

THE SURVIVAL OF A MUTANT GENE UNDER SELECTION

P. A. P. MORAN

(rec. 24 July 1958)

Abstract: The problem of the survival of a single mutant in a haploid genetic population when there exists selection is considered for a type of population model in which the generations are overlapping. The results are compared with the previous work of Fisher and others for other models. The need is stressed for a solution of the same problem in a diploid population with general phenotypic selection coefficients.

R. A. Fisher [3] has discussed the survival of a single mutant in an infinite population of haploid individuals. If p_0, p_1, p_2, \dots are the probabilities of a mutant gene having 0, 1, 2, \dots offspring and $f(z) = p_0 + p_1z + p_2z^2 + \dots$ is the corresponding generating function, a well-known theory due to H. W. Watson and others (see Feller [2]) shows that the mutant will ultimately die out with unit probability if $f'(1) \leq 1$, and will have a probability of survival of $1 - \zeta > 0$ if $f'(1) > 1$, where ζ is the unique non-unit positive root of the equation $z = f(z)$. This solves the problem for an infinite population.

R. A. Fisher has also given an approximate solution for a finite population of $2n$ haploid individuals for which the generations do not overlap and the mutant gene has a small effective selective advantage α . The number of offspring in this model has a distribution which is Poissonian to a good approximation, with a mean $1 + \alpha$, for the offspring of a mutant type and unity for a non-mutant type. Fisher shows that the probability of survival is then $2\alpha(1 - \exp - 4\alpha n)^{-1}$ to a good approximation when $\alpha^2 n$ is small, and if α is made negative so that we have a selective disadvantage the same result holds. When $\alpha = 0$ the probability is exactly $(2n)^{-1}$.

In the present note we give an exact solution starting from any initial number of mutants in a model in which the generations overlap. This model was introduced for other purposes in a previous paper (Moran [6]). We suppose we have a population of $2n$ haploid individuals in which k_0 are initially of the mutant type a and the rest A . The development of the process takes place in continuous time and each individual has a life time with a negative exponential distribution whose parameter depends only on whether the individual is a or A . The state of the system at any time t

is then defined by the number, k , of mutant individuals. When an individual dies it is immediately replaced by a new individual which is the offspring of the population immediately before the death. We can introduce selection into this model in two different ways.

In the first we assume that selection acts by varying the lifetime distribution of the haploid individuals. Suppose that each mutant individual of type a has a lifetime T which is distributed in a negative exponential distribution of the form

$$\lambda_1^{-1} \exp - (\lambda_1^{-1} t) dt$$

whilst the distribution of the non-mutant lifetime is similar with parameter λ_2 . Then $\alpha = (\lambda_1 - \lambda_2)\lambda_2^{-1}$ may be regarded as the selective advantage, positive or negative, of the mutant. When the number of mutants is k , the probability of the death of a mutant in the time interval $(t, t + dt)$ is $k\lambda_1^{-1} dt + o(dt)$, whilst the probability of an A individual dying is $\lambda_2^{-1}(2n - k)dt + o(dt)$. Thus the probability of some death occurring is $\{\lambda_1^{-1}k + \lambda_2^{-1}(2n - k)\}dt + o(dt)$, and given that such a death does occur, the probability that it is the death of an a -individual is

$$\lambda_2 k \{\lambda_2 k + \lambda_1(2n - k)\}^{-1}.$$

The individual born as a replacement is chosen at random from the gametic output of the previously existing generation and so the probability that it is a is $k(2n)^{-1}$, and that it is A is $(2n - k)(2n)^{-1}$.

We may now consider the discrete Markov chain embedded in this continuous process defined for those discrete instants of time at which a birth-death event occurs. This chain has $2n + 1$ states defined by the integer k ($k = 0, 1, \dots, 2n$) and transition probabilities p_{ij} from state i to state j given by the formulae

$$\begin{aligned} p_{k, k-1} &= \lambda_2 k (2n - k) (2n)^{-1} \{\lambda_2 k + \lambda_1(2n - k)\}^{-1}. \\ p_{k, k+1} &= \lambda_1 k (2n - k) (2n)^{-1} \{\lambda_2 k + \lambda_1(2n - k)\}^{-1}. \\ p_{k, k} &= 1 - p_{k, k-1} - p_{k, k+1}. \end{aligned}$$

