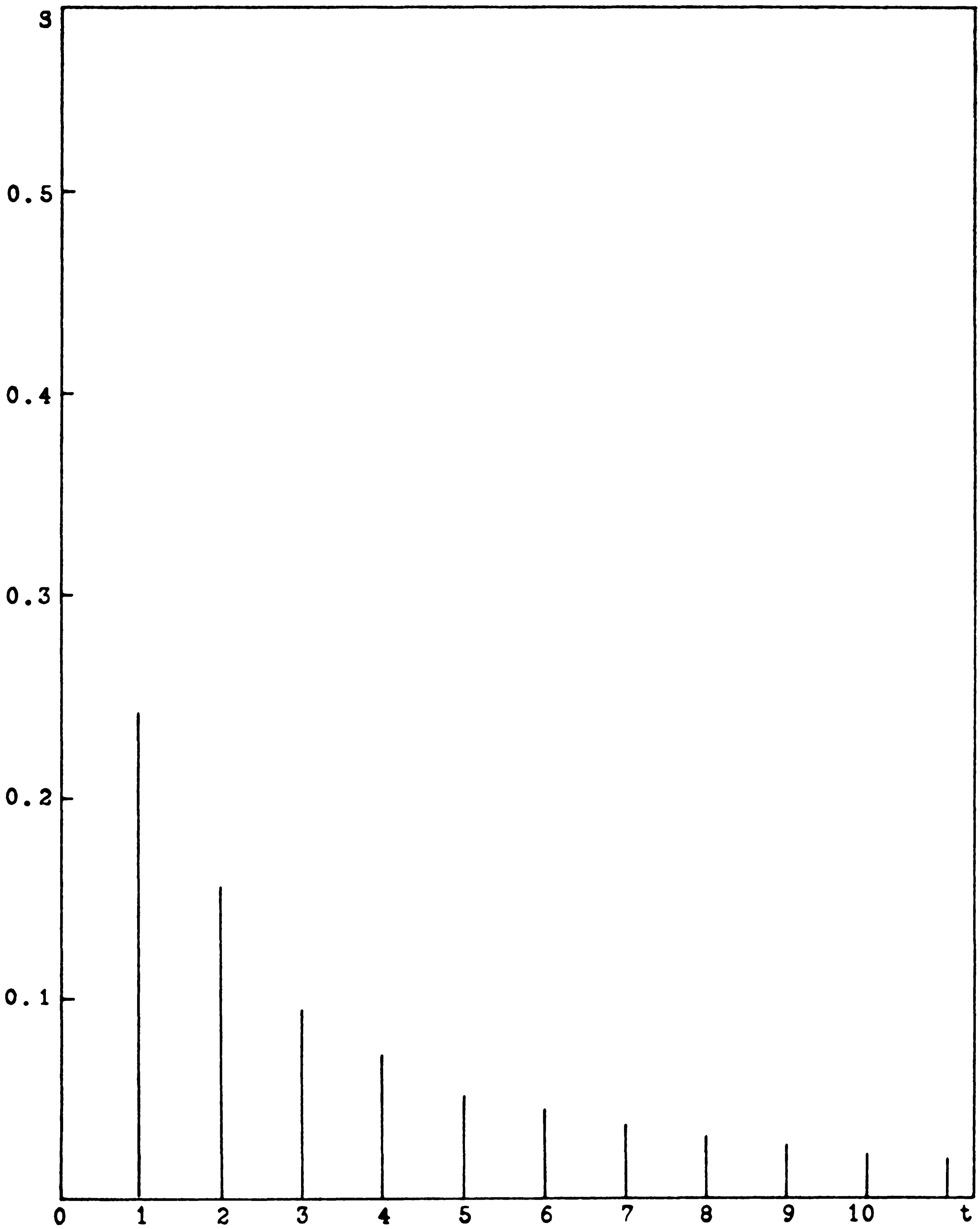

FIGURE 10. The probability function $S_1^{(t)}=S_6^{(t)}$, M=7.

$\mu = 4.96877$, $\sigma^2 = 33.8459$

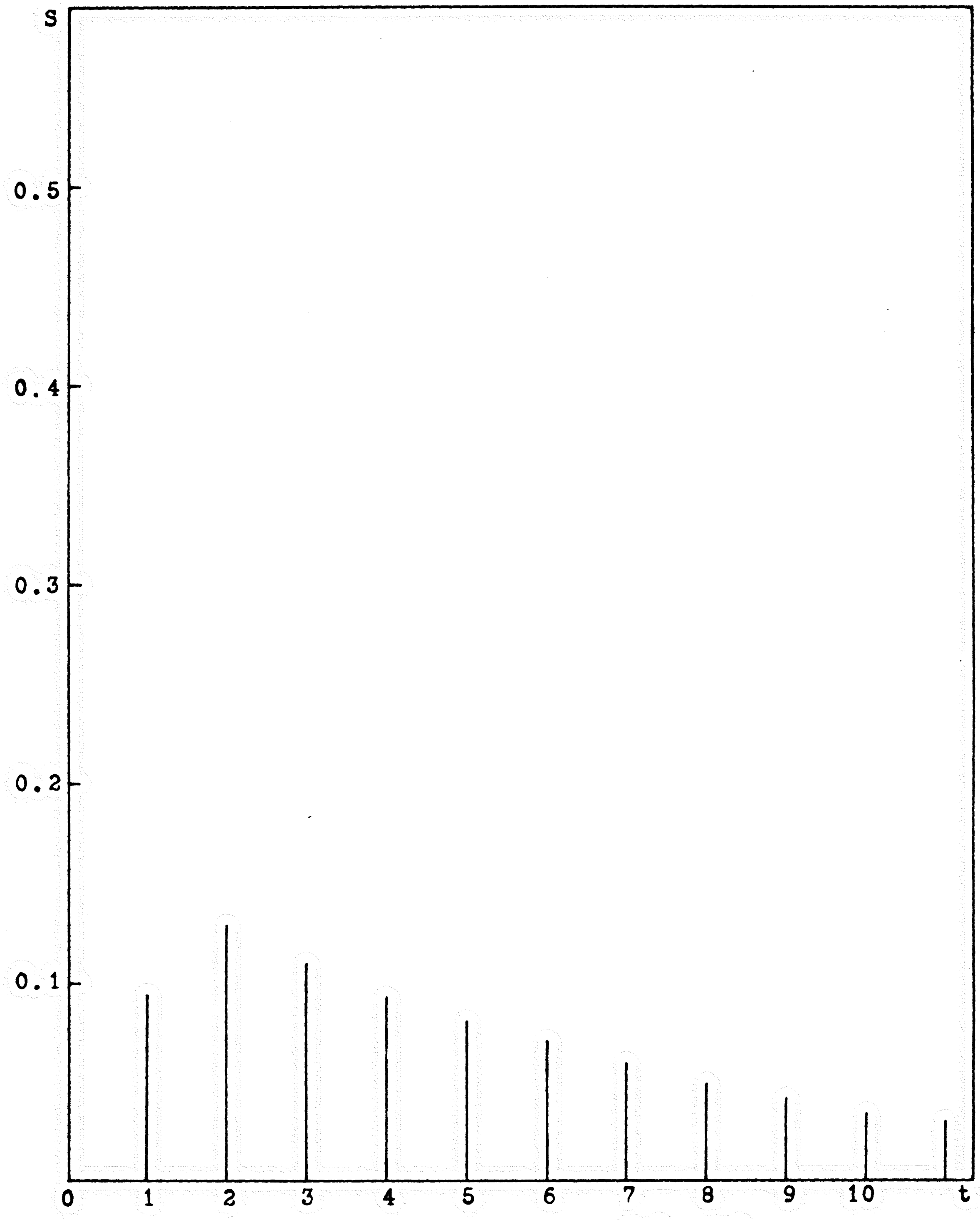

FIGURE 11. The probability function $S_2^{(t)} = S_5^{(t)}$, M=7.

$\mu = 7.31707$, $\sigma^2 = 42.1441$

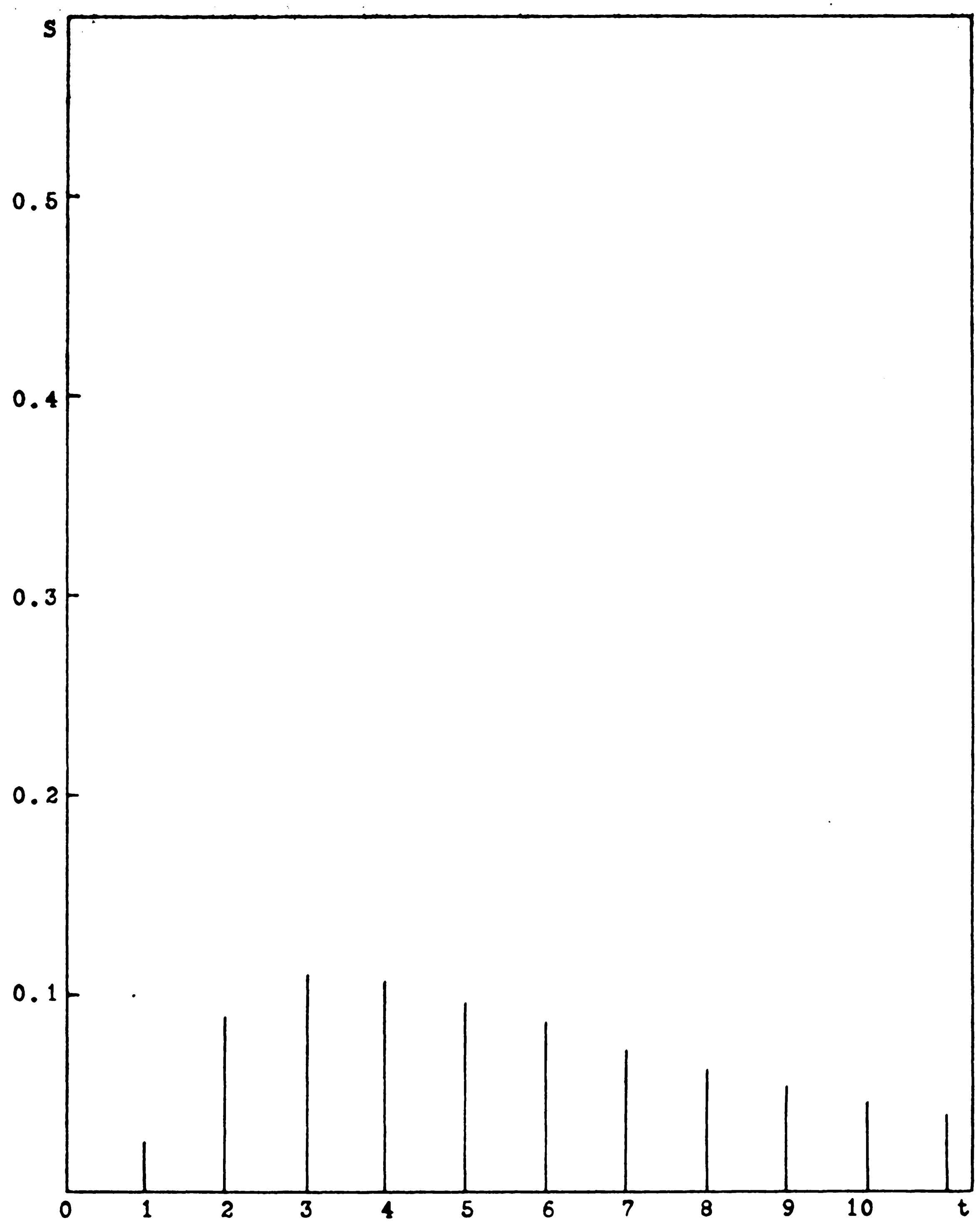

FIGURE 12. The probability function $S_3^{(t)} = S_4^{(t)}$, M=7.

