

COMPARISONS OF MARKOV CHAIN AND DIFFUSION
PROCESS FORMULATIONS OF THE FINITE GENETIC
POPULATION MODEL, WITH EMPHASIS ON COMPARING
RESULTS FOR VERY SMALL POPULATIONS

by: Hameed 'Bayo Lawal
Department of Applied Statistics,
University of Reading.

A Thesis submitted as part of the requirement
for the award of the Degree Master of Science
in Biometry.

August 1974.

ACKNOWLEDGEMENTS

The author is deeply grateful to Dr. S.J. Rogers for suggesting the topic and for her encouragement and guidance, to Professor R.N. Curnow for his helpful suggestions and criticisms. Of course I am solely responsible for errors of fact and judgement that remain.

Thanks are also due to the entire staff of the Department's programming section for their help, guidance and patience.

CHAPTER 1INTRODUCTION

Populations in nature and in the laboratory are finite, since any population maintained for experiment or breeding improvement must be of limited size, and in the natural state effective breeding numbers are limited by geographical locations. Hence any genetic population theory should take that into account.

Gene frequency changes are a consequence of various pressures acting on the population, which can cause radical change in the genetic structure of the population. The various factors by which this change of frequencies is brought about can be conveniently classified into three groups. Here gene frequencies mean the proportions of a particular gene in a given population.

The first consists of factors which cause directed changes, such as mutation and migration occurring at constant rates, and selection of a constant intensity which Wright, (1949) called the systematic evolutionary pressures. The second group consists of factors which produce random fluctuations in gene frequencies, of which two different types may be recognised. One type is the random sampling of gametes in a finite population. This becomes important when the population is very small since the gametes that transmit genes to the next generation carry a sample of the genes in the parent generation, and if this sample is not large, the gene frequencies are liable to change between one generation and the next. This random change of gene frequency is sometimes called the DISPERSIVE PROCESS with its three known important consequences:

(i) Differentiation between sub-populations,

- (ii) Reduction of the genetic variance within the population, and
- (iii) an increase in the frequency of the homozygotes at the expense of the heterozygotes.

The other type is random fluctuation in systematic pressures, of which fluctuation in selection intensity is especially important. There is a third group of factors consisting of events, such as chromosome re-arrangements, duplication or deficiency of a nucleotide sequence or polyploidy, that may occur only once in the history of a species.

This study discusses the second of the three factors, random sampling of gametes as they act to change the genetic make-up of the populations considered in conjunction with the systematic pressure of selection. We will regard the process of change as a stochastic process, where this means the mathematical formulation of a chance event evolving in time.

The purpose of this study is to compare diffusion approximation results with exact results obtained by Markov chain formulations for very small genetic populations. Populations of diploid individuals will be considered, generally under selection of genes of a single autosomal locus (i.e. not on sex chromosome). Two alleles on this locus are assumed. Populations of size two and three are considered. The first consisting of one male and one female, and the second two of one sex and one of the other (two males and one female).

When selection takes place, individuals contribute different numbers of offspring to the next generation and when a gene is subject to selection, its frequency in the offspring even with no sampling of gametes is not the same as in the parents, since parents of different genotypes pass on their genes unequally to the next generation. In this

way selection causes a change of gene frequency and consequently also of genotype frequency. We also have to take account of the degree of dominance shown by the genes in question. Dominance in this connection means dominance with respect to fitness (the proportionate contribution of offspring to the next generation). With no selection, the original number of individuals all contribute equally to the next generation. In this study, the strength of the selection as expressed by the coefficient of selection, S , which is the proportionate increase in the gametic contribution of the A_1A_1 and A_1A_2 genotypes over A_2A_2 genotype is taken as 0.0, 0.04, 0.08 and 0.12. The dominance values, h , used are 0, $\frac{1}{2}$, 1 and $1\frac{1}{4}$, corresponding respectively to "recessive", "additive", "dominance" and "over dominance" selection against A_1A_2 genotype. The method of sampling of gametes before selection is employed. Mating is assumed to be random in that a randomly chosen gamete produced by a male will combine with a randomly chosen gamete produced by a female and only non-overlapping generations are considered.

In what follows in this study, two models have been put forward for the process of change of gene frequency. The first assumes the process is Markovian, that is the probability distribution of population structure at a given time t_1 depends on the structure at a preceding time t_0 but not on the previous history which has led to the structure at t_0 where $t_0 < t_1$. The second considers a continuous stochastic process in gene frequency changes. We further assume that random changes in position in each generation are very small and there are infinitesimal times between generations.

The fundamental equations used to study this continuous Markov process are the Kolmogorov forward and backward equations (Kolmogorov 1931), summarised by Kimura 1964 and that developed by

(Feller, 1954.) The Markov chain formulations (exact approach) assume that there are discrete generations. We examine the diffusion approximations for the following quantities:

- (i) The probability of ultimate fixation of A_1 - gene.
- (ii) The probability of fixation of A_1 - gene by the n^{th} generation,
and
- (iii) the expected mean time until fixation of A_1 - gene.

The diffusion approximations are obtained for these quantities, using where necessary numerical integration, and the exact values by either powering the transition matrix or by a matrix inversion on a high-speed computer.

Since the main reference for this study is on a paper by W.T. Ewens (1963) who found that the diffusion approximation results are remarkably accurate for a population of size six, dominance value 0, and S values (0, 0.02, 0.04, 0.06, 0.08 and 0.10) and with no sex differentiation. His results also suggest that diffusion results are more accurate, in an absolute sense, near the boundaries than in the interior of the gene frequency interval - a negation to the results thought by (Fisher 1930, Kolmogorov 1959) that diffusion methods break down near the boundaries within which the variate under selection lies and that branching process techniques are necessary to examine the behaviour of the process near such points. Ewens also found that for small values of S , the diffusion approximation to the probability of ultimate fixation is very close to the exact value and that the approximation errors in the mean times are relatively greater than those for the probability of ultimate fixation.

With small populations the number of different gene frequencies possible is limited. Thus population of size two transient states initial gene frequencies reduce to only three distinct values instead of seven, while that for a population of size three reduce to seven distinct values instead of thirteen. Also it is unlikely that the diffusion approximation is going to be good in small populations in that there is an underlying assumption of a "continuous" gene frequency (i.e. from a large population) and that with small populations this is manifestly not the case. There must be compensation made for inequality of numbers of two sexes in the diffusion approximation, and a single initial gene frequency is determined by taking the average of the initial gene frequencies in both males and females, rather than one in each of the two sexes.

CHAPTER 2

THEORETICAL DEVELOPMENTS

2.1. Consider a population consisting of a fixed number N_m males and N_f females and one locus with only two alleles A_1 and A_2 .

Let the initial proportion of A_1 gametes from the male population be η_1 and proportion of A_2 gametes be η_2 such that $\eta_1 + \eta_2 = 1$. Similarly, the proportion of A_1 and A_2 gametes in the female population are α_1 and α_2 respectively such that $\alpha_1 + \alpha_2 = 1$. We assume that mating between the males and females is equivalent to random union among their gametes, i.e. random mating. The three possible genotypes therefore have the following frequencies.

$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
$\eta_1 \alpha_1$	$\eta_1 \alpha_2 + \eta_2 \alpha_1$	$\eta_2 \alpha_2$

If the relative fitness of the three genotypes are $1+S$, $1+hS$ and 1 respectively, the frequency of A_1 in offspring (Infinite population of Offspring) is thus

$$P = \frac{\eta_1 \alpha_1 (1+S) + \frac{1}{2}(\eta_1 \alpha_2 + \eta_2 \alpha_1) (1 + hS)}{\eta_1 \alpha_1 (1+S) + (1+hS) (\eta_1 \alpha_2 + \eta_2 \alpha_1) + \eta_2 \alpha_2} \quad (1)$$

To form the next generation, samples of $2N_m$ male gametes and $2N_f$ female gametes are chosen at random. We are therefore assuming that family size is random. If at generation t the male population contains $2N_m \eta_1$ A_1 - genes and thus $2N_m (1 - \eta_1)$ A_2 - genes, then the

probability $P(r_m)$ of the $(t + 1)$ th generation containing exactly r_m A_1 - genes among males is given by the binomial distribution

$$P(r_m) = \binom{2N_m}{r_m} P^{r_m} (1 - P)^{2N_m - r_m} \quad (2)$$

$$r_m = 0, 1, 2, \dots, 2N_m$$

Similarly $P(r_f)$ for the female population is given by

$$P(r_f) = \binom{2N_f}{r_f} P^{r_f} (1 - P)^{2N_f - r_f} \quad (3)$$

$$r_f = 0, 1, 2, \dots, 2N_f$$

where P is given in (1). These probabilities depend on η_1, η_2, α_1 , and α_2 .

Since the number of male and female A_1 - genes are determined by independent sampling, the probability that the "next generation" has r_m male gametes, r_f female gametes before selection, i.e. that the gene frequencies are $\left(\frac{r_m}{2N_m}, \frac{r_f}{2N_f} \right)$ or the new η_1, α_1 at generation $(t+1)$ is then $\underline{P} = P(r_m) P(r_f)$ (4)

η_1 male A_1 - gene frequency is given by

$$\eta_1 \in \left[0, \frac{1}{2N_m}, \frac{2}{2N_m}, \dots, \frac{2N_m - 1}{2N_m}, 1 \right]$$

Similarly for the females we have

$$\alpha_1 \in \left[0, \frac{1}{2N_f}, \frac{2}{2N_f}, \dots, \frac{2N_f - 1}{2N_f}, 1 \right]$$

2.2. Let $X(t) = (\eta_1, \alpha_1)$ be a two dimensional vector representing the number of A_1 - genes in Males and in Females in generation t . Then the stochastic process $(X(0), X(1), \dots)$ is a finite Markov chain with state space given by the two parameters (η_1, α_1) and with transition matrix \underline{P} having (4) as its element. State with parameters $(0, 0)$ i.e. $\eta_1 = 0$ and $\alpha_1 = 0$ represents a fixed population consisting entirely of homozygous A_2A_2 individuals. Likewise, state $(1, 1)$ represents a fixed population consisting only of the homozygous A_1A_1 individuals. Thus states with parameters $(0, 0)$ and $(1, 1)$ are absorbing states and the remaining states are transient. Also the finiteness of the chain implies (Parzen 1962) that the probability of the populations ultimately becoming fixed A_1A_1 or A_2A_2 is 1. The transition states are given by the two parameters η_1, α_1 .

PARTICULAR CASE

Population of size two where $N_m = N_f = 1$.

Thus $\eta_1 = 0, \frac{1}{2}, 1$ and
 $\alpha_1 = 0, \frac{1}{2}, 1$

Since there could be 0, 1 and 2 A_1 -genes in each population.

Thus the states are given in terms of the two parameters as follows:

<u>(η_1, α_1)</u>	corresponding to	<u>Male</u>	<u>Female</u>
$(0, 0)$	"	A_2A_2	A_2A_2
$(0, \frac{1}{2})$	"	A_2A_2	A_1A_2
$(0, 1)$	"	A_2A_2	A_1A_1
$(\frac{1}{2}, 0)$	"	A_1A_2	A_2A_2
$(\frac{1}{2}, \frac{1}{2})$	"	A_1A_2	A_1A_2
$(\frac{1}{2}, 1)$	"	A_1A_2	A_1A_1
$(1, 0)$	"	A_1A_1	A_2A_2

$(1, \frac{1}{2})$	Corresponding to	A_1A_1	A_1A_2
$(1, 1)$	"	A_1A_1	A_1A_1

Thus in this case we have a 9-state Markov system where the states are denoted by 1, 2, 3,.....,9 respectively. Our transition matrix \underline{P} is given by

$$(P_{ij}) = \underline{P} = P(r_m) P(r_f) \quad (5)$$

where P_{ij} is the transition probability from i^{th} state (which defines η_1, α_1) to the j^{th} state (defined by r_m, r_f); r_m and r_f are chosen to be appropriate for the j^{th} state, e.g. for a population of size two.

$\underline{r_m}$	$\underline{r_f}$	<u>State</u>	<u>(η_1, α_1)</u>
0	0	1	(0, 0)
0	1	2	$(0, \frac{1}{2})$
0	2	3	(0, 1)
1	0	4	$(\frac{1}{2}, 0)$
1	1	5	$(\frac{1}{2}, \frac{1}{2})$
1	2	6	$(\frac{1}{2}, 1)$
2	0	7	(1, 0)
2	1	8	$(1, \frac{1}{2})$
2	2	9	(1, 1)

The value of P given in (1) is fixed for a particular i in (5) and the j^{th} state is governed by varying r_m and r_f appropriately. Likewise, for a population of size three, where we have 2 males and 1 female (the case 2 females and 1 male is considered to be symmetrical with the first since we are considering only autosomal loci). Here $N_m = 2$ and $N_f = 1$

Thus $\eta_1 = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$ and

$$\alpha_1 = 0, \frac{1}{2}, 1$$

since we could have initially 0, 1, 2, 3 and 4 A_1 - male genes and 0, 1, and 2 A_1 - female genes. In this case we have a 15-state Markov system and each state governed by P in (1) and r_m and r_f . The 15 states are given consecutively with respect to their parameter values.

<u>State</u>	<u>(η_1, α_1)</u>
1	(0, 0)
2	(0, $\frac{1}{2}$)
3	(0, 1)
4	($\frac{1}{4}$, 0)
5	($\frac{1}{4}$, $\frac{1}{2}$)
6	($\frac{1}{4}$, 1)
7	($\frac{1}{2}$, 0)
8	($\frac{1}{2}$, $\frac{1}{2}$)
9	($\frac{1}{2}$, 1)
10	($\frac{3}{4}$, 0)
11	($\frac{3}{4}$, $\frac{1}{2}$)
12	($\frac{3}{4}$, 1)
13	(1, 0)
14	(1, $\frac{1}{2}$)
15	(1, 1)

2.3. DIFFUSION APPROACH

The diffusion approach makes use of the Kolmogorov forward and backward equations which are briefly described as follows:

We assume in subsequent parts of this discussion that the change in gene frequencies is Markovian, that is, the probability distribution of gene frequencies at a given moment t depends on the gene frequencies at a preceding time t_0 ($t_0 < t$) but not on the previous history which has led to the gene frequencies at t_0 . It is also assumed that the change of gene frequency is a continuous process, that is, for any given positive value, ϵ , the probability that the change in x during the time interval $(t, t + \delta t)$ exceeds ϵ is $O(\delta t)$ i.e. an infinitesimal of higher order than δt where x is the gene frequency of A_1 at a time t . We further assume that there are no overlapping generations.

For the Kolmogorov forward equation, we consider a pair of allele A_1 and A_2 with respective frequencies x and $1-x$. Let $\phi(p, x, t)$ be the conditional probability density that the gene frequency is x at time t , given that the initial frequency is p at time $t = 0$. This gives the transition probability that the gene frequency moves from p to x after time t . With p fixed, $\phi(p, x, t)$ determines a frequency distribution such that when $1/2N_e$ is substituted for dx , $\phi(p, x, t) dx$ gives an approximation to the frequency of the class with gene frequency x ($0 < x < 1$) at time t . This frequency distribution may be denoted by (Kimura, 1964).

$$\bar{f}(x, t) = \phi(p, x, t) / 2N_e \quad (0 < x < 1)$$

 (6)

since p is fixed, $\phi(p, x, t)$ will then be written as $\phi(x, t)$.

Note that the above relation holds only for unfixed classes, i.e.

$0 < x < 1$. Frequencies of classes with $x = 0$ or 1 have to be treated separately.

(Kimura 1964) showed that $\phi(p, x, t)$ satisfied the Kolmogorov forward equation

$$\frac{\partial \phi(p, x, t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} \left\{ V(x, t) \phi(p, x, t) \right\} - \frac{\partial}{\partial x} \left\{ M(x, t) \phi(p, x, t) \right\} \quad (7)$$

where $M(x, t)$ and $V(x, t)$ refer to the first and second moments of δx during the infinitesimal time interval $(t, t + \delta t)$. The derivation of the above equation assumes that

$$\lim_{t \rightarrow 0} \frac{1}{\delta t} (\delta x)^n g(\delta x, x; \delta t, t) d(\delta x) = 0 \quad (8)$$

for $n \geq 3$

i.e. It is assumed that higher order moments are negligible.

Since effects of quantities such as mutation rates, rate of migration, intensity of selection, and effect of random sampling of gametes which determine x are all measured with one generation as a time unit, Kimura (1964) suggested we replace $M(x, t)$ and $V(x, t)$, in the above equation, by $M_{\delta x}$ and $V_{\delta x}$, the mean and variance of the change in gene frequency per generation (δt corresponds to

one generation). Thus we obtain

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} (V_{\delta x} \phi) - \frac{\partial}{\partial x} (M_{\delta x} \phi) \quad (9)$$

which is the Kolmogorov forward equation. The replacement of $V(x, t)$ and $M(x, t)$ by the Variance $V_{\delta x}$ and mean $M_{\delta x}$ suggested by Kimura is based on the consideration that (9) should give the deterministic process correctly when there is no random fluctuation, i.e. in the limit when $V_{\delta x} = 0$.

Since the gene frequency x lies between 0 and 1 in general, the process of change in gene frequency in a population through time is represented as the stochastic movement of a point x on the closed real interval $[0, 1]$. Equation (9) thus describes the movement at least on the open interval $(0, 1)$.

In the above, we have treated gene frequencies after t generations as random variables and initial gene frequencies as fixed. For example, in the expression $\phi(p, x, t)$, x is considered as a random variable and p is assumed fixed. This means that we have considered the process of change in gene frequency in the forward direction in time.

On the other hand, we may regard x as fixed and consider p as a random variable. Namely, we reverse the time sequence and view the process retrospectively. We will assume in what follows, that the process is time homogeneous. That is, if $X(t_1)$ and $X(t_2)$ are

respectively frequencies of a gene at times t_1 and t_2 ($t_2 > t_1$), then the probability distribution of $X(t_2)$ given $X(t_1)$, which in general should be a function of t_1 and t_2 separately, depends only on the difference ($t_2 - t_1$). Then we have

$$\phi(p, x, t + \delta t) = \int g(\delta p, p, \delta t) \phi(p + \delta p, x, t) d(\delta p) \quad (10)$$

The above equation contains g as a function of three variables only, i.e. δp , p and δt . This is because the probability that the gene frequency changes from p to $p + \delta p$ during the time interval of length δt is the same for any t (generation) due to the assumption of time homogeneity. Expanding $\phi(p + \delta p, x, t)$ on the right hand side of the above equation in terms of δp , and neglecting higher order moments other than the first and second, we obtain

$$\frac{\partial \phi(p, x, t)}{\partial t} = \frac{V(x)}{2} \frac{\partial^2 \phi(p, x, t)}{\partial p^2} + M(x) \frac{\partial \phi(p, x, t)}{\partial p} \quad (11)$$

where $V(x)$ and $M(x)$ refer to the second and first moments of δp during the infinitesimal time interval ($t, t + \delta t$) respectively.

Re-writing equation (11) in terms of one generation as unit of time we have

$$\frac{\partial \phi}{\partial t} = \frac{V_{\delta p}}{2} \frac{\partial^2 \phi}{\partial p^2} + M_{\delta p} \frac{\partial \phi}{\partial p} \quad (12)$$

Note here that the initial gene frequency p is the variable and x is assumed to be constant. Equation 11 is the Kolmogorov backward

equation as applied to the time homogeneous case, and is the adjoint form of equation (7).

The initial A_1 - gene frequency is obtained by taking the average of the parameter values η_1 , and α_2 since, if we consider the population at the "Infinite" stage, we must use an average gene frequency, i.e. $\frac{1}{2} (p_m + p_f)$ because in the infinite population any gene has probability (of being from male) = Prob(of being from female) = $\frac{1}{2}$ with conditional probability p_m (when derived from male) and p_f (when derived from female) of being A_1 type. Further, the derivation by (wright, 1931) of

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f} \quad (13)$$

made use of the average gene frequency. Here N_e is the effective population size.

CHAPTER 3

PROBABILITY OF FIXATION BY THE
nth GENERATION

3.1. EXACT APPROACH

Suppose we write $R_j^{(n)}$, the probability that the system is in state j at the n^{th} generation as the row vector $\underline{R}^{(n)}$, the initial distribution being $\underline{R}^{(0)}$. Now the distribution at the first generation is given by

$$R_j^{(1)} = \sum_{i=1}^{\infty} P_{ij} R_i^{(0)} \quad \text{_____ (1)}$$

since the probability that the system is in state i initially is $R_i^{(0)}$, and the probability of a transition from i to j is \underline{P} defined in Chapter 2, equation(5), and we must sum over all transitions leading to state j . In matrix notation, we can express (1) as

$$\underline{R}^{(1)} = \underline{R}^{(0)} \underline{P} \quad \text{where } \underline{P} \text{ is as defined in Chapter 2, equation(5).}$$

$$\text{Similarly } R^{(2)} = \underline{R}^{(1)} \underline{P} = \underline{R}^{(0)} \underline{P}^2$$

$$\text{and in general } \underline{R}^{(n)} = \underline{R}^{(0)} \underline{P}^n \quad \text{_____ (2)}$$

The matrix \underline{P}^n therefore gives the required set of n -step transition probabilities $\left\{ P_{ij}^{(n)} \right\}$. In the numerical calculations we found the first, second, fourth,..... one hundred and twenty eighth powers of \underline{P} , by using the subroutine FOICKF package of the Northingham Algorithm Group, ICL 1900 series. By the 64th generation, the probability that both A_1 and A_2 genes are still present is very small (of order 10^{-8}). These numerical results were obtained for the two populations, when the value of $S = 0$ (no selection.) The numerical

results obtained by powering a transition matrix are subject to the rounding error in the computer. However, these results were checked by adding the elements in each row of each power of the matrix. It was found that the sums thus obtained differed from unity, at the most, in the eighth decimal place, so that since all results are given here to four decimal places, only they may be taken as being correct to this order.

