SIMULATION OF MODELS INVOLVING MIXED SELFING AND RANDOM MATING

I. STOCHASTIC VARIATION IN OUTCROSSING AND SELECTION PARAMETERS

S. K. JAIN and D. R. MARSHALL

Department of Agronomy, University of California, Davis

Received 15.xii.67.

1. INTRODUCTION

SINCE the publication of Darwin's *The Effects of Cross- and Self-fertilization in the Vegetable Kingdom* (1876), the role of random outcrossing has been widely recognised in relation to the breeding structure and evolution of species in the extreme classes, namely, obligate outbreeding versus almost complete or exclusive selfing (Darlington, 1958; Stebbins, 1950). The evolutionary dynamics of a mixed selfing and random mating system has only recently been theoretically and experimentally investigated. These studies showed that the outcrossing parameter must be considered conjointly with selection and other features of the genetic system in order to fully analyse its evolutionary properties (review by Allard, Jain and Workman, 1967). Hayman (1953) had shown that a single diallelic locus could maintain polymorphism in a large population through selection favouring heterozygotes. Further theoretical work using a two-locus, diallelic model illustrated the role of linkage and gene interaction in determining stable equilibria (Jain and Allard, 1966) and similar attributes of a two-locus model optimising selection (Jain and Allard, 1965).

In the series of papers, the theory of mixed selfing and random mating system will be further developed essentially by computer simulation of more general models such as those involving stochastic processes, multiple alleles and selection during gametic as well as zygotic stages. A large body of experimental evidence shows that both the outcrossing rate and the selective values vary markedly between populations in space and time as well as between families within the same population. Yet, our understanding of the effects of such stochastic variation in these parameters is extremely limited. The stochastic equivalent of Hayman's model can be readily investigated by Monte Carlo simulation using randomly varying parameters with some prespecified density function. Analogous to Bateman's (1947, 1948) findings on the leptokurtic form of distribution of certain migration parameters, a review of literature on the variation in outcrossing rate (t) indicated that for α species with low mean outcrossing, the distribution of t -values among families and between generations involved both over-dispersion (high variance/mean ratio) and kurtosis. In this paper we shall briefly review the experimental evidence related to the form of distribution of outcrossing rate and selective values and then examine by computer simulation the effect of random fluctuations from generation to generation on the properties of genotypic equilibria. In particular, we shall compare

the regions allowing stable equilibria under models involving heterozygote advantage and low outcrossing rates.

2. REVIEW OF LITERATURE

(a) *Variation in t among families*

In such multiflowered species as tomato, egg plant and millets there are several reports of outcrossing rates measured by growing progenies of individual plants so that the distribution of rate of outcrossing (t) can be determined within a population. In cases of genetically homogeneous populations this simply provides a measure of environmental variation. Bateman (*loc. cit.*) discussed various factors that might be involved in variation under both insect- and wind-pollination conditions. One might expect under certain assumptions about the reproductive biology that on the basis of binomial law the plant-to-plant variation in the probability of outcrosses would follow nearly a normal distribution for mean values in the range, say 0.1 to 0.9, but as t approached 0 or 1, perhaps a Poisson distribution is obtained. In the following sections we shall describe the observed distribution in terms of deviations from these expected forms using Fisher's k -statistics as measures of dispersion, kurtosis and skewness. Table 1 gives the expected values of $k_1(\mu)$, $k_2(\sigma^2)$, $g_1 = k_3/(k_2)^{3/2}$ and $g_2 = k_4/k_2^2$, for some well-known density functions which are also used in simulation.

TABLE 1

Expected values of mean, variance, g_1 , and g_2 for the normal, Poisson rectangular and exponential distributions

Distribution	k_1	k_2	g_1	g_2
Normal	μ	σ^2	0	0
Poisson	m	m	$m^{-1/2}$	m^{-1}
Rectangular	$c/2$	$c^2/12$	0	0
Exponential	$1/x$	$1/x^2$	$2/x^3$	$9/x^4$

(i) *Cotton*

Plant-to-plant variation in outcrossing has been extensively studied in cotton. From the bulk of experimental data available, six examples with the mean rate of outcrossing ranging from 27.4 to 59.4 per cent. were studied in detail (Simpson, 1954; Innes, 1961; and others). In five out of the six cases studied, t was approximately normally distributed in agreement with the theoretical expectations (*e.g.* fig. 1(a)). However, in one case illustrated in fig. 1 (b), the distribution deviated significantly from normality due to both overdispersion and skewness.

(ii) *Egg plant*

An example of the distribution of outcrossing rate in the egg plant (Kakizaki, 1924) is shown in fig. 2. Since the mean rate of outcrossing (t) is less than 10 per cent., a parametric χ^2 test ($\chi^2_{[n]} = ns^2/m$, where m , s^2 are the sample mean and variance respectively and $n = N-1$, where N is the sample size) was used to test the fit to Poisson distribution. In this case $\chi^2_{[62]} = 531.4$, $P < 0.001$, indicating that the variance is significantly greater than the mean. The estimates of g_1 and g_2 are greater than expected values.

FIG. 1 (a).—Distribution of outcrossing rate (among families) in cotton at Shambat, Sudan (Innes, 1961). No. of families = 134, $k_1 = 31.95$; $k_2 = 34.51$; $g_1 = -0.27$; $g_2 = 0.45$. No significant departure from normal curve ($\chi^2_{(9)} = 8.20$, $0.50 < P < 0.70$).