$\mu = 8.41090$, $\sigma^2 = 43.1802$

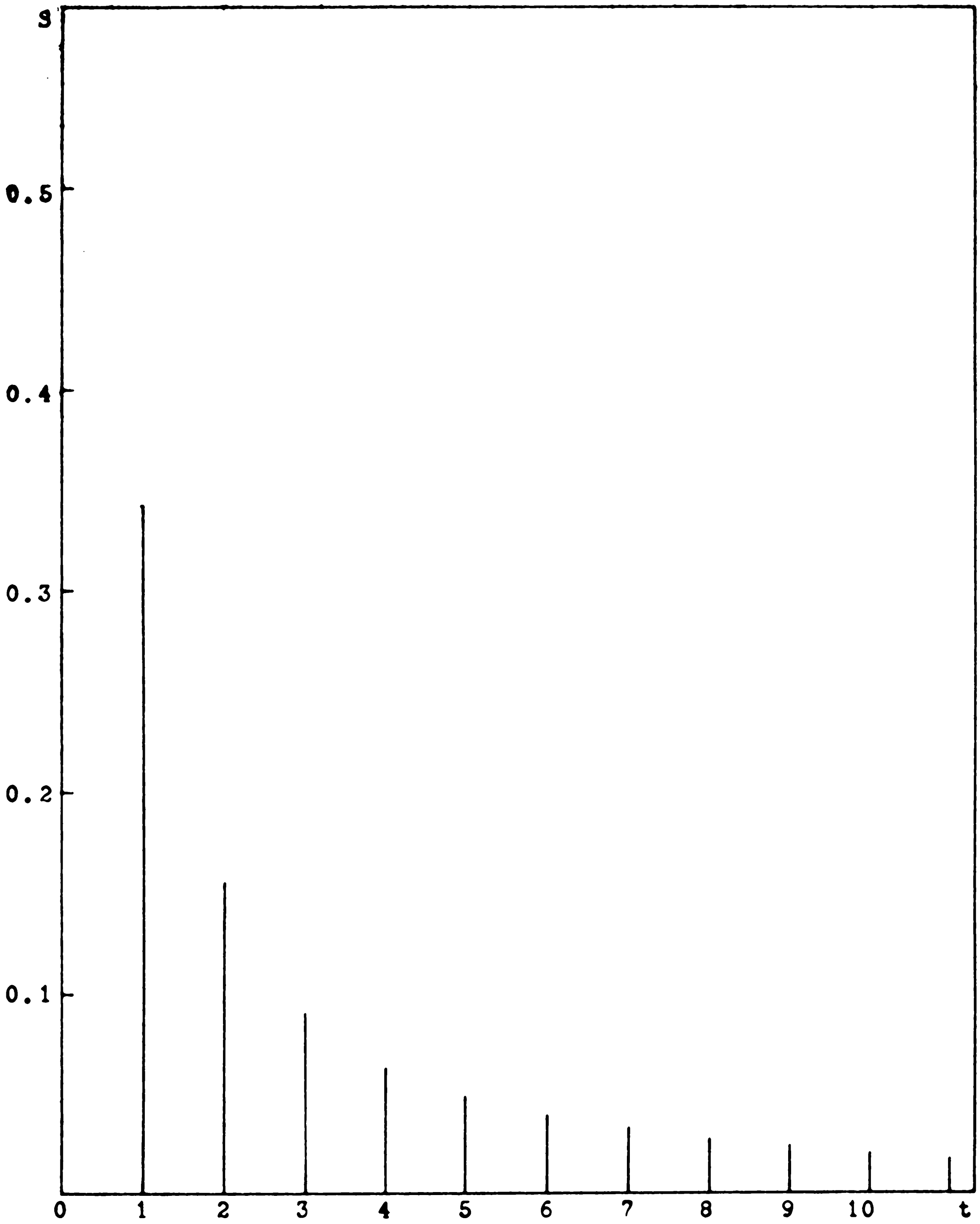

FIGURE 13. The probability function $S_1^{(t)} = S_7^{(t)}$, M=8.

$\mu = 5.26484$, $\sigma^2 = 42.4768$

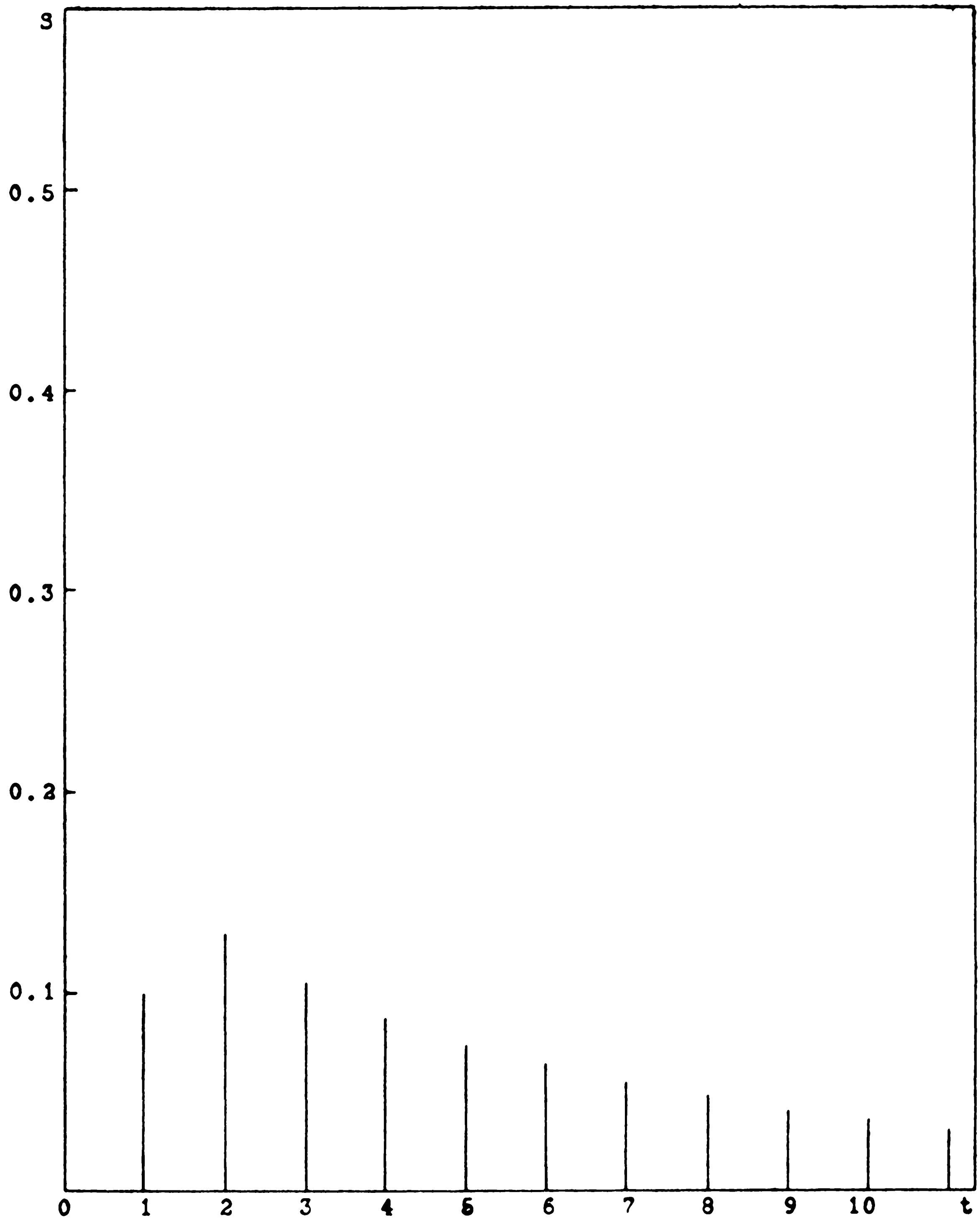

FIGURE 14. The probability function $S_2^{(t)}=S_6^{(t)}$, M=8.

$\mu = 7.93824$, $\sigma^2 = 54.9874$

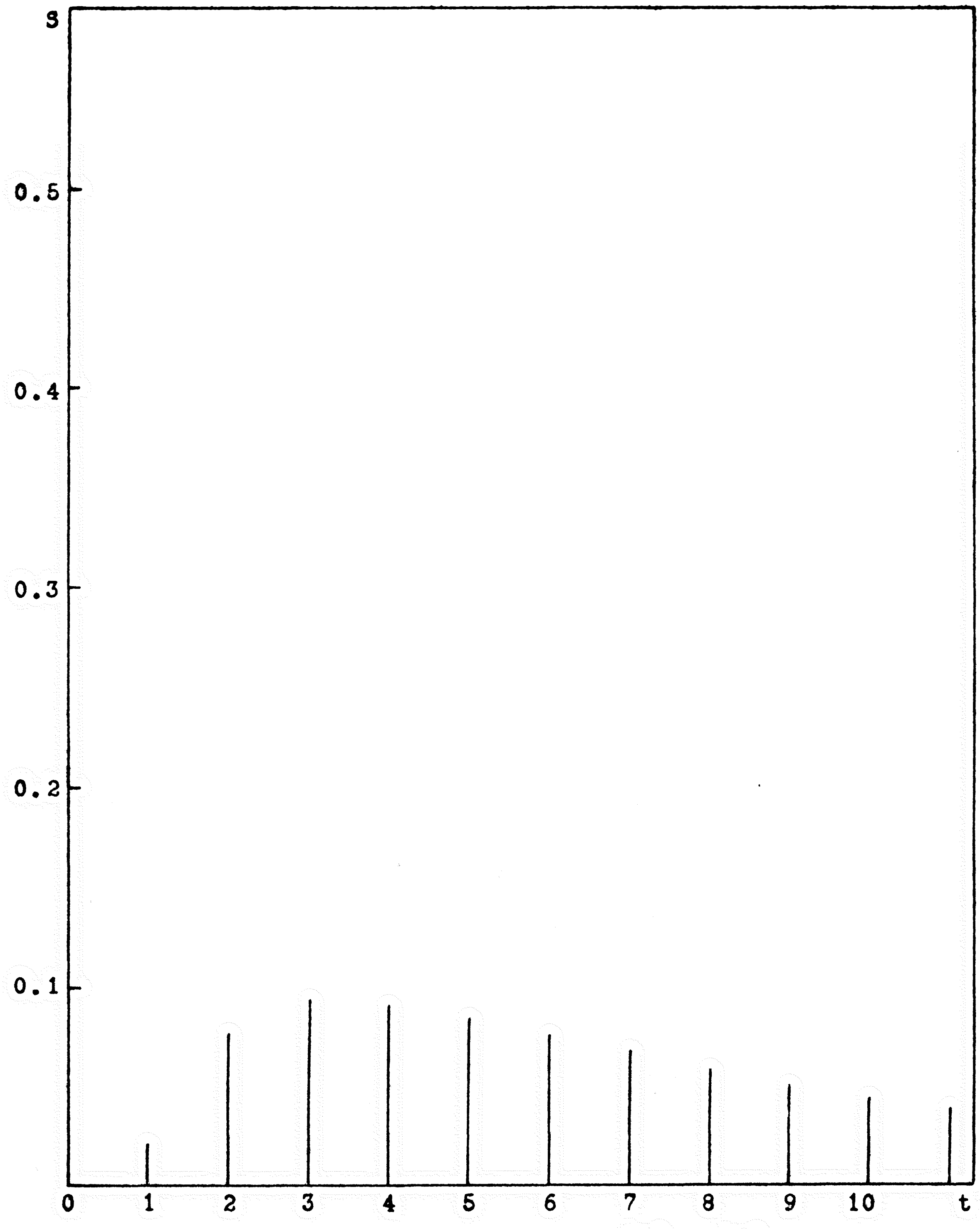

FIGURE 15. The probability function $S_3^{(t)} = S_5^{(t)}$, M=8.

$\mu = 9.41533, \sigma^2 = 57.4626$

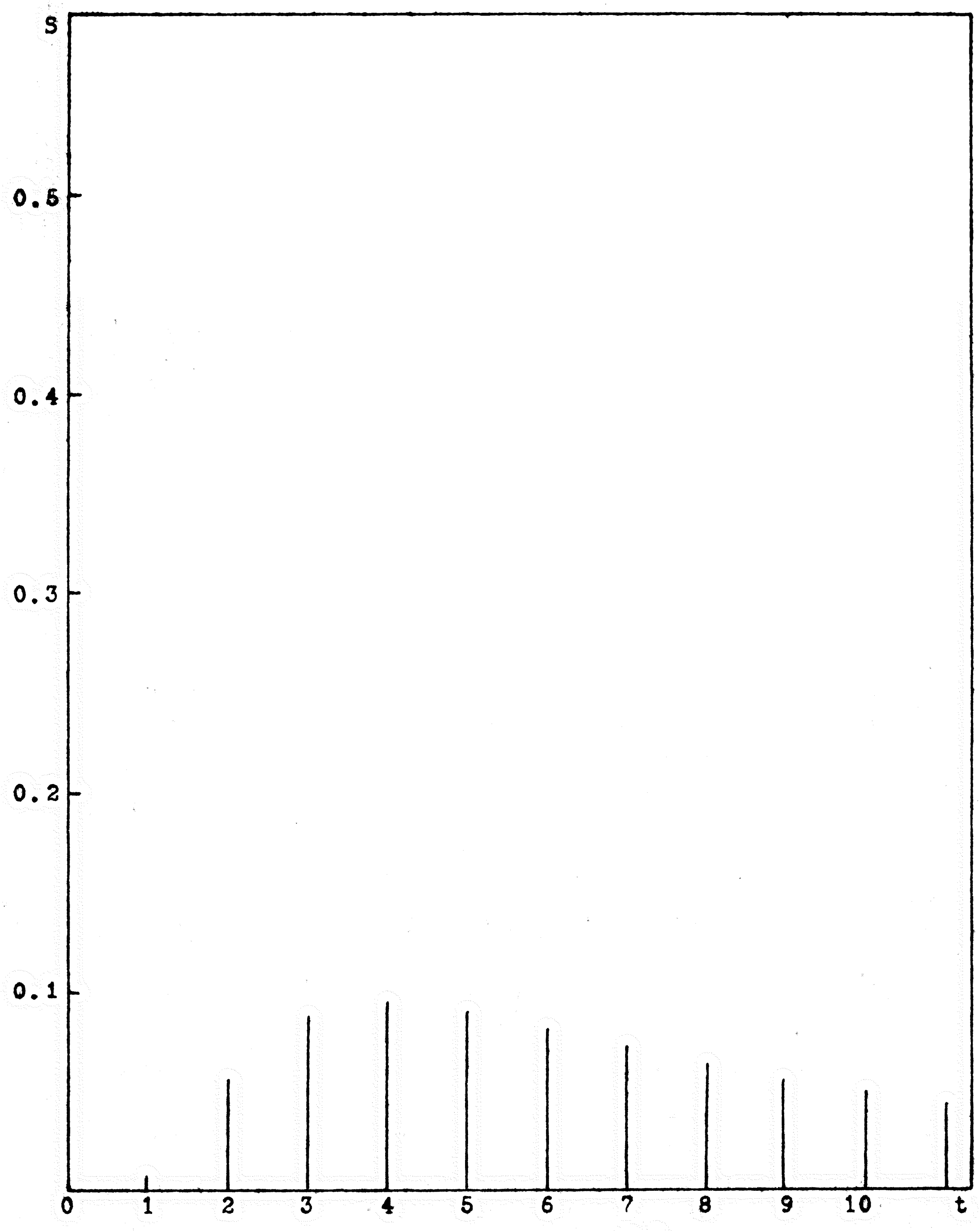

FIGURE 16. The probability function $S_4^{(t)}$, M=8.