3.2. DIFFUSION APPROACH

The diffusion approach makes use of the Kolmogorov forward equation given in Chapter 2 by equation (9).

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} (V \zeta_x \phi) - \frac{\partial}{\partial x} (M \zeta_x \phi) \quad \text{_____ (3)}$$

Let A_1 and A_2 be a pair of alleles with respective frequencies x and $1-x$. We assume that mating is at random and mode of reproduction is such that N_m male and N_f female gametes are drawn as a random sample from the population to form the next generation.

The selective advantages of A_1A_1 and A_1A_2 genotypes over A_2A_2 genotype are S and Sh respectively. The mean and variance of the change of gene frequency x per generation are (see appendix)

$$M \zeta_x = Sx(1-x) \left\{ h + (1-2h)x \right\} \quad \text{_____ (4)}$$

and

$$V \zeta_x = \frac{x(1-x)}{2N_e} \quad \text{_____ (5)}$$

where N_e is the effective population size given by Equation (13) in Chapter 2.

We shall consider the case of $S = 0$ only as solutions for other values of S are difficult.

For $S = 0$

$M_{\phi_x} = 0$, and hence Equation (3) above becomes

$$\frac{\partial \phi}{\partial t} = \frac{1}{4N_e} \frac{\partial^2}{\partial x^2} \left\{ x(1-x) \phi \right\} \quad 0 < x < 1 \quad \text{_____ (6)}$$

where $\phi = \phi(p, x, t)$ is the probability density that the gene frequency becomes x in the t^{th} generation, given that it is p at $t = 0$, i.e. the initial condition is

$$\phi(p, x, t) = \delta(x - p) \quad \text{_____ (7)}$$

where $\delta(\cdot)$ is the Dirac delta function. Solution of (6), is obtained by Kimura (1964) using separation of variables as

$$\phi(p, x, t) = \sum_{i=1}^{\infty} \frac{(2i+1)(1-r^2)}{i(i+1)} T_{i-1}^1(r) T_{i-1}^1(Z) e^{-\frac{i(i+1)t}{4N_e}} \quad \text{_____ (8)}$$

where $r = 1 - 2p$, $Z = 1 - 2x$ and $T_{i-1}^1(r)$ is the Gegenbauer polynomial in r .

The above solution is valid only for our class of A_1 gene frequency in the open interval $(0, 1)$, but considering the rate of change in probability in the terminal classes Crow and Kimura (1970) showed that the frequency of the class in which gene A_1 is fixed or the probability that A_1 becomes fixed by the t^{th} generation is as follows:

$$\begin{aligned}
 f(p, 1, t) &= \frac{1}{4N_e} \int_0^t \phi(p, 1, t) dt \\
 &= p + \frac{1}{2} \sum_{i=1}^{\infty} \frac{(2i+1)(1-r^2)}{i(i+1)} T_{i-1}^1(r) (-1)^i e^{-\frac{i(i+1)t}{4N_e}}
 \end{aligned}
 \tag{9}$$

We have $t=8, 16, 32$ and 64 in our problem, thus for $i=5, N_e = 2$ (population of size 2) and $t=8$, we have

$$e^{-\frac{i(i+1)t}{4N_e}} = e^{-30} \text{ which we see is very small for our}$$

smallest value of t and since for a given $i, T_{i-1}^1(r)$ is a polynomial in p (the initial gene frequency), for $r = 1-2p$ from (8), it follows that $T_{i-1}^1(r)$ is finite and not very large for our class of values of p in the open interval $0 < p < 1$. From the foregoing, it therefore follows intuitively that our summation should be reasonably good for values of i between 1 and 4 only.

Putting $r = 1-2p$ we have

$$1 - r^2 = 4p(1-p)$$

thus (9) can be written as probability (fixation by t^{th} generation)

$$\begin{aligned}
 &= p + \sum_{i=1}^{\infty} (-1)^i \frac{2(2i+1)p(1-p)}{i(i+1)} T_{i-1}^1(2p) e^{-\frac{i(i+1)t}{4N_e}}
 \end{aligned}
 \tag{10}$$

Thus summing over values of i from 1 to 4, we have

$$\begin{aligned}
 &\text{Probability of Fixation} \\
 \underline{P} &= p - \frac{2(3)p(1-p)}{2} T_0^1(1-2p) e^{-2t/4N_e}
 \end{aligned}$$

$$\begin{aligned}
& + \frac{2(5) p(1-p)}{6} T_1^1(1-2p) e^{-6t/4N} \\
& - \frac{2(7) p(1-p)}{12} T_2^1(1-2p) e^{-12t/4N} \\
& + \frac{2(9) p(1-p)}{20} T_3^1(1-2p) e^{-20t/4N}
\end{aligned}$$

$$T_0^1(r) = 1$$

$$T_1^1(r) = 3r$$

$$T_2^1(r) = \frac{3}{2} (5r^2 - 1)$$

$$T_3^1(r) = \frac{5}{2} (7r^3 - 3r)$$

where $r = 1-2p$, thus we have for probability of fixation

$$\begin{aligned}
\underline{f} &= p - 3p(1-p) e^{-2t/4N} + 5p(1-p)(1-2p) e^{-6t/4N} \\
& - 7p(1-p)(1-5p+5p^2) e^{-12t/4N} \\
& + 9p(1-p)(1-2p)(1-7p+7p^2) e^{-20t/4N}
\end{aligned}$$

The probability of fixation by the 8th, 16th, 32nd and 64th generations given by the above expression is then obtained numerically on the computer for the various values of the initial gene frequency. The two populations effective sizes are 2 and $2^2/3$. Here as in all

subsequent discussion in this work, the initial gene frequency is taken as the average of the two parameters η_1 and α_1 . Comparisons are made with the exact solution when there is no selection.

3.3. RESULTS

Tables 1 and 2 give the exact values and diffusion approximations (D.A) for the probability that the A_1 - genes have become fixed in populations of size 3, and 2 respectively, by the j^{th} generation, given initial proportion of A_1 genes as the average of the two parameters η_1 and α_1 .

Figures 1, 2 and 3 refer to graphs of R (probability of fixation by j^{th} generation) against generations for fixed initial A_1 - genes frequencies 0.25, 0.50 and 0.75 respectively. Graphs a and b in each figure refer to populations of sizes 3 and 2 respectively.

TABLE 1.

States (Transient)	Parameters (η_1 α_1)	j = 8		j = 16		j = 32		j = 64	
		EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	(0 $\frac{1}{2}$)	0.1872	0.1297	0.2465	0.2221	0.2500	0.2486	0.2500	0.2500
3	(0 1)	0.4254	0.3327	0.4964	0.4626	0.5000	0.4981	0.5000	0.5000
4	($\frac{1}{4}$ 0)	0.0952	0.0563	0.1236	0.1087	0.1250	0.1242	0.1250	0.1250
5	($\frac{1}{4}$ $\frac{1}{2}$)	0.3137	0.2214	0.3725	0.3400	0.3750	0.3732	0.3750	0.3750
6	($\frac{1}{4}$ 1)	0.5704	0.4649	0.6233	0.5899	0.6250	0.6232	0.6250	0.6250
7	($\frac{1}{2}$ 0)	0.2186	0.1297	0.2491	0.2221	0.2500	0.2486	0.2500	0.2500
8	($\frac{1}{2}$ $\frac{1}{2}$)	0.4563	0.3327	0.4989	0.4626	0.5000	0.4981	0.5000	0.5000
9	($\frac{1}{2}$ 1)	0.7218	0.6193	0.7496	0.7219	0.7500	0.7486	0.7500	0.7500
10	($\frac{3}{4}$ 0)	0.3558	0.2214	0.3748	0.3400	0.3750	0.3732	0.3750	0.3750
11	($\frac{3}{4}$ $\frac{1}{2}$)	0.6046	0.4649	0.6248	0.5899	0.6250	0.6232	0.6250	0.6250
12	($\frac{3}{4}$ 1)	0.8678	0.7972	0.8750	0.8586	0.8750	0.8742	0.8750	0.8750
13	(1 0)	0.4923	0.3327	0.5000	0.4626	0.5000	0.4981	0.5000	0.5000
14	(1 $\frac{1}{2}$)	0.7447	0.6193	0.7500	0.7219	0.7500	0.7486	0.7500	0.7500

TABLE 2.

States (Transient)	Parameters (η_1 α_1)	j = 8		j = 16		j = 32		j = 64	
		EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	(0 $\frac{1}{2}$)	0.2226	0.1750	0.2493	0.2397	0.2500	0.2498	0.2500	0.2500
3	(0 1)	0.4726	0.3985	0.4995	0.4863	0.5000	0.4997	0.5000	0.5000
4	($\frac{1}{2}$ 0)	0.2334	0.2750	0.2497	0.2397	0.2500	0.2498	0.2500	0.2500
5	($\frac{1}{2}$ $\frac{1}{2}$)	0.4817	0.3985	0.4998	0.4963	0.5000	0.4997	0.5000	0.5000
6	($\frac{1}{2}$ 1)	0.7427	0.6727	0.7500	0.7397	0.7500	0.7498	0.7500	0.7500
7	(1 0)	0.4961	0.3985	0.5000	0.4863	0.5000	0.4997	0.5000	0.5000
8	(1 $\frac{1}{2}$)	0.7476	0.6727	0.7500	0.7397	0.7500	0.7498	0.7400	0.7500

