

$(4N-2)/(\sigma^2+2)$  which is equivalent in our terminology to saying that the asymptotic rate of approach to homozygosity is

$$\lambda = 1 - \{(\sigma^2+2)/(8N-4)\}.$$

A brief derivation of this result is given in Li (1955, p. 321), and it was also obtained by Haldane (1939) and Fisher (1939).

For dioecious populations of diploid individuals only particular examples have been studied. Wright (1931, 1933) found that the rate of approach, for a population with  $N_1$  males, and  $N_2$  females practising a mating system to be described in the next section, is

$$\lambda \doteq 1 - N/8N_1N_2,$$

for autosomal genes, where  $N = N_1 + N_2$ , and for sex-linked genes it is

$$\lambda \doteq 1 - (N_2 + 2N_1)/9N_1N_2.$$

In Wright's model, the generations are kept distinct and mating between them is not possible. An alternative model has been proposed for which the generations do overlap (Moran 1958), and the resulting rate of approach for autosomal genes is

$$\lambda \doteq 1 - N/4N_1N_2.$$

Finally, a third model very similar to the previous one, and having overlapping generations, was shown (Watterson, unpublished data) to have a rate identical to Wright's model,

$$\lambda \doteq 1 - N/8N_1N_2.$$

All these dioecious models have specific distributions for the number of offspring per individual of the parent generation, and it is our aim to generalize the results to cases with arbitrary offspring distributions, and also investigate the effect that permanent marriage between sexes has on the population's asymptotic behaviour. Therefore, we first consider populations from the point of view of family structure and size, and then introduce genetic considerations and investigate the effect of family structure on the asymptotic genetic behaviour of the population.

## II. FAMILY STRUCTURE

For a given dioecious population, let  $\pi_1(x)$  be the probability of a male individual surviving to the age  $x$  at least, and  $\pi_2(x)$  be the similar probability for a female individual. For these quantities to be well defined, it is implicitly assumed that no selective effects are present, and that all individuals of same sex are equivalent with respect to lifetime distributions. With the same assumption, we may define two further quantities  $m_1(x)dx$  and  $m_2(x)dx$  to be the probabilities that a male and female of age  $x$  will produce an offspring of either sex in the time interval  $(x, x+dx)$ . We suppose that such events occur independently in different intervals of  $x$ , i.e. we ignore the effect of a finite length of pregnancy and suppose that the probability of producing an offspring at any given age is independent of what has happened before. Then the number of offspring produced by a male which lived

to an age  $x$  and then died in the interval  $(x, x+dx)$ , an event with probability  $-\pi'_1(x) dx$ , has a Poisson distribution with mean

$$M_1(x) = \int_0^x m_1(t) dt,$$

and a probability generating function

$$\exp \left\{ (z-1) \int_0^x m_1(t) dt \right\}.$$

Averaging over all possible ages we therefore get the probability generating function for the total number of offspring from a male individual :

$$P_1(z) = - \int_0^\infty \exp \left\{ (z-1) \int_0^x m_1(t) dt \right\} \pi'_1(x) dx. \quad \dots\dots(1)$$

Similarly for offspring from a female individual the generating function is

$$P_2(z) = - \int_0^\infty \exp \left\{ (z-1) \int_0^x m_2(t) dt \right\} \pi'_2(x) dx. \quad \dots\dots(2)$$

Such generating functions are to play a large part in the next section of the paper, so we shall consider them in somewhat more detail. In the above, we have made several simplifying assumptions which may or may not be valid when an actual population is investigated. Nevertheless, generating functions for the number of offspring per individual are important whether equations (1) and (2) are applicable or not. For example, if the generations of a population neither overlap nor interbreed, then a workable theoretical model can be formulated in which all individuals die simultaneously, and the entire succeeding generation is born at that instant. For this model the previous birth and death rate functions are degenerate, but of course the generating functions may be defined from first principles. Thus if the  $i$ th male has  $x_i$  offspring, and  $Pr\{x_i = n\} = p_n$  (say), then the generating function is

$$P_1(z) = \sum_{n=0}^{\infty} p_n z^n.$$

The simplest way to introduce monogamy ("marriage") into a population model is to assume that the sex numbers are equal and allow permanent random pairing between members of opposite sex. One consequence of this pairing is that both parents in the family have the same number of offspring, and hence the generating functions  $P_1(z)$  and  $P_2(z)$  must be identical.

Typical examples of generating functions which may be of use in practical cases are those corresponding to Poisson, geometric, and binomial distributions for the  $x_i$  variates. The first two of these distributions allow the  $x_i$  to vary over all integers, whilst the latter restricts them to a finite range. Most theoretical models which have been found suitable for mathematical investigation restrict the population size to exactly  $N$  for all generations, the sex numbers being held constant at  $N_1$  and  $N_2$ . For these models the offspring distributions must have finite range, but

it is interesting to consider how such a model may approximate an actual population with arbitrary offspring distributions, perhaps of infinite range.

Let us concentrate on the male parents. Given any generating function  $Q(z)$  for a random variate  $y$  taking integral values, the joint generating function for  $N_1$  independent variables  $y_1, y_2, \dots, y_{N_1}$  (say) is

$$Q(z_1) Q(z_2) \dots Q(z_{N_1}).$$

If the variable  $y_i$  is the number of offspring from the  $i$ th male in a population, then the next generation will have size  $\Sigma y_i$  with expectation  $N_1 Q'(1)$ . To construct a constant size model approximating to this population, the numbers of offspring, per male, must be dependent random variables  $x_1, x_2, \dots, x_{N_1}$  (say) with total  $\Sigma x_i = N$ . Suppose the  $x_i$  have identical distributions with generating function  $P_1(z)$  and write  $p_n = \text{Pr}\{x_i = n\}$ ,  $q_n = \text{Pr}\{y_i = n\}$ . For the model to fit the actual population approximately, we need  $N = N_1 P'_1(1) = N_1 Q'(1)$  and  $p_n \doteq q_n$ . In fact, we shall see that the former condition is sufficient to ensure the latter.