$\mu = 9.88868, \sigma^2 = 57.6050$

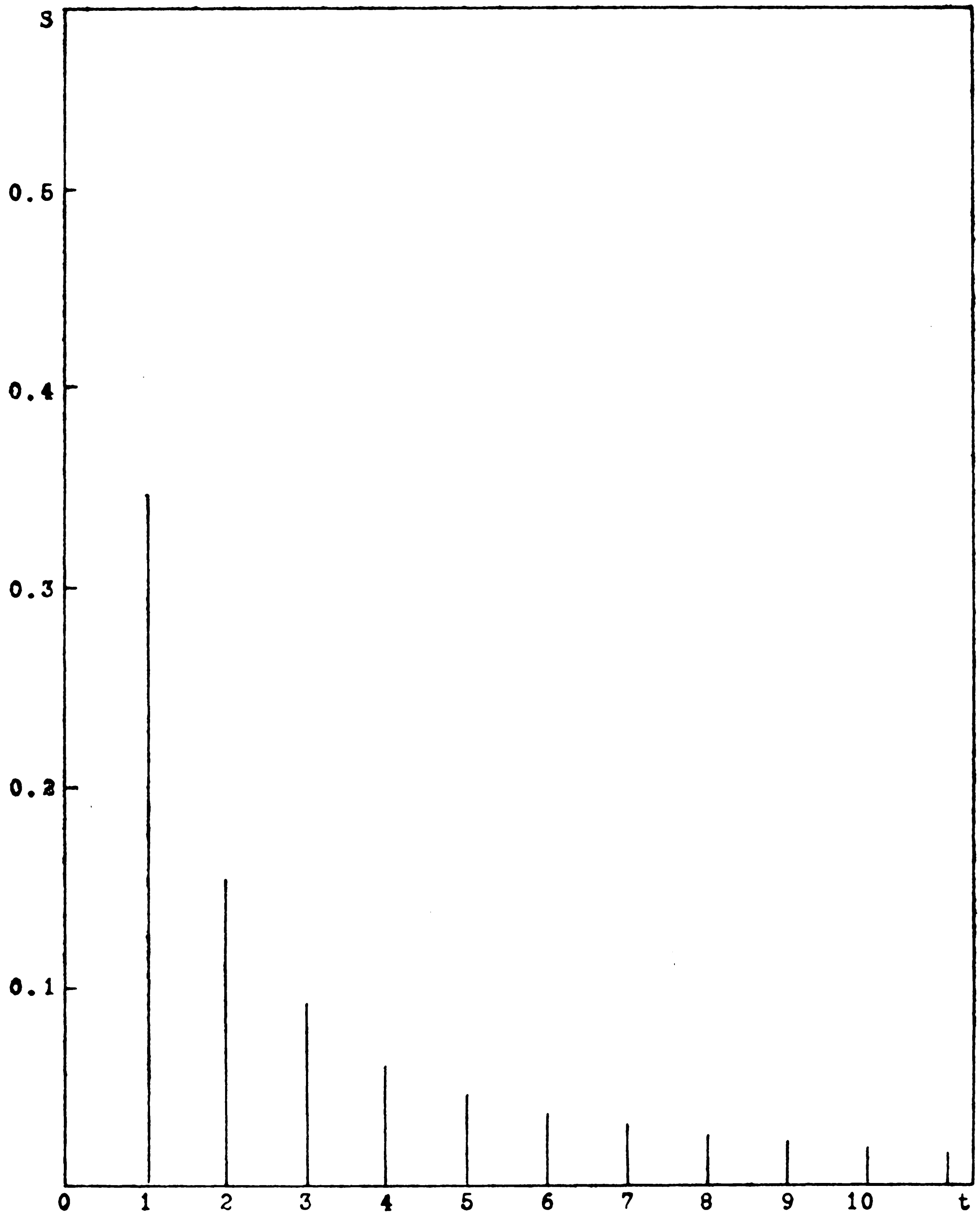

FIGURE 17. The probability function $S_1^{(t)} = S_8^{(t)}$, $M=9$.

$\mu = 5.51784$, $\sigma^2 = 51.5569$

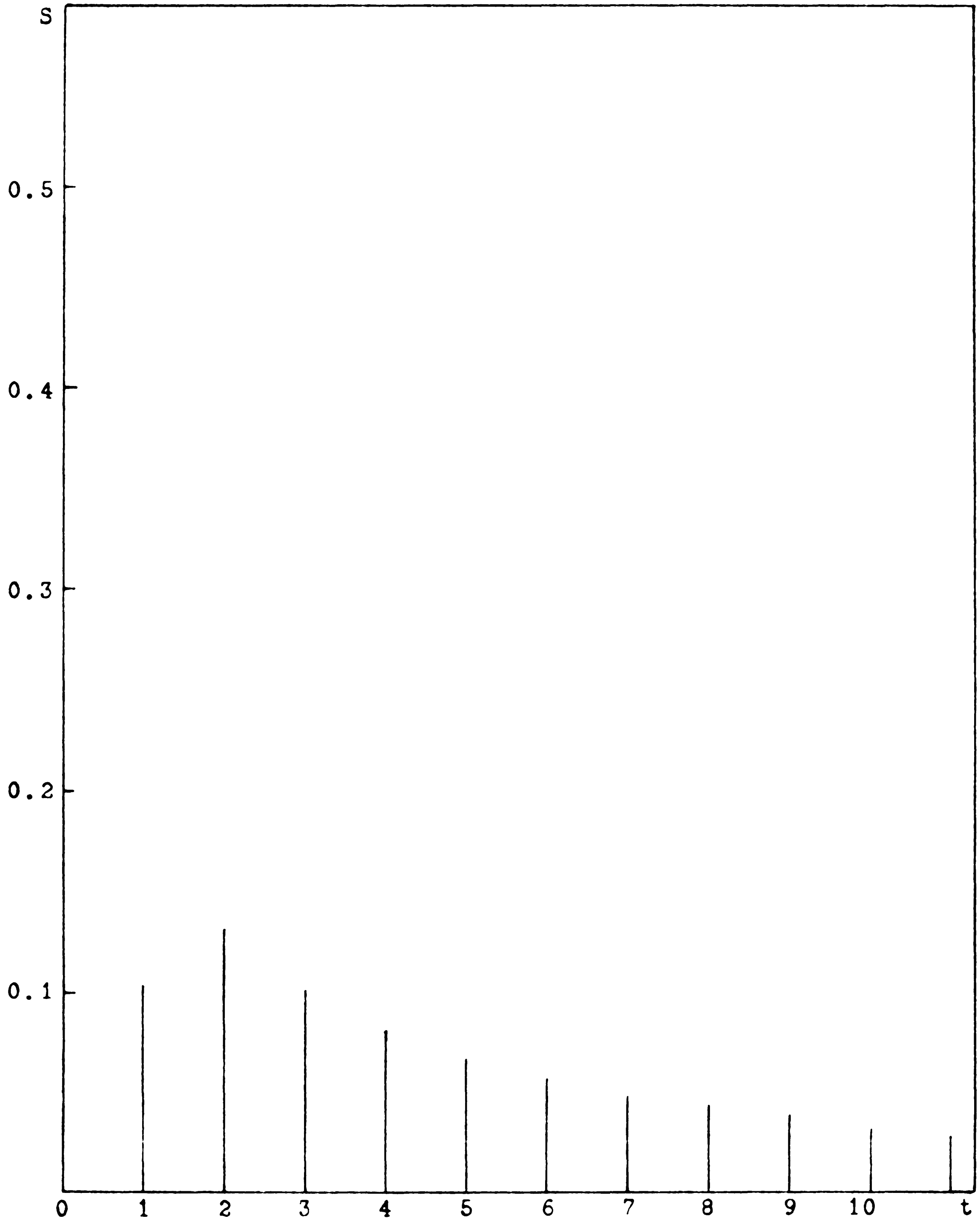

FIGURE 18. The probability function $S_2^{(t)}=S_7^{(t)}$, M=9.

$\mu = 8.48487$, $\sigma^2 = 68.7845$

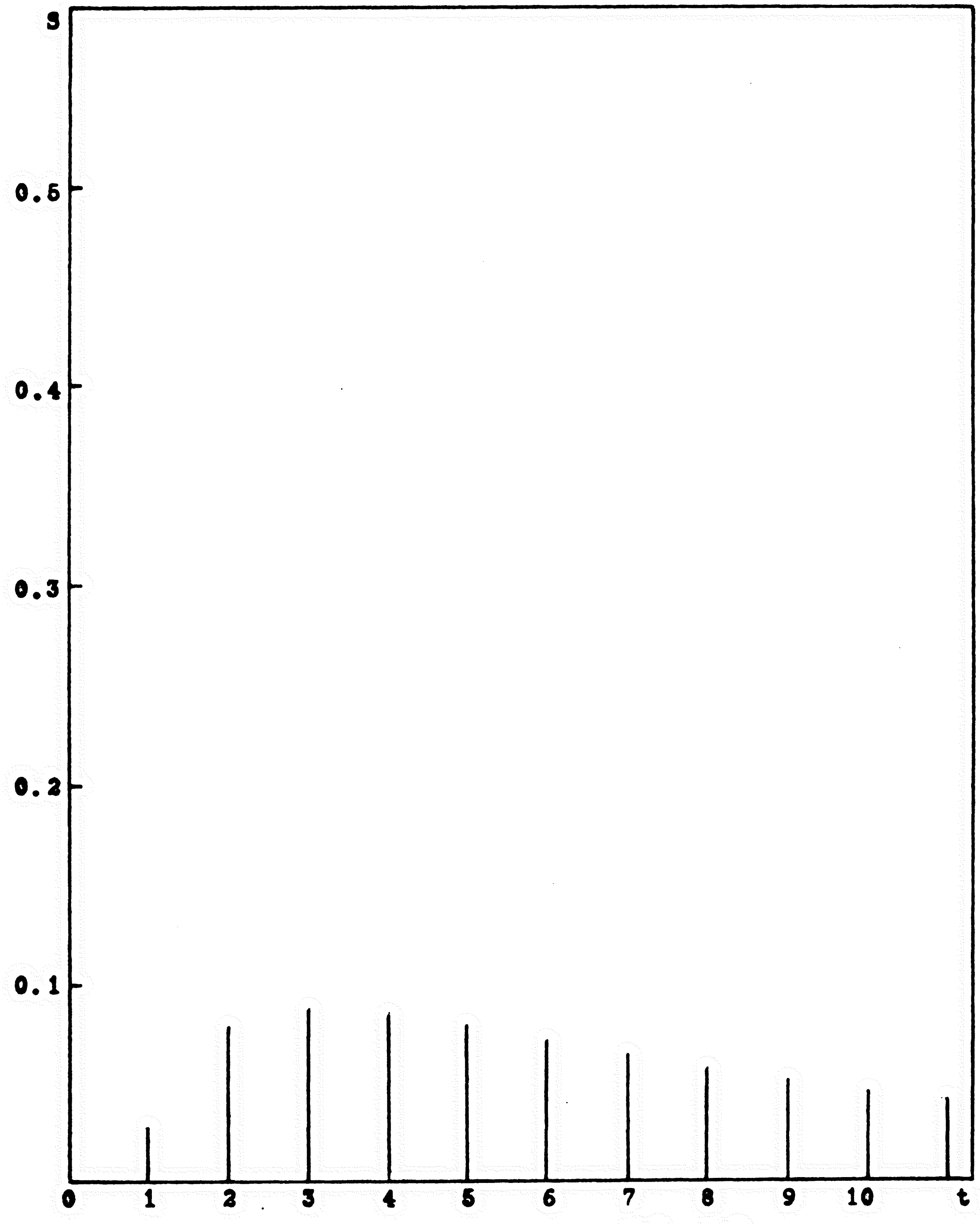

FIGURE 19. The probability function $S_3^{(t)}=S_6^{(t)}$, M=9.