FIG. 1 (b).—Distribution of outcrossing rate (among families) in cotton at Athens, Georgia (Simpson, 1954). No. of families = 274, $k_1 = 29.87$; $k_2 = 263.58$; $g_1 = 0.89$, $g_2 = 0.88$. Significant deviation from normality.

A number of distributions have been devised in which the variance is significantly larger than the mean and overdispersion is assumed to arise from contagion (Anscombe, 1950). Since the negative binomial is the most widely applicable of the unimodal "contagious" distributions, an attempt was made to fit it to the present data. The negative binomial, given by $(q-p)^k$, is completely described by two parameters, the mean μ and the exponent k ; where $p = \mu/k$ and $q = 1+p$. For the egg plant data, the joint estimates of μ and k were 7.44 and 1.45. The χ^2 test for goodness of fit which yielded $\chi^2_{[4]} = 4.68$, $0.30 < P < 0.50$, indicated that the observed data fit the binomial distribution.

(iii) *Tomato*

The distribution of outcrossing in tomato on an individual fruit basis described by Kakizaki (1920) is shown in fig. 3. Here, the parametric χ^2 test for the poisson distribution yielded $\chi^2_{[107]} = 1742$, $P < 0.001$, indicating, as before with egg plant data, that the distribution shows considerable overdispersion. Estimation of parameters of the negative binomial distribution gave $\bar{t} = 1.89$, $\hat{k} = 0.176$. The χ^2 test for goodness of fit ($\chi^2_{[11]} = 6.20$, $0.80 < P < 0.90$) indicated good agreement between the observed and expected frequencies under the negative binomial.

(iv) *Foxtail millet*

The data of Macvicar and Parnell (1941) on the plant to plant variation in outcrossing are illustrated in fig. 4. Again the parametric χ^2 test for the Poisson distribution indicated that the variance was significantly greater than the mean ($\chi^2_{[63]} = 120.6$, $P < 0.001$). Moreover, these data also appeared to fit the negative binomial distribution with $\bar{t} = 0.73$, $k = 0.564$. The χ^2 test for goodness of fit yielded $\chi^2_{[3]} = 5.46$, $0.20 > P > 0.10$.

A further nine examples, representing five species, of plant-to-plant variations were examined in detail. These data are summarised in table 2. It will be noted that in only one case (blackberry) is the mean equal to, or less than, the variance. In the other eight cases the parametric χ^2 test for the Poisson distribution indicated that the variance was significantly greater than the mean. Consequently, all the examples considered are consistent in indicating that in predominantly selfing species the outcrossing parameter is usually not normally distributed. Note that from a similar literature survey it appears that in self-incompatible species with t nearly 1.0, t again approaches Poisson or an overdispersed form.

(b) *Temporal and spatial variation in outcrossing rate*

Another important aspect of variation in the outcrossing rate is the temporal variation in the mean rate of outcrossing of a population. A large number of environmental and biotic factors are known to affect the mean rate of outcrossing of a population (Simpson, 1954; Fryxell, 1957). To adequately define an empirical distribution, a large number of estimates of outcrossing are required for the same, or at least a genetically similar, population grown in different environments. Unfortunately, such extensive data are available for only a few species. Three sets of data seemed to be suitable for the present purpose (fig. 5). These examples show a considerable

FIG. 2.—Distribution of outcrossing rate among families in egg plant (data of kakizaki). No. of families = 63, $k_1 = 7.44$, $k_2 = 62.69$, $g_1 = 2.88$, $g_2 = 9.74$. Fits negative binomial with exponent $k = 1.45$ ($\chi^2_{(4)} = 4.68$).

FIG. 3.—Distribution of outcrossing rate among families in tomato (data of kakizaki). No. of families = 108, $k_1 = 1.89$, $k_2 = 30.94$, $g_1 = 5.72$, $g_2 = 39.82$. Fits negative binomial with exponent $k = 0.176$ ($\chi^2_{(11)} = 6.20$).

FIG. 4.—Distribution of outcrossing among families in foxtail millet (data of MacVicar and Parnell). No. of families = 64; $k_1 = 0.73$; $k_2 = 1.40$; $g_1 = 1.68$; $g_2 = 2.11$. Fits negative binomial with $k = 0.564$ ($\chi^2_{(3)} = 5.46$).

TABLE 2

Plant-to-plant variation in outcrossing. k_1 and k_2 (Fisher's k-statistics) are unbiased estimates of the population mean and variance respectively

Species	Source	Sample size	Mean (k_1)	Var. (k_2)	Parametric χ^2 -test
Castor Bean	Domingo (1944)	7	36.1	175.8	$\chi^2_{[6]} = 25.1$ $P < 0.01$
Blackberry	Haskell (1954)	8	17.4	22.2	$\chi^2_{[7]} = 7.8$ $0.30 < P < 0.50$
Pigeon Pea	Wilsie and Takahashi (1934)	44	14.6	44.8	$\chi^2_{[43]} = 131.9$ $P < 0.001$
Safflower	Claasen (1950)	25	7.9	99.8	$\chi^2_{[24]} = 303.3$ $P < 0.001$
Sorghum var. Double Dwarf 38	Smeltzer (Unpublished)	40	11.8	78.4	$\chi^2_{[39]} = 259.0$ $P < 0.001$
var. DWD	Smeltzer (Unpublished)	38	5.2	18.3	$\chi^2_{[37]} = 130.4$ $P < 0.001$
var. Early Hegari	Smeltzer (Unpublished)	38	5.9	12.9	$\chi^2_{[37]} = 80.8$ $P < 0.001$
var. DW39	Smeltzer (Unpublished)	40	4.9	13.0	$\chi^2_{[39]} = 103.6$
var. Milo	Smeltzer (Unpublished)	36	7.7	38.1	$\chi^2_{[35]} = 173.7$ $P < 0.001$

amount of overdispersion. Indeed, the parametric χ^2 -test indicated that in all cases the variance is significantly greater than the mean. It will be noted, however, that the form of the distribution varies among the cases studied. For example, in cotton the distribution appears to be platykurtic while in lima bean the distribution is markedly leptokurtic.