The joint generating function for the  $x_i$  variables (conditional on  $\Sigma x_i = N$ ) is

$$\frac{\text{coefficient of } w^N \text{ in } Q(z_1 w) Q(z_2 w) \dots Q(z_{N_1} w)}{\text{coefficient of } w^N \text{ in } Q(w)^{N_1}},$$

thus with  $z = z_1, z_2 = z_3 = \dots = z_{N_1} = 1$ ,

$$P_1(z) = \frac{\text{coefficient of } w^N \text{ in } Q(zw) Q(w)^{N_1}}{\text{coefficient of } w^N \text{ in } Q(w)^{N_1}},$$

and so

$$\begin{aligned} p_n &= \text{coefficient of } z^n \text{ in } P_1(z), \\ &= q_n \frac{\text{coefficient of } w^{N-n} \text{ in } Q(w)^{N_1-1}}{\text{coefficient of } w^N \text{ in } Q(w)^{N_1}}. \end{aligned}$$

To prove that  $p_n \rightarrow q_n$  as  $N_1, N \rightarrow \infty$  provided  $N = N_1 P'_1(1) = N_1 Q'(1)$ , we must show that the coefficients of  $w^{N-n}$  in  $Q(w)^{N_1-1}$  and of  $w^N$  in  $Q(w)^{N_1}$  are asymptotically equal. This is obviously true if  $NN_1^{-1}$  is an integer,  $n = NN_1^{-1}$ , and  $Q(w) = w^{NN_1^{-1}}$ . Suppose that  $Q(w)$  has a non-zero finite variance. Then  $Q(w)^{N_1}$  is the generating function of a discrete probability distribution whose ordinates, by a form of the Central Limit Theorem, are asymptotically equal to quantities proportional to the ordinates of a normal distribution with mean  $N$  and variance proportional to  $N_1$ . The coefficient of  $w^N$  will tend to the central ordinate, and since  $n$  is fixed and tends to zero in comparison with the standard deviation, the coefficient of  $w^{N-n}$  in  $Q(w)^{N_1-1}$  will also tend to the central ordinate. Thus under these conditions,  $p_n \rightarrow q_n$ , and this convergence is easily shown to be uniform on  $n$ . The correspondence between the two populations, however, only holds from one generation to the next, and clearly the constant size model will not have the same behaviour after many generations as another population which may die out completely.

As we have seen, for a population to remain at constant size, the means of the generating functions for each sex must be

$$P'_1(1) = NN_1^{-1}, P'_2(1) = NN_2^{-1}. \quad \dots \dots \dots (3)$$

Because the constraint imposes a degree of dependence between the numbers of offspring from two males (or females) there is a covariance  $c_1$  (say) between those for the males, and similarly  $c_2$  between females. We write  $\sigma_1^2$  for the variance of the distribution defined by  $P_1(z)$ , and  $\sigma_2^2$  for the distribution defined by  $P_2(z)$ . Now

because  $\sum_1^{N_1} x_i = N$ , we must have  $N_1\sigma_1^2 + N_1(N_1 - 1)c_1 = 0$ , so that

$$\left. \begin{aligned} c_1 &= -\sigma_1^2(N_1 - 1)^{-1} = -(P_1''(1) + P_1'(1) - P_1'(1)^2)(N_1 - 1)^{-1}, \\ \text{and similarly} \\ c_2 &= -\sigma_2^2(N_2 - 1)^{-1} = -(P_2''(1) + P_2'(1) - P_2'(1)^2)(N_2 - 1)^{-1}. \end{aligned} \right\} \dots (4)$$

The first two moments of the offspring distribution are critical quantities for subsequent theory. For a population with non-degenerate birth and death rates, we have from (1) and (2) that

$$\begin{aligned} P_1'(1) &= \int_0^\infty m_1(x) \pi_1(x) dx, \\ P_2'(1) &= \int_0^\infty m_2(x) \pi_2(x) dx, \end{aligned}$$

found by integrating by parts and noting that  $\pi_i(x)$  tends to zero as  $x$  tends to infinity. For constant expected population size these means must equal  $NN_1^{-1}$  and  $NN_2^{-1}$  respectively. Similarly we find the second derivatives

$$\begin{aligned} P_1''(1) &= 2 \int_0^\infty m_1(x) \pi_1(x) \left\{ \int_0^x m_1(t) dt \right\} dx = 2 \int_0^\infty m_1(x) \pi_1(x) M_1(x) dx, \\ P_2''(1) &= 2 \int_0^\infty m_2(x) \pi_2(x) \left\{ \int_0^x m_2(t) dt \right\} dx = 2 \int_0^\infty m_2(x) \pi_2(x) M_2(x) dx. \end{aligned}$$