$\mu = 10.2805$, $\sigma^2 = 73.4074$

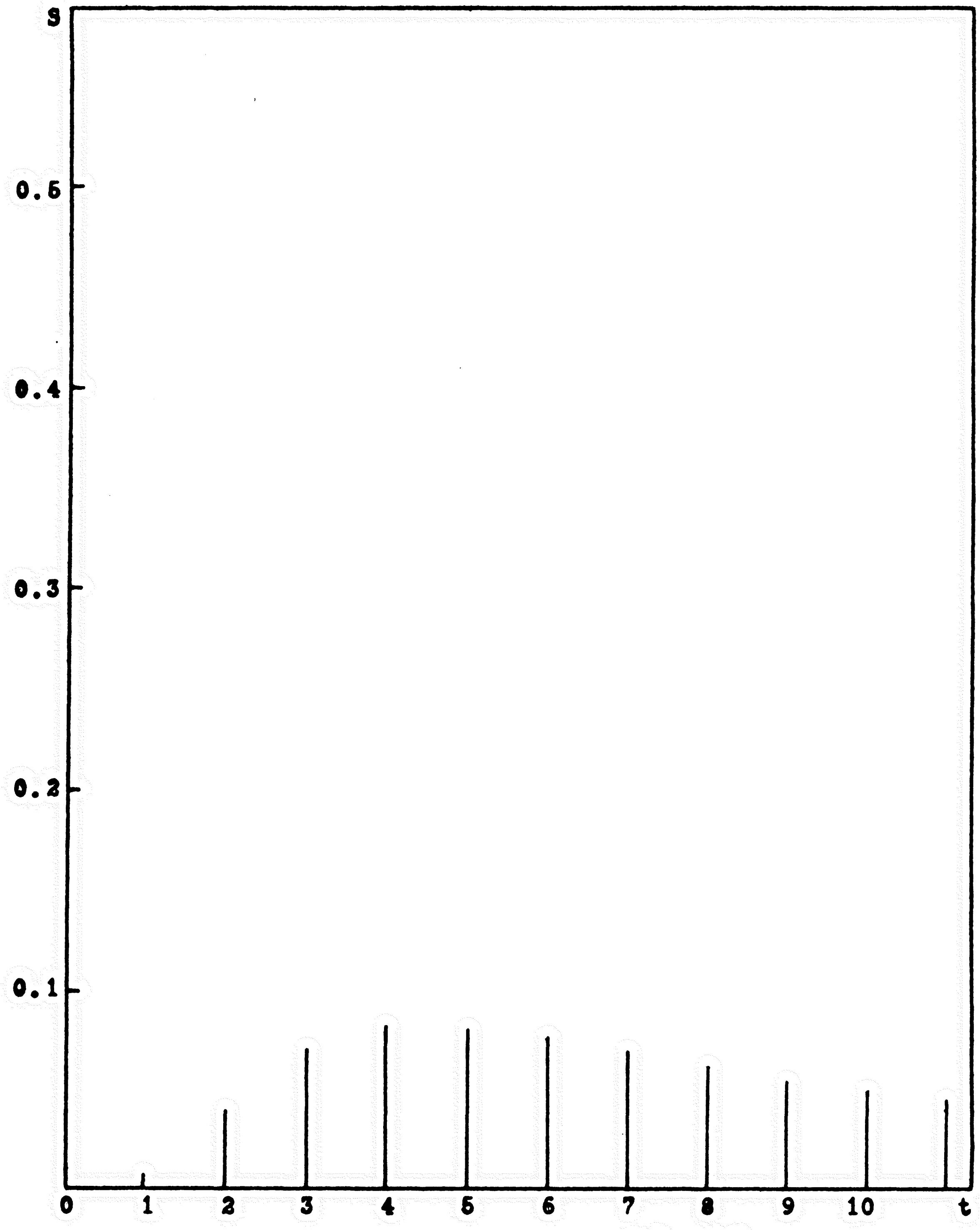

FIGURE 20. The probability function $S_4^{(t)}=S_5^{(t)}$, M=9.

$\mu = 11.1276,\ \sigma^2 = 74.2076$

IV. THE MOMENT APPROACH

4.1. Alternate method. Since the attempt in the previous chapter to find the distribution of time to absorption for general M was not successful, let us turn our attention to a study of some moments of the distribution. For a large population size M, it is possible to approximate the Markov chain by means of a diffusion process continuous in space and time. This approximation will be considered in Chapter V, but, in order to judge the accuracy of the approximation, exact results will be needed for the comparison. The present chapter will thus be concerned with obtaining the first and second moments of absorption time for specific values of M.

Because of the difficulties in numerical computation encountered by using the eigenvector approach, an alternative approach in which the initial state i is assumed to be transient will be used. The expressions for the first and second moments stated in that which follows have been obtained by Watterson [17] and by Kemeny and Snell [12] using different notations and derivations. In general, the method of derivation of Watterson will be used.

4.2. Moment theory. For convenience, let us consider the canonical form (see |12|) of the transition matrix P, (2.2.2), in an aggregated version in which the two absorbing sets and the M-1 transient sets are united. Thus, the canonical form of P may be written

$$P = \left[\begin{array}{c:c} I & 0 \\ \hdashline R & Q \end{array}\right].$$

Note that the matrix Q is obtained by deleting the first and last rows and columns of P and hence is not stochastic since some non-zero elements of the stochastic matrix P have been removed. Letting

$$\underline{s}_Q^{(t)} = \begin{bmatrix} s_1^{(t)} \\ s_2^{(t)} \\ \vdots \\ s_{M-1}^{(t)} \end{bmatrix},$$

we find from (3.1.4) that

$$\underline{s}_Q^{(0)} = \underline{0},$$

(4.2.1)
$$\underline{s}_Q^{(1)} = \begin{bmatrix} p_{10} + p_{1M} \\ p_{20} + p_{2M} \\ \vdots \\ p_{M-1,0} + p_{M-1,M} \end{bmatrix}$$

that is, $\underline{s}_Q^{(1)} = (I-Q)\underline{j}$, where $\underline{j}$ is a column vector of ones and $\underline{j}$ and the identity matrix I are both of order M-1.

Then corresponding to (3.1.6) we have

(4.2.2)
$$\begin{aligned} \underline{s}_Q^{(t)} &= Q^{t-1}\, \underline{s}_Q^{(1)} \\ &= Q^{t-1}\, (i-Q)\underline{j}. \end{aligned}$$

The probability generating function may be written

$$\underline{G}_Q(z) = \sum_{t=0}^{\infty} z^t \, \underline{S}_Q^{(t)},$$

or, since $\underline{S}_Q^{(0)} = \underline{0}$,

$$\underline{G}_Q(z) = \sum_{t=1}^{\infty} z^t \, \underline{S}_Q^{(t)},$$

which by (4.2.2) is

$$\underline{G}_Q(z) = z(I-Q)\underline{j} + z \sum_{t=2}^{\infty} z^{t-1} \, \underline{S}_Q^{(t)}$$

$$\begin{aligned} &= z(I-Q)\underline{j} + z \sum_{t=2}^{\infty} (zQ)^{t-1}(I-Q)\underline{j} \\ (4.2.3) \qquad &= z(I-Q)\underline{j} + z^2 Q(I-zQ)^{-1}(I-Q)\underline{j} \\ &= (z^{-1}I-Q)^{-1}(I-Q)\underline{j}. \end{aligned}$$

The first moment of absorption time may now be obtained by differentiating (4.2.3) with respect to z and evaluating the result at z=1. This gives

$$\begin{aligned} \underline{\mu} &= \frac{d}{dz} \underline{G}_Q \, (z)\Big]_{z=1} \\ (4.2.4) \qquad &= \left[z^{-2}(z^{-1}I-Q)^{-2}\right]_{z=1}(I-Q)\underline{j} \\ &= (I-Q)^{-1}\underline{j}. \end{aligned}$$

In [12], the matrix $(I-Q)^{-1}$ is called the "fundamental matrix", and is denoted by the symbol N.

Evaluating the second derivative of (4.2.3) with respect to z at z=1 gives the second factorial moment which is found to be

$$\begin{aligned}\frac{d^2}{dz^2}\ \underline{G}_Q(z)]_{z=1} &= [2z^{-4}(z^{-1}I-Q)^{-3}-2z^{-3}\\ &\qquad (z^{-1}I-Q)^{-2}]_{z=1}(I-Q)\underline{j}\\ &= [2(I-Q)^{-3}-2(I-Q)^{-2}](I-Q)\underline{j}\\ &= 2[(I-Q)^{-1}-I](I-Q)^{-1}\underline{j}\\ &= 2[(I-Q)^{-1}-I]\underline{\mu}.\end{aligned}\tag{4.2.5}$$

Using the well-known relation between factorial moments and ordinary moments of the second order, we obtain from (4.2.4) and (4.2.5) the following expression for the variance of time to absorption.

$$\begin{aligned}\underline{\sigma}^2 &= 2[(I-Q)^{-1}-I]\underline{\mu} + \underline{\mu} - D_{\underline{\mu}}\,\underline{\mu}\\ &= [2(I-Q)^{-1}-I-D_{\underline{\mu}}]\underline{\mu},\end{aligned}\tag{4.2.6}$$

where $D_{\underline{\mu}}$ denotes a diagonal matrix with the elements of $\underline{\mu}$ in the diagonal. Higher moments of the distribution may be obtained in a similar manner.

4.3. Numerical results. In order to obtain the means and variances of the distributions for particular values of M and the various initial states by (4.2.4) and (4.2.6) it is necessary that one know the inverse of I-Q. The process of inversion of I-Q requires lengthy computations, but, fortunately, the form of this matrix is such that the blocking transformation described in Section 3.2 may be applied to give

$$C^{-1}(I-Q)C = \left[\begin{array}{c|c} D & 0 \\ \hline 0 & S \end{array}\right],$$

or

$$(I-Q)^{-1} = C\left[\begin{array}{c|c} D^{-1} & 0 \\ \hline 0 & S^{-1} \end{array}\right]C^{-1}$$

where D and S denote matrices quite similar to those derived in Chapter III.

Using the above and assuming M to be odd, (4.2.4) may be written

$$\begin{aligned} \underline{\mu} &= (I-Q)^{-1}\underline{j} \\ &= C\left[\begin{array}{c|c} D^{-1} & 0 \\ \hline 0 & S^{-1} \end{array}\right]C^{-1}\underline{j} \\ &= C\left[\begin{array}{c|c} D^{-1} & 0 \\ \hline 0 & S^{-1} \end{array}\right]\left[\begin{array}{c} \underline{0} \\ \hline \sqrt{2}\,\underline{j} \end{array}\right] \\ &= \left[\begin{array}{c|c} I & J \\ \hline -J & I \end{array}\right]\left[\begin{array}{c} \underline{0} \\ \hline S^{-1}\underline{j} \end{array}\right] \\ &= \left[\begin{array}{c} JS^{-1}\underline{j} \\ \hline S^{-1}\underline{j} \end{array}\right], \end{aligned}$$

where $JS^{-1}\underline{j}$ and $S^{-1}\underline{j}$ are both of order $\frac{M-1}{2}$. Similarly, if M is even we have

$$\underline{\mu} = C\begin{bmatrix} D^{-1} & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & S^{-1} \end{bmatrix}\begin{bmatrix} \underline{0} \\ \cdots\cdots \\ 1 \\ \cdots\cdots \\ \sqrt{2}\,\underline{j} \end{bmatrix}$$

$$= CS^{-1}\begin{bmatrix} 1 \\ \cdots\cdots \\ \sqrt{2}\,\underline{j} \end{bmatrix}$$

$$= \begin{bmatrix} JS_{\Delta}^{-1}\begin{bmatrix} 1/\sqrt{2} \\ \cdots\cdots \\ \underline{j} \end{bmatrix} \\ \cdots\cdots\cdots\cdots \\ (\underline{s}_1^{-1})'\begin{bmatrix} 1/\sqrt{2} \\ \cdots\cdots \\ \underline{j} \end{bmatrix} \\ \cdots\cdots\cdots\cdots \\ S^{-1}\begin{bmatrix} 1/\sqrt{2} \\ \cdots\cdots \\ \underline{j} \end{bmatrix} \end{bmatrix}$$

where S_{Δ}^{-1} denotes the matrix S^{-1} with the first row, $(\underline{s}_1^{-1})'$, deleted and is of order $\frac{M}{2}$. We see then that only the inverse of S need be determined in order to obtain $\underline{\mu}$, and hence, $\underline{\sigma}^2$, both of which, it is noted, are symmetric column vectors.

The theory just presented has been applied to obtain the means and variances which are given in Table 4.1. For simplicity of programming the IBM 650 a slight variation of the blocking transformation was used and a Runcible I program was written to generate the matrix corresponding to S for M = 10,20,30,40,50. A standard matrix inversion program was then used and the means and variances of time to absorption calculated for the different states. For the convenience of the reader, the means and variances for M = 2, 3,...,9 have also been tabulated. These may be obtained from the results of Chapter III or by the method demonstrated in Chapter VIII. As previously mentioned, these values will be compared with the diffusion approximation obtained in Chapter V.