FIG. 5.—Distribution of outcrossing in cotton and lima bean over different years (data of Simpson, 1954; Allard, 1954; Harding and Tucker, 1964).

TABLE 3

Equilibria under deterministic model

<i>t</i> = 0.20				<i>t</i> = 0.05			
<i>x</i>	<i>y</i>	<i>p</i> _∞	<i>F</i> _∞	<i>x</i>	<i>y</i>	<i>p</i> _∞	<i>F</i> _∞
0.75	0.56	1.0	1.0	0.75	0.71	1.0	1.0
0.75	0.57	0.9866	0.4309	0.75	0.72	0.8903	0.8206
0.75	0.58	0.9682	0.4340	0.75	0.73	0.7678	0.8232
0.75	0.59	0.9491	0.4373	0.75	0.74	0.6379	0.8258
0.75	0.60	0.9292	0.4406	0.75	0.75	0.500	0.8284
0.75	0.75	0.500	0.5000	—	—	—	—

For $x = 0.75$, the values of y_e are 0.5625 ($t = 0.20$) and 0.7105 ($t = 0.05$).

(c) *Variation in selective values*

Here again, data on many consecutive generations are needed in order to obtain an empirical distribution of the selective values. In *Drosophila* and several other random mating species, such estimates are available showing a marked variation in the selective values between generations. In predominantly selfing species, the most extensive data are those of Allard and Workman (1963), who found that the selective values fluctuated widely between years more or less cyclically with near-normal distribution (*c.f.* Author's table 3). Outcrossing rate (t) was assumed to be constant in

estimating selective values (x, y) so that variation in t and (x, y) were confounded. However, as stated by authors, the observed fluctuations in t alone were not large enough to yield these variations. In other studies (Jain and Allard, 1960; Allard, Jain and Workman, 1967) there is some evidence for variation in selective values although these studies provided insufficient sets of estimates to determine the form of distribution. In this paper we shall study random variations in selective values using only the normal distribution.

3. CONCLUSIONS FROM EXPERIMENTAL DATA

From the review of experimental work on the variation in outcrossing rate there appears to be significant plant-to-plant variation following a distribution function, more or less as expected, governed by the mean value. Overdispersion and leptokurtosis are frequently involved for low values of t . In genetically variable materials involving some form of genotype control of the mating system, such variation might have a selective pressure in favour of the families with higher self-fertilisation unless counteracted by some other form of selection in the more outbred families (*cf.* Drayner, 1959). In finitely small populations, such variation can also result in the fluctuation of t between generations and would require a finite population model to be analysed. Our primary concern here is the effect of random fluctuations between generations so that simulation involves the use of a random variable to determine the value of t in each successive generation. Similarly, input selective values are taken using appropriate random variables with the successive values varying between generations.

4. THE MODEL AND SIMULATION PROCEDURE

Consider a single locus with two alleles, A_1 and A_2 , and let (P, R, Q) and (P', R', Q') denote the genotypic proportions of (A_1A_1, A_1A_2, A_2A_2) in generations n and $(n+1)$. Assume that each individual has probability of selfing (s) and of random outcrossing ($t = 1 - s$), and let the relative viabilities of the three genotypes be in the ratio $(x:1:y)$. Under a model in which the census for genotypic frequencies is assumed to occur just before mating, the recursion relations are:

$$P'\alpha x\{s(P+\frac{1}{4}R)+t(P+\frac{1}{2}R)^2\}$$

$$Q'\alpha y\{s(Q+\frac{1}{4}R)+t(Q+\frac{1}{2}R)^2\}$$

$$R'\alpha \frac{1}{2}sR+2t(P+\frac{1}{2}R)(Q+\frac{1}{2}R)$$

If we denote $p = P + \frac{1}{2}R$, $q = Q + \frac{1}{2}R$, and fixation index $F = 1 - R/2pq$ at equilibrium, where $\Delta P = 0$, and $\Delta F = 0$, the solutions for p_∞ and F_∞ describe the equilibrium (Workman and Jain, 1966).

Hayman's phase diagrams (fig. 6) describe the conditions, in terms of the parameters (t, x, y) , which allow stable nontrivial equilibria (regions C, D), the fixation of allele A_1 (region A) or the fixation of allele A_2 (region B). Lewontin (1958) obtained the condition for the maintenance of a stable

nontrivial equilibria in terms of (F, x, y) , viz. $\frac{1-F}{F} > \frac{x}{1-x} + \frac{y}{1-y}$, where F is given by the equation

$$2F^2(1-x)(1-y) + F\{(x+y-2) - t(x+y-2xy)\} + (x+y-2xy)(1-t) = 0.$$

Regions C and D are defined in terms of the level of heterozygosis as com-

FIG. 6.—Phase diagrams showing regions A-D for $t = 0.05$ and $t = 0.20$ under Hayman's model and two areas around the boundary A-C investigated in our stochastic model.

compared with the expectation from Hardy-Weinberg rule, $\hat{R} = 2pq$ (*i.e.* $F = 0$), such that $F < 0$ in region D and $F > 0$ in region C. In this paper we are interested in the critical bounds on (x, y) only in terms of fixation *versus* polymorphic regions so that boundary between C and D will be ignored. Two main features of phase diagrams should be pointed out here; namely (1) as the level of outcrossing decreases, regions C, D become smaller so that maintenance of polymorphisms would require greater amounts of heterozygote advantage, and (2) the homozygote fitnesses need to be increasingly closer ($x - y \rightarrow 0$) which thus becomes highly restrictive condition under a predominantly selfing system.