One other aspect of this problem deserves mention (see Fisher 1939). Consider a population in which the generations do not overlap and for which each individual has a large number of offspring only a few of which survive. Then the model we use will depend on the stage at which the population is enumerated. If the population of sexually mature adults is of size  $N$ , and for simplicity the sex numbers are equal,  $N_1 = N_2 = \frac{1}{2}N$ , and if each individual has a probability distribution of producing immature individuals with generating function  $R(z)$ , then the correct generating functions to use in the model will be

$$P_1(z) = P_2(z) = R(1 - \pi + \pi z), \quad \dots \dots \dots (5)$$

where  $\pi$  is the probability of an immature individual surviving to maturity. If on the other hand we enumerate the immature individuals the correct generating functions to use will be

$$P_1(z) = P_2(z) = 1 - \pi + \pi R(z), \quad \dots \dots \dots (6)$$

together with the much larger population size  $\frac{1}{2}NR'(1)$ . Since the genetic results must be the same we have set up an interesting equivalence between populations of different size and different distributions of the numbers of offspring. As a simple

example, the expected population size will be constant only if

$$P'_1(1) = \pi R'(1) = N(\frac{1}{2}N)^{-1} = 2,$$

in the first case, and in the second if

$$P'_1(1) = \pi R'(1) = \frac{1}{2}NR'(1)\{\frac{1}{4}NR'(1)\}^{-1} = 2.$$

Both conditions are satisfied if  $R'(1) = 2\pi^{-1}$ .

To conclude this section, some examples will be given of generating functions applying to population models previously studied. Perhaps the most important is the following.

(a) *Wright's Model*

Wright (1931) studied a population having non-overlapping generations, and assumed that  $N$  offspring resulted from matings formed by random sampling, with replacement, amongst the  $N_1$  male and  $N_2$  female parents. Thus the number of offspring per individual is a binomial variate, and the generating functions are

$$\left. \begin{aligned} P_1(z) &= (1 - N_1^{-1} + N_1^{-1}z)^N, \\ P_2(z) &= (1 - N_2^{-1} + N_2^{-1}z)^N. \end{aligned} \right\} \dots\dots\dots (7)$$

(b) *Degenerate Model*

If all male individuals have exactly  $NN_1^{-1}$  offspring, all females exactly  $NN_2^{-1}$ , these being integers, then the variances  $\sigma_1^2$  and  $\sigma_2^2$  are both zero, and the generating functions are

$$\left. \begin{aligned} P_1(z) &= z^{NN_1^{-1}}, \\ P_2(z) &= z^{NN_2^{-1}}. \end{aligned} \right\} \dots\dots\dots (8)$$

(c) *Overlapping Generation Model I*

In a previous paper (Moran 1958) a model has been constructed in which the generations are overlapping. Individuals die at random and are replaced by new individuals formed by the mating of gametes chosen from the output of the population before the death took place. The probability of any given individual dying at such a birth-death event is  $N^{-1}$ , and of having a life time of exactly  $n$  units is  $N^{-1}(1 - N^{-1})^{n-1}$ ,  $n = 1, 2, 3, \dots$ . This may be expressed by saying that  $n-1$  has a geometric distribution; the expected life time is  $N$  and is the number of time units corresponding to one generation in a non-overlapping model.

At each instant at which a death occurs, including its own, the probability of the individual becoming a parent is  $N_1^{-1}$  if male,  $N_2^{-1}$  if female, so that the probability generating function of the number of offspring per male is

$$\left. \begin{aligned} P_1(z) &= \sum_{n=1}^{\infty} N^{-1}(1 - N^{-1})^{n-1}(1 - N_1^{-1} + N_1^{-1}z)^n \\ &= (N_1 - 1 + z)\{N + N_1 - 1 - (N - 1)z\}^{-1}, \\ \text{and similarly for the females} \\ P_2(z) &= (N_2 - 1 + z)\{N + N_2 - 1 - (N - 1)z\}^{-1}. \end{aligned} \right\} \dots\dots (9)$$

(d) *Overlapping Generation Model II*

Watterson (unpublished data) has considered an overlapping model slightly different from that above by assuming that individuals die at random, and each is replaced by a new individual with one parent being the dying individual, the other being chosen at random from the opposite sex. In this case the probability of an individual living for exactly  $n$  time units is  $N^{-1}(1-N^{-1})^{n-1}$ ,  $n = 1, 2, 3, \dots$ , as before, but the probability of its being a parent at any stage is altered. It must always be a parent at its own death. Consider an event not involving its death, but being the death of one of the  $N-1$  remaining individuals. If the original individual was male, then for it to be a parent at a death other than its own, the dying individual must be of opposite sex (probability  $N_2(N-1)^{-1}$ ) and the original individual must be chosen as the second parent from  $N_1$  similar males. Therefore the required probability is  $N_2(N-1)^{-1}N_1^{-1}$ , and the generating function becomes

$$\begin{aligned}
 P_1(z) &= \sum_{n=1}^{\infty} N^{-1}(1-N^{-1})^{n-1} \{1 - N_2(N-1)^{-1}N_1^{-1} + N_2(N-1)^{-1}N_1^{-1}z\}^{n-1}z, \\
 &= zN_1(N-N_2z)^{-1}.
 \end{aligned}$$

Similarly

$$P_2(z) = zN_2(N-N_1z)^{-1}.$$

$$\left. \begin{array}{l} \\ \\ \end{array} \right\} \dots\dots\dots (10)$$

The latter two models have generating functions closely akin to those for geometrically distributed variates. Model I differs from this by altering the probability of an individual having no offspring, whilst model II defines geometric variates taking the values 1, 2, 3, . . . . By their definition, it is clear that the population remains constant in size for all the models mentioned, and it is easy to verify that condition (3) is satisfied.