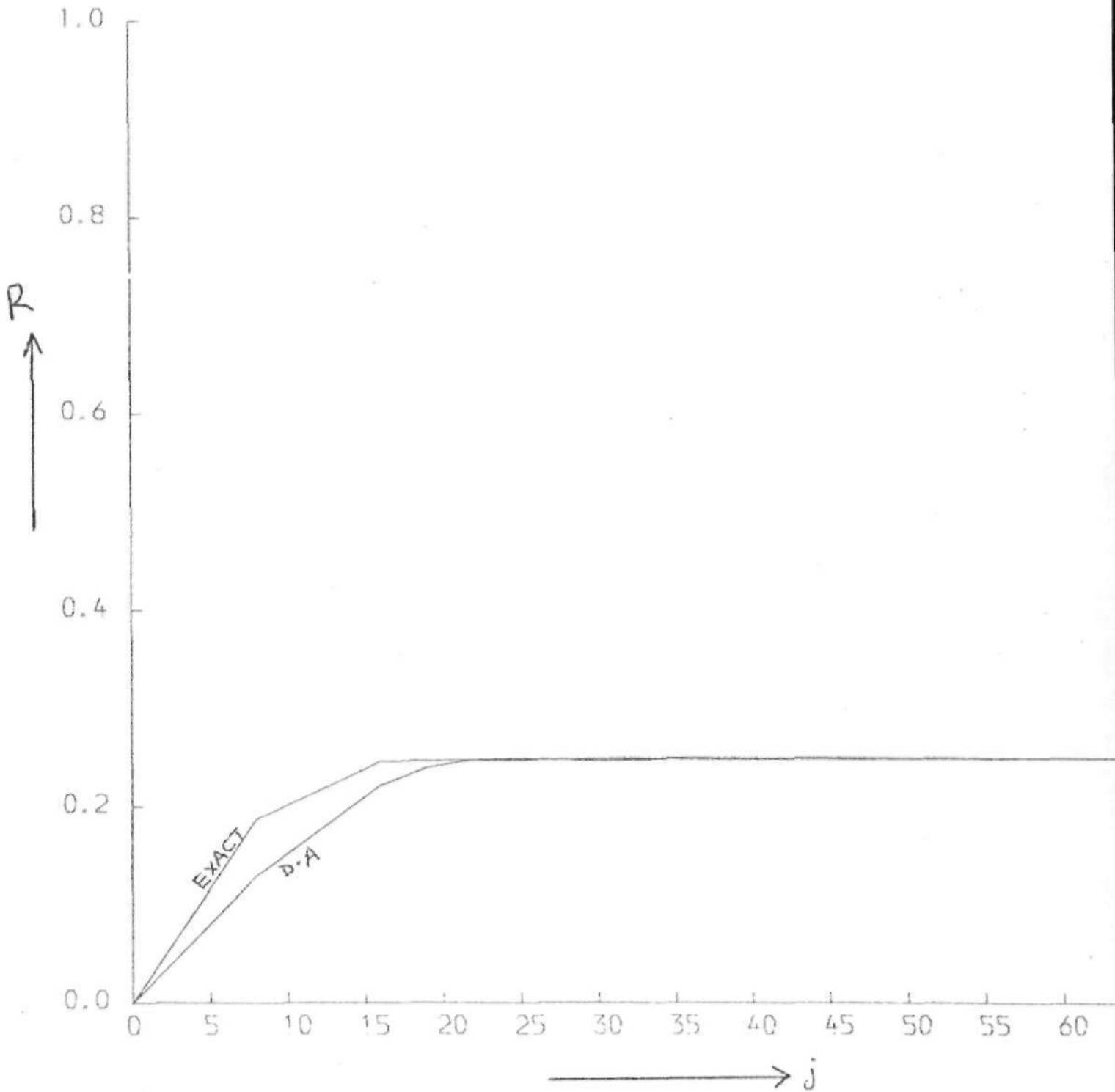
FIGURE 1a

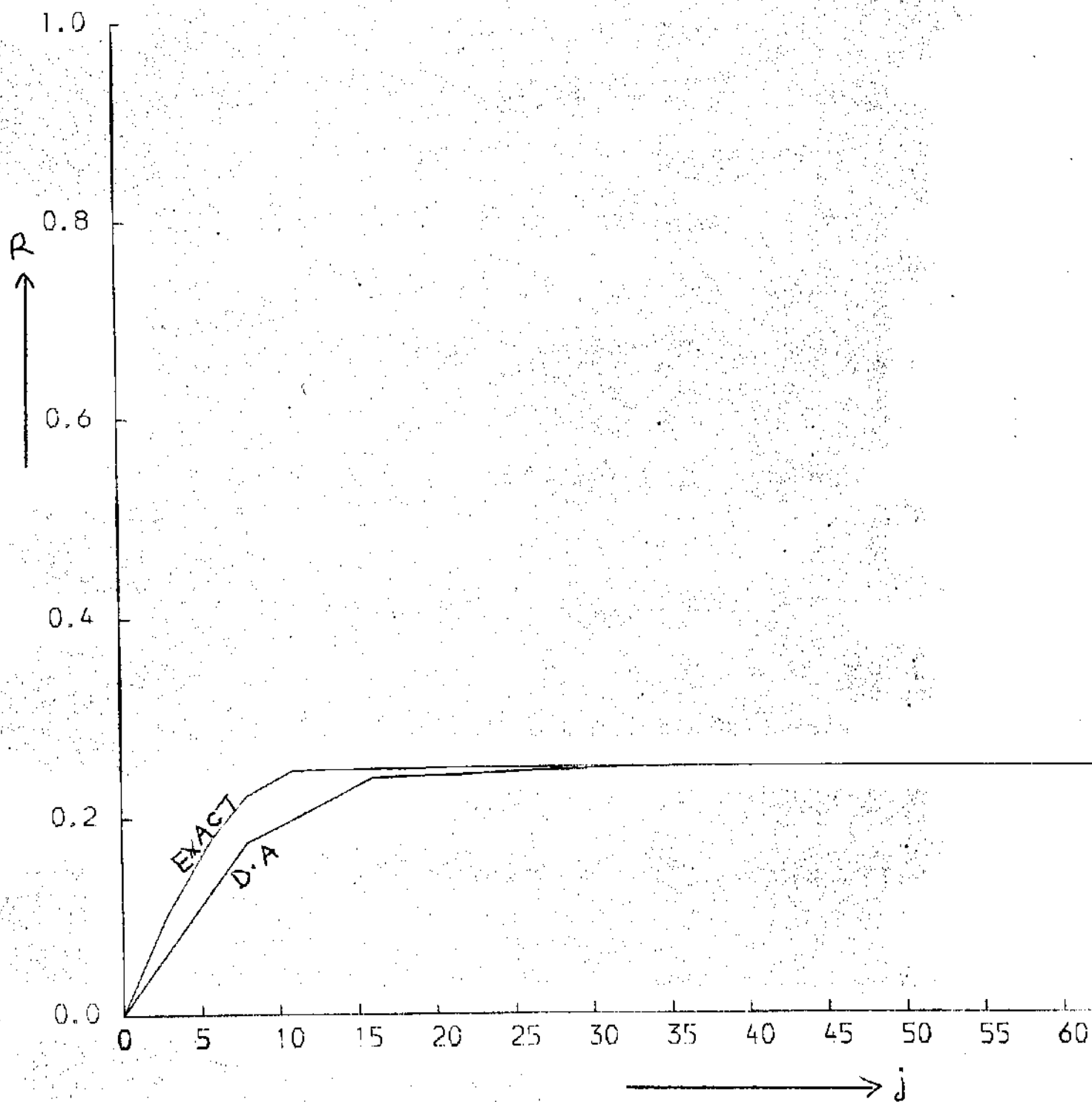
FIGURE 1b

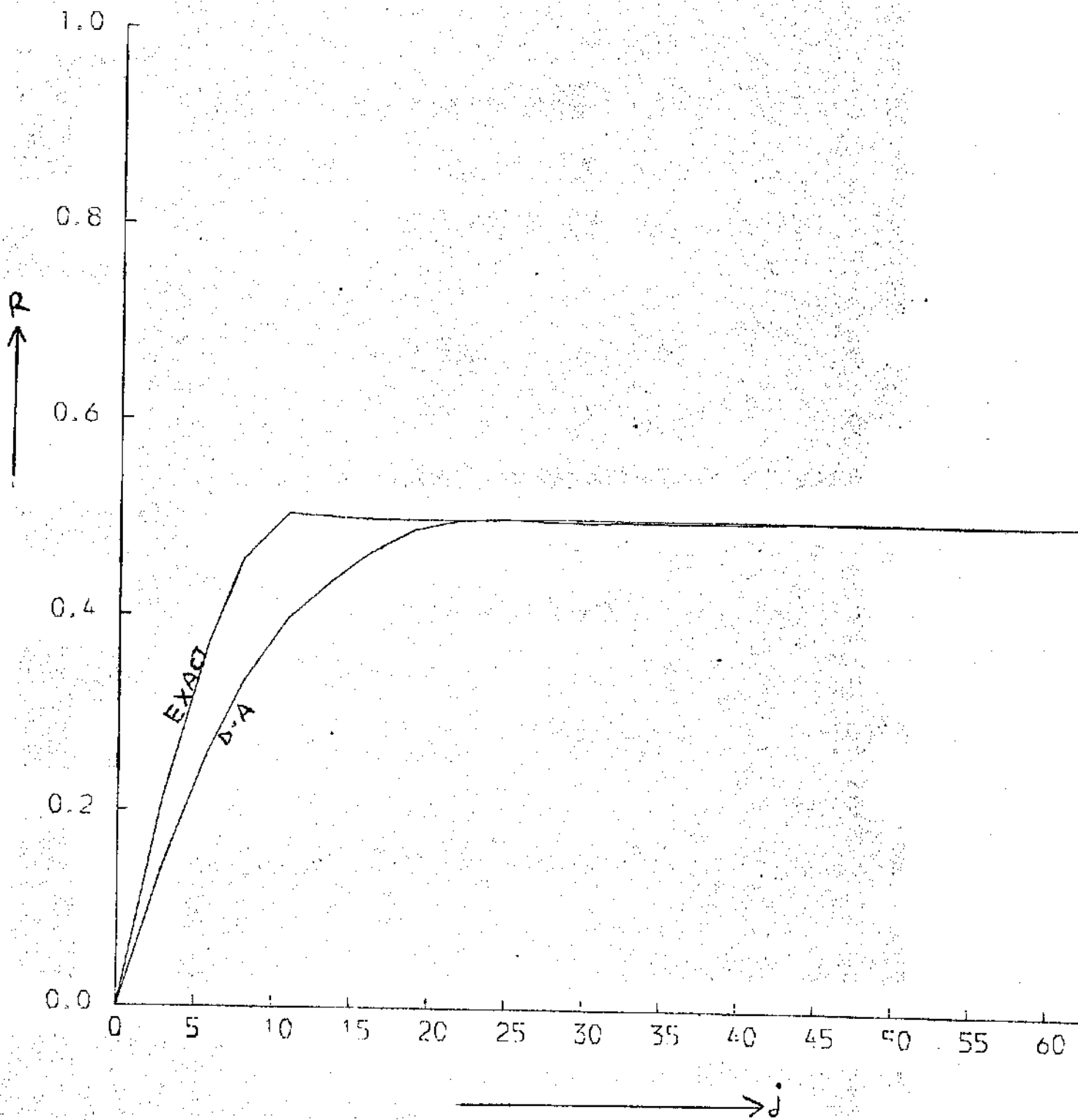
FIGURE 2a

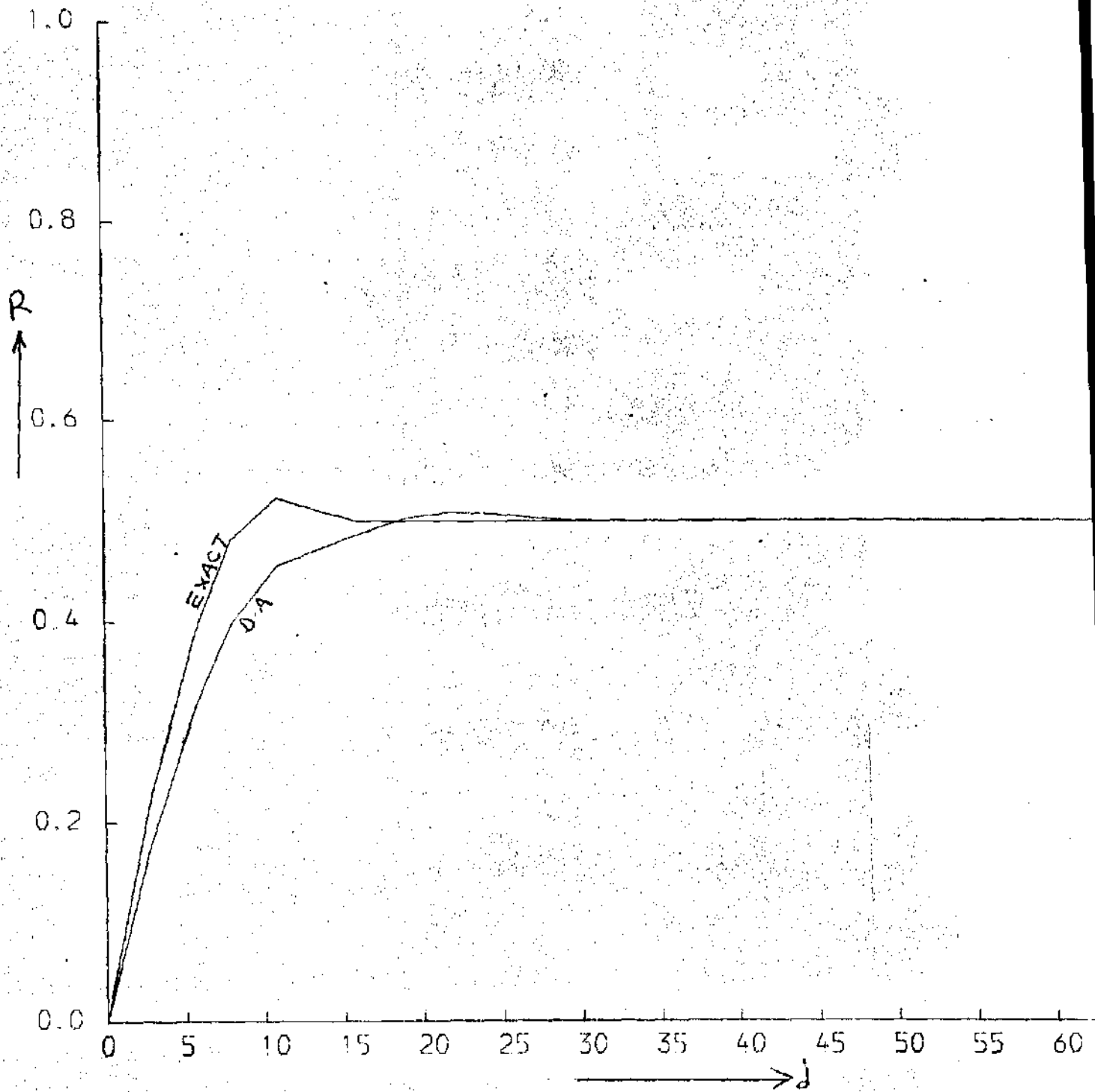
FIGURE 2b

FIGURE 3a

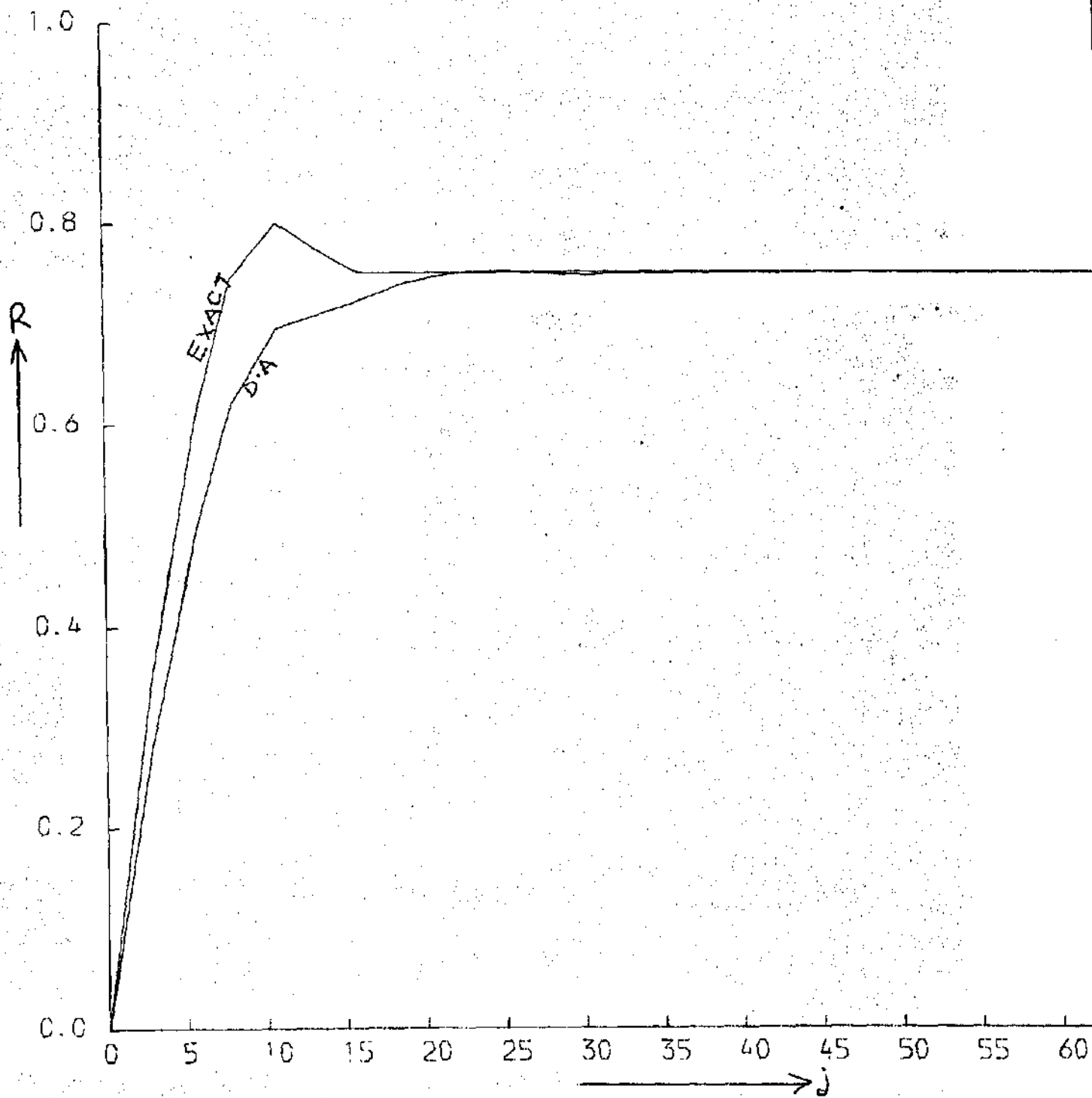
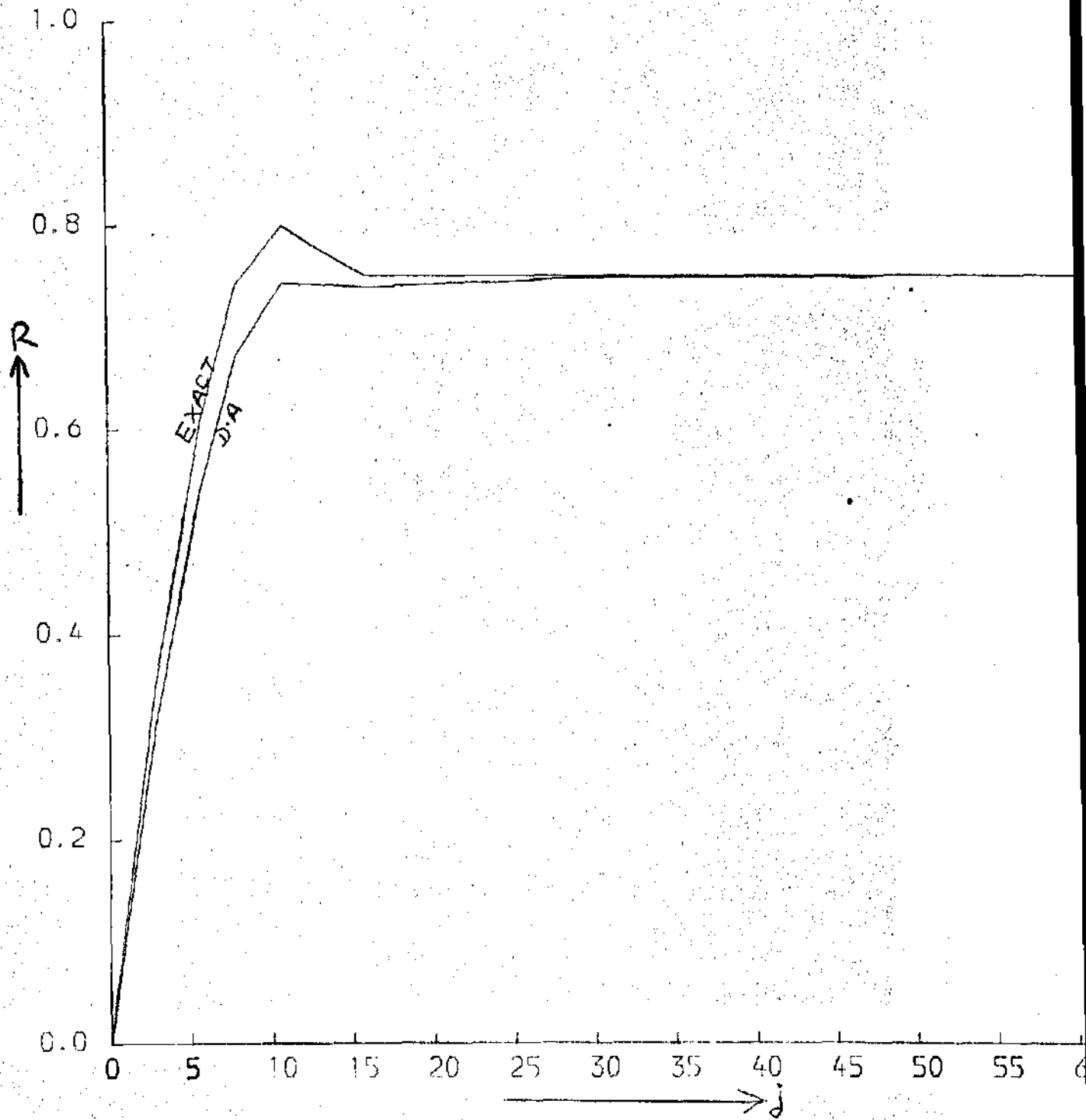


FIGURE 3b

3.4. DISCUSSION

It was found that the diffusion approximation becomes more accurate as j increases as would be expected. For $j = 64$ the approximation gives the exact value to four decimal places considered in this study. For j less than eight, the diffusion approximation is very poor. From the tables and graphs it will be noticed that the diffusion approximation underestimates the true probability for all initial gene frequencies and all values of j considered for both populations. The results also suggest that the diffusion approximation errors are less in a population of size 2 than in a population of size 3. (This could be better appreciated by comparison of any pair of the figures labelled a and b drawn to the same scale).

Further, the Diffusion approximation gave the same results for all those states with the same initial gene frequency (average of the two parameters) even with states with different parameters, while the exact results differentiate between these states. Thus states 3 and 8 in Table 1 give the same diffusion result, while the exact approach give slightly different results. The same holds for all other states whose average frequencies are the same, but whose initial parameter values are different. The convergence of the diffusion results to the exact results as j increases also suggests that the approximation used for our summation in 3.2 is reasonably good.

CHAPTER 4

PROBABILITY OF ULTIMATE FIXATION OF
AND ALLELE (A₁)

4.1. EXACT APPROACH

Consider here the case of population of size 3 with 2 males and 1 female. As stated in Chapter 2, we thus have a 15-state Markov System with the first and last states as the absorbing states

Let U_j be the probability of ultimate fixation of A_1 - allele from state j to state 15.

$$\begin{aligned} \text{Then } U_j &= \text{Prob}(\text{from } j^{\text{th}} \text{ state to } 15^{\text{th}} \text{ state}) \\ &= \sum_{K=1}^{15} \text{Prob}(\text{from } j^{\text{th}} \text{ state to } K^{\text{th}} \text{ state in one step}) \\ &\quad \times \text{Prob}(\text{from } K^{\text{th}} \text{ state ultimately to the } 15^{\text{th}} \text{ state}) \\ \text{i.e. } U_j &= \sum_{K=1}^{15} P_{jK} U_K \quad \text{_____ (1)} \end{aligned}$$

where P_{jK} is the probability from j^{th} state in one step to the K^{th} state and U_K is the probability from K^{th} state ultimately to the 15th state.

$U_0 = 0$ since it is impossible to go to the 15^{th} state from the first state.

Similarly $U_{15} = 1$

Thus from (1)

$$U_j = \sum_{K=2}^{14} P_{jK} U_K + P_{j, 15}$$

$j = 2, 3, 4, \dots, 14$

Let
$$\begin{pmatrix} U_2 \\ \vdots \\ U_{14} \end{pmatrix} = \underline{U} \quad \text{a } 13 \times 1 \text{ column vector}$$

$$\begin{pmatrix} P_{2,2} & P_{2,3} & \dots & P_{2,14} \\ P_{14,2} & P_{14,3} & \dots & P_{14,14} \end{pmatrix} = \underline{Q}$$

and
$$\begin{pmatrix} P_{2,15} \\ \vdots \\ P_{14,15} \end{pmatrix} = \underline{r} \quad \text{a } 13 \times 1 \text{ column vector}$$

Thus our system of equation reduced to the matrix form

$$\underline{U} = \underline{Q} \underline{U} + \underline{r} \quad \text{_____ (2)}$$

where \underline{Q} as defined above is a 13×13 matrix of transient state probabilities from equation (2)

$$\underline{U} = (\underline{I} - \underline{Q})^{-1} \underline{r} \quad \text{_____ (3)}$$

where \underline{I} is a 13×13 unit matrix.

The matrix $(\underline{I} - \underline{Q})^{-1} = \sum_{n=0}^{\infty} \underline{Q}^n$ does exist since \underline{Q} is the matrix of transient state probabilities and $\lim_{n \rightarrow \infty} \underline{Q}^n = \underline{0}$, the null matrix (Cox and Miller 1965).

Likewise an equation similar to (3) also governs the probability of ultimate fixation of A_1 - gene for the population of size 2. Here however, \underline{Q} reduces to a 7×7 matrix of transient state probabilities. Equation (3) was solved for various values of the parameters S and h . For the two populations, solutions were obtained numerically on 1904 Computer system by using the subroutine FOIAAF (Inversion of a non-symmetric matrix) package of the Northingham Algorithm Group ICL 1900 series, after performing the operation $\underline{I} - \underline{Q}$.

4.2. DIFFUSION APPROXIMATION TO THE PROBABILITY OF ULTIMATE FIXATION

Approximation results are obtained by employing the Kolmogorov backward equation (Chapter 2) in terms of one generation as unit of time

$$\frac{\partial \phi}{\partial t} = \frac{V\sigma_p}{2} \frac{\partial^2 \phi}{\partial p^2} + M\sigma_p \frac{\partial \phi}{\partial p} \quad \text{_____} (4)$$

When $x = 1$, ϕ in (4) gives the probability that the gene whose initial frequency was p has become fixed in the population by the t^{th} generation. We will denote this probability by $U(p, t)$, for which we have

$$\frac{\partial U(p, t)}{\partial t} = \frac{V\sigma_p}{2} \frac{\partial^2 U(p, t)}{\partial p^2} + M\sigma_p \frac{\partial U(p, t)}{\partial p} \quad \text{_____} (5)$$

The probability of fixation by a given time t , is the solution of (5) with the boundary conditions

$$U(0, t) = 0, \quad U(1, t) = 1$$

The probability of ultimate fixation defined by

$$U(p) = \lim_{t \rightarrow \infty} U(p, t)$$

Since $\frac{\partial U(p)}{\partial t} = 0$ for this quantity, the above equation reduces to the ordinary differential equation

$$\frac{V\sigma_p}{2} \frac{d^2 U(p)}{dp^2} + M\sigma_p \frac{d U(p)}{dp} = 0 \quad \text{_____} (6)$$

with boundary conditions

$$U(0) = 0, \quad U(1) = 1 \quad \text{_____ (7)}$$

From Chapter 3 and equations (4) and (5) and Chapter 2, equation (13)

we have

$$M_{\sigma x} = Sx(1-x) \left\{ h + (1-2h)x \right\}$$

$$V_{\sigma x} = \frac{x(1-x)}{2N_e} \quad \text{and}$$

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f}$$

Equation (6) thus becomes, after a change of notation

$$\frac{x(1-x)}{4N_e} \frac{d^2 U}{dp^2} + Sx(1-x) \left\{ h + (1-2h)x \right\} \frac{dU}{dp} = 0$$

$$\text{or } \frac{d^2 U}{dp^2} \Big/ \frac{dU}{dp} = -S \left\{ h + (1-2h)x \right\} 4N_e \quad \text{_____ (8)}$$

$$\text{i.e. } \frac{dU}{dp} = e^{-\int 4N_e S \left\{ h + (1-2h)x \right\} dx}$$

$$\text{Let } G(x) = e^{-\int 4N_e S \left\{ h + (1-2h)x \right\} dx}$$

$$\text{i.e. } G(x) = \exp \left[-4N_e S \left\{ hx + (1-2h) \frac{x^2}{2} \right\} \right] \quad \text{_____ (9)}$$

Therefore

$U(p) = \int_0^p G(x) dx$ and with the boundary conditions, we have

$$U(p) = \frac{\int_0^p G(x) dx}{\int_0^1 G(x) dx} \quad (10)$$

where $G(x)$ is as defined in (9).

Numerical solutions of (10) using Simpson's rule (with 100 sub divisions and accuracy is 1×10^{-6}) were obtained for various values of S , h and p (the Initial gene frequency) for the two populations.

ultimate fixation as ordinates, versus Selection Intensity, S , and increasing dominance values 0 , $\frac{1}{2}$, 1 and $1\frac{1}{4}$ in a population of size 3 with initial gene frequencies fixed at 0.375 , 0.50 and 0.75 respectively. Graphs of Figures 9b, 10b and 11b represent the same quantities in a population of size 2 except that 9b has a fixed initial gene frequency at 0.25 (since there is no initial gene frequency of 0.375 in a population of size 2).

TABLE 3a.

S = 0

States	Parameters η_1 α_1	All H Values	
		EXACT	D.A
2	0 $\frac{1}{2}$	0.2500	0.2500
3	0 1	0.5000	0.5000
4	$\frac{1}{4}$ 0	0.1250	0.1250
5	$(\frac{1}{4} \frac{1}{2})$	0.3750	0.3750
6	$(\frac{1}{4} 1)$	0.6250	0.6250
7	$\frac{1}{2}$ 0	0.2500	0.2500
8	$\frac{1}{2} \frac{1}{2}$	0.5000	0.5000
9	$\frac{1}{2}$ 1	0.7500	0.7500
10	$\frac{3}{4}$ 0	0.3750	0.3750
11	$\frac{3}{4} \frac{1}{2}$	0.6250	0.6250
12	$\frac{3}{4}$ 1	0.8750	0.8750
13	1 0	0.5000	0.5000
14	1 $\frac{1}{2}$	0.7500	0.7500

TABLE 3b.

States	Parameters		All H Values	
	η_1	α_1	EXACT	D.A
2	0	$\frac{1}{2}$	0.2500	0.2500
3	0	1	0.5000	0.5000
4	$\frac{1}{2}$	0	0.2500	0.2500
5	$\frac{1}{2}$	$\frac{1}{2}$	0.5000	0.5000
6	$\frac{1}{2}$	1	0.7500	0.7500
7	1	0	0.5000	0.5000
8	1	$\frac{1}{2}$	0.7500	0.7500

S = 0

TABLE 4a.