TABLE 4.1

Means and variances of time to absorption

M	i	$\mu_{i,M}$	$\sigma^2_{i,M}$
2	1	2.00000	2.00000
3	1	3.00000	6.00000
4	1	3.68966	11.5291
	2	4.55172	12.3615
5	1	4.20792	18.2041
	2	5.69307	20.5421
6	1	4.62300	25.7099
	2	6.58484	30.5728
	3	7.20503	30.8424
7	1	4.96877	33.8459
	2	7.31707	42.1441
	3	8.41090	43.1802
8	1	5.26484	42.4768
	2	7.93824	54.9874
	3	9.41533	57.4626
	4	9.88869	57.6050
9	1	5.41784	51.5569
	2	8.48487	68.7845
	3	10.2805	73.4074
	4	11.1276	74.2076
10	1	5.75328	60.8664
	2	8.95376	83.6956
	3	11.0311	91.0090
	4	12.2084	92.5574
	5	12.5905	92.6548
20	1	7.23122	164.992
	2	11.9807	261.689

TABLE 4.1 (continued)

M	i	$\mu_{i,M}$	$\sigma^2_{i,M}$
20	3	15.7303	320.090
	4	18.7351	354.926
	5	21.1380	374.672
	6	23.0246	385.159
	7	24.4486	389.918
	8	25.4446	391.697
	9	26.0342	392.032
	10	26.2295	392.058
30	1	8.07599	278.556
	2	13.6922	467.616
	3	18.3442	601.425
	4	22.2921	697.309
	5	25.6837	765.746
	6	28.6096	813.884
	7	31.1316	846.861
	8	33.2860	868.886
	9	35.1292	882.626
	10	36.6200	891.018
	11	37.8369	895.463
	12	38.7717	897.540
	13	39.4336	898.303
	14	39.8283	898.477
	15	39.9595	898.489
40	1	8.66930	396.453
	2	14.8895	686.828
	3	20.1624	908.373
	4	24.7502	1080.22
	5	28.8019	1214.14
	6	32.4098	1318.30
	7	35.6355	1398.80
	8	38.5227	1460.37
	9	41.1037	1506.81
	10	43.4032	1541.20
	11	45.4406	1566.07
	12	47.2313	1583.53
	13	48.7874	1595.32
	14	50.1188	1602.89
	15	51.2334	1607.43
	16	52.1375	1609.90
	17	52.8360	1611.07
	18	53.3324	1611.50
	19	53.6292	1611.59
	20	53.7280	1611.60

TABLE 4.1 (continued)

M	i	$\mu_{i,M}$	$\sigma^2_{i,M}$
50	1	9.12677	516.920
	2	15.8108	914.023
	3	21.5574	1230.90
	4	26.6297	1488.05
	5	31.1774	1698.25
	6	35.2936	1870.43
	7	39.0404	2011.34
	8	42.4623	2126.26
	9	45.5925	2219.49
	10	48.4566	2294.58
	11	51.0747	2354.49
	12	53.4633	2401.75
	13	55.6358	2438.52
	14	57.6031	2466.64
	15	59.3745	2487.70
	16	60.9578	2503.09
	17	62.3593	2513.98
	18	63.5847	2521.39
	19	64.6382	2526.17
	20	65.5238	2529.06
	21	66.2444	2530.63
	22	66.8024	2531.37
	23	67.1998	2531.65
	24	67.4377	2531.71
	25	67.5169	2531.72

Here, as in the preceding chapter, inferences relative to the results of this chapter will be made in Chapter VI.

V. THE DIFFUSION PROCESS

5.1. Probability of absorption. Under the assumption of large population size M, we can approximate the Markov chain by means of a diffusion process continuous in space and time. Let us make the time scale transformation $u=M^{-1}t$ and the state transformation $y(u) = M^{-1}x(Mu)$ where the random variable $x(Mu=t)$, $t=0,1,2,\dots$, denotes the frequency of A-genes. Since $x = 0,1,2,\dots,M$, we have $y=0,M^{-1},2M^{-1},\dots,1$, and letting $M\to\infty$ but keeping u fixed, it can be shown that the distribution of y(u) approaches a distribution function which has jumps at y=0 and y=1 but is differentiable in the open interval (0,1). Similar examples may be seen in [1], [16], and [17]. Simply stated, we may say that, for sufficiently large M, the discrete variable y(u) has an approximately continuous distribution within (0,1). Denoting the derivative of this distribution by f(y,u), it may be shown (see [1]) that

$$\frac{\partial f(y,u)}{\partial u} = \frac{1}{2}\,\frac{\partial^2 y(1-y)f(y,u)}{\partial y^2}, \quad 0<y<1,$$

and f(y,u) will behave as an approximate density for y(u) with the exceptions of the accumulations of probability at y=0 and y=1. This is a special case of the Kolmogorov or Fokker-Planck diffusion equation and requires a specification of the initial function f(y,0) for its unique solution.

Using the approach of [1], [13], or [17], the probability that the diffusing state is absorbed at $y=1$ at or before time u may be found by solving the backward equation

$$(5.1.1) \qquad \frac{\partial G(p,u)}{\partial u} = \frac{1}{2}\, p(1-p)\, \frac{\partial^2 G(p,u)}{\partial p^2} \,,$$

where $p=y(0)=M^{-1}x(0)=iM^{-1}$, with the boundary conditions $G(0,u)=0$, $G(1,u)=1$ being satisfied for all $0< u$ and $G(p,0)=0$ for $0< p< 1$. By [13], the solution of this equation is

$$\begin{aligned} G(p,u) &= \Pr\{\text{absorption in state } M \text{ at or before } u\} \\ (5.1.2) \qquad &= p + \sum_{i=1}^{\infty} (2i+1)\, p(1-p)(-1)^{i} \\ &\qquad F(1-i,i+2,2;p)\, e^{-\frac{1}{2}i(i+1)u} \,, \end{aligned}$$

where $F(a,b,c;z)$ is the hypergeometric function defined by the series

$$F(a,b,c;z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}$$

with $(x)_0=1$, $(x)_k=x(x+1)\cdots(x+k-1)$.

By symmetry, the probability of absorption in either state is given by

$$G(p,u)+G(1-p,u).$$

Thus we may find the approximate probability of absorption at exactly time $t=Mu$ by using

$$\begin{aligned} \Pr\{T_i=t\} &= s_i^{(t)} \\ (5.1.3) \qquad &\doteqdot G(p,M^{-1}t) + G(1-p,M^{-1}t) \\ &\qquad -G[p,M^{-1}(t-1)] - G[1-p,M^{-1}(t-1)]. \end{aligned}$$

5.2. Moments. We may approximate the moments of the first absorption time T_i by means of the distribution given in (5.1.3). These moments are calculated, however, only with considerable difficulty. A simpler procedure for determining the mean first passage time makes use of the following. Let U(p) be the expected value of the time, measured on the u-scale, to first reach one or the other absorbing state, that is,

$$U(p) = \int_0^\infty u\,d[G(p,u) + G(1-p,u)].$$

Following Feller [8], letting

$$g(p,u) = \frac{\partial G}{\partial u}$$

and differentiating (5.1.1) with respect to u, we have

$$\frac{\partial^2 G}{\partial u^2} = \frac{1}{2}p(1-p)\,\frac{\partial^2}{\partial p^2}\left(\frac{\partial G}{\partial u}\right)$$

$$= \frac{1}{2}p(1-p)\,\frac{\partial^2 g}{\partial p^2},$$

or

$$(5.2.1)\qquad \frac{\partial g}{\partial u} = \frac{1}{2}p(1-p)\,\frac{\partial^2 g}{\partial u^2}\ .$$

Multiplying both members of (5.2.1) by u and integrating from 0 to ∞ gives

$$-1 = \frac{1}{2}p(1-p)\,\frac{\partial^2}{\partial p^2}[U(p)],$$

which has, under the boundary conditions U(0) = U(1) = 0, the solution

$$U(p) = 2 \log[p^{-p}(1-p)^{-(1-p)}],$$

or, measured on the t-scale with $p = iM^{-1}$,

$$(5.2.2) \qquad \mu_{i,M} = E(T_i) \doteq 2M \log[p^{-p}(1-p)^{-(1-p)}] = \overset{*}{\mu}_{i,M}.$$

Similarly, the second moment about the origin of the distribution, say $V(p)$, may be found by multiplying both sides of (5.2.1) by u^2 and integrating from 0 to ∞ to obtain

$$(5.2.3) \qquad -2U(p) = \frac{1}{2}\, p(1-p)\, \frac{\partial^2 V}{\partial p^2}$$

with boundary conditions $V(0) = V(1) = 0$ and where $U(p)$ is as given above. The solution of (5.2.3) is

$$V(p) = 4\left\{\frac{\pi^2}{3} - 2\sum_{n=1}^{\infty} \frac{p^{n+1} + (1-p)^{n+1}}{n^2} - 2 \log[p^{-p}(1-p)^{-(1-p)}]\right\},$$

or measured on the t-scale,

$$M^2 V(p) = 4M^2\left\{\frac{\pi^2}{3} - 2\sum_{n=1}^{\infty} \frac{p^{n+1} + (1-p)^{n+1}}{n^2} - 2 \log[p^{-p}(1-p)^{-(1-p)}]\right\}.$$

The variance of the first passage time may now be approximated by making the mean correction on the above which gives

$$\sigma^2_{i,M} = \mathrm{Var}(T_i)$$

$$(5.2.4) \qquad \doteq 4M^2\left\{\frac{\pi^2}{3} - 2\sum_{n=1}^{\infty} \frac{p^{n+1} + (1-p)^{n+1}}{n^2} - 2 \log[p^{-p}(1-p)^{-(1-p)}]\right\} - [E(T_i)]^2 = \overset{*}{\sigma}{}^2_{i,M}$$

where the approximation given by (5.2.2) is used for $E(T_i)$.

5.3. Numerical results and comparisons. The question now arises concerning the accuracy of the approximations given by the diffusion process. This question of accuracy may be partially answered by calculating the values of $E(T_i)$ and $Var(T_i)$ as given by (5.2.2) and (5.2.4) for the values of i and M appearing in Table 4.1 and comparing results. These approximations for the means and variances are tabulated in Table 5.1 along with the percentage error, $\frac{\mu^*_{i,M} - \mu_{i,M}}{\mu_{i,M}} \cdot 100\% = E_\mu\%$ and $\frac{\sigma^{*2}_{i,M} - \sigma^2_{i,M}}{\sigma^2_{i,M}} \cdot 100\% = E_\sigma\%$, when compared with the entries in Table 4.1. By noting in Table 5.1 the rapidity of

TABLE 5.1

Diffusion approximations and per cent error.

M	i	$\mu^*_{i,M}$	$\sigma^{*2}_{i,M}$	$E_\mu\%$	$E_\sigma\%$
2	1	2.77259	4.13824	38.63	106.9
3	1	3.81908	9.23465	27.30	53.91
4	1	4.49868	15.8361	21.93	37.36
	2	5.54518	16.5529	21.82	33.91
5	1	5.00403	23.4588	18.92	28.87
	2	6.73012	25.8370	18.22	25.78
6	1	5.40673	31.8072	16.95	23.72
	2	7.63817	36.9386	16.00	20.82
	3	8.31777	37.2441	15.44	20.75
7	1	5.74163	40.6994	15.55	20.25
	2	8.37577	49.5328	14.47	17.53
	3	9.56071	50.6797	13.67	17.37

TABLE 5.1 (continued)

M	i	$\mu^*_{i,M}$	$\sigma^{*2}_{i,M}$	$E_\mu\%$	$E_\sigma\%$
8	1	6.02832	50.0150	14.50	17.75
	2	8.99736	63.3445	13.34	15.20
	3	10.5850	66.0422	12.42	14.94
	4	11.0904	66.2118	12.15	14.94
9	1	6.27898	59.7603	13.79	15.74
	2	9.53471	78.1678	12.37	13.64
	3	11.5473	83.1119	11.45	13.22
	4	12.3653	83.7910	11.12	12.91
10	1	6.50166	69.6053	13.01	14.36
	2	10.0080	93.8351	11.77	12.11
	3	12.2173	101.671	10.75	11.71
	4	13.4602	103.348	10.25	11.66
	5	13.8629	103.456	10.11	11.66
20	1	7.94061	177.859	9.81	7.80
	2	13.0033	278.421	8.54	6.39
	3	16.9084	339.150	7.49	5.95
	4	20.0161	375.341	6.83	5.75
	5	22.4934	395.903	6.41	5.64
	6	24.4346	406.685	6.12	5.59
	7	25.8979	411.610	5.92	5.56
	8	26.9205	413.392	5.80	5.54
	9	27.5256	413.797	5.73	5.55
	10	27.7259	413.824	5.70	5.55
30	1	8.76868	294.076	8.58	5.57
	2	14.6958	488.795	7.33	4.53
	3	19.5050	626.447	6.33	4.16
	4	23.5605	724.962	5.69	3.97
	5	27.0337	795.181	5.26	3.84
	6	30.0241	844.516	4.94	3.76
	7	32.5964	878.370	4.71	3.72
	8	34.9749	900.827	4.53	3.68
	9	36.6519	915.041	4.33	3.67
	10	38.1908	923.465	4.29	3.64
	11	39.4295	928.010	4.20	3.63
	12	40.3807	930.133	4.14	3.63
	13	41.0539	930.912	4.10	3.63
	14	41.4554	931.091	4.08	3.63
	15	41.5888	931.103	4.08	3.63
40	1	9.35255	413.984	7.88	4.42
	2	15.8812	711.432	6.66	3.58
	3	21.3108	938.084	5.70	3.27
	4	26.0066	1113.68	5.08	3.10
	5	30.1416	1250.37	4.65	2.98

TABLE 5.1 (continued)

M	i	$\mu^*_{i,M}$	$\sigma^{*2}_{i,M}$	$E_\mu\%$	$E_\sigma\%$
40	6	33.8167	1356.60	4.34	2.91
	7	37.0981	1438.64	4.10	2.85
	8	40.0322	1501.36	3.92	2.81
	9	42.6531	1548.65	3.77	2.78
	10	44.9868	1583.61	3.65	2.75
	11	47.0535	1608.97	3.55	2.74
	12	48.8691	1626.74	3.47	2.73
	13	50.4465	1638.74	3.40	2.72
	14	51.7957	1646.44	3.35	2.72
	15	52.9251	1651.05	3.30	2.71
	16	53.8409	1653.57	3.27	2.71
	17	54.5484	1654.75	3.24	2.71
	18	55.0511	1655.19	3.22	2.71
	19	55.3517	1655.29	3.21	2.71
	20	55.4518	1655.29	3.21	2.71
50	1	9.80391	536.068	7.42	3.70
	2	16.7944	941.404	6.22	3.00
	3	22.6968	1264.43	5.29	2.72
	4	27.8769	1526.29	4.68	2.56
	5	32.5083	1740.13	4.27	2.47
	6	36.6925	1915.17	3.96	2.39
	7	40.4963	2058.32	3.73	2.34
	8	43.9670	2175.03	3.54	2.29
	9	47.1393	2269.67	3.39	2.26
	10	50.0402	2345.88	3.27	2.24
	11	52.6908	2406.67	3.16	2.22
	12	55.1080	2454.61	3.08	2.20
	13	57.3057	2491.90	3.00	2.19
	14	59.2953	2520.42	2.94	2.18
	15	61.0864	2541.78	2.88	2.17
	16	62.6870	2557.38	2.84	2.17
	17	64.1035	2568.42	2.80	2.17
	18	65.3418	2575.93	2.76	2.16
	19	66.4064	2580.78	2.74	2.16
	20	67.3012	2583.70	2.71	2.16
	21	68.0292	2585.30	2.69	2.16
	22	68.5930	2586.05	2.68	2.16
	23	68.9944	2586.33	2.67	2.16
	24	69.2347	2586.39	2.66	2.16
	25	69.3147	2586.40	2.66	2.16

convergence of the diffusion approximations to the true means and variances it appears that the diffusion process proves adequate for all but comparatively small populations. Additional inferences of this nature will be found in the following chapter.