In the second method of introducing selection we suppose that all individuals have the same negative exponential lifetime distribution so that when a death occurs the probability that it is the death of an a -individual is $k(2n)^{-1}$, and of an A individual $(2n - k)(2n)^{-1}$. If the relative reproductive powers of the two kinds of gamete have the ratio $\mu_1 \mu_2^{-1}$ the probability of the birth of an a -individual is $\mu_1 k \{\mu_1 k + \mu_2(2n - k)\}^{-1}$, and of an A individual $\mu_2(2n - k) \{\mu_1 k + \mu_2(2n - k)\}^{-1}$. The transition probabilities of the embedded Markov chain are then

$$\begin{aligned} p_{k, k-1} &= \mu_2 k (2n - k) (2n)^{-1} \{\mu_1 k + \mu_2(2n - k)\}^{-1}. \\ p_{k, k+1} &= \mu_1 k (2n - k) (2n)^{-1} \{\mu_1 k + \mu_2(2n - k)\}^{-1}. \\ p_{k, k} &= 1 - p_{k, k-1} - p_{k, k+1}. \end{aligned}$$

We notice that in both this and the former case the ratio $\phi_{k, k+1}(\phi_{k, k-1})^{-1}$ is constant, being $\mu_1\mu_2^{-1}$ and $\lambda_1\lambda_2^{-1}$ respectively. This suggests that we now embed another chain in this one by considering only those jumps of the process at which k changes its value. Writing $\pi = \lambda_1(\lambda_1 + \lambda_2)^{-1}$ or $\mu_1(\mu_1 + \mu_2)^{-1}$ we thus get a process in which the transition probabilities are

$$(1) \quad \begin{cases} \phi_{k, k-1} = 1 - \pi. \\ \phi_{k, k} = 0. \\ \phi_{k, k+1} = \pi. \end{cases}$$

If we now start the process with an initial value k_0 the probabilities of ultimate absorption P_0, P_{2n} , in the two absorbing states ($k = 0, 2n$) will be the same as in the original continuous process. P_0 and P_{2n} are in fact the probabilities of the mutant form not surviving, and surviving, respectively. When $\pi > \frac{1}{2}$, a has a selective advantage and when $\pi < \frac{1}{2}$, a selective disadvantage.

The process is now identical with that of the gambler's ruin problem (Feller [2]) from which we see that if $\pi \neq \frac{1}{2}$ the probability of a not surviving is

$$(2) \quad P_0 = \frac{\left(\frac{1-\pi}{\pi}\right)^{2n} - \left(\frac{1-\pi}{\pi}\right)^{k_0}}{\left(\frac{1-\pi}{\pi}\right)^{2n} - 1} = 1 - P_{2n}$$

where k_0 is the initial value of k . When $\pi = \frac{1}{2}$ ($\lambda_1 = \lambda_2$ or $\mu_1 = \mu_2$) it is known that

$$(3) \quad P_{2n} = k_0(2n)^{-1} = 1 - P_0.$$

(3) also follows at once when we observe that $E(k_{t+1} - k_t | k_t) = 0$ and so

$$k_0 = E(k_t) = (2n)P_{2n}.$$

In fact this type of argument on the moments may be used to solve the gambler's ruin problem without using difference equations. Let X be a real number and consider

$$\begin{aligned} E(X^{k_{t+1}} - X^{k_t} | k_t) &= (\pi X + (1 - \pi)X^{-1} - 1)X^{k_t} (k_t \neq 0, 2n) \\ &= 0, (k_t = 0, 1) \end{aligned}$$

Thus if we write $X = (1 - \pi)\pi^{-1}$ we find that

$$E(X^{k_{t+1}} - X^{k_t} | k_t) = 0$$

whatever the value of k_t . However, ultimately k_t will be either 0 or $2n$ since these are the only two absorbing states and so

$$P_{2n}X^{2n} + P_0 = X^{k_0}$$

and therefore, since $P_{2n} = 1 - P_0$,

$$P_0 = (X^{2n} - X^{k_0}) / (X^{2n} - 1)$$

which is the above solution. In terms of the selective advantage α we have

$$\pi = (1 + \alpha)(2 + \alpha)^{-1}$$

and therefore

$$(4) \quad P_0 = \frac{1 - (1 + \alpha)^{2n - k_0}}{1 - (1 + \alpha)^{2n}}$$

$$(5) \quad P_{2n} = \frac{(1 + \alpha)^{2n - k_0} - (1 + \alpha)^{2n}}{1 - (1 + \alpha)^{2n}}$$

and for $\alpha = O(n^{-1})$, P_{2n} is given approximately for n large by

$$(1 - e^{-\alpha k_0})(1 - e^{-2\alpha n})^{-1}$$

or

$$(6) \quad \alpha k_0 (1 - e^{-2\alpha n})^{-1}$$

when αk_0 is small and positive. In particular this is so if $k_0 = 1$ when we get

$$\alpha (1 - e^{-2\alpha n})^{-1}$$

which is of the same form as Fisher's result but with his α multiplied by $\frac{1}{2}$.