States	Parameters η_1 α_1	H = 0.0		H = 0.5		H = 1.0		H = 1.25		
		EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A	
		2	(0	$\frac{1}{2}$)	0.2599	0.2667	0.2649	0.2703	0.2699	0.2740
3	(0	1)	0.5162	0.5264	0.5163	0.5266	0.5164	0.5269	0.5165	0.5270
4	($\frac{1}{4}$	0)	0.1301	0.1338	0.1342	0.1370	0.1383	0.1402	0.1404	0.1418
5	($\frac{1}{4}$	$\frac{1}{2}$)	0.3920	0.3978	0.3948	0.4002	0.3975	0.4025	0.3988	0.4037
6	($\frac{1}{4}$	1)	0.6453	0.6516	0.6420	0.6498	0.6389	0.6479	0.6373	0.6470
7	($\frac{1}{2}$	0)	0.2599	0.2667	0.2649	0.2703	0.2699	0.2740	0.2723	0.2758
8	($\frac{1}{2}$	$\frac{1}{2}$)	0.5212	0.5264	0.5212	0.5266	0.5212	0.5269	0.5212	0.5270
9	($\frac{1}{2}$	1)	0.7692	0.7726	0.7645	0.7696	0.7599	0.7666	0.7576	0.7651
10	($\frac{3}{4}$	0)	0.3888	0.3978	0.3923	0.4002	0.3956	0.4025	0.3973	0.4037
11	($\frac{3}{4}$	$\frac{1}{2}$)	0.6471	0.6516	0.6443	0.6498	0.6418	0.6474	0.6405	0.6470
12	($\frac{3}{4}$	1)	0.8874	0.8890	0.8838	0.8863	0.8801	0.8837	0.8782	0.8823
13	(1	0)	0.5162	0.5264	0.5163	0.5266	0.5164	0.5269	0.5165	0.5270
14	(1	$\frac{1}{2}$)	0.7692	0.7726	0.7645	0.7696	0.7599	0.7666	0.7576	0.7651

S = 0.04

TABLE 4b.

S = 0.04

States	Parameters η_1 α_1		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
			EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	0	$\frac{1}{2}$	0.2559	0.2625	0.2599	0.2652	0.2638	0.2679	0.2658	0.2692
3	0	1	0.5098	0.5198	0.5098	0.5200	0.5098	0.5201	0.5098	0.5202
4	$\frac{1}{2}$	0	0.2559	0.2625	0.2599	0.2652	0.2638	0.2679	0.2658	0.2692
5	$\frac{1}{2}$	$\frac{1}{2}$	0.5148	0.5198	0.5147	0.5200	0.5146	0.5201	0.5146	0.5202
6	$\frac{1}{2}$	1	0.7636	0.7671	0.7597	0.7648	0.7559	0.7625	0.7540	0.7613
7	1	0	0.5098	0.5198	0.5098	0.5200	0.5098	0.5201	0.5098	0.5202
8	1	$\frac{1}{2}$	0.7636	0.7671	0.7597	0.7648	0.7559	0.7625	0.7540	0.7613

TABLE 5a.

S = 0.08

States	Parameters η_1 α_1	H = 0.0		H = 0.5		H = 1.0		H = 1.25		
		EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A	
		2	0	$\frac{1}{2}$	0.2694	0.2834	0.2797	0.2913	0.2897	0.2992
3	0	1	0.5316	0.5522	0.5320	0.5531	0.5324	0.5541	0.5325	0.5545
4	$\frac{1}{4}$	0	0.1350	0.1426	0.1435	0.1495	0.1520	0.1562	0.1566	0.1602
5	$\frac{1}{4}$	$\frac{1}{2}$	0.4084	0.4205	0.4141	0.4257	0.4194	0.4309	0.4219	0.4335
6	$\frac{1}{4}$	1	0.6645	0.6771	0.6581	0.6739	0.6522	0.6708	0.6494	0.6692
7	$\frac{1}{2}$	0	0.2694	0.2834	0.2797	0.2913	0.2897	0.2992	0.2945	0.3003
8	$\frac{1}{2}$	$\frac{1}{2}$	0.5416	0.5522	0.5416	0.5531	0.5415	0.5541	0.5415	0.5545
9	$\frac{1}{2}$	1	0.7868	0.7939	0.7779	0.7885	0.7694	0.7830	0.7652	0.7802
10	$\frac{3}{4}$	0	0.4020	0.4205	0.4091	0.4257	0.4157	0.4309	0.4188	0.4335
11	$\frac{3}{4}$	$\frac{1}{2}$	0.6680	0.6771	0.6627	0.6739	0.6578	0.6708	0.6553	0.6692
12	$\frac{3}{4}$	1	0.8986	0.9017	0.8917	0.8971	0.8849	0.8922	0.8815	0.8897
13	1	0	0.5316	0.5522	0.5320	0.5531	0.5324	0.5541	0.5325	0.5545
14	1	$\frac{1}{2}$	0.7868	0.7939	0.7779	0.7885	0.7694	0.7830	0.7652	0.7802

TABLE 5b.

S = 0.08

States	Parameters η_1 α_1		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
			EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	0	$\frac{1}{2}$	0.2615	0.2751	0.2697	0.2808	0.2774	0.2865	0.2811	0.2893
3	0	1	0.5191	0.5394	0.5192	0.5399	0.5193	0.5404	0.5193	0.5407
4	$\frac{1}{2}$	0	0.2615	0.2751	0.2697	0.2808	0.2774	0.2865	0.2811	0.2893
5	$\frac{1}{2}$	$\frac{1}{2}$	0.5291	0.5394	0.5288	0.5399	0.5286	0.5404	0.5284	0.5407
6	$\frac{1}{2}$	1	0.7764	0.7834	0.7688	0.7792	0.7615	0.7748	0.7580	0.7727
7	1	0	0.5191	0.5394	0.5192	0.5399	0.5193	0.5404	0.5193	0.5407
8	1	$\frac{1}{2}$	0.7764	0.7834	0.7688	0.7792	0.7615	0.7748	0.7580	0.7727

TABLE 6a.

S = 0.12

States	Parameters η_1 α_1		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
			EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
			2	0	$\frac{1}{2}$	0.2784	0.3001	0.2943	0.3128	0.3092
3	0	1	0.5462	0.5773	0.5470	0.5793	0.5478	0.5814	0.5481	0.5825
4	$\frac{1}{4}$	0	0.1396	0.1515	0.1529	0.1626	0.1659	0.1742	0.1724	0.1802
5	$\frac{1}{4}$	$\frac{1}{2}$	0.4244	0.4428	0.4328	0.4514	0.4406	0.4599	0.4442	0.4642
6	$\frac{1}{4}$	1	0.6825	0.7014	0.6734	0.6974	0.6651	0.6935	0.6612	0.6916
7	$\frac{1}{2}$	0	0.2784	0.3001	0.2943	0.3128	0.3092	0.3257	0.3163	0.3322
8	$\frac{1}{2}$	$\frac{1}{2}$	0.5612	0.5773	0.5611	0.5793	0.5610	0.5814	0.5609	0.5825
9	$\frac{1}{2}$	1	0.8031	0.8137	0.7904	0.8064	0.7784	0.7991	0.7727	0.7955
10	$\frac{3}{4}$	0	0.4145	0.4428	0.4254	0.4514	0.4352	0.4599	0.4397	0.4642
11	$\frac{3}{4}$	$\frac{1}{2}$	0.6876	0.7014	0.6800	0.6974	0.6729	0.6935	0.6695	0.6916
12	$\frac{3}{4}$	1	0.9086	0.9134	0.8990	0.9071	0.8895	0.9006	0.8848	0.8972
13	1	0	0.5462	0.5773	0.5470	0.5793	0.5478	0.5814	0.5481	0.5825
14	1	$\frac{1}{2}$	0.8031	0.8137	0.7904	0.8064	0.7784	0.7991	0.7727	0.7955

TABLE 6b.

States	Parameters		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
	η_1	α_1	EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	0	$\frac{1}{2}$	0.2670	0.2876	0.2792	0.2966	0.2906	0.3058	0.2960	0.3103
3	0	1	0.5281	0.5585	0.5283	0.5597	0.5284	0.5609	0.5284	0.5615
4	$\frac{1}{2}$	0	0.2670	0.2876	0.2792	0.2966	0.2906	0.3058	0.2960	0.3153
5	$\frac{1}{2}$	$\frac{1}{2}$	0.5429	0.5585	0.5424	0.5597	0.5418	0.5609	0.5415	0.5615
6	$\frac{1}{2}$	1	0.7884	0.7990	0.7773	0.7931	0.7669	0.7871	0.7619	0.7841
7	1	0	0.5281	0.5585	0.5283	0.5597	0.5284	0.5609	0.5284	0.5615
8	1	$\frac{1}{2}$	0.7884	0.7990	0.7773	0.7931	0.7669	0.7871	0.7619	0.7841

S = 0.12

FIGURE 3a

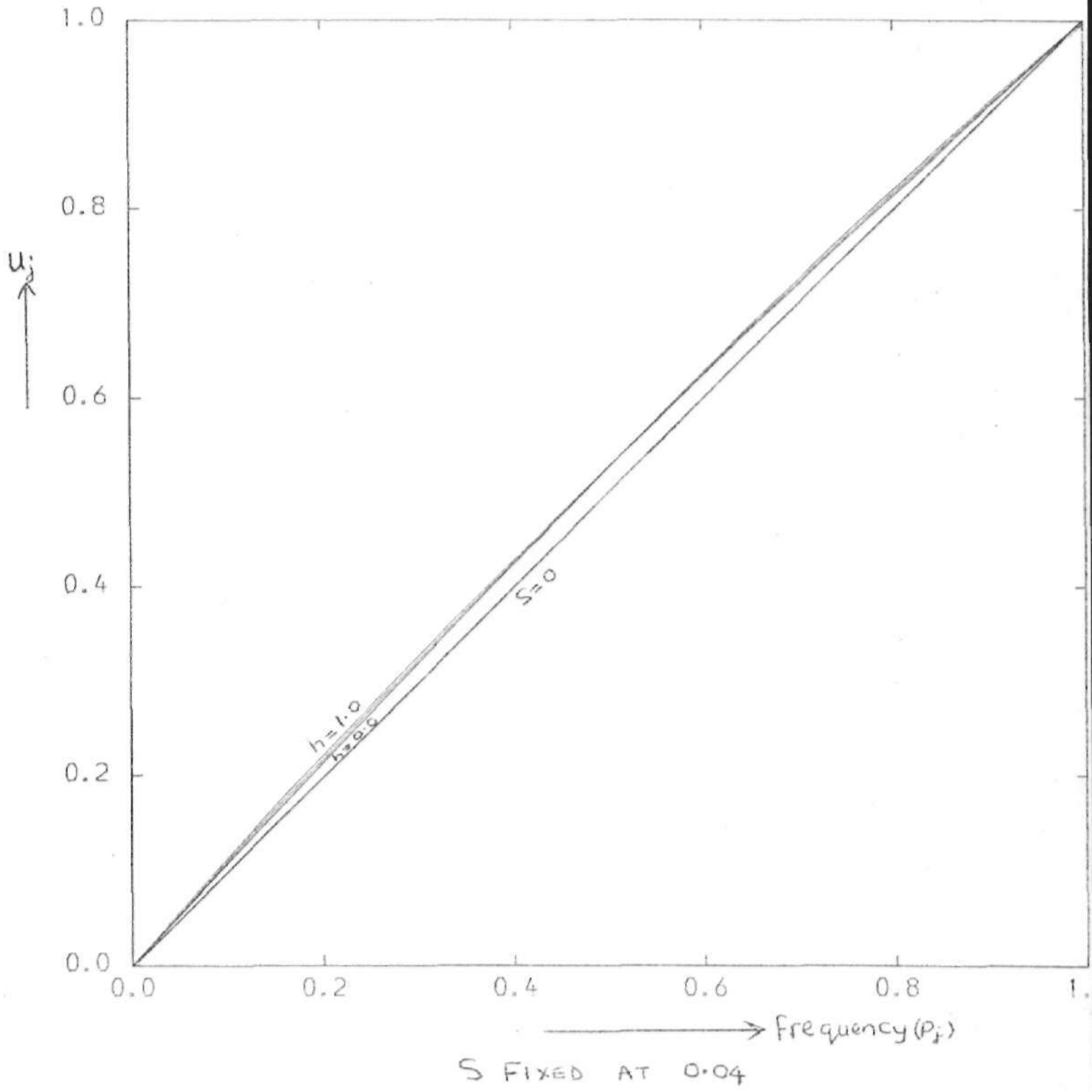
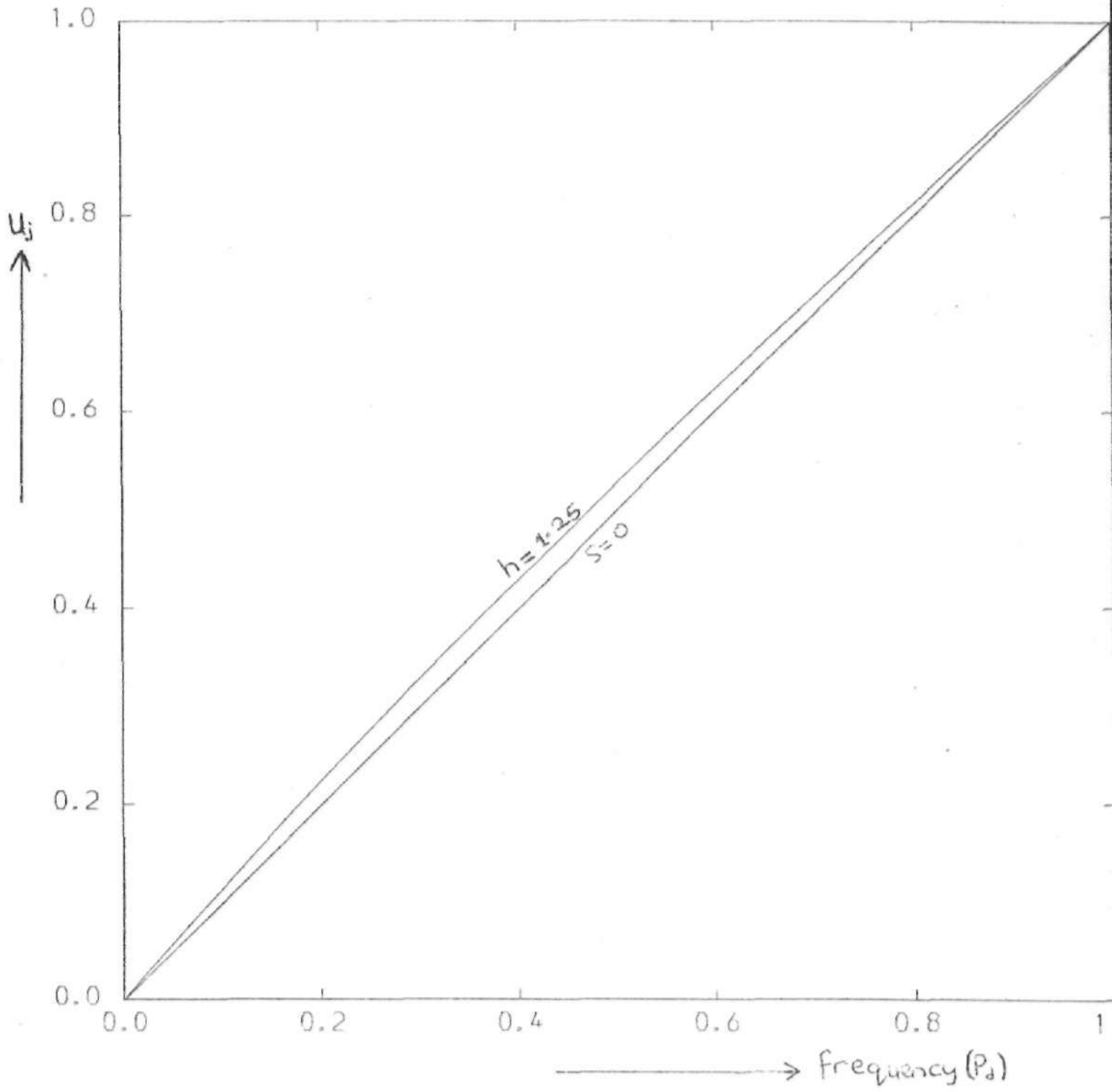
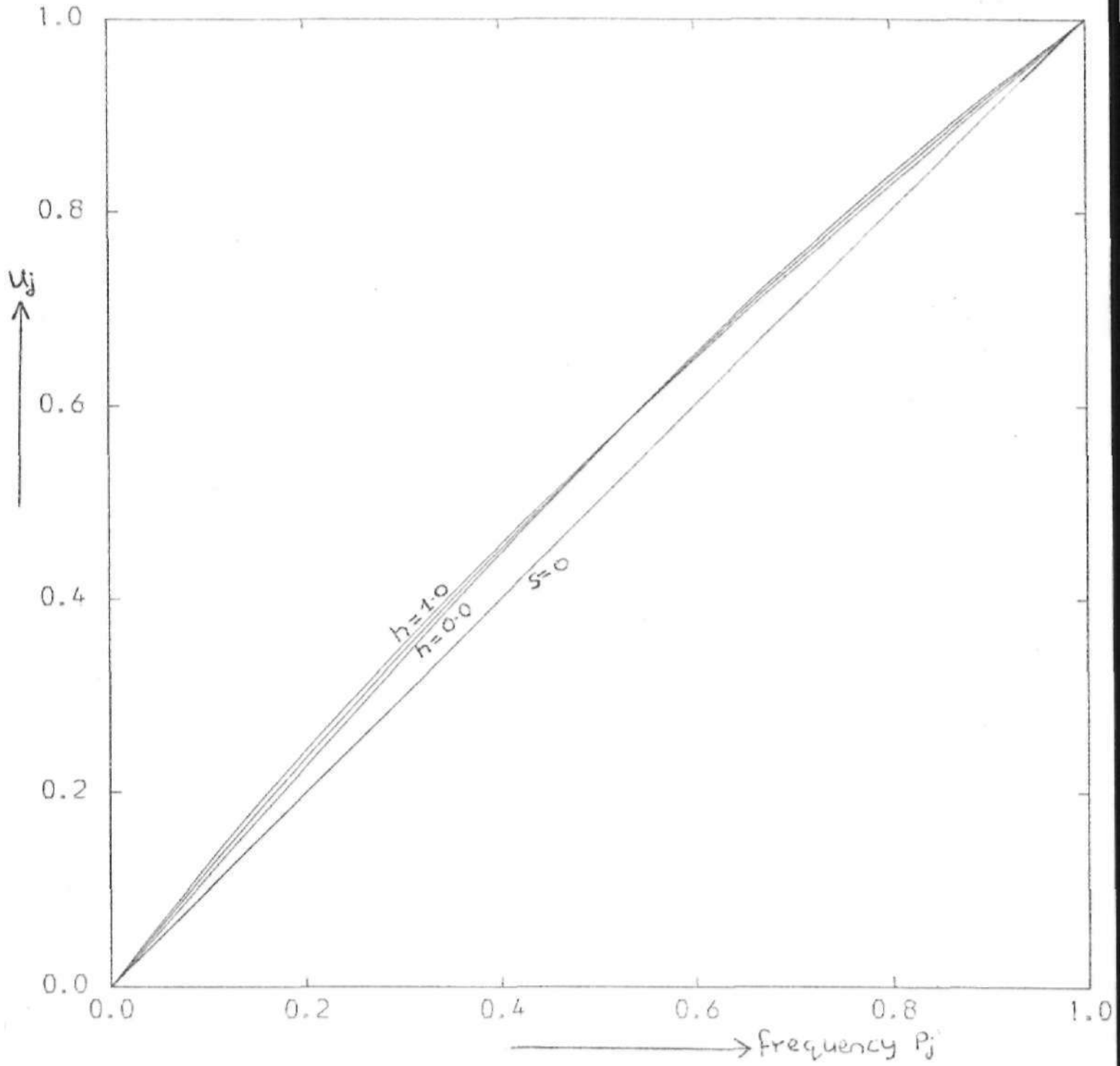


FIGURE 3b



S FIXED AT 0.04

FIGURE 4a.