It should be noted here that in computing the entries for the preceding table a slightly different form of (5.2.4) was used. The infinite series which occurs in the approximation for $\mathrm{Var}(T_1)$ converges rather slowly. Fortunately, however, this series may be written in a form in which the rate of convergence is increased. First, note that we may write

$$\sum_{n=1}^{\infty} \frac{p^{n+1}}{n^2} = -p \int_0^p \frac{\log(1-x)}{x}\, dx, \tag{5.3.1}$$

and

$$\sum_{n=1}^{\infty} \frac{(1-p)^{n+1}}{n^2} = -(1-p) \int_0^{1-p} \frac{\log(1-y)}{y}\, dy,$$

or by letting $y = 1-x$ in the last integral we may write

$$\sum_{n=1}^{\infty} \frac{(1-p)^{n+1}}{n^2} = -(1-p) \int_p^1 \frac{\log x}{1-x}\, dx. \tag{5.3.2}$$

Integrating by parts in (5.3.2) we have

$$\int_p^1 \frac{\log x}{1-x}\, dx = \log(p)\log(1-p) + \int_p^1 \frac{\log(1-x)}{x}\, dx,$$

but

$$\int_0^1 \frac{\log(1-x)}{x}\, dx = -\frac{\pi^2}{6},$$

so

$$\int_p^1 \frac{\log(1-x)}{x}\, dx = -\frac{\pi^2}{6} - \int_0^p \frac{\log(1-x)}{x}\, dx.$$

Thus, by adding (5.3.1) and (5.3.2), we obtain

$$\sum_{n=1}^{\infty} \frac{p^{n+1} + (1-p)^{n+1}}{n^2} = (1-p)\,\frac{\pi^2}{6} - (1-p)\log(p)$$

$$\cdot\log(1-p) - (1-2p)\cdot\sum_{n=1}^{\infty} \frac{p^n}{n^2},$$

which may be substituted in (5.2.4) to give

$$\mathrm{Var}(T_1) \doteq 4M^2\Big\{p\cdot\frac{\pi^2}{3} + 2(1-p)\log(p)\log(1-p)$$

$$+ 2\log[p^p(1-p)^{(1-p)}]$$

$$+ 2(1-2p)\cdot\sum_{n=1} \frac{p^n}{n^2}\Big\} - [E(T_1)]^2.$$

It is interesting to note that the infinite series in the form above is a special case of

$$\phi(s,x) = \sum_{n=1}^{\infty} \frac{x^n}{n^s}$$

$$= \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{xz^{s-1}dz}{e^z - x}$$

which may be found in Whittaker and Watson [18].[2]

---

[2] The preceding simplification was the result of a suggestion by Dr. H. A. David.

VI. INTERPRETATIONS

6.1. <u>From moments and graphs</u>. By considering the graphs of the distributions presented at the end of Chapter III and the values of the means and variances given in Table 4.1, one notes that the distributions tend to "spread out" with increasing M. This "spreading out" process proceeds rather slowly for small or large i, but at a more rapid pace for i near $\frac{M}{2}$.

If, however, we assume that M is fixed rather than increasing, a further investigation of the graphs and moments indicates that the distribution is concentrated near zero for i large or small but not for the case in which the two different types of genes are approximately equal in number.

6.2. <u>From diffusion approximations</u>. Results much less obvious than the ones just stated concern the behavior of the diffusion approximations to the means and variances for fixed i and increasing M. Recall that the approximation to the mean as derived in Chapter V was given by

$$\overset{*}{\mu}_{i,M} = 2M \log[(\tfrac{i}{M})^{-\frac{i}{M}}(1- \tfrac{i}{M})^{-(1- \frac{i}{M})}],$$

where p has been replaced by $\frac{i}{M}$. This may be written

$$\overset{*}{\mu}_{i,M} = 2i \log(\tfrac{M}{i}) + 2(1- \tfrac{i}{M})\log(\tfrac{M}{M-i}),$$

and for $i = 1,2,\dots,M-1$, the second term of the right-hand member is seen to be always positive. Hence

$$\overset{*}{\mu}_{i,M} > 2i\cdot\log(\tfrac{M}{i})$$

and thus, letting $M\to\infty$ while keeping i fixed, we see that

$$\lim_{M\to\infty} \overset{*}{\mu}_{i,M} \geq \lim_{M\to\infty} 2i\cdot\log(\tfrac{M}{i}) \to\infty,$$

that is, $\overset{*}{\mu}_{i,M}$ becomes infinite. The behavior of the variance approximation under the assumption of fixed i and increasing M may be determined by considering the following form of (5.2.4),

$$\overset{*}{\sigma}{}^2_{i,M} = \frac{4\pi^2}{3}\cdot M^2-8M^2\cdot\sum_{n=1}^{\infty}\frac{(\frac{i}{M})^{n+1}+(1-\frac{i}{M})^{n+1}}{n^2}$$

$$-8M^2\log[(\tfrac{i}{M})^{-\frac{i}{M}}(1-\tfrac{i}{M})^{-(1-\frac{i}{M})}]$$

$$-4M^2\{\log[(\tfrac{i}{M})^{-\frac{i}{M}}(1-\tfrac{i}{M})^{-(1-\frac{i}{M})}]\}^2.$$

First let us consider the infinite series appearing in this expression. Since $0<\frac{i}{M}<1$, it is easy to show that the numerator of each term of the infinite series attains its maximum value when $i=\frac{M}{2}$. Thus

$$\sum_{n=1}^{\infty}\frac{(\frac{i}{M})^{n+1}+(1-\frac{i}{M})^{n+1}}{n^2}\leq\sum_{n=1}^{\infty}\frac{1}{n^2}(\tfrac{1}{2})^n,$$

but

$$\sum_{n=1}^{\infty} \frac{1}{n^2} (\tfrac{1}{2})^n = \frac{1}{2} + \frac{1}{16} + \frac{1}{72} + \sum_{n=4}^{\infty} \frac{1}{n^2} (\tfrac{1}{2})^n$$

$$< \frac{83}{144} + \sum_{n=4}^{\infty} (\tfrac{1}{2})^n$$

$$= \frac{83}{144} + \frac{1}{8}$$

$$= \frac{101}{144},$$

and therefore

$$\sum_{n=1}^{\infty} \frac{(\frac{i}{M})^{n+1} + (1- \frac{i}{M})^{n+1}}{n^2} < \frac{101}{144}.$$

It is also easy to show that

$$\log[(\tfrac{i}{M})^{-\frac{i}{M}} (1- \tfrac{i}{M})^{-(1- \frac{i}{M})}]$$

attains its maximum when $\frac{i}{M} = \frac{1}{2}$, and thus

$$\log[(\tfrac{i}{M})^{-\frac{i}{M}} (1- \tfrac{i}{M})^{-(1- \frac{i}{M})}] \leq \log 2.$$

Hence

$$\overset{*}{\sigma}{}^2_{i,M} > [\frac{4\pi^2}{3} - 8(\frac{101}{144}) - 8\log 2 - 4(\log 2)^2]M^2,$$

and since $\frac{4\pi^2}{3} < 13.159$, $\frac{101}{144} > 0.702$, $\log 2 > 0.694$, and

$(\log 2)^2 > 0.482$, we have

$$\overset{*}{\sigma}{}^2_{i,M} > [13.159 - 8(0.702) - 8(0.694) - 4(0.482)]M^2$$

or

$$\overset{*}{\sigma}{}^2_{i,M} > 0.063M^2.$$

Therefore $\overset{*}{\sigma}{}^2_{i,M}$ also becomes infinite as $M \to \infty$.

Of all interpretations that might be made, perhaps the ones of greatest significance are those based on the percentage errors given in Table 5.1. A point that has been made earlier, but one that seems worthy of being repeated, is the adequacy of the diffusion approximation for comparatively small populations. In research for this study no references were found on which one might base an estimate for population size sufficient for the application of diffusion theory. That such a small percentage error would be encountered for a population size of fifty or even less was not anticipated.

It has long been known that diffusion theory works as M becomes infinite but with $p = \frac{i}{M}$ kept fixed. A rigorous demonstration of this for general genetical problems is given by Watterson [16]. It is also in common agreement that diffusion theory works better near $p = \frac{1}{2}$ than in the tails. The entries of Table 5.1 demonstrate both of these points. Of primary importance, however, is the indication that diffusion theory works even when i remains fixed. One would not expect the absolute error of moments to decrease, but here we have an indication that the percentage error does, which is especially useful if one wishes to determine the chances of survival of a single mutant introduced into an otherwise pure line in which case $p = \frac{1}{M}$.

Table 5.1 also demonstrates that the percentage error decreases at a faster rate for p fixed than for i fixed. It seems to, as if the percentage error decreases like $\frac{1}{M} + \frac{1}{\log M}$ for fixed i. Here, however, there is no theory available and with values of M ranging only from two to fifty this should be taken purely as a guess. There appears to be no doubt that diffusion theory consistently over-estimates the absorption time moments.

PART II

A METHOD OF MATRIX INVERSION

VII. DESCRIPTION OF THE METHOD

7.1. Introduction. Many diverse methods for determining the adjoint, that is, the matrix of cofactors, and the inverse of a matrix have been devised. The most efficient method depends upon the properties of the matrix under consideration and the mechanical devices available for computation. The technique presented here has been found to be especially useful to those who work with desk calculators and matrices of small order.

The approach which is developed in this chapter and applied in the next is based on the pivotal condensation method of multiplication and subtraction with exact division which is a refinement of the work of Dodgson [4]. The basic operational unit used is of the form (ab-cd)/e in which the division is exact if the elements are integers, and thus a semi-check is inherent in the scheme. The pivotal elimination method is thoroughly described by Dwyer [5] and may be briefly explained as follows. In order to evaluate a determinant, a pivotal element is chosen, usually the upper left-hand element, and all second order determinants which can be formed having the pivotal element in the upper left-hand corner are evaluated. These values are written in a square array of order one less than the original. A pivotal element is chosen in this new array and again the second order determinants are evaluated, but here, and in the remaining computations, each new value is divided by the preceding pivot. This procedure is continued until a single element remains, this element being the value of the determinant.

Using the notation $\binom{i\ j\ \cdots\ k}{m\ n\ \cdots\ p}$ to denote the value of the determinant formed from the elements in the i-th, j-th, ..., k-th rows and m-th, n-th, ..., p-th columns of the original array, the evaluation of a fourth order determinant would appear as in Table 7.1. The upper left-hand elements

TABLE 7.1

Schematic of the pivotal condensation method for evaluating a fourth order determinant.

$\left[{1 \atop 1}\right]$	$\binom{1}{2}$	$\binom{1}{3}$	$\binom{1}{4}$
$\binom{2}{1}$	$\binom{2}{2}$	$\binom{2}{3}$	$\binom{2}{4}$
$\binom{3}{1}$	$\binom{3}{2}$	$\binom{3}{3}$	$\binom{3}{4}$
$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$
	$\left[{12 \atop 12}\right]$	$\binom{12}{13}$	$\binom{12}{14}$
	$\binom{13}{12}$	$\binom{13}{13}$	$\binom{13}{14}$
	$\binom{14}{12}$	$\binom{14}{13}$	$\binom{14}{14}$
		$\left[{123 \atop 123}\right]$	$\binom{123}{124}$
		$\binom{124}{123}$	$\binom{124}{124}$
			$\binom{1234}{1234}$

have been chosen as pivots and for clarity have been enclosed in brackets rather than parentheses. To further clarify the scheme, some typical elements are

$$\binom{13}{14} = \genfrac{[}{]}{0pt}{}{1}{1}\binom{3}{4} - \binom{3}{1}\binom{1}{4},$$

$$\binom{123}{124} = \frac{\genfrac{[}{]}{0pt}{}{123}{123}\binom{124}{124} - \binom{124}{123}\binom{123}{124}}{\genfrac{[}{]}{0pt}{}{12}{12}} .$$

7.2. *Determination of the adjoint and inverse.* The pivotal condensation scheme described above is basically nothing more than successive ordered groupings of four numbers which form the elements of the second order determinants. For example, consider the set consisting of the first four positive integers, $\{1,2,3,4\}$. Pair the first of these with each element that follows, creating the set $\{12,13,14\}$. Again pair the first element of this set with each element that follows to obtain $\{123,124\}$, noting that integers which occur in both elements which are paired are written only once. Performing this pairing once more, the set $\{1234\}$ is obtained. We note that this is exactly the manner in which the rows and columns are grouped to form the second order determinants which lead us to the value of the determinant. This leads us, however, to the value of only one determinant. If we wish to obtain the adjoint of a matrix, or the inverse if it exists, we must generate additional determinants and hence the scheme must be amended.