It might be expected that introducing random fluctuations of t or (x, y) might change these conditions for stable equilibria and the relative position of the boundaries. Without loss of generality, we may choose several points on the boundaries AC (or BC, by virtue of symmetry) to test this expectation. Table 3 gives equilibria for a range of values of (x, y) corresponding to the regions under test near boundary AC in fig. 6. The critical bound on $y(y_c)$ for any given t and x can be obtained from solutions given above. Thus,

theoretically, the effects of stochastic variations on these bounds and y_c can be determined by finding equilibrium values of p and F . Using random variables for varying t or (x, y) between generations allows the recursions to be represented by a Markov chain in which the equilibria have to be defined in terms of a steady state. The probability distribution of allelic frequency is expected to approach a limiting distribution independent of initial conditions, at least to the second order, and thus it attains an unchanging mean and variance in relation to the time axis (see Bartlett, 1955; Papoulis, 1965; for a discussion of stationary processes). As pointed out by Shreider (1966), the choice of time for describing the probability distribution of repeated trials as well as the required degree of accuracy in decisions about fixed *versus* unfixed state have to remain arbitrarily within certain prespecified limits. The cumulative round-off errors in any such simulation work as well as the extreme slowness of change around the equilibrium region pose further problems. These and other problems associated with numerical solutions of stochastic processes have been discussed in more detail by Leslie and Gower (1960) and Bartlett *et al.* (1960).

Within the region where none of the alleles attain fixation, we are interested in the ergodicity of the mean as well as the distribution of the replicate runs for any given set of input parameters so that the process can be described by the time averages approaching certain expected values. Therefore, the properties of stochastic models are described here in terms of (1) the stable distribution of replicate runs at some arbitrary time t , (2) the changes in mean allelic frequency ($\bar{p}$) over time and approach to some steady value from two or more sets of initial frequencies, and (3) the determination of critical bound on $y(y_c)$ either from the values of $\bar{p}$, or on the basis of the proportion of replicates fixed. In all cases, only a sample of results are presented to illustrate these aspects of the process.

The density function for t was chosen to be one of the following forms: (i) normal with input $\bar{t}$, σ_t , (ii) rectangular, $0 \leq \bar{t} \leq 2\bar{t}$, (iii) negative binomial with input $\bar{t}$ and k , and (iv) exponential with input $\bar{t}$. The selective values were simulated using a normal distribution with input $(\bar{x}, \bar{y})$ and σ_x, σ_y . Schreider (1966) described the methods of generating a set of random numbers (S_i) with specified density function, say $f(x)$, given a set of pseudo-random numbers R_i , uniformly distributed over the interval (0, 1). For the exponential distribution, a method described by Clark and Holz (1960) was used, and for the normal distribution, we used a programme kindly provided by Dr D. G. Burnell. Other distribution functions, including a more general representation by Gamma function, $f(x) = \frac{1}{\beta(\alpha+1) \Gamma(\alpha+1)} x^\alpha e^{-x/\beta}$, $0 < x < \infty$, where (a) $\alpha = 1$, $\beta = 0$, gives exponential, (b) $\alpha = -\frac{1}{2}$, $\beta = 1$, gives leptokurtic, and (c) $\alpha = 10$, $\beta = 0.2$ gives near-normal distribution, can be generated following Schreider's method based on the piecewise approximation of $f(x)$. The output distribution was checked for several parametric values by testing the observed values of μ , σ^2 , g_1 , and g_2 against their expectations (cf. table 1).

5. STOCHASTIC VARIATION IN OUTCROSSING PARAMETER (t)

Taking selective values (x, y) constant between generations and in the two regions around $y_c = 0.5625$ for $t = 0.20$ and $y_c = 0.7105$ for $t = 0.05$

(fig. 6), the value of t was varied using random variables having one of the four density functions. In the majority of the cases studied, 20 replicate runs were obtained for two initial allelic frequency sets $p = 0.50$ and $p = 0.99$ over 1000 generations. In some critical cases, 100 replicates were run for as many as 3000 generations. For the normal distribution, two different values of input variance (σ^2_t) were used in order to obtain a rather high-peaked distribution (coefficient of variation, C.V. = 0.05) in contrast to

TABLE 4

Values of $\bar{p}$ and $\bar{F}$ (based on rep means), and proportion of replicate runs fixed in 1000th generation for stochastic t ($x = 0.75$, $\sigma_x = \sigma_y = 0$)