S FIXED AT 0.03

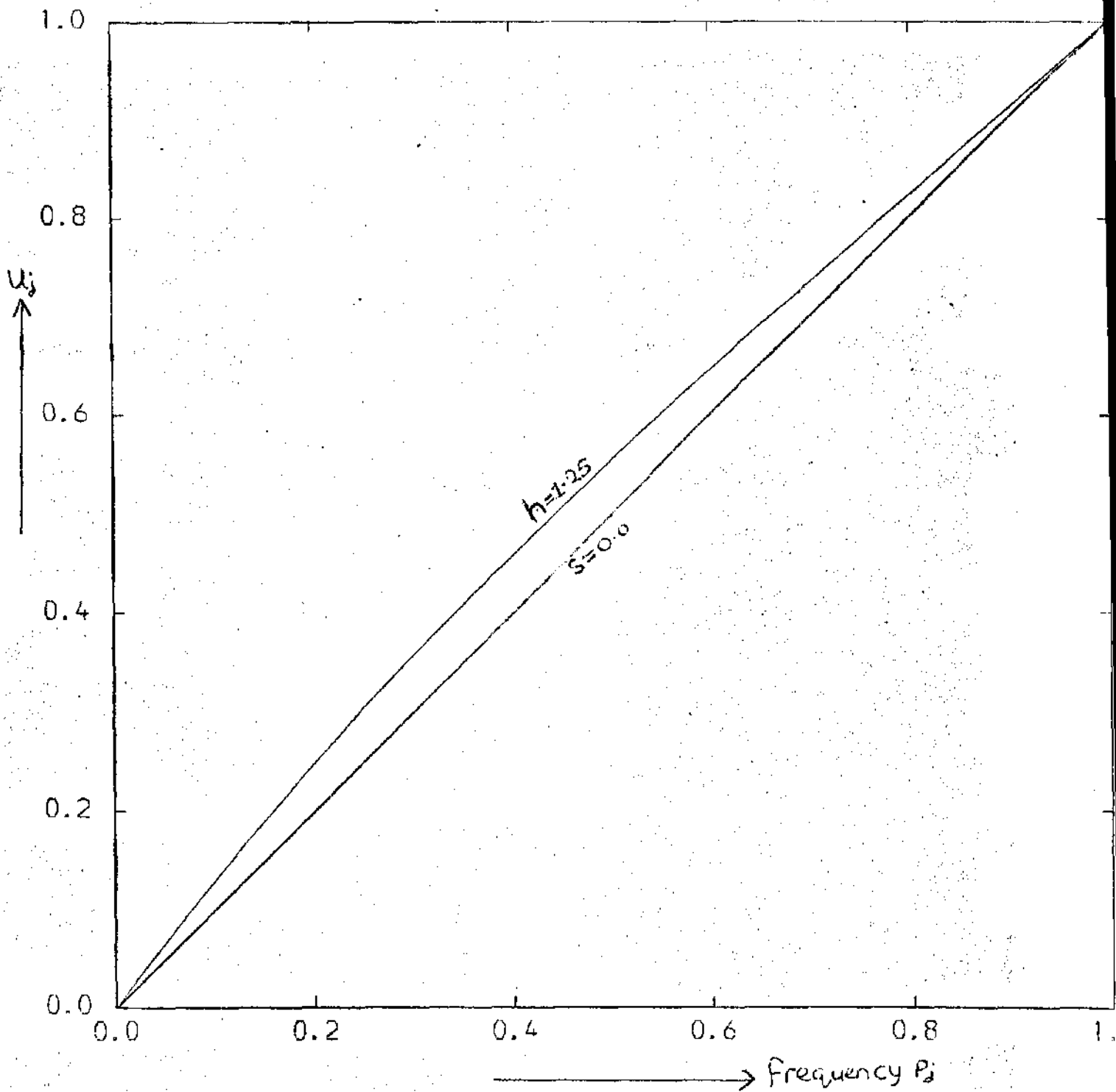
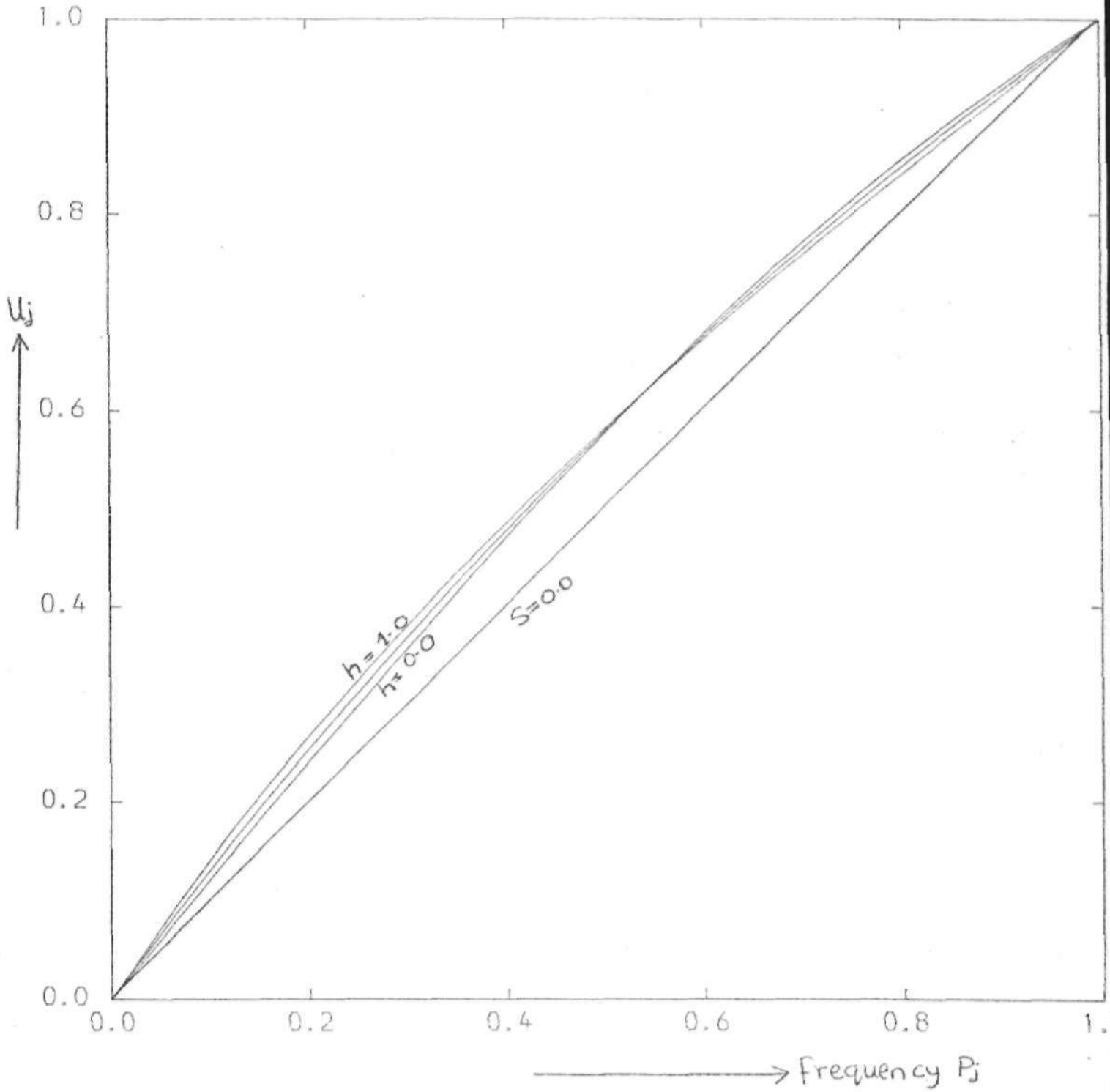
FIGURE 4bS FIXED AT 0.03

FIGURE 5a



S FIXED AT 0.12

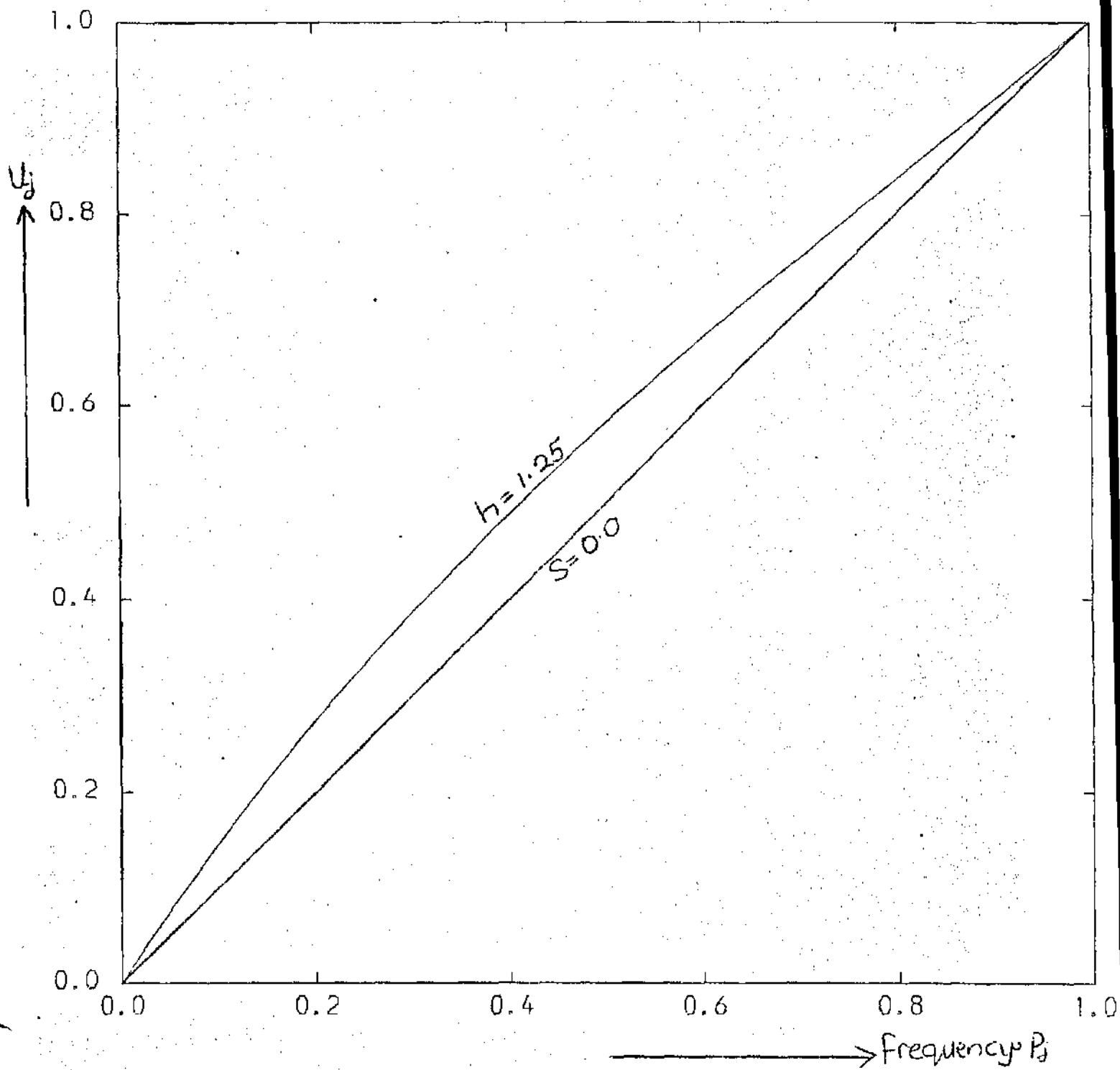
FIGURE 56S FIXED AT 0.12

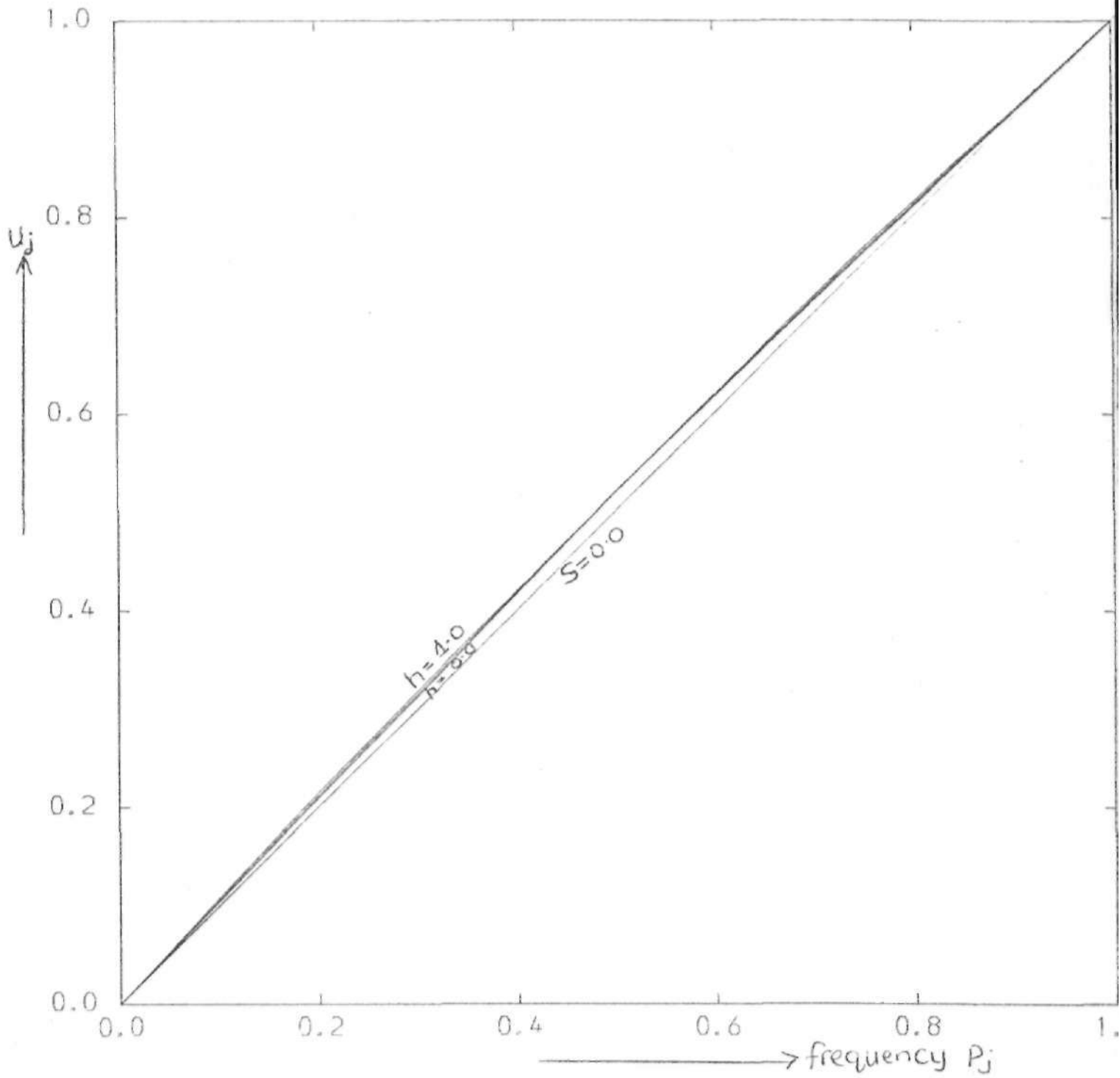
FIGURE 6aS FIXED AT 0.04

FIGURE 6b

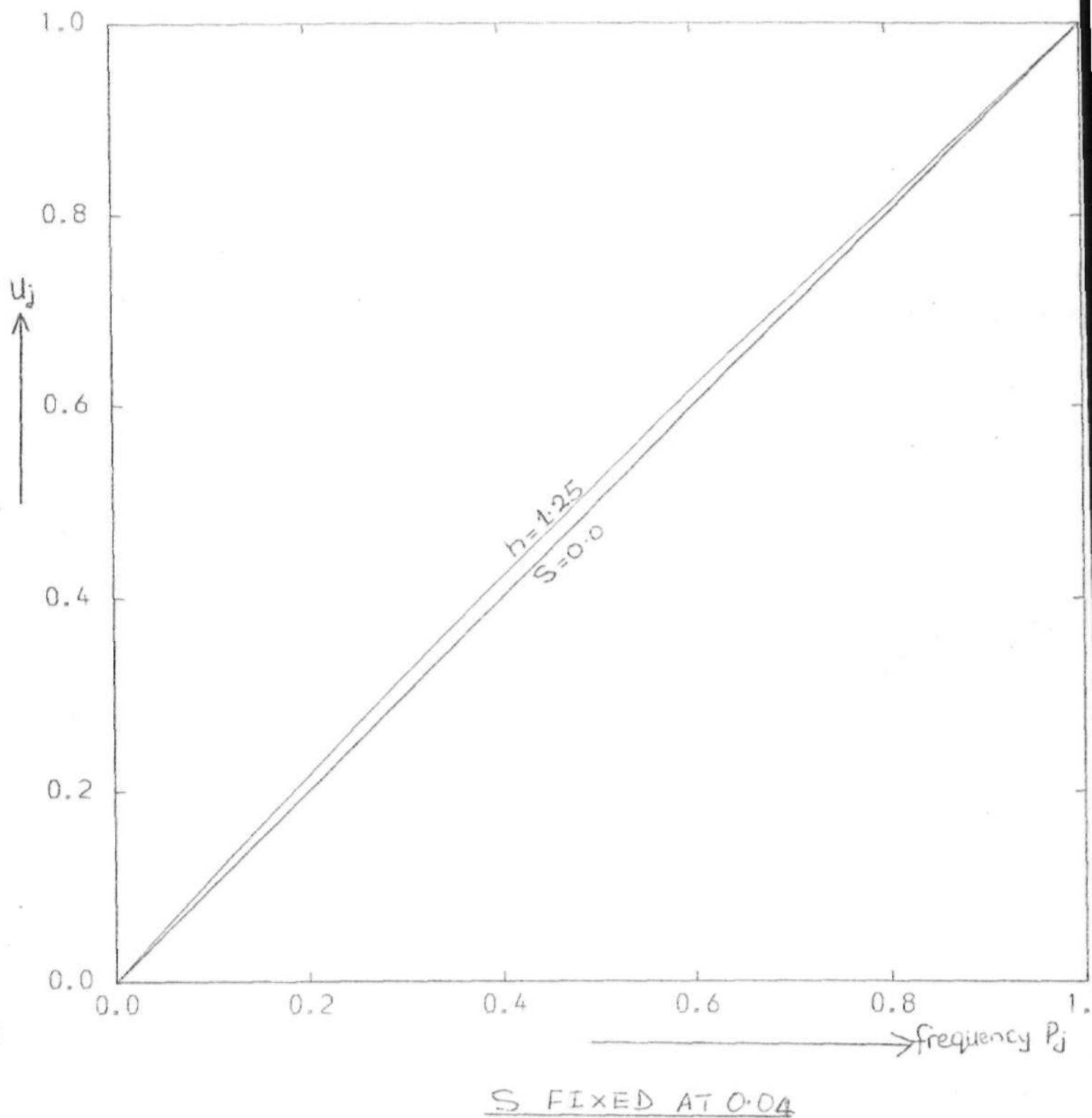
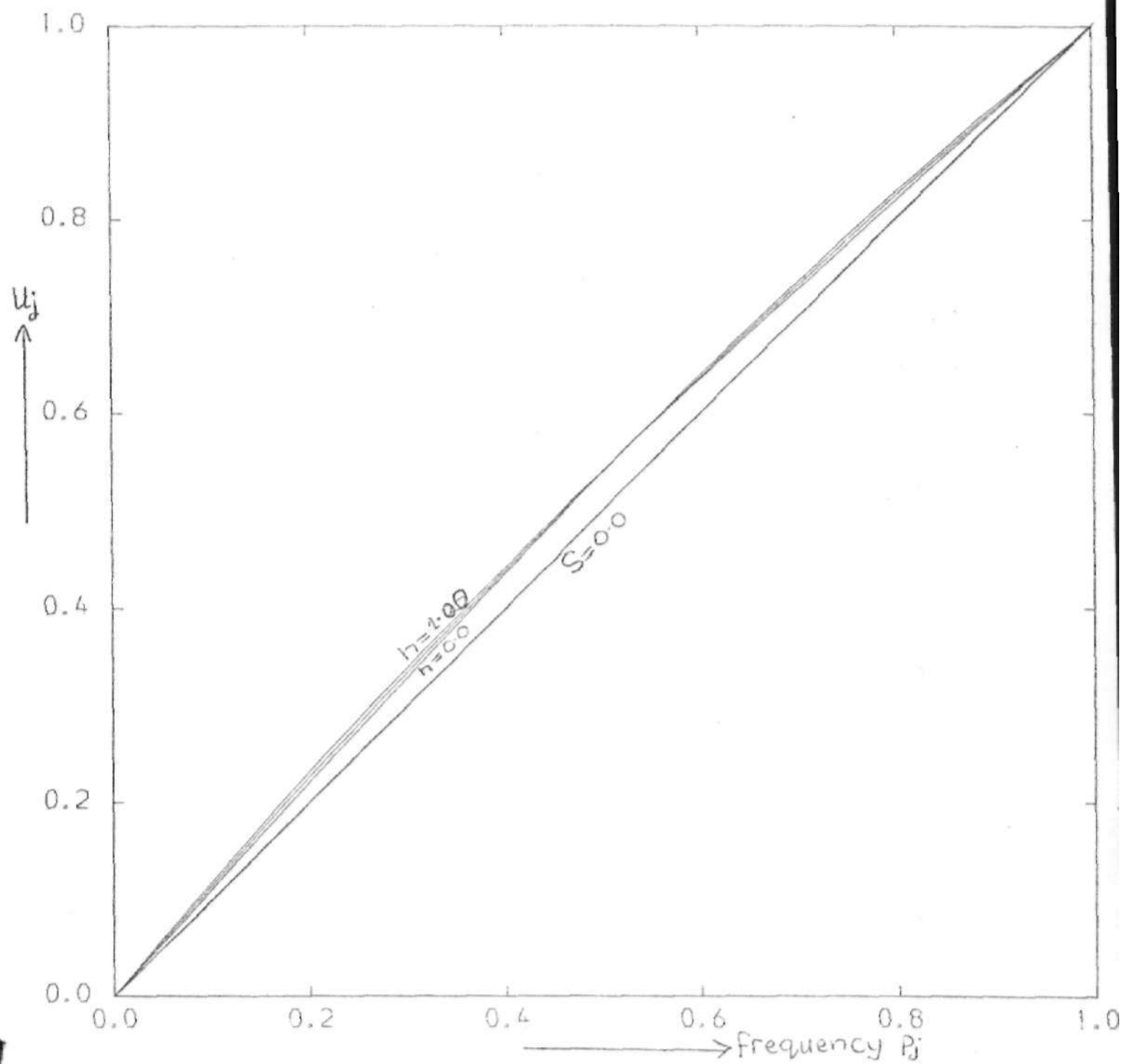
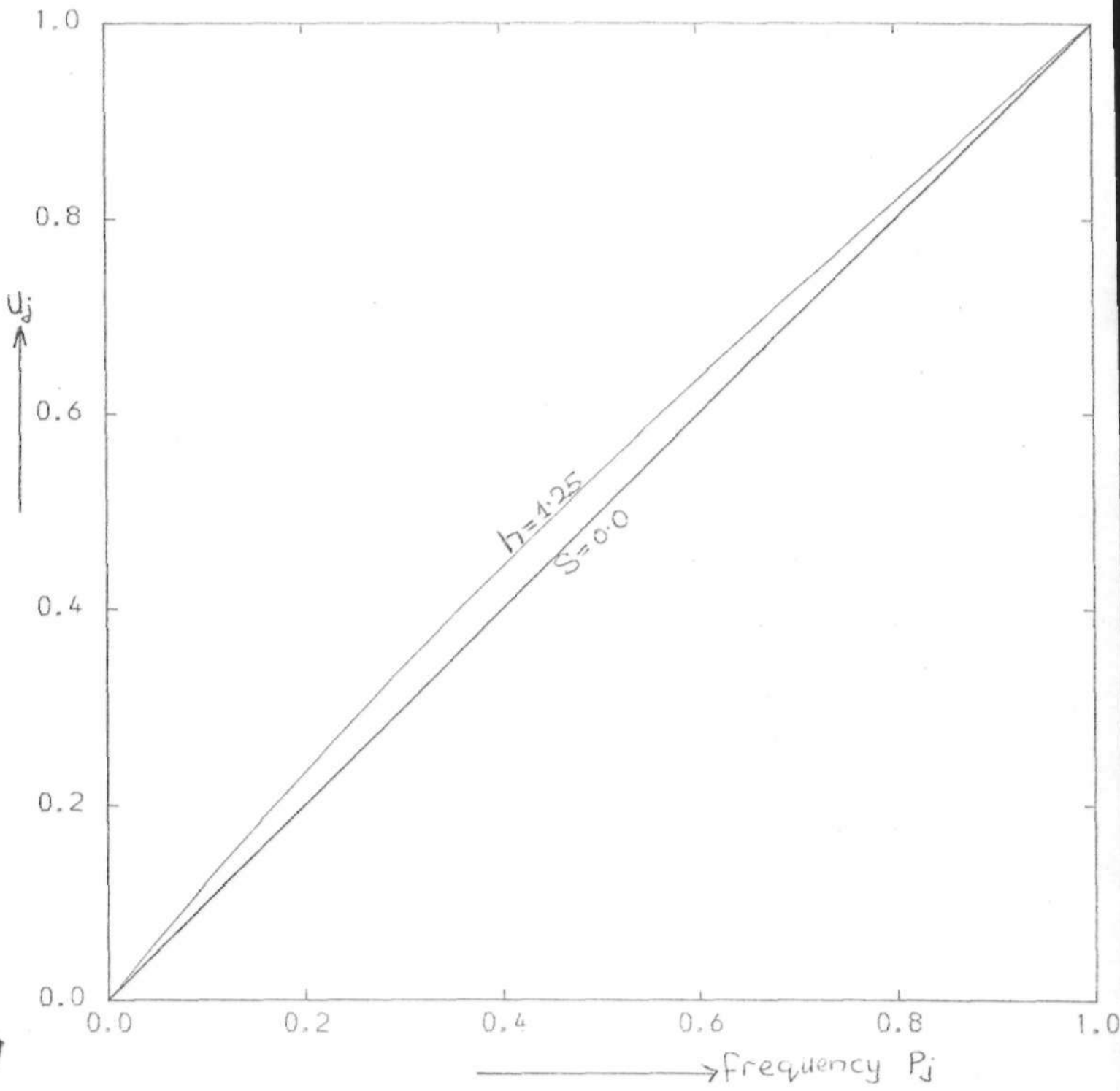