Let us consider two operators, A and B, over an ordered set defined respectively as the ordered pairing of the first element of the set with each that follows, and, the ordered pairing of the second element of the set with each that follows. Let I denote the set consisting of the first n positive integers and perform the following series of operations on I. First, operate on I by A and then apply operator B, thus forming the set $\{I_A, I_B\}$ where $I_A = \{12, 13, \ldots, 1n\}$ and $I_B = \{23, 24, \ldots, 2n\}$. Next, operate by A and then B on the elements of I_A and by A only on the elements of I_B to form the set $\{I_{AA}, I_{BA}, I_{AB}\}$ where $I_{AA} = \{123, 124, \ldots, 12n\}$, $I_{BA} = \{134, 135, \ldots, 13n\}$, and $I_{AB} = \{234, 235, \ldots, 23n\}$. Again applying A and then B on I_{AA} and A only on I_{BA} and I_{AB} we form $\{I_{AAA}, I_{BAA}, I_{ABA}, I_{AAB}\}$ where the elements in each group are determined as before. If this scheme is continued for a total of (n-2) times the result is the set $\{123\ldots(n-1), 12\ldots(n-2)(n), \ldots, 23\ldots(n)\}$, the n ordered combinations of the n integers taken n-1 at a time. Applying this pairing technique to the rows and columns of a square array, rather than a line array, and using the pivotal condensation method of evaluating a determinant, we are able to generate the values of the n^2 determinants needed for the adjoint.

To illustrate this process, the technique just described is presented in Table 7.2 for a square matrix of fifth order. The bracket notation again is used to indicate the pivotal elements. The dotted lines have been used to partition the

TABLE 7.2

Schematic for determining the adjoint and inverse of a fifth order square matrix.

$\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}1\\2\end{bmatrix}$	$\binom{1}{3}$	$\binom{1}{4}$	$\binom{1}{5}$		
$\begin{bmatrix}2\\1\end{bmatrix}$	$\begin{bmatrix}2\\2\end{bmatrix}$	$\binom{2}{3}$	$\binom{2}{4}$	$\binom{2}{5}$		
$\binom{3}{1}$	$\binom{3}{2}$	$\binom{3}{3}$	$\binom{3}{4}$	$\binom{3}{5}$		
$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$	$\binom{4}{5}$		
$\binom{5}{1}$	$\binom{5}{2}$	$\binom{5}{3}$	$\binom{5}{4}$	$\binom{5}{5}$		
$\begin{bmatrix}12\\12\end{bmatrix}$	$\begin{bmatrix}12\\13\end{bmatrix}$	$\binom{12}{14}$	$\binom{12}{15}$	$\begin{bmatrix}12\\23\end{bmatrix}$	$\binom{12}{24}$	$\binom{12}{25}$
$\begin{bmatrix}13\\12\end{bmatrix}$	$\begin{bmatrix}13\\13\end{bmatrix}$	$\binom{13}{14}$	$\binom{13}{15}$	$\begin{bmatrix}13\\23\end{bmatrix}$	$\binom{13}{24}$	$\binom{13}{25}$
$\binom{14}{12}$	$\binom{14}{13}$	$\binom{14}{14}$	$\binom{14}{15}$	$\binom{14}{23}$	$\binom{14}{24}$	$\binom{14}{25}$
$\binom{15}{12}$	$\binom{15}{13}$	$\binom{15}{14}$	$\binom{15}{15}$	$\binom{15}{23}$	$\binom{15}{24}$	$\binom{15}{25}$
$\begin{bmatrix}23\\12\end{bmatrix}$	$\begin{bmatrix}23\\13\end{bmatrix}$	$\binom{23}{14}$	$\binom{23}{15}$	$\begin{bmatrix}23\\23\end{bmatrix}$	$\binom{23}{24}$	$\binom{23}{25}$
$\binom{24}{12}$	$\binom{24}{13}$	$\binom{24}{14}$	$\binom{24}{15}$	$\binom{24}{23}$	$\binom{24}{24}$	$\binom{24}{25}$
$\binom{25}{12}$	$\binom{25}{13}$	$\binom{25}{14}$	$\binom{25}{15}$	$\binom{25}{23}$	$\binom{25}{24}$	$\binom{25}{25}$

TABLE 7.2 (continued)

$\left[{123 \atop 123}\right]$	$\left[{123 \atop 124}\right]$	$\binom{123}{125}$	$\left[{123 \atop 134}\right]$	$\binom{123}{135}$	$\left[{123 \atop 234}\right]$	$\binom{123}{235}$
$\left[{124 \atop 123}\right]$	$\left[{124 \atop 124}\right]$	$\binom{124}{125}$	$\left[{124 \atop 134}\right]$	$\binom{124}{135}$	$\left[{124 \atop 234}\right]$	$\binom{124}{235}$
$\binom{125}{123}$	$\binom{125}{124}$	$\binom{125}{125}$	$\binom{125}{134}$	$\binom{125}{135}$	$\binom{125}{234}$	$\binom{125}{235}$
$\left[{134 \atop 123}\right]$	$\left[{134 \atop 124}\right]$	$\binom{134}{125}$	$\left[{134 \atop 134}\right]$	$\binom{134}{135}$	$\left[{134 \atop 234}\right]$	$\binom{134}{235}$
$\binom{135}{123}$	$\binom{135}{124}$	$\binom{135}{125}$	$\binom{135}{134}$	$\binom{135}{135}$	$\binom{135}{234}$	$\binom{135}{235}$
$\left[{234 \atop 123}\right]$	$\left[{234 \atop 124}\right]$	$\binom{234}{125}$	$\left[{234 \atop 134}\right]$	$\binom{234}{135}$	$\left[{234 \atop 234}\right]$	$\binom{234}{235}$
$\binom{235}{123}$	$\binom{235}{124}$	$\binom{235}{125}$	$\binom{235}{134}$	$\binom{235}{135}$	$\binom{235}{234}$	$\binom{235}{235}$
$\left[{1234 \atop 1234}\right]$	$\binom{1234}{1235}$	$\binom{1234}{1245}$	$\binom{1234}{1345}$	$\binom{1234}{2345}$		
$\binom{1235}{1234}$	$\binom{1235}{1235}$	$\binom{1235}{1245}$	$\binom{1235}{1345}$	$\binom{1235}{2345}$		
$\binom{1245}{1234}$	$\binom{1245}{1235}$	$\binom{1245}{1245}$	$\binom{1245}{1345}$	$\binom{1245}{2345}$		
$\binom{1345}{1234}$	$\binom{1345}{1235}$	$\binom{1345}{1245}$	$\binom{1345}{1345}$	$\binom{1345}{2345}$		
$\binom{2345}{1234}$	$\binom{2345}{1235}$	$\binom{2345}{1245}$	$\binom{2345}{1345}$	$\binom{2345}{2345}$		
$\binom{12345}{12345}$						

the arrays in order to facilitate reference to the preceding pivot. Examples of the evaluation of some typical elements appearing in the array are

$$\binom{15}{14} = \left[{1 \atop 1}\right]\binom{5}{4} - \binom{5}{1}\binom{1}{4},$$

$$\binom{13}{25} = \left[{1 \atop 2}\right]\binom{3}{5} - \binom{3}{2}\binom{1}{5},$$

$$\binom{125}{134} = \frac{\genfrac{[}{]}{0pt}{}{12}{13}\binom{15}{14} - \binom{15}{13}\binom{12}{14}}{\genfrac{[}{]}{0pt}{}{1}{1}} ,$$

$$\binom{235}{125} = \frac{\genfrac{[}{]}{0pt}{}{23}{12}\binom{25}{15} - \binom{25}{12}\binom{23}{15}}{\genfrac{[}{]}{0pt}{}{2}{1}} ,$$

and

$$\binom{1245}{1235} = \frac{\genfrac{[}{]}{0pt}{}{124}{123}\binom{125}{125} - \binom{125}{123}\binom{124}{125}}{\genfrac{[}{]}{0pt}{}{12}{12}} .$$

The adjoint is found by reflecting the five by five array of fourth order determinants in the secondary diagonal and applying the regular rule of signs, these determinants being the minors rather than the cofactors. If the determinant of the matrix, $\binom{12345}{12345}$, is non-zero, then the inverse will exist and may be found by dividing this value into each element of the adjoint.

If the original matrix is symmetric, the number of computations is reduced as is common in many adjoint methods. Also, if additional rows and columns are annexed the original computations may be retained and only those computations involving the new elements need be made. Applications of this method are found in the next chapter along with comparisons with other schemes.

VIII. APPLICATION OF THE METHOD

8.1. <u>Some examples</u>. One approach to the study of an absorbing Markov chain uses the method of moments as presented in Chapter IV. Basic to this method is the fundamental matrix, that is, the inverse of I-Q where Q is the submatrix of the transition matrix which concerns the Markov process so long as it remains in transient states.

To illustrate the inversion technique described in the preceding chapter, let us consider the calculation of the fundamental matrix of the Markov chain studied in Part I. For the first example, consider the case for which M=4. The transition matrix may be found by using (2.2.2) and is

$$P = \frac{1}{256}\begin{bmatrix} 256 & 0 & 0 & 0 & 0 \\ 81 & 108 & 54 & 12 & 1 \\ 16 & 64 & 96 & 64 & 16 \\ 1 & 12 & 54 & 108 & 81 \\ 0 & 0 & 0 & 0 & 256 \end{bmatrix},$$

from which we obtain

$$I-Q = \frac{1}{256}\begin{bmatrix} 148 & -54 & -12 \\ -64 & 160 & -64 \\ -12 & -54 & 148 \end{bmatrix}.$$

Basically, then, our problem is reduced to that of determining

$$\begin{bmatrix} 148 & -54 & -12 \\ -64 & 160 & -64 \\ -12 & -54 & 148 \end{bmatrix}^{-1}. \qquad (8.1.1)$$

Table 8.1 presents the method of Chapter VII as applied to this matrix. In this and the following examples, the pivotal elements will be enclosed in parentheses. Since the matrix considered has a special pattern, some entries are not necessary but have been included for completeness of the illustration. The adjoint is obtained by reflection in the secondary diagonal and application of the rule of signs. This gives

TABLE 8.1

Illustration for a third-order matrix.

(148)	(-54)	-12
(-64)	(160)	-64
12	-54	148
(20224)	-10240	5376
-8640	21760	-8640
5376	-10240	20224
2375680		

$$(8.1.2)\qquad \begin{bmatrix} 20224 & 8640 & 5376 \\ 10240 & 21760 & 10240 \\ 5376 & 8640 & 20224 \end{bmatrix},$$

and the inverse, (8.1.1), is found by dividing each element of (8.1.2) by 2375680, or the fundamental matrix may be found by multiplying each element of (8.1.2) by $\frac{256}{2375680}$ $= \frac{1}{9280}$ which gives

$$(I-Q)^{-1} = \frac{1}{9280}\begin{bmatrix} 20224 & 8640 & 5376 \\ 10240 & 21760 & 10240 \\ 5376 & 8640 & 20224 \end{bmatrix}.$$

By evaluating the row sums we see that these results agree with the values given for the means in Table 4.1. As a second example, we find that for $M = 5$,

$$I-Q = \frac{1}{3125}\begin{bmatrix} 1845 & -640 & -160 & -20 \\ -810 & 2045 & -720 & -240 \\ -240 & -720 & 2045 & -810 \\ -20 & -160 & -640 & 1845 \end{bmatrix},$$

and the computing scheme for this case is presented in Table 8.2. Here, however, we take advantage of the pattern of the

TABLE 8.2

Illustration for a fourth-order matrix.

(1845)	(-640)	-160	-20	
(-810)	(2045)	-720	-240	
-240	-720	2045	-810	
-20	-160	-640	1845	
(3254625)	(-1458000)	-459000	(788000)	194500
-1482000	3734625	-1499250	-1424000	504000
-3080000	-1184000	3403625	384000	-1184000
*	*	*	*	*
*	*	*	*	*
(5416828125)	-3013406250	2113875000	-1053312500	
-2332000000	5927453125	-2984250000	1574500000	
*	*	*	*	
*	*	*	*	
7706181640625				

matrix and only those entries necessary for the calculation of the inverse are listed. The partioning of the arrays by the dotted lines is for reference to the preceding pivot. $(I-Q)^{-1}$ may now be found by performing the reflection and change of sign and multiplying each element by

$$\frac{3125}{7706181640625} = \frac{1}{2465978125},$$

which gives approximately

$$(I-Q)^{-1} \doteq \begin{bmatrix} 2.1966246 & 0.9456694 & 0.6384890 & 0.4271378 \\ 1.2219923 & 2.4036925 & 1.2101689 & 0.8572156 \\ 0.8572156 & 1.2101689 & 2.4036925 & 1.2219923 \\ 0.4271378 & 0.6384890 & 0.9456694 & 2.1966246 \end{bmatrix}.$$

For a more general example, let us consider the calculation of the ajoint and inverse of a matrix for which

TABLE 8.3

Illustration for a fifth-order matrix.