		Normal distribution (a) C.V. = 1/20			Normal distribution (b) C.V. = 1/3			Rectangular		
t	y	$\bar{p}$	$\bar{F}$	% reps fixed	$\bar{p}$	$\bar{F}$	% reps fixed	$\bar{p}$	$\bar{F}$	% reps fixed
0.20	0.550	1.0	1.0	100	1.0	1.0	100	1.0	1.0	100
	0.560	0.9987+	0.4246	0	0.9988+	0.4350	10	0.9993+	0.4449	40
	0.565	0.9975+	0.4285	—	0.9958+	0.4353	0	0.9984+	0.4305	0
	0.570	0.9861 ^e	0.4318	0	0.9890 ^e	0.4279	0	0.9905 ^e	0.4475	0
	0.580	—	—	—	0.9700 ^e	0.4364	0	0.9775 ^e	0.4320	0
		Exponential			Negative binomial					
t	y	$\bar{p}$	$\bar{F}$	% reps fixed	$\bar{p}$	$\bar{F}$	% reps fixed			
0.20	0.560	1.0	1.0	100	1.0	1.0	100			
	0.570	0.9982+	0.4521	40	0.9986+	0.4609	2			
	0.575	—	—	—	0.9932+	0.4608	0			
	0.580	0.9901 ^e	0.4573	0	0.9842 ^e	0.4598	0			
	0.590	0.9786	0.4794	0	0.9670	0.4488	0			
0.05	0.700	1.0	1.0	100	1.0	1.0	100			
	0.710	0.9987+	0.8221	0	0.9951+	0.8405	0			
	0.715	—	—	—	0.9546+	0.8333	0			
	0.720	0.8946 ^e	0.8196	0	0.9030 ^e	0.8230	0			
	0.730	0.7707 ^e	0.8263	0	0.7711	0.8148	0			

+ Values of p and F increasing steadily and therefore given for 1000th generation.

^e Steady state distribution by nearly 1000th generation; values given for rep means over 900 to 1000th generation interval.

one with C.V. = 0.333. In all runs, the quantities of interest, p and F for individual replicate as well as mean $\bar{p}$ and $\bar{F}$ over reps, were output at regular intervals of 25 or 100 generations. Table 4 gives a sample of data on the output values in the 1000th generation. At this time the distribution of various gene frequency classes seemed more or less stationary except in the cases very near the y_c values due to slow change (Δp less than 10^{-6} per generation). The proportion of replicate runs with $p = 1$ (actually, with 8-digit accuracy, p greater than $1 \cdot 10^{-8}$) are also given in the last column which is often based on 20 replicates and therefore, is merely an approximate value. Using y at intervals of 0.01, the value of y_c was found to be within the same range as given by the corresponding deterministic cases for both normal and rectangular functions. From the values of $\bar{p}$ and $\bar{F}$, it appears

in general that y_e becomes slightly greater, thus narrowing the region of stable equilibria (C, D) in the case of the stochastic models.

This trend is further seen in the cases of both exponential and negative binomial distribution having relatively greater C.V. and kurtosis. The values of y_e are shifted upward narrowing the C region. For $t = 0.20$, y_e changed from the expected 0.5625 to somewhere between 0.57 and 0.575. It can not be overemphasised that exact determination of y_e with any greater accuracy using Monte Carlo simulation would entail a great deal more digital accuracy, more computer time or perhaps, as pointed out by Schreider (1966), some new developments in terms of microprogramming, etc. Figs. 7*a-c* illustrate the distribution of 100 reps at time $n = 100, 500, 1000$ $n = 2000$ for $y = 0.71, 0.715$ and 0.720 , with $t = 0.05$ varied as a negative binomial variable. For $y = 0.715$, the distribution of the replicates approaches a steady state (with no case of fixation) by the 1000th generation and this value appears to allow a non-trivial stationary gene frequency distribution. On the other hand, for $y = 0.71$ the value of $\bar{p}$ approached fixation in a majority of the replicate runs. This point is also emphasised by the data given in table 4. From both the nature of the distribution and the proportion of replicate runs fixed, it is possible to locate approximately the critical value of y , which lies on the boundary between region A ($\bar{p} = 1$) and region C ($\bar{p} < 1$). The mean allelic frequency ($\bar{p}$) over reps is shown in fig. 8 for various values of y and initial $p = 0.5$. Further, note from fig. 9 that the two initial sets converged fairly rapidly toward such equilibria. Also, it will be noted that fluctuations in replicate mean ($\bar{p}$) are relatively small. Thus, y_e shows a small change upward in a model involving stochastic variation in t . The boundaries were further checked in a similar manner for several other points using values of x other than 0.75. For $x = 0.85$, for instance, $y_e = 0.759$ for $t = 0.2$ and $y_e = 0.831$ for $t = 0.05$ in the deterministic case. The corresponding stochastic models did not appear to significantly alter these critical bounds.

TABLE 5
Values of $\bar{p}$, $\bar{\sigma}_p^2$, $\bar{F}$ and proportion of replicate runs fixed in 1000th generation
(based on rep means) for stochastic (x, y)
(a) Normal distribution $\sigma_x = \sigma_y = 0.2$

t	x	y	$\text{Cov}_{xy} = 0$				$\text{Cov}_{xy} = 1.0$			
			$\bar{p}$	$\bar{\sigma}_p^2$	$\bar{F}$	% reps fixed	$\bar{p}$	$\bar{\sigma}_p^2$	$\bar{F}$	% reps fixed
0.20	0.75	0.510	0.9860 ⁺	—	—	5	0.9908 ⁺	—	—	5
		0.515	0.9849 ⁺	—	—	0	—	—	—	—
		0.520	0.9715 ^e	0.00295	0.3617	0	0.9823 ^e	0.00354	0.3738	0
		0.530	0.9673 ^e	0.00334	0.3788	0	—	—	—	—
		0.540	0.9527 ^e	0.00548	0.3811	0	—	—	—	—
		0.550	0.9490 ^e	0.00510	0.3812	0	0.9599 ^e	0.00186	0.3632	0
		0.560	0.9323 ^e	0.00623	0.3837	0	—	—	—	—
		0.570	0.9081 ^e	0.00891	0.3921	0	0.9170 ^e	0.00150	0.3986	0
		0.580	0.9071 ^e	0.00922	0.3972	0	—	—	—	—

⁺ and ^e explained in the footnotes to table 4.