FIGURE 7a



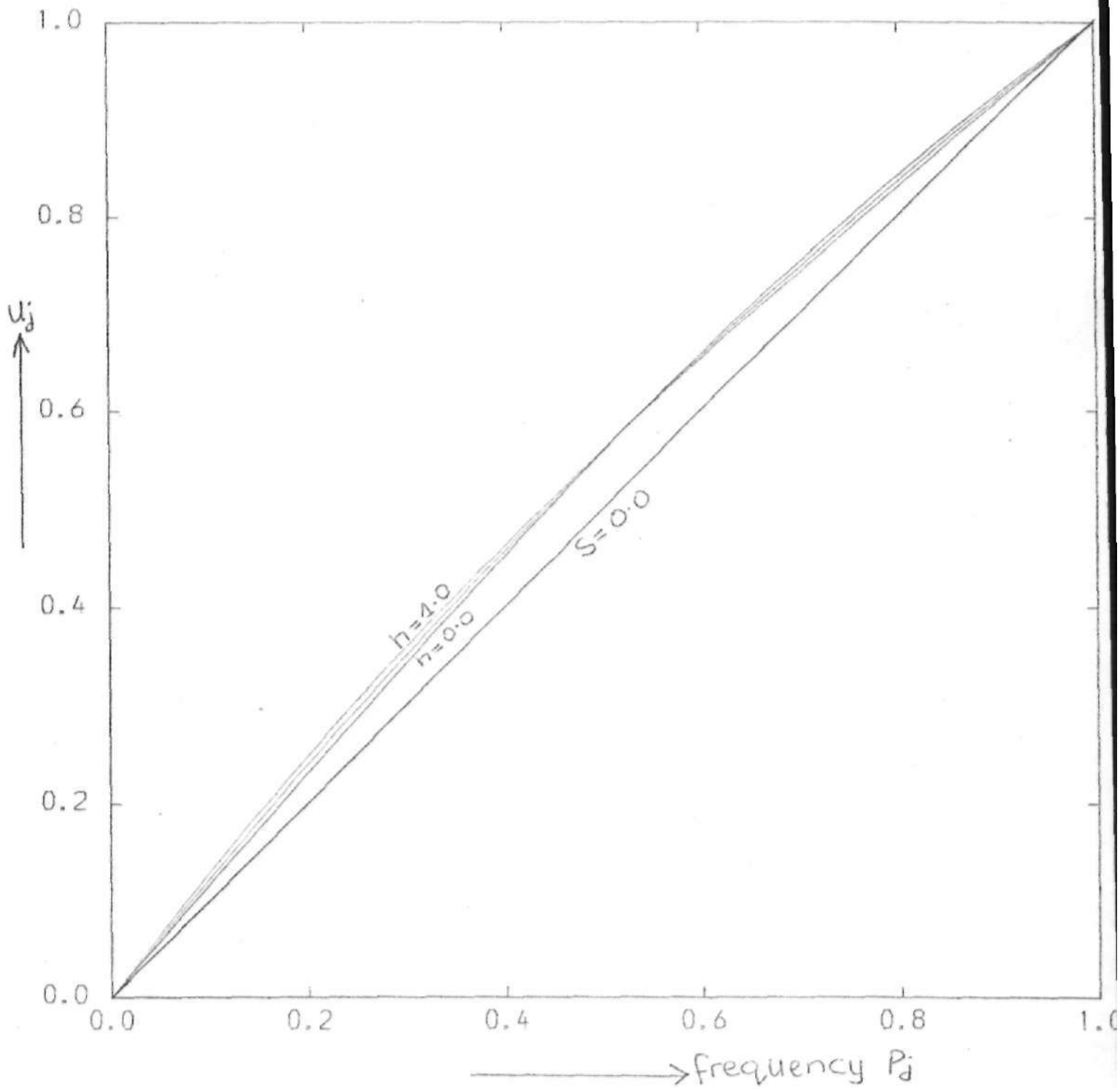
S FIXED AT 0.08

FIGURE 7b



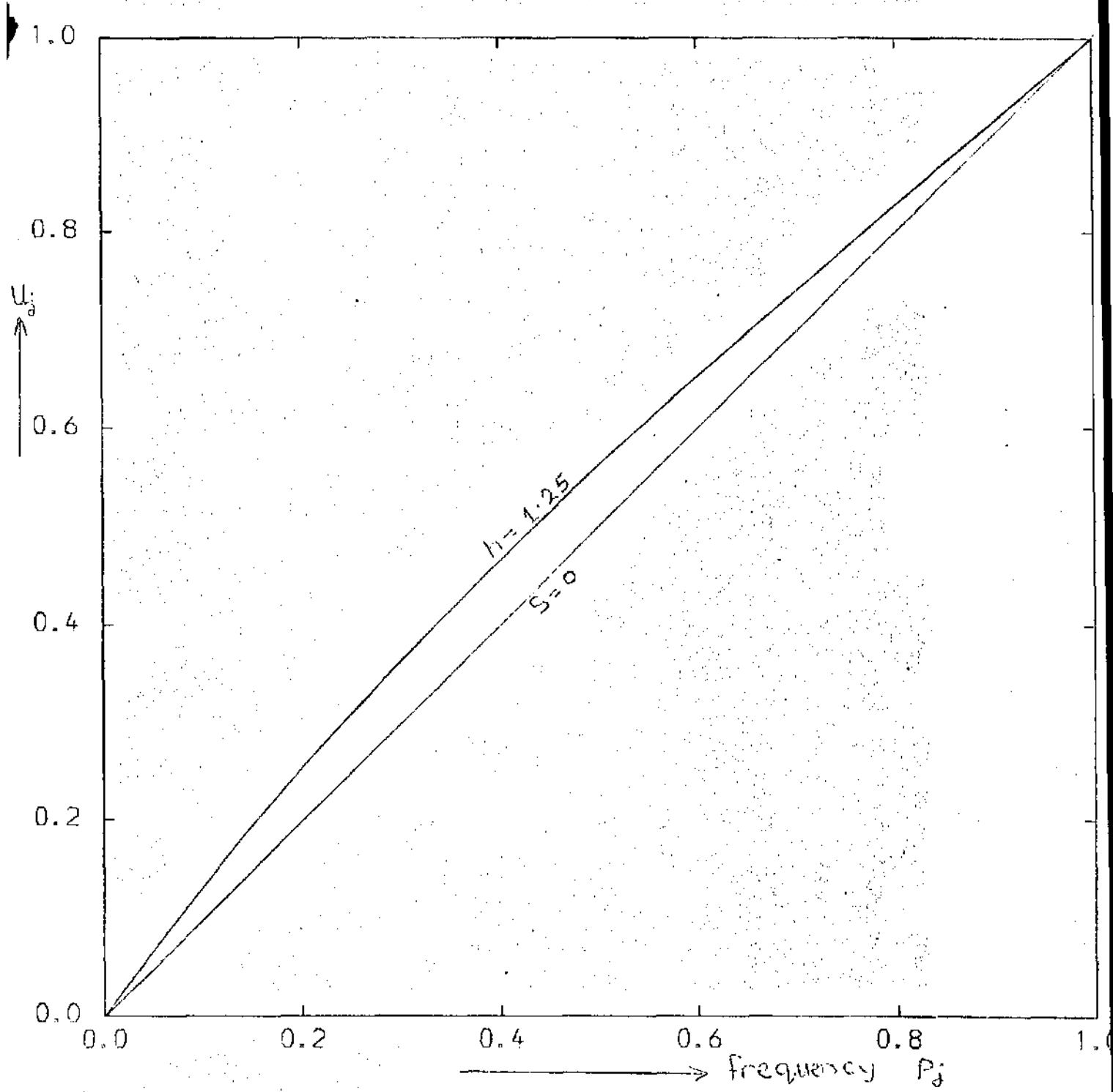
S FIXED AT 0.08

FIGURE 8a

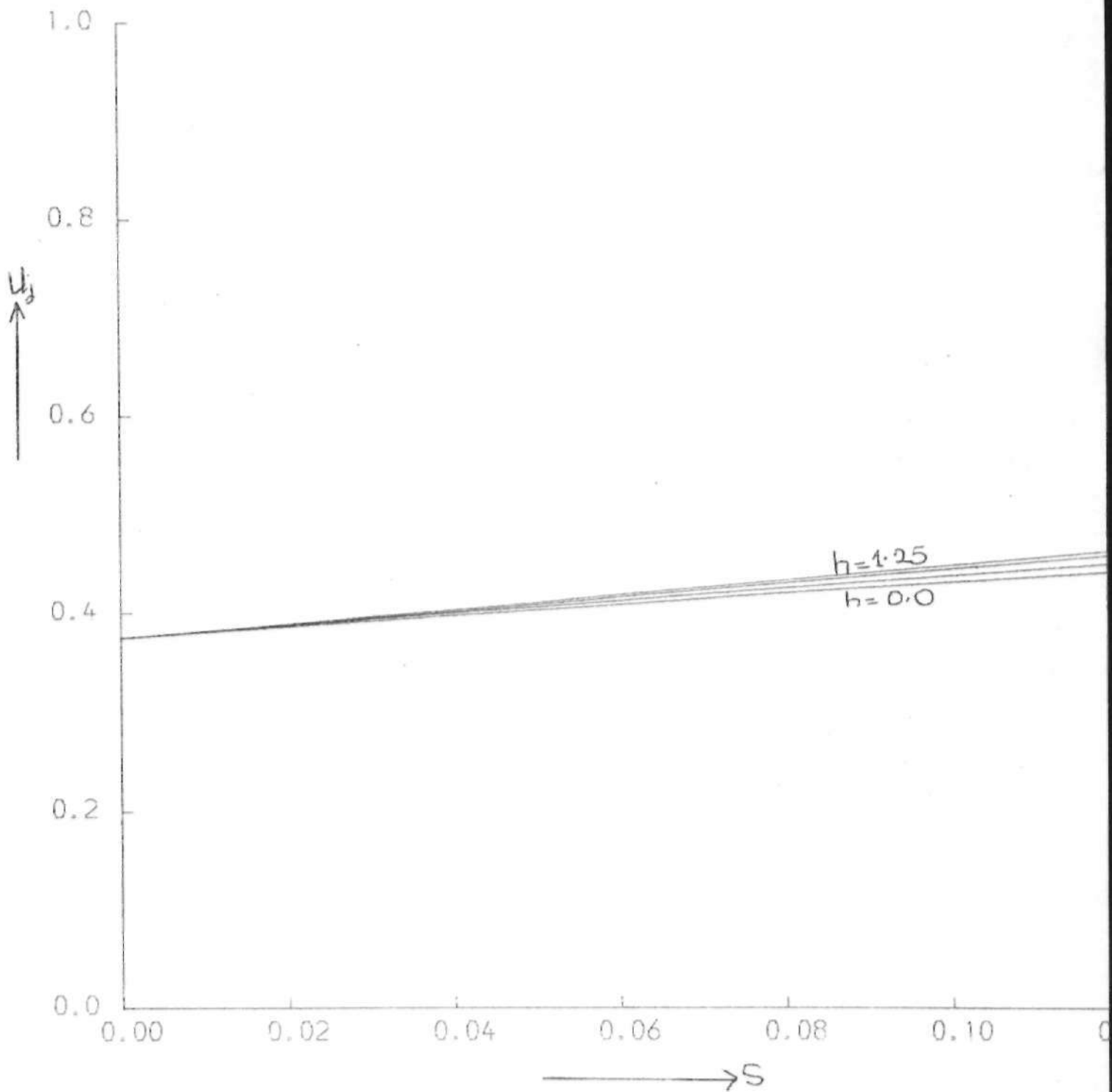


S FIXED AT 0.12

FIGURE 8b

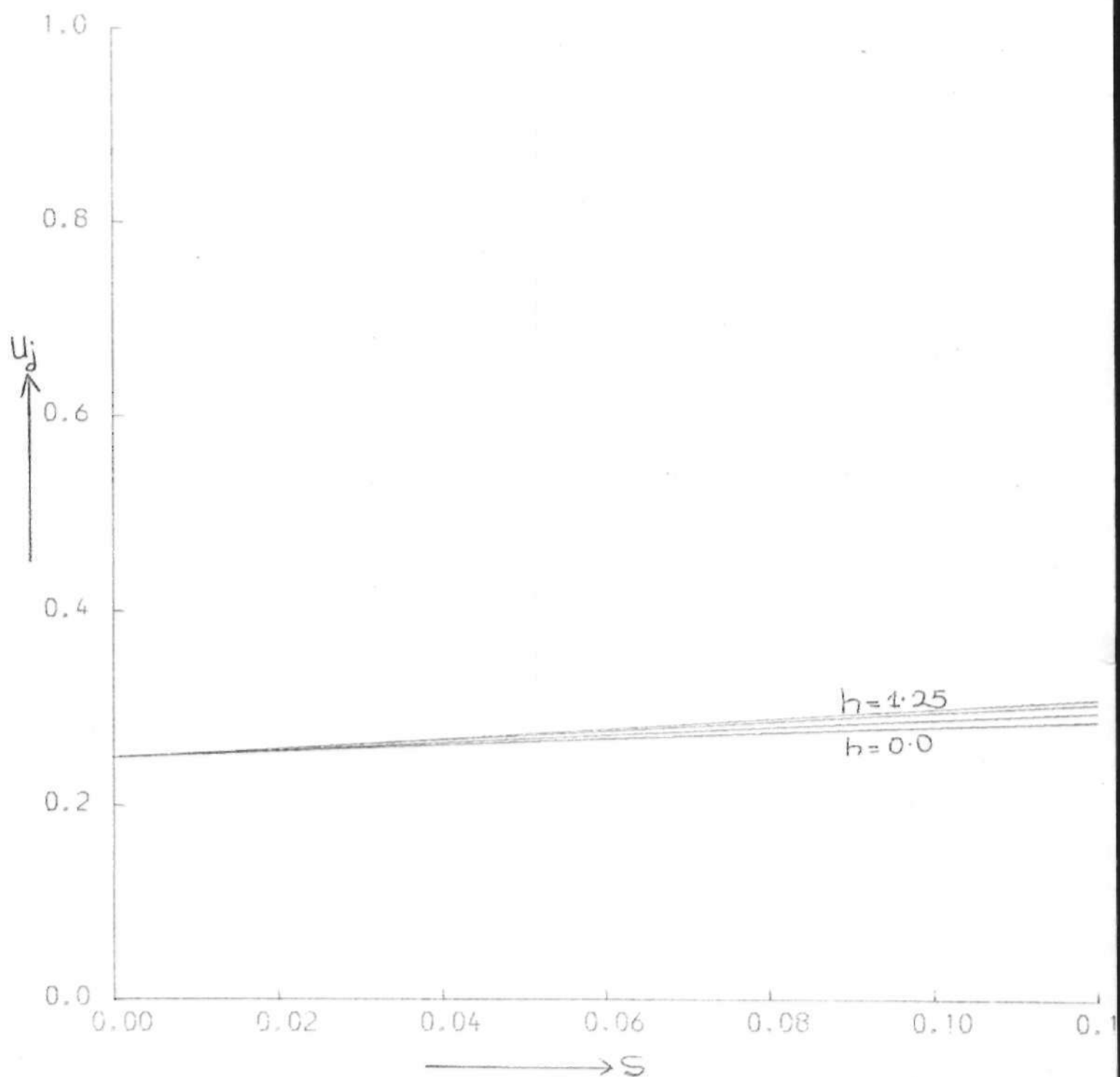


S FIXED AT 0.12

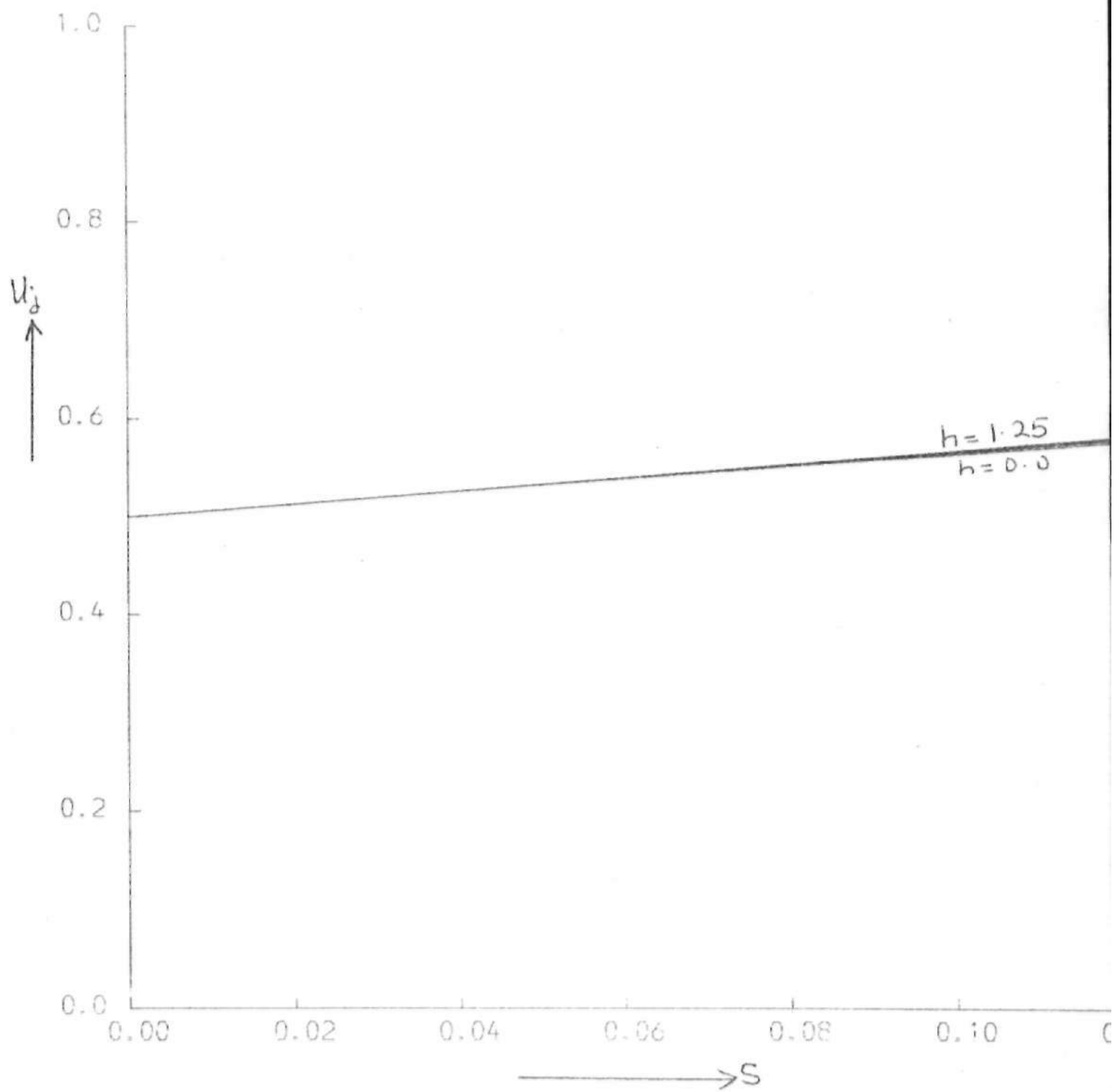
FIGURE 9a

FIXED VALUE OF P_j AT 0.375

FIGURE 9b

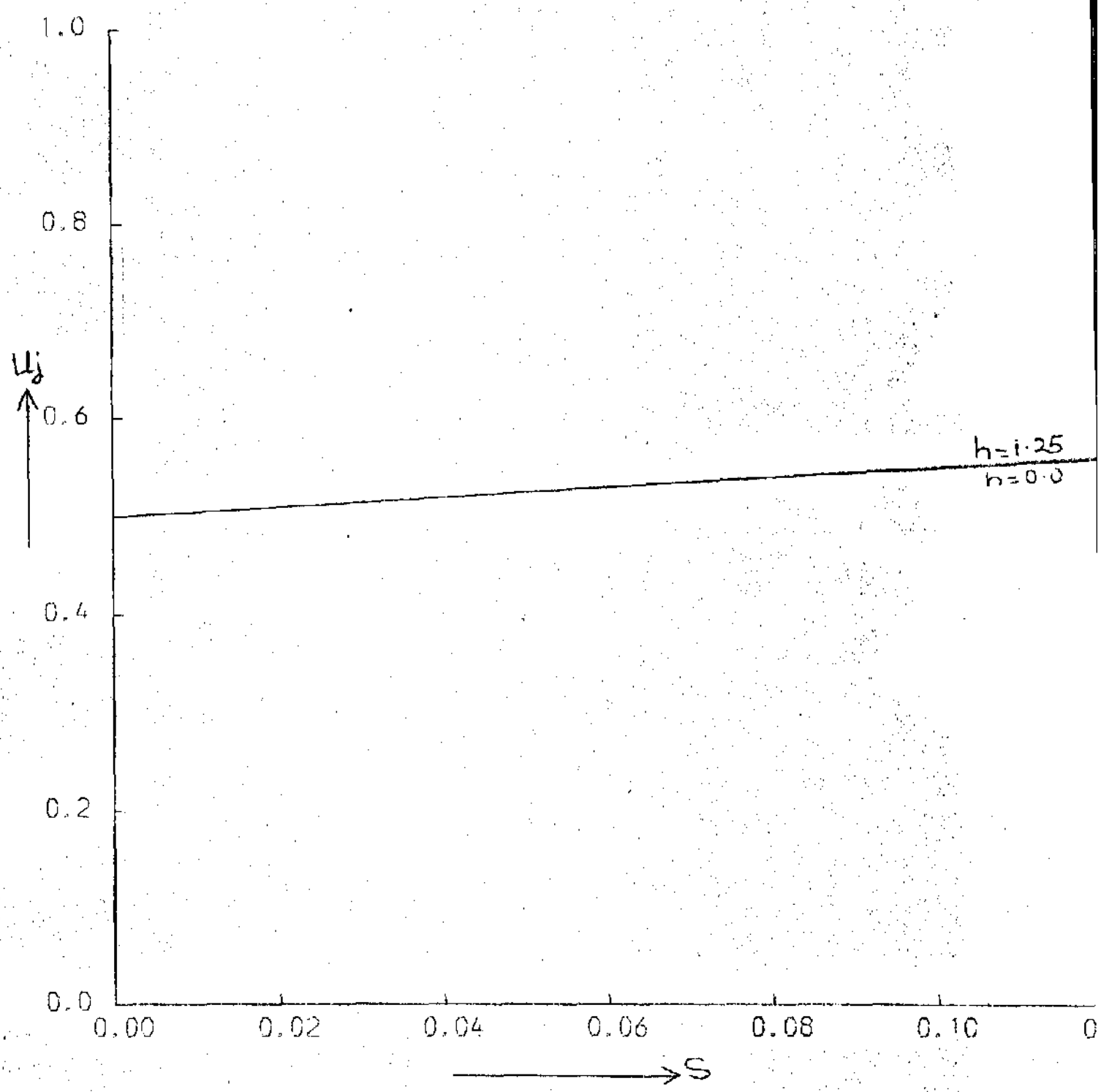


FIXED VALUE OF P_j AT 0.25

FIGURE 10a

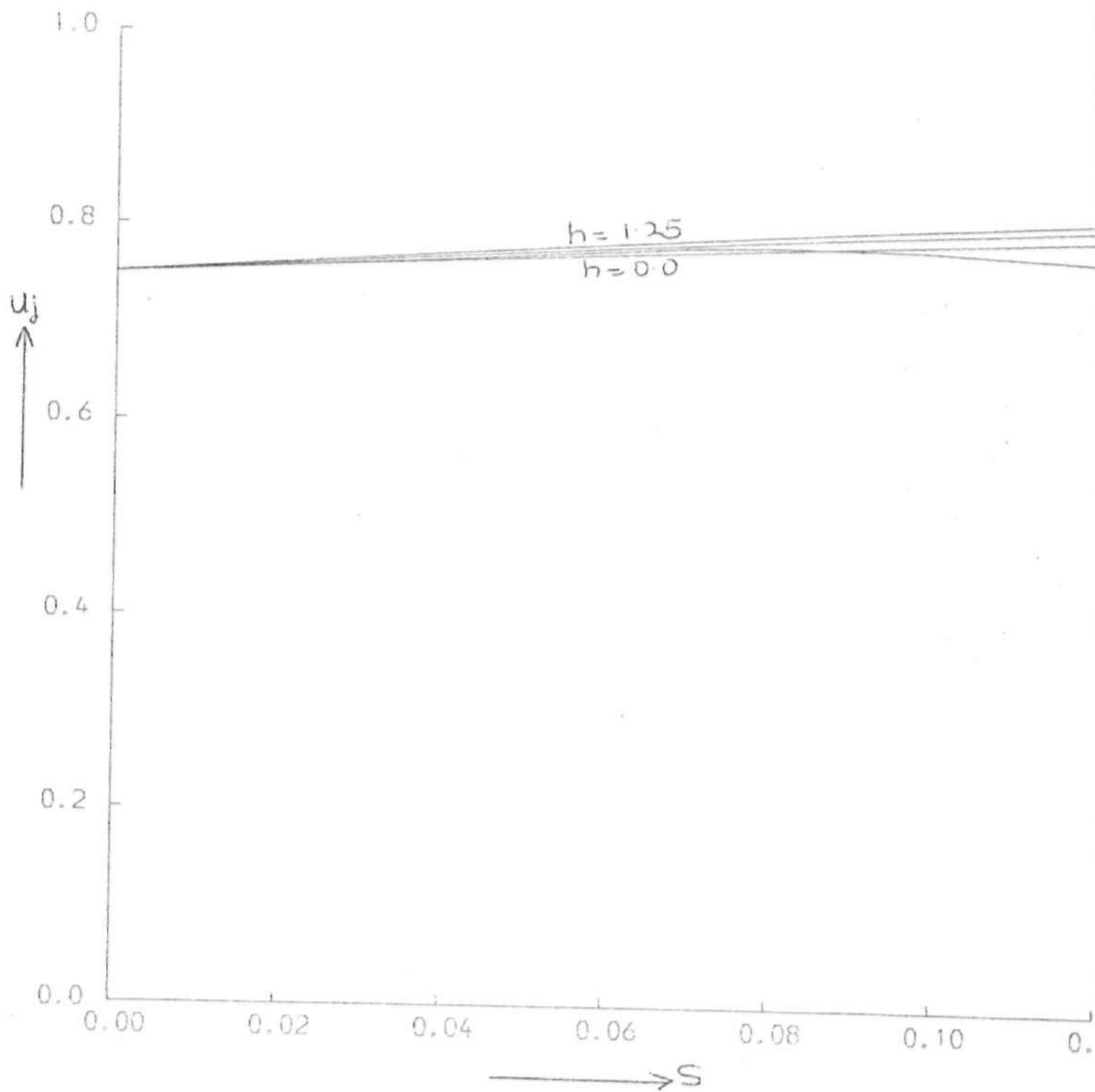
FIXED VALUE OF P_d AT 0.50

FIGURE 10b



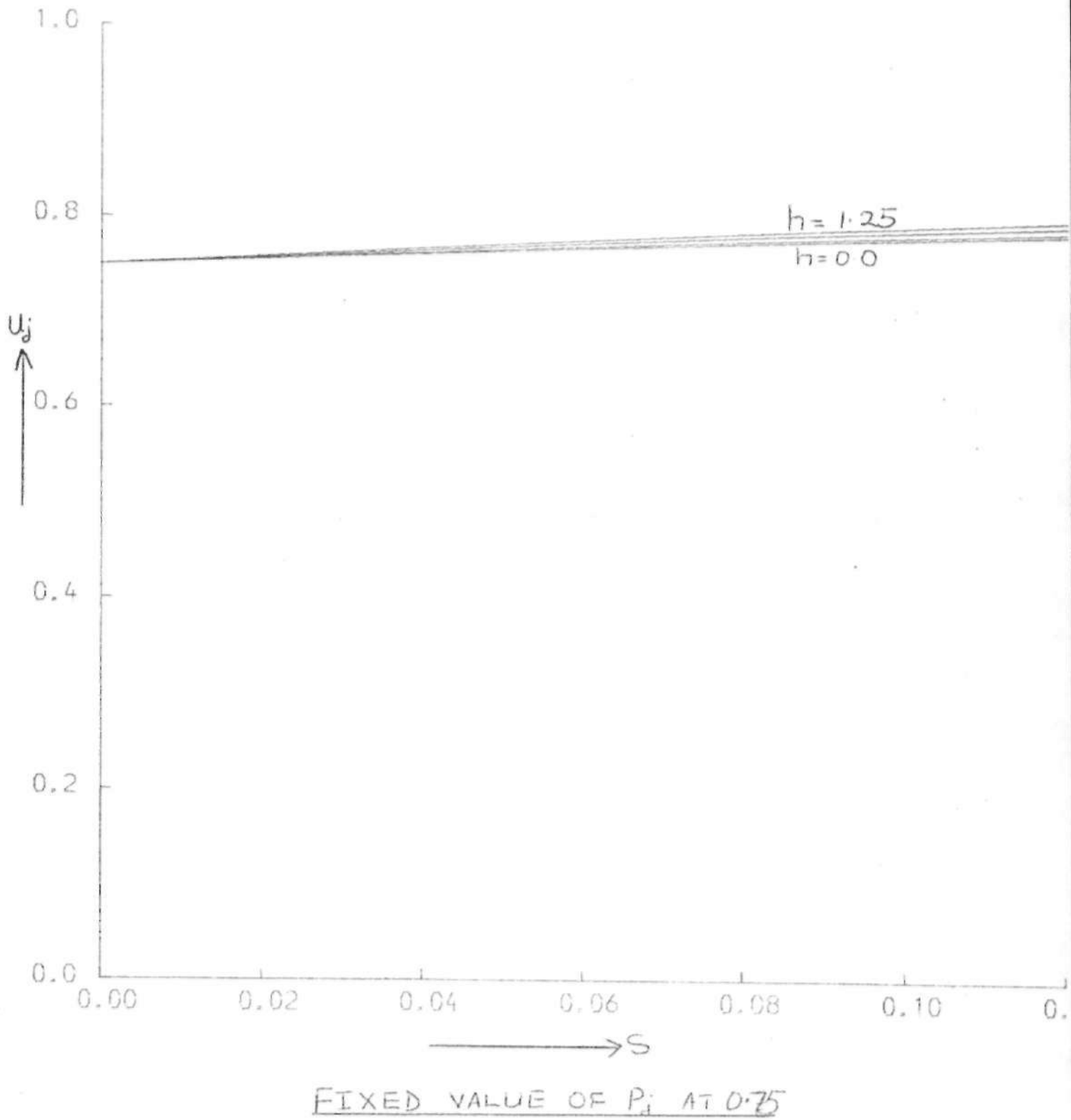
FIXED VALUE OF P_j AT 0.50

FIGURE 11a



FIXED VALUE OF P_j AT 0.75

FIGURE 11 b



4.4. DISCUSSION

From all the tables, it is clear that for small values of S at least, the diffusion approximation is very close to the true value. It will be noted that in all cases the diffusion approximation always exceeds the true probability, that this will always be so for positive S has been proved by (Moran 1960). Moran obtains a lower and upper bound for the exact probability, of which the diffusion approximation is the upper bound. However, the difference was not so great (0.0331 at most) and was smaller for high than for low initial gene frequencies. For both populations, for any $N_e S$ value, the agreement between the exact and diffusion results was greater for the larger N_e and smaller S (compare $N_e = 2$, $S = 0.08$ tables to $N_e = 2^2/3$, $S = 0.04$). This is not surprising since changes in gene frequency are expected to approach a continuous process as N_e increases and S decreases.

Considering the population of size 3, it will be observed that due to duplications of initial gene frequencies, the diffusion results did not take into account the initial state in which the process started. Thus the diffusion approach give the same results for states 6 ($\frac{1}{4}, 1$) and 11 ($\frac{2}{4}, \frac{1}{2}$) while in tables 4a, 5a and 6a, the exact approach gave slightly different results for both states for all dominance values considered (though both states have same average initial frequency). The same holds for the pairs of states 3(0, 1), 8 ($\frac{1}{2}, \frac{1}{2}$) and 5($\frac{1}{4}, \frac{1}{2}$), 10 ($\frac{2}{4}, 0$). Thus it could be said from both results that the exact results differentiate for pairs of initial states with no symmetrical initial gene parameters (η_1, α_1) while the diffusion approximations gave the same results. The exception

to this are the state pairs $3(0, 1)$, $13(1, 0)$; $2(0, \frac{1}{2})$, $7(\frac{1}{2}, 0)$; and $9(\frac{1}{2}, 1)$, $14(1, \frac{1}{2})$; that is, states with symmetrical initial gene frequency parameters.

It was observed from figures 3a, 4a, 5a, 6a, 7a and 8a that for additive ($h = \frac{1}{2}$), recessive ($h = 0$) and full dominance ($h = 1$) gene effects, selection increased U_j , the probability of Ultimate fixation as compared with no selection ($S = 0$) for all initial gene frequencies P_j , U_j was a strictly increasing function of P_j . It was also noted that there was a point P^1 (gene frequency at which curves intersect) below which U_j was an increasing function of h and above which it was a decreasing function of h .

The probability of Ultimate fixation, U_j for dominance ($h = 0, \frac{1}{2}, 1$) and over dominance ($h = 1.25$) was found to be an increasing function of N_e and S (by comparing Figures 10a and 10b, 11a and 11b) but not a function of the product $N_e S$ as was found for large N_e and small S by Kimura (1955). While for both dominance and overdominance U_j was an increasing function of P_j , the rate of increase was larger for overdominance ($h = 1.25$) than for dominance ($h = 0, \frac{1}{2}, 1$) when the initial gene frequency, P_j , was low and smaller when P_j was high.

CHAPTER 5EXPECTED MEAN TIME UNTIL FIXATION5.1. EXACT APPROACH

Here again we are considering this for a population of size 3. Similar arguments may be employed to obtain the corresponding equations satisfied by the population of size 2.