(3)	(2)	-1	-1	2		
(1)	(4)	1	2	2		
1	2	-2	3	1		
4	1	3	-2	1		
1	-3	4	5	3		
(10)	(4)	7	4	(6)	8	-4
(4)	(-5)	10	1	(-2)	8	-2
-5	13	-2	-5	7	-3	0
-11	13	16	7	5	7	12
(-2)	(-3)	1	-1	(-10)	8	0
-15	-1	-10	-7	11	-10	2
-7	3	3	1	19	26	18

TABLE 8.3 (continued)

(-22)	(24)	-2	(25)	8	(32)	-10
(50)	(5)	-10	(-33)	-24	(-37)	14
58	79	38	-9	-8	1	46
(9)	(14)	-5	(-40)	4	(-25)	7
-1	58	13	-70	-16	-27	-7
(-43)	(35)	-1	(31)	20	(3)	-5
-27	1	-9	-12	0	-103	-45
(-131)	32	-23	-84	13		
-313	-72	107	-32	247		
366	248	98	12	-286		
134	28	118	-184	-182		
-451	-180	157	-80	65		
-884						

there is no immediately discernible pattern. The schematic for a fifth-order matrix is presented in Table 8.3 in which the form given in Table 7.2 is followed. Performing the required transformations we find the inverse to be

$$-\frac{1}{884}\begin{bmatrix} 65 & 182 & 286 & -247 & 13 \\ 80 & -184 & -12 & -32 & 84 \\ 157 & -118 & 98 & -107 & -23 \\ 180 & 28 & -248 & -72 & -32 \\ -451 & -134 & 366 & 313 & -131 \end{bmatrix}.$$

Rounding off procedures may be used with this method if desired, however, if such are employed the exact division property is no longer retained. Also, as mentioned earlier, we may take advantage of the symmetry of a given matrix.

Both of these procedures are demonstrated in Table 8.4 where, in each step of the process, the entries are rounded off to three decimal places. The asterisks are used to indicate that the corresponding entries need not be recorded. Using the results of Table 8.4 the inverse is found to be approximately

$$\begin{bmatrix} 1.400 & -0.089 & 0.496 & -1.030 \\ -0.089 & 1.889 & -0.719 & -0.607 \\ 0.496 & -0.719 & 3.230 & -2.370 \\ -1.030 & -0.607 & -2.370 & 3.807 \end{bmatrix}.$$

TABLE 8.4

Illustration for symmetric matrix.

(1.000)	(0.313)	0.280	0.495	
*	(1.000)	0.625	0.650	
*	*	1.000	0.803	
*	*	*	1.000	
(0.902)	(0.564)	0.495	(-0.076)	-0.292
*	(0.922)	0.664	(0.130)	-0.071
*	*	0.755	0.069	-0.009
*	*	*	(0.575)	0.379
*	*	*	*	0.578
(0.514)	0.320	-0.082	0.139	
*	0.436	0.097	0.067	
*	*	0.255	0.012	
*	*	*	0.189	
0.135				

8.2. Some comparisons. Since the technique presented in this paper is an exact method, only comparisons with methods which lead to the exact value of the adjoint will be made, with one exception. The exception will be a comparison with the Gauss-Doolittle method for symmetric matrices [5], p.191. The bases of the comparisons will be the number of entries required in addition to the elements of the original matrix, the number of multiplications, the number of divisions, and the number of additions or subtractions.

Numerous methods for obtaining the exact adjoint have been proposed and thus it is necessary that we restrict our comparisons to representative methods. Those chosen to

TABLE 8.5

Comparison of adjoint methods.

Order	Method	Entries	Mult.	Div.	Add.-Sub.
3	This Study	9	18	0	9
	Determinants	24	27	13	15
	Bingham	22	37	1	35
4	This Study	41	82	16	41
	Determinants	50	72	31	42
	Bingham	54	163	2	144
5	This Study	123	246	74	123
	Determinants	90	150	61	90
	Bingham	108	456	3	402

compare with the method of this study are the Method of Determinants of Dwyer [5], and the Bingham Method [2] which is quite similar to the ones proposed by Hotelling [11] and by Frame [10]. Results for third, fourth, and fifth order matrices are given in Table 8.5. This table does not give the complete picture, however, for the amount of information yielded by each method varies. For example, the Method of Determinants produces the adjoint only while the two other methods will give the coefficients of the characteristic equation with a few additional calculations. It should also be noted that the Bingham Method does not provide a check. Furthermore, the method of this study requires the retention of the preceding pivots for only one additional step while the Method of Determinants and the Bingham Method require that all entries be retained.

Now let us compare the proposed method with the Gauss-Doolittle Method [5] for obtaining the approximate inverse

TABLE 8.6

Comparison of inverse methods.

Order	Method	Entries	Mult.	Div.	Add.-Sub.
3	This Study	13	14	7	7
	Gauss-Doolittle	33	13	18	13
4	This Study	42	52	27	26
	Gauss-Doolittle	54	22	36	34

of a symmetric matrix. The results of this comparison are found in Table 8.6, but here again one should recall that the proposed method provides a check and, with few additional calculations, the characteristic equation while the Gauss-Doolittle Method does not.

IX. SUMMARY

Since Darwin first effectively presented evolution from the point of view of being primarily a statistical process in which random hereditary variation furnished the raw material much time has been devoted to population studies. One point of particular interest is the study of the stochastic process which describes the change in the frequency of a given gene type in the population as a function of time. For example, one result of such a study might be the expected number of generations for extinction of gene type.

One population that has received considerable attention is a random-mating haploid type of constant size. Much of the early study in this area was done by R. A. Fisher and Sewall Wright and later studies have been carried on by William Feller and others. One of the more powerful methods applied to the study of a genetic population uses the theory of Markov chains. Gustav Malécot was the first to express the population in terms of a Markov chain in which the transition probabilities are found to be $p_{ij} = \binom{M}{j}(\frac{i}{M})^j (1- \frac{i}{M})^{M-j}, i, j = 0, 1, \ldots, M$, for a population of size M.

This paper has been primarily concerned with the time taken for the chain to first reach one or the other absorbing state from a given initial state. Two approaches have been used in this study, one leading to the distribution of first passage time by means of the eigenvectors of the transition matrix and the other leading directly to the moments of the distribution. Although neither of the methods led to a general solution, contributions were made to the advancement of the study and numerical results of interest were obtained.

Since the direct approaches to the problem did not prove successful, diffusion theory was applied in order to obtain an approximation to the distribution. It has long been known that diffusion theory works as $M \to \infty$ but with $p = \frac{i}{M}$ kept fixed. The results of this paper indicate, however, that the population size need not be too large before the diffusion approximation can be applied. Furthermore, it is also indicated that diffusion theory may be applied even when the gene frequency ratio is not kept constant, that is, diffusion theory seems to work even in the extreme tails of the distribution.

Matrix inversion plays an important role in the study of Markov chains as well as in other areas of probability and statistics. Thus, any simplification of the computation of the inverse of a matrix should not be overlooked. Hence, it was considered worthwhile to include in this thesis a second part which introduces a new inversion technique.

Basically, this technique is one which generates the inverse of a matrix with a minimum number of determinantal operations. This new method has been described and applications of the method have been given. By the comparisons which have been made with typical well-known methods presently used, it is seen that the method presented in this paper is especially efficient for matrices of small order.

X. ACKNOWLEDGEMENTS

The author takes this opportunity to thank his many excellent teachers over the years. In particular, he would like to express his sincere appreciation to the director of his dissertation, Dr. Geoffrey A. Watterson, for his invaluable guidance and advice during the course of this work.

The author also gratefully acknowledges his deep indebtedness to Dr. Boyd Harshbarger for his counsel and encouragement and to Dr. Rudolf J. Freund for his advice concerning computations for the tables.

Finally, an expression of thanks is due my wife, Dorothy, for her patient and expert typing help.

XI. BIBLIOGRAPHY

[1] A. T. Bharucha-Reid, Elements of the Theory of Markov Processes and Their Applications, McGraw-Hill, New York, 1960.

[2] M. D. Bingham, "A new method for obtaining the inverse matrix," J. Amer. Stat. Assoc., Vol. 36 (1941), pp. 530-534.

[3] Harald Cramér, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1951, p. 257.

[4] C. L. Dodgson, "Condensation of determinants," Proc. Royal Soc., Vol. 15 (1866), pp.150-155.

[5] P. S. Dwyer, Linear Computations, Wiley, New York, 1951.

[6] William Feller, An Introduction to Probability Theory and its Applications, Vol. I, 2nd ed., Wiley, New York, 1950.

[7] William Feller, "Diffusion processes in genetics," Second Berkeley Symposium on Math. Stat. and Prob., Edited by J. Neyman, University of Cal. Press, Berkeley, 1951, pp.227-246.

[8] William Feller, "Diffusion processes in one dimension," Trans. Amer. Math. Soc., Vol. 77 (1954), pp. 1-3.

[9] R. A. Fisher, The Gentical Theory of Natural Selection, Oxford University Press, Oxford, 1930.

[10] J. S. Frame, A Simple Recursion Formula for Inverting a Matrix. Presented to the Amer. Math. Soc. at Boulder, Colorado, on Sept. 1, 1949.

[11] H. Hotelling, "Some new methods in matrix calculation," Ann. Math. Stat., Vol. 14 (1943), pp. 1-34.

[12] John G. Kemeny and J. Laurie Snell, Finite Markov Chains, Van Nostrand, Princeton, 1959.

[13] Motoo Kimura, "Some problems of stochastic processes in genetics," Ann. Math. Stat., Vol. 28 (1957), pp. 882-901.

[14] Gustave Malécot, "Sur un problème de probabilités en chain que pose la génétique," C. R. Acad. Sci. Pris, Vol. 219 (1944), pp. 379-381.

[15] C. H. Richardson, Introduction to the Calculus of Finite Differences, Van Nostrand, New York, 1954.

[16] G. A. Watterson, "Probability theory applied to genetic populations," Ph.D. thesis, Australian National University, (1960).

[17] G. A. Watterson, "Markov chains with absorbing states: A genetic example," Ann. Math. Stat., Vol. 32 (1961), pp. 716-729.

[18] E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4-th ed., University Press, Cambridge, 1951. 5/n W6

[19] Sewall Wright, "Evolution in Mendelian populations." Genetics, Vol. 16 (1931), pp. 97-159.

ABSTRACT

KNOX, SAMUEL R. (Virginia Polytechnic Institute, Blacksburg, Virginia, U.S.A.)
A study of a random-mating population of fixed size - In English
112 pages (19 references, 20 figures, 10 tables)

This paper is divided into two parts. The first part is devoted to a study of absorption in a genetic population model and the second part to introducing a new method of matrix inversion.

The author begins Part I by presenting the haploid population model under consideration, a constant-size population in which mutation is absent. Using the approach of Malécot, the problem is formulated in terms of an absorbing Markov chain. A brief review of earlier work by Malécot, Fisher, Wright, and Feller is given before proceding to the study of the distribution of time taken for the population to consist of only one gene type.

The first method used by the author to determine the probability function for first passage time requires the eigenvectors of the transition matrix. A blocking transformation is used to show that only the eigenvectors corresponding to the even eigenvalues are needed to find the distributions. It is also shown that these eigenvectors are symmetric. In order to simplify computation of the needed vectors, a transformation which triangularizes the transition

matrix is presented. This transformation also leads to a simple derivation of the eigenvalues and is used to derive the distributions for populations of size two through nine. Although the general solution is not obtained, expressions for the first seven eigenvectors are listed along with general results concerning the triangularized form.

Next, the author attacks the problem by developing the theory of moments of the distribution and applies this theory to determine the means and variances. These are tabulated for population sizes 2(1)9 and 10(10)50. Then, by assuming a large population size, a diffusion process continuous in space and time is used to approximate the Markov chain. Expressions for the mean and variance are derived and are tabulated for the same values as in the moment approach. A comparison of the entries of these two tables gives support to the theory that the diffusion process works as the population size becomes infinite but with the gene frequency ratio kept fixed. Also, it demonstrates that diffusion theory works better for a gene frequency ratio near one-half than for values near zero or one and further that this theory consistently over-estimates the absorption time moments. The author notes, however, that the point of primary importance is the indication that diffusion theory works even with fixed gene number, that is, even in the tails of the distribution. He also states that the percentage error decreases at a faster rate for fixed gene frequency ratio than for fixed gene number and concludes the first part by suggesting a rate of decrease of percentage error for constant gene frequency.

In Part II, the author begins by noting the importance of the inverse matrix in Markov chain theory. The new inversion technique given is basically one which generates the inverse with a minimum number of determinantal operations. The method is thoroughly described and it is applied to some problems encountered in the first part of the paper. The paper ends with a comparison of the new method with well-known techniques presently in use which shows that the method proposed is especially efficient for small order matrices.