6. STOCHASTIC VARIATION IN SELECTIVE VALUES (x, y)

For the same sets of (x, y) as used earlier, and following the same general procedure, the values of x and y were determined in each generation using

FIG. 7.—Histograms showing the distribution of allelic frequency (p) among replicate runs for input $\bar{t} = 0.05$ (with negative binomial distribution) and $x = 0.75$, $y = 0.710$ (a, left) 0.715 (b, middle) and 0.720 (c, right). Note the estimates of $\bar{p}$ and σ_p^2 in the n th generation.

FIG. 8.—Changes in mean allelic frequency ($\bar{p}$) for the same model as used in fig. 7.

random variables having normal distribution function and with two values of input variance (σ_x , $\sigma_x = 0.1$ and 0.2). In one set of runs, we assumed $\text{cov}_{xy} = 0$, that is, x and y are independently distributed whereas in another parallel set of runs, $\text{cov}_{xy} = 1$ was obtained by using the same value of random number in any generation for varying both x and y . Allard and Workman (1963) presented evidence for both cases; in one population x and y varied more or less independently between years, but in another population there was a significant covariance between x and y . Table 6 gives a sample of the replicate means and the proportion of replicate runs fixed for various sets of input values. Again, from the values of $\bar{p}$ and the proportion of replicate runs fixed, y_c could be approximately estimated. In this case, the estimated values of y_c were significantly less than expected under the deterministic model, thus extending the region C for stable, nontrivial equilibria. In case of $t = 0.20$ and $\sigma_x = \sigma_y = 0.2$, for instance, y_c changed from 0.5625 to somewhere between 0.51 and 0.52 . A similar, but relatively smaller, change was obtained with a smaller variance in selective values ($\sigma_x, \sigma_y = 0.1$) or in the case of $\text{cov}_{xy} = 1$.

Figs. 10a-c show the distribution of replicates for $y_c = 0.51, 0.515$ and 0.52 at time $n = 1000$ and 2000 . The distribution becomes more or less stationary by the 1000th generation, although as one approaches y_c the time involved becomes increasingly larger. Also, the mean passage time is so large in the subterminal frequency classes that the frequency of replicate runs with p greater than 0.99 becomes disproportionately high. Fig. 11 further illustrates the changes in mean p over replicate runs for $\sigma_x, \sigma_y = 0.2$. In all these cases $\bar{p}$ from two initial sets converged fairly well toward an average stationary value. Further checks on stability were made for some

FIG. 9.—Changes in mean allelic frequency ($\bar{p}$) for the same model as used in fig. 7, to show the convergence from two initial values of p (I, 0.50; II, 0.99).

FIG. 10.—Histograms showing the distribution of allelic frequency (p) among replicate runs for input $t = 0.20$, $\bar{x} = 0.75$, $\bar{y} = 0.510$ (left), 0.515 (middle), 0.520 (right) and $\sigma_x = \sigma_y = 0.2$, $\text{cov}_{xy} = 0$. Note the estimates of p and σ_p^2 in the n th generation.

sets of parameters by run lengths of up to 10,000 generations as well as by disturbing the final steady state distribution and letting it return to the original form. Again, although y_c can only be approximately determined by this method, there was clearly a consistent significant decrease for a wide range of input parameters.

TABLE 6
Some examples of equilibria with $x = y$

t	distrib.	$x = y$	$\sigma_x = \sigma_y$	$\bar{p}$	σ_p	$\bar{F}$	σ_F^2	Exp. F
(a) Stochastic t								
0.20	exponential	1.0	0	0.50	0	0.6662	0.03248	0.6667
0.20	exponential	0.75	0	0.50	0	0.5241	0.04104	0.5000
0.05	exponential	1.0	0	0.50	0	0.9044	0.00264	0.9048
0.05	exponential	0.75	0	0.50	0	0.8224	0.00756	0.8284
0.05	neg. bin.	0.75	0	0.50	0	0.8199	0.01543	0.8284
(b) Stochastic (x, y)								
0.20	$\sigma_t = 0$	1.0	normal	0.5006	0.15120	0.6484	0.00273	0.6667
			$\sigma_x = \sigma_y = 0.2$					
0.20	$\sigma_t = 0$	0.75	$\sigma_x = \sigma_y = 0.2$	0.5002	0.03090	0.4621	0.00633	0.5000
0.20	$\sigma_t = 0$	1.0	$\sigma_x = \sigma_y = 0.1$	0.5004	0.06621	0.6588	0.00109	0.6667
0.20	$\sigma_t = 0$	0.75	$\sigma_x = \sigma_y = 0.1$	0.5009	0.01088	0.4897	0.00148	0.5000

FIG. 11.—Changes in the mean allelic frequency ($\bar{p}$) for the same model as used in fig. 10.