## III. GENETIC BEHAVIOUR OF POPULATIONS

We come now to the effect of family structure on the genetic behaviour of a population, in particular after a large number of generations. The population is assumed to be diploid, and the genetic factor of interest is diallelic, so that the individuals are either of genotype  $aa$ ,  $Aa$ , or  $AA$ . This is the formulation for autosomal genes, but we shall later deal with sex-linked genes, and then the male genotypes are either  $a$  or  $A$ . For the autosomal case, write  $k_t$ ,  $u_t$ ,  $l_t$  as the number of male individuals at the  $t$ th generation whose genotypes are  $aa$ ,  $Aa$ , and  $AA$  respectively. Similarly, we define the numbers  $r_t$ ,  $v_t$ ,  $s_t$  for the females. If the sex numbers are  $N_1$  and  $N_2$ , then the number of heterozygotes in either sex are  $u_t = N_1 - k_t - l_t$ ,  $v_t = N_2 - r_t - s_t$ . The genetic state of the population at any time is determined by the four variates ( $k_t$ ,  $l_t$ ,  $r_t$ ,  $s_t$ ).

We consider a non-overlapping generation model for which all individuals are produced simultaneously at the death of the preceding generation and their genotypes are determined as if mating occurred between the sexes of that generation. For this model, the genetic state forms a Markov chain with points determined by the coordinates ( $k, l, r, s$ ). There are two absorbing states ( $N_1, 0, N_2, 0$ ) and ( $0, N_1, 0, N_2$ )

corresponding to homozygosity and gene fixation, and the transition probabilities from any state to any other state could be written down. By the general theory of Markov chains it can be shown that the probability that the system has not reached an absorbing state at time  $t$  ( $t = 0, 1, 2, \dots$ ) is asymptotically  $C\lambda^t$ , where  $t$  is measured in units of one generation,  $C$  is a constant depending on the initial state, and  $\lambda$  is the largest non-unit root of the matrix of transition probabilities. This matrix is too complicated to be written down explicitly, but by calculating the first and second moments of  $k, l, r, s$  at time  $t+1$  in terms of those at time  $t$  it can be shown that  $\lambda$  is also the largest characteristic root governing these difference equations. We proceed by the latter method; the determination of the constant  $C$  is more difficult, and has been carried out only for the simpler haploid models. It will therefore not be attempted here.

First we take the case when permanent marriage does not occur, and assume random mating subject to the restriction that the generating functions for the number of offspring per individual are  $P_1(z)$  and  $P_2(z)$  for males and females respectively. At each generation the sex numbers will be constant at  $N_1$  and  $N_2$ , so equations (3) must hold. We write  $x_{ij}$  for the numbers of offspring produced by all matings between individuals of specified genotypes according to the scheme:

		Males				
		$aa$	$Aa$	$AA$		
Females	$aa$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{1.}$	....(11)
	$Aa$	$x_{21}$	$x_{22}$	$x_{23}$	$x_{2.}$	
	$AA$	$x_{31}$	$x_{32}$	$x_{33}$	$x_{3.}$	
		$x_{.1}$	$x_{.2}$	$x_{.3}$	$N$	

Thus  $x_{.j} = \sum_i x_{ij}$  and so on. If the parent generation is the  $t$ th, then we obtain from the offspring distributions, equation (3), that

$$\left. \begin{aligned} E(x_{.1}) &= k_t N N_1^{-1}, & E(x_{1.}) &= r_t N N_2^{-1}, \\ E(x_{.3}) &= l_t N N_1^{-1}, & E(x_{3.}) &= s_t N N_2^{-1}. \end{aligned} \right\} \dots\dots\dots(12)$$

Furthermore, using the values of  $c_1$  and  $c_2$  from equation (4) we get

$$\left. \begin{aligned} \text{var}(x_{.1}) &= k_t \sigma_1^2 \{1 - (k_t - 1)(N_1 - 1)^{-1}\}, \\ \text{var}(x_{.3}) &= l_t \sigma_1^2 \{1 - (l_t - 1)(N_1 - 1)^{-1}\}, \\ \text{var}(x_{1.}) &= r_t \sigma_2^2 \{1 - (r_t - 1)(N_2 - 1)^{-1}\}, \\ \text{var}(x_{3.}) &= s_t \sigma_2^2 \{1 - (s_t - 1)(N_2 - 1)^{-1}\}, \\ \text{cov}(x_{.1}, x_{.3}) &= -k_t l_t \sigma_1^2 (N_1 - 1)^{-1}, \\ \text{cov}(x_{1.}, x_{3.}) &= -r_t s_t \sigma_2^2 (N_2 - 1)^{-1}. \end{aligned} \right\} \dots\dots(13)$$

Of these offspring exactly  $N_1$  are to be male,  $N_2$  female, but apart from these restrictions, we assume random sex distributions. If we write  $x_{ij} = m_{ij} + f_{ij}$ , where  $m_{ij}$  and  $f_{ij}$  are the numbers of male and female offspring from the mating type  $(i, j)$ , we can express this condition by representing the  $m_{ij}$  and  $f_{ij}$  in a  $2 \times 9$  contingency table as follows:

$m_{11}$	$m_{12}$	$m_{13}$	$m_{21}$	$m_{22}$	$m_{23}$	$m_{31}$	$m_{32}$	$m_{33}$	$N_1$
$f_{11}$	$f_{12}$	$f_{13}$	$f_{21}$	$f_{22}$	$f_{23}$	$f_{31}$	$f_{32}$	$f_{33}$	$N_2$
$x_{11}$	$x_{12}$	$x_{13}$	$x_{21}$	$x_{22}$	$x_{23}$	$x_{31}$	$x_{32}$	$x_{33}$	$N$