Let T_i be expected time to fixation and P_{ij} the probability from i^{th} state to the j^{th} state in one transition.

$T_1 = 0, T_{15} = 0$ since fixation has already occurred in these states

then the T_i values satisfy the equation

$$T_i = \sum_{j=2}^{14} P_{ij} T_j + 1 \quad i = 2, 3, \dots, 14$$

and following the arguments developed earlier in Chapter 4, section 4.1, we have vector

\underline{T} given by

$$\underline{T} = (\underline{I} - \underline{Q})^{-1} \underline{I} \quad \text{_____} (1)$$

where \underline{Q} is as defined in Chapter 4.1 and \underline{I} is a 13×1 vector of 1's.

Equation 1 thus gives the set of mean times until fixation is reached, where each element in the vector \underline{T} corresponds to a certain initial value of the number of A_1 - genes. Thus exact values for these mean times may be obtained numerically by matrix Inversion (4.1).

Likewise Equation 1 also governs the expected mean times until fixation of A_1 - genes for the population of size 2. Here \underline{Q} is a 7×7 matrix of transient state probabilities. Equation 1 was solved for various values of the parameters S and h for the two populations.

5.2. DIFFUSION APPROXIMATION

In order to find the diffusion approximation to the mean time we let $M_{\delta x}$ denote the mean increase in the proportion of A_1 - individuals from one generation where the proportion is x to the next generation, and $V_{\delta x}$ the variance of this increase. Then the diffusion equation satisfied by the mean time $T(x)$ until fixation for initial proportion x of A_1 - genes is given by Feller (1954).

$$M_{\delta x} \frac{dT(x)}{dx} + \frac{1}{2} V_{\delta x} \frac{d^2T(x)}{dx^2} = -1 \quad \text{_____} (2)$$

with the following boundary conditions

$$T(0) = T(1) = 0 \quad \text{_____} (3)$$

Solution of (2) is obtained by numerical integration using the central difference approximations to derivatives (C.F. Gerald, 1970).

The derivatives were approximated by finite difference quotients. Thus replacing the derivatives of the differential equation above by these difference quotients, convert it to a difference equation whose solution is an approximation to the solution of the differential equation.

The central difference approximations to derivatives replace the derivatives with

$$\left. \frac{dT}{dx} \right|_{x=x_i} = \frac{T_{i+1} - T_{i-1}}{2K} + O(K^2)$$

$$\left. \frac{d^2T}{dx^2} \right|_{x=x_i} = \frac{T_{i+1} - 2T_i + T_{i-1}}{K^2} + O(K^2)$$

where $K = x_{i+1} - x_i = \text{constant}$

Substituting these equivalences into equation (2) and rearranging, we have

$$(V_{\sigma x} - M_{\sigma x} K) T_{i-1} - 2V_{\sigma x} T_i + (M_{\sigma x} K + V_{\sigma x}) T_{i+1} = -2K^2 \quad (4)$$

From Chapters 2 and 3

$$M_{\sigma x} = Sx(1-x) [h + (1-2h)x]$$

$$V_{\sigma x} = \frac{x(1-x)}{2N_e}$$

Putting these values in (4) and rearranging we have

$$\left\{ 1 - 2N_e SK [h + (1-2h)x] \right\} T_{i-1} - 2T_i + \left\{ 1 + 2N_e SK [h + (1-2h)x] \right\} T_{i+1} = \frac{-4N_e K^2}{x(1-x)} \quad (5)$$

which reduces to

$$\left[1 - QK(H + Rx_i) \right] T_{i-1} - 2T_i + \left[1 + QK(H + Rx_i) \right] T_{i+1} = \frac{-4N_e K^2}{x_i(1-x_i)} \quad (6)$$

where $2N_e S = Q$ and

$$1 - 2h = R$$

In equation (6) we have replaced $T(x)$ by T_i , and x by x_i , since these values correspond to the point at which the difference quotients represent the derivatives.

Our problem now reduces to solving (6) at points in the interval from $x = 0$ to $x = 1$. Let us subdivide the interval into a number of equal sub-intervals. Here we choose $K = 1/8$ to correspond with our initial gene frequencies in a population of size 3. Thus the points $x_1 = 0$, $x_2 = 1/8$, $x_3 = 2/8$, $x_4 = 3/8$, $x_5 = 4/8$, $x_6 = 5/8$, $x_7 = 6/8$, $x_8 = 7/8$ and $x_9 = 1$, subdivide the interval into 8 sub-intervals.

We write the difference equations in (6) for each of these values of x with the conditions $T_1 = T_9 = 0$ corresponding to $T(0) = T(1) = 0$ (Boundary conditions) and given x_i values $i = 1, 2, \dots, 9$.