7. DISCUSSION

The effect of stochastic variation in the outcrossing rate and selective values was determined primarily in terms of the conditions for maintaining

polymorphisms under selection favouring heterozygotes. The interrelationships among the parameters of the mating system and selection, as first shown by Hayman (1953), are critical in terms of the amount of heterozygote advantage and the relative fitnesses of two homozygotes. These conditions are quite restrictive for low levels of outcrossing. Stochastic variation in the rate of outcrossing (t) caused a very small change in bounds on the selective values of homozygotes (x, y) in an upward direction, the magnitude depending upon the form of density function. On the other hand, the stochastic fluctuations in (x, y) between generations markedly increased the region of stable equilibria. Thus, under such a process the homozygote fitnesses can depart from equality farther than in the equivalent deterministic case, especially with low t values. The amount of genetic variation at equilibria in the newly generated region of course varies with the actual values of p and F , such that it increases proportionately as one moves away from y_c , more or less in the same way as seen in the deterministic case (table 3).

Thus, both the direction and magnitude of change in bounds are different for random variation in t versus (x, y). The explanation may reside in the fact that variation in t changes the amount of effective heterozygote advantage in terms of variance in F but by virtue of symmetry leaves allelic frequency unchanged; variations in x, y result in variance in both p and F since both heterozygosis and allelic frequency are affected. Since F approaches 1 at a slower rate than p near the critical bounds, the fixation process might be largely governed by the trends of changing p and σ_p^2 rather than F and σ_F^2 both in terms of attaining the steady state distribution or quasi-fixation. From runs with no selection ($x = y = 1$) or symmetrical selection ($x = y < 1$), it is shown in table 6 that variations in t result in effectively the same or even larger F -value (greater inbreeding) than expected from deterministic solutions whereas stochastic x, y tend to yield effectively less inbreeding, as measured by the level of heterozygosis. One would therefore expect the changes in critical regions to occur in opposite directions as observed in our study.

These findings on the equilibrium properties of the stochastic mixed mating model are directly relevant to certain aspects of our experimental work on predominantly self-fertilising populations. Evidence for heterozygote advantage has now been reported in several species (Allard, Jain and Workman, 1967). Selective values and t are all likely to vary between different seasons and stages of population history; further work is needed to determine their empirical distribution over time as well as variation between families of individual genotypes. Clearly, the estimation of parameters of outcrossing and selection should take this variation into account both in terms of the estimates of genetic variation as well as prediction of equilibria. Ignoring fluctuations in t would tend to underestimate (x, y), *i.e.* overestimate overdominance, and the reverse would hold for the case in which variations in selective values are ignored. Further, it should be emphasised that the mechanisms for maintaining variability in predominantly selfing species would not be helped except in relation to certain unique events, by occasional outbursts of outcrossing at sporadic time intervals which might occur characteristically in some species or groups of species. In fact, such an overdispersed distribution of t would restrict, or at best, not significantly alter, the conditions for maintaining polymorphisms. Autocorrelation between values of t in successive generations might, on the

other hand, influence these conditions such as to increase the region of polymorphisms. Simulation of models involving rate of outcrossing as a function of gene frequency seemed to change only the trajectories of genotypic frequencies and not the final equilibrium values of p and F . An extensive literature review on variation in mating systems indicates that there are many interesting modes of genetic or environmental variation and diverse evolutionary mechanisms involved in the stochastic component of mixed mating systems. It is hoped that some generalised features would emerge from further theoretical work that would help extend our simple models discussed above.

The effect of randomly varying selection intensities in random mating population was investigated by Kimura (1958), who showed that this led to quasi-fixation instead of fixation as might be expected from directional selection in favour of one allele (see also Barker and Butcher, 1963). Therefore it would appear that, in general, stochastic selective processes allow wider range of conditions for nontrivial equilibria. In fact, a similar problem on the relative conditions for stability of nontrivial equilibria might be of general interest in relation to the evolutionary models involving interspecific competition, predator-prey or host-pathogen systems, etc. However, Leslie (1958, 1962) and Leslie and Gower (1958) have simulated two-species competition models largely to investigate the rates of change or outcome of stochastic models, rather than the conditions on birth rate and death rate coefficients. Leslie and Gower (1958) found that the properties of stochastic model and its deterministic equivalent differed mostly in the case of a system with an unstable stationary state (initial frequency-dependent outcome). In case of predator-prey system in fact, Leslie and Gower (1960) concluded that variable risks of predation ("heterogeneity" of environment) would lead to, provided variability in parameter k is limited, wider conditions for the coexistence of two species (k = fraction of the prey population exposed to the risk of predation). Numerical work on these stochastic processes from the viewpoint of finding conditions for various types of equilibria would appear worth while.

8. SUMMARY

1. The experimental evidence for stochastic variation of outcrossing rate and selective values was reviewed briefly to illustrate their form of distribution over space, time and individual plant families. In the case of low outcrossing rates, the distribution of t often involves overdispersion and kurtosis.

2. A single, diallelic locus model involving mixed selfing and random mating was simulated by Monte Carlo procedure on a digital computer in which the rate of outcrossing was varied generation to generation with the use of random variable having a specified distribution (normal, rectangular, negative binomial and exponential) with various sets of input parametric values. The effect of stochastic t was specifically investigated in terms of change in the bounds on selective values that would yield nontrivial genotypic equilibria. Similarly, in another set of runs, t was held constant but selective values (x, y) of homozygotes were varied among successive generations.

3. The most significant outcome of these studies was that fluctuating t , even with overdispersion characteristic of the negative binomial or exponential distribution, resulted in very small change in the critical bounds on

(x, y) whereas stochastic selective values resulted in significant shifts towards producing a larger region of equilibria with polymorphisms.

4. The implications of these properties of a stochastic mixed mating model were discussed briefly in terms of the conditions allowing the maintenance of long-term variability, estimation of selective and mating systems parameters, and the role of outcrossing rates in relation to the distribution of intrapopulation variation.