The  $x_{11}$  offspring from matings of type  $(aa \times aa)$  are all of genotype  $aa$ , the  $x_{12}$  offspring from matings  $(aa \times Aa)$  are a mixture of genotypes  $aa$  and  $Aa$ , each formed with probability  $\frac{1}{2}$ , and similar results hold for each of the other mating types. We write

$$\begin{aligned} m_{ij} &= k_{ij} + u_{ij} + l_{ij}, \\ f_{ij} &= r_{ij} + v_{ij} + s_{ij}, \end{aligned}$$

where  $k_{ij}$  is the number of male offspring of type  $aa$  from matings of type  $(i, j)$  and so on. If we write  $p_{ij}$ ,  $q_{ij}$  for the probabilities of a single offspring from a mating type  $(i, j)$  being of genotype  $aa$  or  $AA$ , then the total genotypic outputs in males and females have the trinomial distributions

$$\left. \begin{aligned} Pr\{k_{ij}, u_{ij}, l_{ij}\} &= \frac{m_{ij}!}{k_{ij}! u_{ij}! l_{ij}!} p_{ij}^{k_{ij}} (1-p_{ij}-q_{ij})^{u_{ij}} q_{ij}^{l_{ij}}, \\ Pr\{r_{ij}, v_{ij}, s_{ij}\} &= \frac{f_{ij}!}{r_{ij}! v_{ij}! s_{ij}!} p_{ij}^{r_{ij}} (1-p_{ij}-q_{ij})^{v_{ij}} q_{ij}^{s_{ij}}, \end{aligned} \right\} \dots (14)$$

where  $p_{ij}$  and  $q_{ij}$  are given in the following table:

$i, j$	11	12	13	21	22	23	31	32	33
$p_{ij}$	1	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{4}$	0	0	0	0
$q_{ij}$	0	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	0	$\frac{1}{2}$	1

Finally the genotype numbers in the offspring generation are  $k_{t+1} = \Sigma k_{ij}$ ,  $l_{t+1} = \Sigma l_{ij}$ ,  $r_{t+1} = \Sigma r_{ij}$ ,  $s_{t+1} = \Sigma s_{ij}$ . We now consider the first- and second-order moments:

$$\begin{aligned} p_{t+1} &= N_1^{-1} E(k_{t+1} + l_{t+1}), \\ q_{t+1} &= N_2^{-1} E(r_{t+1} + s_{t+1}), \\ v_{t+1} &= N_1^{-1} N_2^{-1} E(k_{t+1} - l_{t+1})(r_{t+1} - s_{t+1}), \\ a_{t+1} &= N_1^{-2} E(k_{t+1} - l_{t+1})^2, \\ b_{t+1} &= N_2^{-2} E(r_{t+1} - s_{t+1})^2. \end{aligned}$$

We have to express these quantities in terms of the same quantities at generation  $t$ .



First consider expectations conditional on the values of  $m_{ij}$  and  $f_{ij}$ . From the distributions (14) we have

$$\begin{aligned} E(k_{ij}) &= m_{ij} p_{ij}, & E(k_{ij}^2) &= m_{ij} p_{ij} + m_{ij}(m_{ij}-1)p_{ij}^2, \\ E(l_{ij}) &= m_{ij} q_{ij}, & E(l_{ij}^2) &= m_{ij} q_{ij} + m_{ij}(m_{ij}-1)q_{ij}^2, \\ E(k_{ij} l_{ij}) &= m_{ij}(m_{ij}-1)p_{ij}q_{ij}, & E(k_{ij} l_{lm}) &= m_{ij} m_{lm} p_{ij} p_{lm}, \end{aligned}$$

and  $k_{ij}$  and  $k_{lm}$  are independent for  $(i, j) \neq (l, m)$ . Similar relations hold for the female variates. In this way we find

$$\begin{aligned} p_{t+1} &= N_1^{-1} E(m_{11} + \frac{1}{2}m_{12} + \frac{1}{2}m_{21} + \frac{1}{2}m_{22} + \frac{1}{2}m_{23} + \frac{1}{2}m_{32} + m_{33}), \\ &= \frac{1}{2} + \frac{1}{2}N_1^{-1} E(m_{11} + m_{33} - m_{13} - m_{31}), \end{aligned}$$

and similarly

$$q_{t+1} = \frac{1}{2} + \frac{1}{2}N_2^{-1} E(f_{11} + f_{33} - f_{13} - f_{31}).$$

Since the distributions of genotype frequencies in males and females are independent so long as the  $m_{ij}$  and  $f_{ij}$  are kept fixed we have

$$\begin{aligned} v_{t+1} &= N_1^{-1} N_2^{-1} E(m_{11} + \frac{1}{2}m_{12} + \frac{1}{2}m_{21} - \frac{1}{2}m_{23} - \frac{1}{2}m_{32} - m_{33}) \\ &\quad \times (f_{11} + \frac{1}{2}f_{12} + \frac{1}{2}f_{21} - \frac{1}{2}f_{23} - \frac{1}{2}f_{32} - f_{33}), \end{aligned}$$

and

$$\begin{aligned} a_{t+1} &= N_1^{-2} E\{(m_{11} + \frac{1}{2}m_{12} + \frac{1}{2}m_{21} - \frac{1}{2}m_{23} - \frac{1}{2}m_{32} - m_{33})^2 \\ &\quad + \frac{1}{4}m_{12} + \frac{1}{4}m_{21} + \frac{1}{2}m_{22} + \frac{1}{4}m_{23} + \frac{1}{4}m_{32}\}, \\ b_{t+1} &= N_2^{-2} E\{(f_{11} + \frac{1}{2}f_{12} + \frac{1}{2}f_{21} - \frac{1}{2}f_{23} - \frac{1}{2}f_{32} - f_{33})^2 \\ &\quad + \frac{1}{4}f_{12} + \frac{1}{4}f_{21} + \frac{1}{2}f_{22} + \frac{1}{4}f_{23} + \frac{1}{4}f_{32}\}. \end{aligned}$$