Equation (6) was solved numerically on a high-speed computer using the Gaussian elimination method in solving a set of linear equations for the two effective population sizes, and the various values of the two parameters h and S . Solutions of (6) at x_3 , x_5 , and x_7 respectively, correspond to the expected mean times until fixation in a population of size 2 with effective population size $N_e = 2$. The results presented in the tables are in terms of generations.

5.3. RESULTS

Tables 7a, 8a, 9a and 10a give the exact values and diffusion approximations (D.A) for the mean time until fixation in a population of size 3 for fixed selection intensity values 0.0, 0.04, 0.08 and 0.12 respectively. The initial proportion of A_1 genes in the diffusion approximation being the average of the two parameters η_1 and α_1 for all dominance values $(0, \frac{1}{2}, 1, 1\frac{1}{4})$ considered.

Similarly Tables 7b, 8b, 9b and 10b give the exact values and diffusion approximations (D.A) for the mean time until fixation in a population of size 2 for fixed selection intensity values 0.0, 0.04, 0.08 and 0.12 respectively.

The mean times in the tables are in terms of generations.

TABLE 7a.

S = 0

States	Parameters		All H Values	
	η_1	α_1	EXACT	D.A
2	0	$\frac{1}{2}$	5.2689	5.3905
3	0	1	6.5241	6.7682
4	$\frac{1}{4}$	0	3.5843	-3.4571
5	$\frac{1}{4}$	$\frac{1}{2}$	6.2178	6.4349
6	$\frac{1}{4}$	1	6.2178	6.4349
7	$\frac{1}{2}$	0	5.2689	5.3905
8	$\frac{1}{2}$	$\frac{1}{2}$	6.5241	6.7682
9	$\frac{1}{2}$	1	5.2689	5.3905
10	$\frac{3}{4}$	0	6.2178	6.4349
11	$\frac{3}{4}$	$\frac{1}{2}$	6.2178	6.4349
12	$\frac{3}{4}$	1	3.5843	3.4571
13	1	0	6.5241	6.7682
14	1	$\frac{1}{2}$	5.2689	5.3905

TABLE 7b.

S = 0

States	Parameters		All H Values	
	η_1	α_1	EXACT	D.A
2	0	$\frac{1}{2}$	3.6896	4.0429
3	0	1	4.5517	5.0762
4	$\frac{1}{2}$	0	3.6896	4.0429
5	$\frac{1}{2}$	$\frac{1}{2}$	4.5517	5.0762
6	$\frac{1}{2}$	1	3.6896	4.0429
7	1	0	4.5517	5.0762
8	1	$\frac{1}{2}$	3.6896	4.0429

TABLE 8a.

S = 0.04

States	Parameters η_1 α_1	H = 0.0		H = 0.5		H = 1.0		H = 1.25		
		EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A	
		2	0	$\frac{1}{2}$	5.2256	5.3696	5.3261	5.4720	5.4268	5.5773
3	0	1	6.4328	6.6568	6.5225	6.7625	6.6125	6.8712	6.6577	6.9266
4	$\frac{1}{4}$	0	3.5634	3.4556	3.6443	3.5344	3.7261	3.6156	3.7672	3.6572
5	$\frac{1}{4}$	$\frac{1}{2}$	6.1647	6.3752	6.2571	6.4814	6.3497	6.5906	6.3962	6.6463
6	$\frac{1}{4}$	1	6.0907	6.2731	6.1866	6.3784	6.2826	6.4865	6.3306	6.5417
7	$\frac{1}{2}$	0	5.2256	5.3696	5.3261	5.4720	5.4268	5.5773	5.4773	5.6311
8	$\frac{1}{2}$	$\frac{1}{2}$	6.4314	6.6568	6.5211	6.7625	6.6112	6.8712	6.6564	6.9266
9	$\frac{1}{2}$	1	5.1098	5.2025	5.2097	5.3025	5.3099	5.4054	5.3601	5.4579
10	$\frac{3}{4}$	0	6.1502	6.3752	6.2456	6.4814	6.3211	6.5906	6.3890	6.6463
11	$\frac{3}{4}$	$\frac{1}{2}$	6.0812	6.2731	6.1739	6.3784	6.2669	6.4865	6.3135	6.5417
12	$\frac{3}{4}$	1	3.4463	3.3025	3.5247	3.3778	3.6041	3.4556	3.6441	3.4954
13	1	0	6.4328	6.6568	6.5225	6.7625	6.6125	6.8712	6.6577	6.9266
14	1	$\frac{1}{2}$	5.1098	5.2025	5.2097	5.3025	5.3099	5.4054	5.3601	5.4579

TABLE 8b.

 $S = 0.04$

States	Parameters η_1 α_1		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
			EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
			2	0	$\frac{1}{2}$	3.6694	4.0315	3.7167	4.0892	3.7635
3	0	1	4.5134	5.0140	4.5513	5.0738	4.5888	5.1347	4.6073	5.1657
4	$\frac{1}{2}$	0	3.6694	4.0315	3.7167	4.0892	3.7635	4.1481	3.7866	4.1780
5	$\frac{1}{2}$	$\frac{1}{2}$	4.5128	5.0140	4.5507	5.0738	4.5882	5.1347	4.6067	5.1657
6	$\frac{1}{2}$	1	3.6147	3.9372	3.6621	3.9938	3.7090	4.0517	3.7323	4.0811
7	1	0	4.5134	5.0140	4.5513	5.0738	4.5888	5.1347	4.6073	5.1657
8	1	$\frac{1}{2}$	3.6147	3.9372	3.6621	3.9938	3.7090	4.0517	3.7323	4.0811

TABLE 9a.

States	Parameters		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
	η_1	α_1	EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	0	$\frac{1}{2}$	5.1822	5.3432	5.3792	5.5466	5.5771	5.7616	5.6763	5.8737
3	0	1	6.3426	6.5379	6.5179	6.7454	6.6946	6.9645	6.7834	7.0787
4	$\frac{1}{4}$	0	3.5424	3.4513	3.7023	3.6091	3.8654	3.7771	3.9481	3.8652
5	$\frac{1}{4}$	$\frac{1}{2}$	6.1101	6.3080	6.2903	6.5175	6.4718	6.7386	6.5629	6.8538
6	$\frac{1}{4}$	1	5.9655	6.1065	6.1534	6.3121	6.3420	6.5293	6.4365	6.6425
7	$\frac{1}{2}$	0	5.1822	5.3432	5.3792	5.5466	5.5771	5.7616	5.6763	5.8737
8	$\frac{1}{2}$	$\frac{1}{2}$	6.3375	6.5379	6.5128	6.7454	6.6896	6.9645	6.7784	7.0787
9	$\frac{1}{2}$	1	4.9564	5.0147	5.1512	5.2087	5.3427	5.4140	5.4457	5.5212
10	$\frac{3}{4}$	0	6.0828	6.3080	6.2690	6.5175	6.4558	6.7386	6.595	6.8538
11	$\frac{3}{4}$	$\frac{1}{2}$	5.9461	6.1065	6.1275	6.3121	6.3102	6.5293	6.4020	6.6425
12	$\frac{3}{4}$	1	3.3173	3.1523	3.4679	3.2968	3.6220	3.4509	3.7002	3.5317
13	1	0	6.3426	6.5379	6.5179	6.7454	6.6946	6.9645	6.7834	7.0787
14	1	$\frac{1}{2}$	4.9564	5.0147	5.1512	5.2087	5.3472	5.4140	5.4457	5.5212

S = 0.08

TABLE 9b.

$S = 0.08$

States	Parameters		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
	η_1	α_1	EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	0	$\frac{1}{2}$	3.6492	4.0178	3.7423	4.1326	3.8332	4.2524	3.8779	4.3142
3	0	1	4.4754	4.9486	4.5501	5.0665	4.6231	5.1893	4.6590	5.2527
4	$\frac{1}{2}$	0	3.6492	4.0178	3.7423	4.1326	3.8332	4.2524	3.8779	4.3142
5	$\frac{1}{2}$	$\frac{1}{2}$	4.4731	4.9486	4.5478	5.0665	4.6209	5.1893	4.6568	5.2527
6	$\frac{1}{2}$	1	3.5417	3.8314	3.6352	3.9422	3.7268	4.0579	3.7719	4.1176
7	1	0	4.4754	4.9486	4.5501	5.0665	4.6231	5.1893	4.6590	5.2527
8	1	$\frac{1}{2}$	3.5417		3.6352	3.9422	3.7268	4.0579	3.7719	4.1176

TABLE 10a.

S = 0.12

States	Parameters η_1 α_1		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
			EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	0	$\frac{1}{2}$	5.1388	5.3116	5.4283	5.6138	5.7195	5.9418	5.8656	6.1164
3	0	1	6.2539	6.4128	6.5107	6.7172	6.7706	7.0473	6.9015	7.2228
4	$\frac{1}{4}$	0	3.5212	3.4443	3.7583	3.6811	4.0019	3.9408	4.1260	4.0801
5	$\frac{1}{4}$	$\frac{1}{2}$	6.0544	6.2342	6.3180	6.5431	6.5842	6.8777	6.7182	7.0556
6	$\frac{1}{4}$	1	5.8427	5.9365	6.1187	6.2368	6.3963	6.5627	6.5356	6.7362
7	$\frac{1}{2}$	0	5.1388	5.3116	5.4283	5.6138	5.7195	5.9418	5.8656	6.1164
8	$\frac{1}{2}$	$\frac{1}{2}$	6.2428	6.4128	6.4998	6.7172	6.7598	7.0473	6.8908	7.2228
9	$\frac{1}{2}$	1	4.8091	4.8284	5.0935	5.1097	5.3810	5.4160	5.5257	5.5795
10	$\frac{3}{4}$	0	6.0158	6.2342	6.2884	6.5431	6.5622	6.8777	6.6996	7.0556
11	$\frac{3}{4}$	$\frac{1}{2}$	5.8130	5.9365	6.0792	6.2368	6.3482	6.5627	6.4837	6.7362
12	$\frac{3}{4}$	1	3.1969	3.0071	3.4137	3.2146	3.6381	3.4428	3.7530	3.5656
13	1	0	6.2539	6.4128	6.5107	6.7172	6.7706	7.0473	6.9015	7.2228
14	1	$\frac{1}{2}$	4.8091	4.8284	5.0935	5.1097	5.3810	5.4160	5.5257	5.5795

TABLE 10b.

States	Parameters		H = 0.0		H = 0.5		H = 1.0		H = 1.25	
	η_1	α_1	EXACT	D.A	EXACT	D.A	EXACT	D.A	EXACT	D.A
2	0	$\frac{1}{2}$	3.6291	4.0018	3.7664	4.1731	3.8991	4.3553	3.9638	4.4508
3	0	1	4.4379	4.8804	4.5481	5.0545	4.6550	5.2397	4.7071	5.3366
4	$\frac{1}{2}$	0	3.6291	4.0018	3.7664	4.1731	3.8991	4.3553	3.9638	4.4508
5	$\frac{1}{2}$	$\frac{1}{2}$	4.4328	4.8804	4.5432	5.0545	4.6502	5.2397	4.7024	5.3366
6	$\frac{1}{2}$	1	3.4708	3.7260	3.6089	3.8883	3.7431	4.0614	3.8087	4.1522
7	1	0	4.4379	4.8804	4.5481	5.0545	4.6550	5.2397	4.7071	5.3366
8	1	$\frac{1}{2}$	3.4708	3.7260	3.6089	3.8883	3.7431	4.0614	3.8087	4.1522

S = 0.12

5.4. DISCUSSION

It is clear from the tables that the diffusion approximation in these small population situations studied is not very good. In almost all cases, the diffusion approximation over-estimates the exact mean time until fixation. Results for a population of size 3 showed that the exception to the above are states with parameters $(\frac{1}{4}, 0)$ and $(\frac{3}{4}, 1)$ where we have under-estimation by the diffusion approximations.

As expected, selection decreased the mean time until fixation and for a fixed selection intensity, the mean time was a strictly increasing function of initial gene frequencies. It was also noted that there was a frequency value below which for a fixed selection intensity the mean time was an increasing function of gene frequency and above which it was a decreasing function of gene frequency for any fixed dominance value.

In all cases, the mean time is an increasing function of dominance $(0, \frac{1}{2}, 1, 1\frac{1}{4})$ values for a fixed selection intensity. Results similar to those obtained in Chapter 4 are observed here too, i.e. the exact approach differentiates between pairs of states with the same average initial gene frequency but different parameters, while the diffusion approximation gives the same result for such pairs of states. Thus in table 8a, states 3 and 8 give different exact results though both have the same average initial gene frequency $(\frac{1}{2})$ while the diffusion approximation give the same result for both states. The only exception however is the obvious case with no selection ($S = 0$ Tables 7a, 7b) where states with the same average initial gene frequencies give the same exact results irrespective of the two parameters involved.

More interesting, however, is the fact that the diffusion approximation errors are not as much as Ewens (1963) observed. (Ewens studied a diploid population of size 6 with no sex differentiation). He observed that the mean error is approximately unity and suggested that subtracting unity from the diffusion approximation for the mean time will increase the accuracy of the approximation. However in this study the diffusion approximations to the mean times until fixation, differ by at most 0.5164 in a population of size 2 to about 0.1924 in a population of size 3, and not by a mean error of approximately unity as suggested by Ewens.

For both populations, the diffusion approximation errors are less in a population of size 3 than in a population of size 2; this is expected since changes in gene frequency are expected to approach a continuous process as N_e (effective population size) increases.

The results, as expected, suggest that A_1 - genes are expected to be fixed earlier in a population of size 2 than in a population of size 3 for the same selection intensity, same dominance values and the same initial gene parameters η_1, α_1 .

CHAPTER 6CONCLUSIONS

On the basis of the results obtained, some conclusions are possible.

Firstly, that the diffusion approximation method underestimates the exact probability of fixation by the j^{th} generation for all the initial gene frequencies, and all generation values considered in these small populations studied. The diffusion approximation is, however, good for large values of j . Similar results were obtained by Kimura (1955a) and Ewens (1963). Results obtained from the diffusion approximation for increasing j suggest that the approximation of an infinite series by a finite series used in 3.2 was reasonably good. Certainly further investigation is needed to reach a final conclusion on the extent to which the Diffusion Approximation could be used as an approximation to the exact results for non-zero selection intensity and various dominance values.

As expected with no selection, the probability of fixation by any generation is always less than the initial gene frequency at which the process started. However, graphs of Figures 2a, 2b, 3a and 3b are misleading since they suggest that this is not necessarily the case. The numerical results suggest this is so, but this should not in any way affect the conclusion above.

On the other hand, the diffusion approximation usually exceeds (overestimates) the exact (true) probability of ultimate fixation and the expected mean times until fixation for all cases considered. Exceptions to the latter being those cases discussed in 5.3. The diffusion approximation to the probability of ultimate fixation is

very good for small selection intensity values studied, and the errors are smaller for high than for low initial gene frequencies. The agreement between the diffusion approximations and the exact results was greater for the population of size 3 and smaller s (selection intensity) than for the population of size 2 and bigger s .

Interesting is the fact that the diffusion approximation errors to the expected mean times until fixation are not as much as Ewens (1963) observed for slightly larger population. This does not mean that Ewens' results were not correct, since various work by Ewens (1962), Moran (1958) and Wright (1931) suggested that subtracting unity from the diffusion approximation for the mean times will increase the accuracy of the approximation. However, the difference between their results and mine could arise due to the different methods used in the solutions of the equation. As mentioned in 5.2, numerical integration using the central difference approximations to derivatives method was employed in this study for the solution to the expected mean time until fixation of an allele (A_1), whereas the other authors used numerically integration using Simpson's rule.

Perhaps the most striking difference between the exact and diffusion approaches is that the exact results differentiate between pairs of initial states with non-symmetrical initial gene parameters (i_1, α_1) while the diffusion approximations used always give the same results for all cases considered in the two populations.

APPENDIX

For the proofs in this section we shall refer to Chapter 2 Section 2.1. There we have the frequency of A_1 in offspring (infinite population of offspring) as

$$P = \frac{\eta_1 \alpha_1 (1+S) + \frac{1}{2} (\eta_1 \alpha_2 + \eta_2 \alpha_1) (1 + hS)}{\eta_1 \alpha_1 (1+S) + (1+hS) (\eta_1 \alpha_2 + \eta_2 \alpha_1) + \eta_2 \alpha_2} \quad \text{_____ (1)}$$

The effective population size N_e is given by Wright (1931) as

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f} \quad \text{_____ (2)}$$

Let the two alleles A_1 and A_2 have frequencies x and $1-x$ respectively in the population. Thus probability (of A_1 gene in the offspring population) = P where P is given by (1), suppose that after one generation of selection, the frequency of A_1 in the population is $x + \delta x$, then the expected change in frequency of A_1 after one generation is given by

$$M_{\delta x} = E (x + \delta x) - x \quad \text{_____ (3)}$$

Since sampling is binomial, R has the sampling variance $\frac{P(1-P)}{2N_e}$

where R is the number of A_1 - genes in sample after one generation of selection.

$$\text{Thus } x + \delta x = \frac{R}{2N} \quad \text{_____ (4)}$$

since there are N individuals in the population and hence $2N$ genes at each locus.

$$\begin{aligned} \text{Therefore } M_{\sigma_x} &= E(x + \sigma_x) - x \\ &= \frac{E(R)}{2N} - x \end{aligned}$$

$$\text{But } E(R) = 2NP$$

$$\text{therefore } M_{\sigma_x} = \frac{2NP}{2N} - x = P - x$$

$$\text{i.e. } M_{\sigma_x} = P - x \text{ (with } \eta_1 = \alpha_1 = x)$$

which after some manipulations becomes

$$\begin{aligned} M_{\sigma_x} &= \frac{2x(1-x)(1+hS) + x^2(1+S)}{x^2(1+S) + 2(1+hS)x(1-x) + (1-x)^2} - x \\ &= \frac{Sx(1-x) [h + (1-2h)x]}{1 + [x^2(1-2h) + 2xh]s} \\ &= Sx(1-x) [h + (1-2h)x] \left\{ 1 + [x^2(1-2h) + 2xh]s \right\}^{-1} \end{aligned}$$

But the denominator can be expanded as

$$1 - [x^2(1-2h) + 2xh]s + [x^2(1-2h) + 2xh]^2 s^2 - \dots$$

Thus to first order in S

$$M_{\sigma_x} = Sx(1-x) [h + (1-2h)x]$$

which is the required proof.

Similarly, the variance of change in gene frequency in one generation is given by

$$\begin{aligned} V_{\sigma_x} &= V(x) = V(x + \sigma_x) \\ &= V(R/2N) \end{aligned}$$

but $V(R/2N) =$ variance of a binomial proportion with parameters $2N_e$ and P

$$= 2N_e P (1 - P)$$

$$\begin{aligned} \text{Thus } V_{\sigma_x} &= \frac{2N_e \cdot P (1 - P)}{4N_e^2} \\ &= \frac{P (1 - P)}{2N_e} \quad \text{with } \eta_1 = \alpha_1 = x \end{aligned}$$

which if we neglect terms of order S/N_e , i.e. $S = 0$, we have

$$\begin{aligned} V_{\sigma_x} &= \frac{(1-x) [x^2 + x(1-x)]}{2N_e [x^2 + (1-x)(1+x)]^2} \\ &= \frac{x(1-x)}{2N_e} \end{aligned}$$

REFERENCES

1. COX, D.R. and MILLER, H.D. (1965). The theory of stochastic processes.
2. CROW, J.F., and KIMURA, M. Introduction to population genetics theory. (1970).
3. EWENS, W.J. (1962). The mean time for absorption in a genetic process. Submitted to Aust. Math. J.
4. EWENS, W.J. (1963). Numerical results and diffusion approximations in a genetic process. Biometrika 3 and 4.
5. FELLER, W. (1954) Diffusion processes in one dimension. Trans. American Math. Soc. 77, 1-31.
6. FISHER, R.A. (1930). The genetical theory of Natural Selection. Oxford University Press.
7. GERALD, C.F. (1970). Applied Numerical Analysis.
8. KIMURA, (1955a). Process of random genetic drift with a continuous model. Proc. Nat. Acad. Sci., Wash. 41, 144-50.
9. KIMURA, M. (1955). Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symp. 20, 33-53.
10. KIMURA M. (1964). Diffusion models in population genetics. Journal of Applied Probability.
11. KOLMOGOROV, A.N. (1931). Uber die analytischen methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. 104, 415-458.
12. KOLMOGOROV, A.N. (1959). Transition of branching processes into diffusion processes and some problems in genetics. Teor. Veroyatnost. i Primenem, 4, 233-6.

13. MORAN, P.A.P. (1958). Random processes in genetics. Proc. Camb. Phil. Soc. 54, 60-71.
14. MORAN, P.A.P. (1960). The survival of a mutant gene under Selection II. Austr. Math. J. 1, 485-91.
15. PARZEN, E. (1962). Stochastic Processes. Holden-Day, Inc., San Francisco.
16. WRIGHT, S. (1931). Evolution in Mendelian populations Genetics, 16, 97-159.
17. WRIGHT, S. (1949). Adaptation and Selection in Genetics, Paleontology and Evolution. Jepsen et.al. (ed.), Princeton Univ. Press.