Acknowledgments.—This work was supported by a research grant (GM10476) from the U.S. Public Health Service. We would like to thank Dr D. G. Smeltzer for permission to cite his unpublished data on sorghum populations, and the Computer Center at the Davis campus for the use of its many facilities.

9. REFERENCES

- ALLARD, R. W. 1954. Natural hybridization in lima beans in California. *Proc. Amer. Soc. Hort. Sci.*, 64, 410-416.
- ALLARD, R. W., JAIN, S. K., AND WORKMAN, P. L. 1967. The population genetics of inbreeding species. *Adv. Genetics*. (In press.)
- ALLARD, R. W., AND WORKMAN, P. L. 1963. Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. *Evolution*, 18, 470-480.
- ANScombe, F. J. 1950. Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, 37, 358-382.
- BARKER, J. S. F., AND BUTCHER, J. C. 1966. A simulation study of quasi-fixation of genes due to random fluctuation of selection intensities. *Genetics*, 53, 261-268.
- BARTLETT, M. S., GOWER, J. C., AND LESLIE, P. H. 1960. A comparison of theoretical and empirical results for some stochastic populations models. *Biometrika*, 47, 1-11.
- BATEMAN, A. J. 1947. Contamination of seed crops. III. Relation with isolation distance. *Heredity*, 1, 303-336.
- BATEMAN, A. J. 1950. Is gene dispersion normal? *Heredity*, 4, 353-364.
- BARTLETT, M. S. 1955. *An Introduction to Stochastic Processes*. Cambridge Univ. Press, London.
- BLISS, C. I., AND FISHER, R. A. 1953. Fitting the negative binomial distribution to biological data. *Biometrics*, 9, 173-200.
- CLAASEN, C. E. 1950. Natural and controlled crossing in safflower, *Carthamus tinctorius* L. *Agron. J.*, 42, 381-384.
- CLARK, C. E., AND HOLZ, B. W. 1960. *Exponentially distributed Random Numbers*. The Johns Hopkins Press, Baltimore.
- DOMINGO, W. E. 1944. Amount of natural out-crossing in the castor oil plant. *Agron. J.*, 36, 360-361.
- DRAYNER, J. M. 1959. Self- and cross-fertility in field beans (*Vicia faba*). *J. Agric. Sci.*, 53, 387-403.
- FRYXELL, P. A. 1957. Mode of reproduction of higher plants. *Bot. Rev.*, 23, 135-233.
- HARDING, J. A., AND TUCKER, C. L. 1964. Quantitative studies on mating systems. I. Evidence for the non-randomness of outcrossing in *Phaseolus lunatus*. *Heredity*, 19, 369-381.
- HAYMAN, B. I. 1953. Mixed selfing and random mating when homozygotes are at a disadvantage. *Heredity*, 7, 185-192.
- INNES, N. L. 1961. Natural crossing in cotton in the Sudan. *Empire Cotton Growing Review*, 38, 17-21.
- JAIN, S. K., AND ALLARD, R. W. 1960. Population studies in predominantly self-pollinated species. I. Evidence for heterozygote advantage in a closed population of barley. *Proc. Natl. Acad. Sci. (U.S.)*, 46, 1371-1377.
- JAIN, S. K., AND ALLARD, R. W. 1965. The nature and stability of equilibria under optimizing selection. *Proc. Natl. Acad. Sci. (U.S.)*, 54, 1436-1443.
- JAIN, S. K., AND ALLARD, R. W. 1966. The effects of linkage, epistasis and inbreeding on population changes under selection. *Genetics*, 53, 633-659.
- KAKIZAKI, Y. 1924. The flowering habit and natural crossing in the egg plant. *Jap. Jour. Genet.*, 3, 29-38.
- KAKIZAKI, Y. 1929. Natural crossing in the tomato. *Jap. Jour. Genet.*, 4, 81-85.

- KIMURA, M. 1954. Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics*, 39, 280-295.
- KIMURA, M. 1964. *Diffusion Models in Population Genetics*. Methuen, London.
- LESLIE, R. H. 1962. A stochastic model for two competing species of *Tribolium* and its application to some experimental data. *Biometrika*, 49, 1-25.
- LESLIE, P. H., AND GOWER, J. C. 1960. The properties of a stochastic model for the predator-prey type of interaction between two species. *Biometrika*, 47, 219-234.
- MACVICAR, R. M., AND PARNELL, H. R. 1941. The inheritance of plant colour and the extent of natural crossing in foxtail millet. *Sci. Agric.*, 22, 80-84.
- MORTON, N. E., AND YASUDA, N. 1962. The genetical structure of human populations. In: *Les Déplacements Humains*, Entertien de Monaco en Science Humanes, 185-203.
- PAPOULIS, A. 1965. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, New York.
- SHREIDER, YU A. (Ed.). 1964. *Method of Statistical Testing: Monte Carlo Method*. Elsevier Publishing Co., Amsterdam.
- SIMPSON, D. M. 1954. Natural cross-pollination in cotton. U.S. Dept. of Agric. Technical Bulletin No. 1094.
- WILSIE, C. P., AND TAKAHASHI, M. 1934. Natural crossing in the pigeon pea. *J. Agric. Res.*, 49, 923-927.
- WORKMAN, P. L., AND JAIN, S. K. 1966. Zygotic selection under mixed random mating and self-fertilization: Theory and problems of estimation. *Genetics*, 54, 159-171.