We now have to calculate these expectations conditional on the values of the  $x_{ij}$ . Using standard results for the conditional expectations of the entries in a contingency table (e.g. see Wilks 1946, p. 216) we have

$$\begin{aligned} E(m_{ij}) &= x_{ij} N_1 N^{-1}, \\ E(f_{ij}) &= x_{ij} N_2 N^{-1}, \\ E(m_{ij}^2) &= x_{ij}(x_{ij}-1)N_1(N_1-1)N^{-1}(N-1)^{-1} + x_{ij}N_1N^{-1}, \\ E(f_{ij}^2) &= x_{ij}(x_{ij}-1)N_2(N_2-1)N^{-1}(N-1)^{-1} + x_{ij}N_2N^{-1}, \\ E(m_{ij} f_{ij}) &= x_{ij}(x_{ij}-1)N_1N_2N^{-1}(N-1)^{-1}, \end{aligned}$$

and for  $(i, j) \neq (l, m)$

$$\begin{aligned} E(m_{ij} m_{lm}) &= x_{ij} x_{lm} N_1(N_1-1)N^{-1}(N-1)^{-1}, \\ E(f_{ij} f_{lm}) &= x_{ij} x_{lm} N_2(N_2-1)N^{-1}(N-1)^{-1}, \\ E(m_{ij} f_{lm}) &= x_{ij} x_{lm} N_1N_2N^{-1}(N-1)^{-1}. \end{aligned}$$

Using these results and also the fact that

$$x_{11} + \frac{1}{2}x_{12} + \frac{1}{2}x_{21} - \frac{1}{2}x_{23} - \frac{1}{2}x_{32} - x_{33} = \frac{1}{2}(x_{1.} - x_{3.} + x_{.1} - x_{.3}),$$

we find

$$\left. \begin{aligned} p_{t+1} &= q_{t+1} = \frac{1}{2} + \frac{1}{2}N^{-1}E(x_{11} + x_{33} - x_{13} - x_{31}), \\ v_{t+1} &= N^{-1}(N-1)^{-1}E\{(x_{11} + \frac{1}{2}x_{12} + \frac{1}{2}x_{21} - \frac{1}{2}x_{23} - \frac{1}{2}x_{32} - x_{33})^2 \\ &\quad - (x_{11} + \frac{1}{4}x_{12} + \frac{1}{4}x_{21} + \frac{1}{4}x_{23} + \frac{1}{4}x_{32} + x_{33})\}, \\ &= N^{-1}(N-1)^{-1}E\{\frac{1}{4}(x_{1.} - x_{3.} + x_{.1} - x_{.3})^2 - \frac{1}{4}(x_{1.} + x_{3.} + x_{.1} + x_{.3}) \\ &\quad - \frac{1}{2}(x_{11} + x_{33} - x_{13} - x_{31})\}, \\ a_{t+1} &= (1 - N_1^{-1})v_{t+1} + N_1^{-1}p_{t+1}, \\ b_{t+1} &= (1 - N_2^{-1})v_{t+1} + N_2^{-1}q_{t+1}. \end{aligned} \right\} \dots (15)$$

The last two equations are obtained by comparing the values of  $a_{t+1}$  and  $b_{t+1}$  expressed as expectations conditional on fixed values of  $x_{ij}$  with those obtained for the other three quantities. Now regarding the  $x_{ij}$  as variates of a contingency table (11) with fixed row and column totals, we have the conditional expectations  $E(x_{ij}) = x_{i.}x_{.j}N^{-1}$ , and then taking expectations according to (12), (13), we get

$$\begin{aligned} E(x_{11} + x_{33} - x_{13} - x_{31}) &= E(k_t - l_t)(r_t - s_t)NN_1^{-1}N_2^{-1}, \\ E\{(x_{1.} - x_{3.} + x_{.1} - x_{.3})^2 - (x_{1.} + x_{3.} + x_{.1} + x_{.3})\} \\ &= E(k_t - l_t)^2(N^2N_1^{-2} - \sigma_1^2(N_1 - 1)^{-1}) + E(r_t - s_t)^2(N^2N_2^{-2} - \sigma_2^2(N_2 - 1)^{-1}) \\ &\quad + 2E(k_t - l_t)(r_t - s_t)N^2N_1^{-1}N_2^{-1} - E(k_t + l_t)(NN_1^{-1} - \sigma_1^2N_1(N_1 - 1)^{-1}) \\ &\quad - E(r_t + s_t)(NN_2^{-1} - \sigma_2^2N_2(N_2 - 1)^{-1}). \end{aligned}$$