If we now keep α fixed and let n tend to infinity in (4) and (5) we find $P_0 = 1$ for $\alpha < 0$ and $P_0 = (1 + \alpha)^{-k_0}$ for $\alpha > 0$. This result can be checked by considering the non-unit positive root of the equation $z = f(z)$ where $f(z)$ is the generating function of the limiting probability distribution of the number of offspring generated by a single individual.

To obtain this consider the process as defined for discrete time intervals at the end of each of which a birth death event occurs, and suppose that the population size $2n$ is large compared with the number of mutants. Then the probability of a particular mutant gene having a life time exactly equal to t ($t = 1, 2, \dots$) is $(1 - A)A^{t-1}$, where

$$A = 1 - \lambda_2 (2n\lambda_1)^{-1}$$

in the first model of selection. If it does survive for exactly t units of time and then dies it has t chances of being a parent, each with probability $(2n)^{-1}$ and the total probability generating function of the number of its offspring is therefore

$$\lambda_2 (2n\lambda_1)^{-1} (2n - 1 + z) (2n)^{-1} \left\{ 1 - \left(1 - \frac{\lambda_2}{2n\lambda_1} \right) \left(\frac{2n - 1 + z}{2n} \right) \right\}^{-1},$$

and letting $2n$ tend to infinity this becomes

$$\begin{aligned} f(z) &= (1 + \lambda_1 \lambda_2^{-1} (1 - z))^{-1} \\ &= (1 + (1 + \alpha)(1 - z))^{-1}. \end{aligned}$$

The positive root of $f(z) = z$ not equal to unity is then $(1 + \alpha)^{-1}$ and so the probability of a single mutant not surviving is $(1 + \alpha)^{-1}$. Thus the probability of k_0 initial mutants not surviving is equal to $(1 + \alpha)^{-k_0}$ as before.

It has long been known that the form of the distribution of the number of offspring has an effect on the rate of progress to homozygosity in the absence of mutation and on the probability of survival of a mutant (Haldane [5], Fisher [4]). The above theory provides an exact solution for the latter problem in a special case but unfortunately this method of embedding one Markov chain in another does not provide any method of determining the asymptotic rate of progress to homozygosity when there are selective effects, a problem which has been discussed by Wright and Kerr [7] (see also Crow and Kimura [1]).

It would be of considerable importance to obtain solutions for the probability of survival of a mutant when the individuals concerned are diploid but this appears to be very difficult in general. Consider first the case where the mutant has no selective advantage or disadvantage. Then for random mating and most types of non-random mating it is easy to obtain the probability of survival in a finite population. If the mutant is dominant and selectively advantageous or disadvantageous, and the population is very large, it would seem that the above theory and Fisher's theory hold in their respective cases. If the mutant is not dominant these theories should hold for large population in which the issue of survival or not is decided before the number of mutants is large enough to give an appreciable probability of their being any diploid homozygous mutants. If finally the mutant is recessive and the population large it would appear that the probability of survival would be close to that obtained by supposing no selection since the number of mutants would have to be large before dominants appeared and selection operated. Phenotypic assortative mating would not change this but inbreeding based on family relationships would have a powerful effect although one difficult to treat mathematically.

I am indebted to Mr. G. A. Watterson for some helpful criticism.

References

- [1] Crow, J. F. and Kimura, M., Some genetic problems in natural populations. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press. (1956).
- [2] Feller, W., *An Introduction to Probability Theory and its Applications*. J. Wiley. New York (1957).
- [3] Fisher, R. A., *The Genetical Theory of Natural Selection*. Clarendon Press. Oxford (1930).

- [4] Fisher, R. A., Stage of developments as a factor influencing the variance in the number of offspring, frequency of mutants and related quantities. *Ann. Eugenics*. 9 (1939), 406–408.
- [5] Haldane, J. B. S., The equilibrium between mutation and random extinction. *Ann. Eugenics* 9 (1939), 400–405.
- [6] Moran, P. A. P., Random processes in genetics. *Proc. Cam. Phil. Soc.* 54 (1958), 60–71.
- [7] Wright, S. and Kerr, W. E. Experimental studies of the distribution of gene frequencies in very small populations of *Drosophila melanogaster*, II. *Bar. Evolution* 8 (1954), 225–240.

Australian National University, Canberra.