Thus our difference equations may be written

$$\left. \begin{aligned} p_{t+1} &= q_{t+1} = \frac{1}{2} + \frac{1}{2}v_t, \\ v_{t+1} &= N^{-1}(N-1)^{-1}\{\frac{1}{4}a_t(N^2 - N_1^2\sigma_1^2(N_1 - 1)^{-1}) + \frac{1}{4}b_t(N^2 - N_2^2\sigma_2^2(N_2 - 1)^{-1}) \\ &\quad - \frac{1}{4}p_t(N - N_1^2\sigma_1^2(N_1 - 1)^{-1}) - \frac{1}{4}q_t(N - N_2^2\sigma_2^2(N_2 - 1)^{-1}) \\ &\quad + \frac{1}{2}N(N-1)v_t\}, \\ a_{t+1} &= (1 - N_1^{-1})v_{t+1} + N_1^{-1}p_{t+1}, \\ b_{t+1} &= (1 - N_2^{-1})v_{t+1} + N_2^{-1}q_{t+1}. \end{aligned} \right\} \dots (16)$$

Substituting for  $q_t$ ,  $a_t$ , and  $b_t$  in the right-hand side of  $v_{t+1}$  we get finally two difference equations for the quantities  $p$  and  $v$ , the other moments being simple functions of these.

$$\begin{aligned} p_{t+1} &= \frac{1}{2} + \frac{1}{2}v_t, \\ v_{t+1} &= \frac{1}{4}N^{-1}(N-1)^{-1}\{(N^3N_1^{-1}N_2^{-1} - 2N + N_1\sigma_1^2 + N_2\sigma_2^2)p_t \\ &\quad + (4N^2 - 2N - N^3N_1^{-1}N_2^{-1} - N_1\sigma_1^2 - N_2\sigma_2^2)v_t\}. \end{aligned}$$

The solution of these equations has the form

$$\begin{aligned} p_t &= 1 - \alpha_1 \lambda_1^t - \alpha_2 \lambda_2^t, \\ v_t &= 1 - \beta_1 \lambda_1^t - \beta_2 \lambda_2^t, \end{aligned}$$

where the  $\lambda$ 's are the characteristic roots of the matrix of coefficients in the equations, and the  $\alpha$ 's and  $\beta$ 's are constants chosen to satisfy the initial conditions. The characteristic equation is

$$\begin{aligned} \lambda^2 - \lambda \frac{1}{4} N^{-1} (N-1)^{-1} (4N^2 - 2N - N^3 N_1^{-1} N_2^{-1} - N_1 \sigma_1^2 - N_2 \sigma_2^2) - \frac{1}{8} N^{-1} (N-1)^{-1} \\ \times (N^3 N_1^{-1} N_2^{-1} - 2N + N_1 \sigma_1^2 + N_2 \sigma_2^2) = 0, \end{aligned}$$

and the larger root, which governs the rate of approach to homozygosity, is approximately

$$\begin{aligned} \lambda &\doteq 1 - \frac{1}{8} N^{-2} (N^3 N_1^{-1} N_2^{-1} - 2N + N_1 \sigma_1^2 + N_2 \sigma_2^2), \\ &= 1 - \frac{1}{8} N^{-2} (N_1 P_1''(1) + N_2 P_2''(1)). \end{aligned} \quad \left. \vphantom{\lambda} \right\} \dots (17)$$

Thus we have related the rate of progress to homozygosity to the offspring distributions whose generating functions are  $P_1(z)$  and  $P_2(z)$ .

By considering the expectations of  $k_{t+1}$ ,  $l_{t+1}$ ,  $r_{t+1}$ , and  $s_{t+1}$  in terms of those at generation  $t$ , we can easily find the probabilities of ultimate absorption in the two absorbing states  $(N_1, 0, N_2, 0)$  and  $(0, N_1, 0, N_2)$ . Thus the probability of absorption in the former is the ultimate value of  $N_1^{-1} E(k_{t+1})$  which may be denoted  $N_1^{-1} E(k_\infty)$ . We know already that  $N_1^{-1} E(k_\infty + l_\infty) = 1$ . It may be shown, by successively taking conditional expectations as before, that

$$E\{N_1^{-1}(k_\infty - l_\infty) + N_2^{-1}(r_\infty - s_\infty)\} = N_1^{-1}(k_0 - l_0) + N_2^{-1}(r_0 - s_0),$$

where the subscript zero denotes the initial value, so we have

$$N_1^{-1} E(k_\infty - l_\infty) = \frac{1}{2} N_1^{-1} (k_0 - l_0) + \frac{1}{2} N_2^{-1} (r_0 - s_0).$$

Therefore the probability of fixation of the  $a$  gene is

$$N_1^{-1} E(k_\infty) = \frac{1}{2} + \frac{1}{4} N_1^{-1} (k_0 - l_0) + \frac{1}{4} N_2^{-1} (r_0 - s_0),$$

which is the mean of the relative frequencies of the  $a$  gene in the males and females of the initial population. We notice in particular that it is only equal to the overall initial frequency of  $a$  when  $N_1 = N_2$ .

In the above model each offspring is the result of an independent random mating between the adults of the previous generation, subject to the restriction imposed by the probability distribution of the numbers of offspring. Consider now a model in which "marriage" occurs, i.e. permanent matings between pairs of the opposite sex. We now take  $N_1 = N_2 = \frac{1}{2}N$  so that each individual has one mate. We must then have  $\sigma_1^2 = \sigma_2^2 (= \sigma^2 \text{ say})$  since the male and female parents of a family must have the same number of offspring. To construct such a model we proceed as before but introduce an additional set of variates. We suppose that  $y_{ij}$  are the numbers of permanent matings between parents of genotypes  $(i, j)$  in accordance with the following scheme:



variables  $(k_i, r_i, s_i)$  and the two possible absorbing states are  $(N_1, N_2, 0)$  and  $(0, 0, N_2)$ . It can be shown that the rate of approach to one or other of these states is

$$\lambda \doteq 1 - \frac{1}{3}N^{-2}(N_1P_1''(1) + 2N_2P_2''(1)), \quad \dots\dots\dots(18)$$

where the symbols have the same meanings as before. The probability of ultimate fixation of the  $a$  genes is now

$$\frac{1}{2} + \frac{1}{6}(2k_0 - N_1)N_1^{-1} + \frac{1}{3}(r_0 - s_0)N_2^{-1}.$$

We may verify that the rates found here substantiate the equivalence between two populations of different size and offspring distribution considered in the previous section. For a population of size  $N$  and generating functions (5), the rates are

$$\lambda \doteq 1 - \frac{1}{8}N^{-1}\pi^2R''(1)$$

for autosomal genes, and

$$\lambda \doteq 1 - \frac{1}{6}N^{-1}\pi^2R''(1)$$

for sex-linked genes. Alternatively, for the population of size  $\frac{1}{2}NR'(1)$  and generating functions (6), we get

$$\lambda \doteq 1 - \frac{1}{4}N^{-1}[R'(1)]^{-1}\pi R''(1)$$

for autosomal genes, and

$$\lambda \doteq 1 - \frac{1}{3}N^{-1}[R'(1)]^{-1}\pi R''(1)$$

for sex-linked genes. Because these pairs of rates may apply to the same population under two methods of enumeration, they must be identical. For this to be true, we need  $R'(1) = 2\pi^{-1}$ , which is just the condition required for a population to have constant size.

The above diploid models assume non-overlapping generations. As particular cases, consider first Wright's model having generating functions (7). For these, the second derivatives are

$$\begin{aligned} P_1''(1) &= N(N-1)N_1^{-2}, \\ P_2''(1) &= N(N-1)N_2^{-2}, \end{aligned}$$

and so the rate of approach to homozygosity is approximately

$$\lambda \doteq 1 - N/8N_1N_2$$

for autosomal genes, and

$$\lambda \doteq 1 - (N_2 + 2N_1)/9N_1N_2$$

for sex-linked genes. These two results had been found previously by Wright (1931, 1933) by a different method.

For a model in which all males have exactly  $NN_1^{-1}$  offspring, all females exactly  $NN_2^{-1}$ , the second derivatives are, from (8),

$$\begin{aligned} P_1''(1) &= NN_1^{-1}(NN_1^{-1} - 1), \\ P_2''(1) &= NN_2^{-1}(NN_2^{-1} - 1), \end{aligned}$$

so that the rate of approach to homozygosity is

$$\begin{aligned} \lambda &\doteq 1 - \frac{1}{8}N^{-1}(N^2N_1^{-1}N_2^{-1} - 2), \\ \lambda &\doteq 1 - \frac{1}{9}N^{-1}\{N(N_2 + 2N_1)N_1^{-1}N_2^{-1} - 3\}, \end{aligned}$$

for autosomal and sex-linked genes, respectively.

In the preceding section, two other models were considered, and their generating functions obtained. Although both models were for populations with overlapping generations, it is interesting to see the effect this has by comparing the results obtained in previous papers for these models with those derived here for non-overlapping generations. Moran (1958) found for his overlapping generation model that the rate of approach to homozygosity for autosomal genes was

$$\lambda \doteq 1 - N/4N_1N_2.$$

Compare this with the result for a population having the same generating functions (9) but having non-overlapping generations. From (17) we get

$$\begin{aligned}\lambda &\doteq 1 - \frac{1}{8}N^{-2}(N_12N(N-1)N_1^{-2} + N_22N(N-1)N_2^{-2}), \\ &\doteq 1 - N/4N_1N_2,\end{aligned}$$

which is identical with the previous expression. Thus it would seem that a population with overlapping generations would have asymptotically the same rate of progress to homozygosity as one with non-overlapping generations and the same offspring distributions. That this, however, cannot be true in general is shown by the theory of the other overlapping generation model II. In this case, Watterson (unpublished data) found the rate of approach to homozygosity for autosomal genes as

$$\lambda \doteq 1 - N/8N_1N_2$$

for random mating, whereas for a non-overlapping population with the same generating functions (10), the present theory gives the rate as

$$\lambda \doteq 1 - (N_1^2 + N_2^2)/4NN_1N_2.$$

This root is not identical with the previous one except when  $N_1 = N_2$ . From this we see that (17) cannot always be extended to overlapping generation models and the determination of the conditions under which this is possible must await the development of a general theory to cover the latter. It appears that this will be a difficult task, for the state of such a general population needs the specification of all individuals ages as well as their genotypes, and such random processes present theoretical difficulties not yet overcome.

#### IV. REFERENCES

- FISHER, R. A. (1939).—Stage of development as a factor influencing the variance in the number of offspring, frequency of mutants and related quantities. *Ann. Eugen., Lond.* **9**: 406–8.
- HALDANE, J. B. S. (1939).—The equilibrium between mutation and random extinction. *Ann. Eugen., Lond.* **9**: 400–5.
- LI, C. C. (1955).—“Population Genetics.” (University of Chicago Press.)
- MORAN, P. A. P. (1958).—Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**: 60–71.
- WILKS, S. S. (1946).—“Mathematical Statistics.” (Princeton University Press.)
- WRIGHT, S. (1931).—Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- WRIGHT, S. (1933).—Inbreeding and homozygosis. *Proc. Nat. Acad. Sci., Wash.* **19**: 411–19.