

# THE PROBABILITY OF SURVIVAL OF A MUTANT

W. J. EWENS

*Australian National University*

Received 20.xii.66

## 1. INTRODUCTION

THE formula giving the probability of survival of a new mutant allele at a locus homozygous for some other allele has been known since the pioneering work of Fisher (1922). Whether or not the new mutant survives will depend almost exclusively on the behaviour of its frequency during the critical early generations; for this reason we ignore the possibility of formation of homozygous mutants and refer then to heterozygous mutants as mutants. If a mutant produces  $i$  viable mutant offspring with probability  $f_i$  and if  $f(s) = \sum s^i f_i$ , then the smallest positive solution of

$$s = f(s) \tag{1.1}$$

gives the probability of extinction of the mutant. This is less than unity if and only if  $\sum i f_i$  exceeds unity.

When more than one allele is present at the locus when the mutation first occurs the situation is more complicated. The survival probability will depend not only on the fitnesses of the various genotypes but also on the frequencies of the various alleles present and on the genotype initially formed by the mutant. In this note we derive survival probabilities in such a situation.

## 2. SURVIVAL PROBABILITIES

Consider a locus at which, at time zero, occur alleles  $A_1, \dots, A_k$ . Assume that the fitness of  $A_i A_j$  is  $w_{ij}$  and that the frequencies of the various genotypes occur in Hardy-Weinberg proportions, the frequency of  $A_i$  being  $p_i$ . We denote the mean fitness of the population at the time the mutation occurs by  $W$ , where

$$W = \sum w_{ij} p_i p_j.$$

Suppose now that at time zero a new allele  $B$  is introduced at the locus in question; let the fitness of  $A_i B$  be  $\mu_i$ . Ignoring for the moment stochastic fluctuations, the frequency  $x$  of  $B$  in successive generations changes according to the formula

$$x(t+1) = [(\sum p_i \mu_i) / W] x(t), \tag{2.1}$$

a result given by Bodmer and Parsons (1960). Clearly the frequency of  $B$  will increase deterministically only when  $\sum p_i \mu_i$  exceeds  $W$ , that is when the mean fitness of heterozygotes formed by the mutant exceeds the mean fitness of the original population.

We now consider the case where the exceedingly small frequency of  $B$  requires a stochastic treatment to be made. For this purpose we shall suppose that the population is in size equilibrium (i.e.  $W = 1$ ), although this convention is made essentially for notational convenience and can easily be relaxed.

Suppose that after  $t$  generations the number of  $A_iB$  individuals is  $n_i(t)$  ( $i = 1, \dots, k$ ). Then the gene  $B$  will be transmitted, on an average, to  $\mu_i$  offspring from each  $A_iB$  individual and each such gene will combine with an  $A_j$  gene, with probability  $p_j$ , to form an  $A_jB$  individual. Clearly

$$E\{n_j(t+1)\} = p_j \sum \mu_i n_i(t), \quad (j = 1, \dots, k) \quad (2.2)$$

where  $E\{\cdot\}$  denotes the expectation operator conditional on given values at generation  $t$ . In matrix terms the system of equations (2.2) may be written

$$E\{\mathbf{n}(t+1)\} = M\mathbf{n}(t), \quad (2.3)$$

where the matrix  $M$  is given by

$$M = \mathbf{p}\boldsymbol{\mu}', \quad \mathbf{p}' = (p_1, \dots, p_k), \quad \boldsymbol{\mu}' = (\mu_1, \dots, \mu_k).$$

To obtain survival probabilities it is necessary to know not only expected values, as given by (2.2), but the complete distribution of the number of viable  $B$  offspring from each  $A_iB$  parent. If the generating function of this distribution is denoted  $f_i(s)$ , then we require that  $f_i(1) = 1$ ,  $f_i'(1) = \mu_i$ . Any such offspring from an  $A_iB$  parent is  $A_jB$  with probability  $p_j$ ; if then we use the dummy  $s_j$  to refer to  $A_jB$ , the probability generating function of the number of such offspring from an  $A_iB$  parent, where account is taken of the type of offspring, is

$$f_i(p_1s_1 + \dots + p_ks_k). \quad (2.4)$$

The interpretation of this function is that the coefficient of  $s_1^{a_1} \dots s_k^{a_k}$  in its Taylor expansion is the probability that an  $A_iB$  parent produces  $a_1$  offspring which are  $A_1B$ ,  $\dots$ ,  $a_k$  which are  $A_kB$ . We can now use the following theorem given in Harris (1963, page 41).

*Theorem.* The probability that the new mutation  $B$  survives exceeds zero only when the largest eigenvalue of  $M$  (in equation (2.3)) exceeds 1. In this case, if  $s_i$  is the probability of ultimate extinction of  $B$  when the initial mutant is  $A_iB$ , then the quantities  $s_1, \dots, s_k$  are the unique positive solutions (less than unity) of the set of equations

$$s_i = f_i(p_1s_1 + \dots + p_ks_k), \quad (i = 1, \dots, k). \quad (2.5)$$

Because of the special nature of  $M$  and the generating functions (2.4) it is easy to find the consequences of this result. Because of the form of  $M$  (i.e.  $M = \mathbf{p}\boldsymbol{\mu}'$ ),  $M$  has rank 1 and hence only one non-zero eigenvalue. This eigenvalue must then be identical to the sum of all eigenvalues, that is to the trace (sum of main diagonal elements) of  $M$ . But

$$\text{trace } M = \sum p_i \mu_i$$

so that the new mutation can survive only when  $\sum p_i \mu_i$  exceeds unity (or more generally, exceeds  $W$ ). If this is the case, we can solve the system of equations (2.5) quickly by multiplying the  $i$ th equation by  $p_i$  and adding. Putting  $s = \sum p_i s_i$ ,  $s$  then solves

$$s = \sum p_i f_i(s). \quad (2.6)$$

Often  $s$  is the quantity of interest, since it may not be known what individual was the initial mutant, so that the weighted survival probability  $\sum p_i s_i$  ( $p_i$  being the probability that the initial mutant is  $A_iB$ ) is required. In any case individual  $s_i$  values can be obtained immediately from (2.5), i.e.

$$s_i = f_i(s), \quad (2.7)$$

where now  $s$  is the required solution of (2.6).

In any particular case, equation (2.6) provides a satisfactory formula for numerical evaluation of  $s$ . It does not seem possible to draw general conclusions from this equation, although some fairly general remarks can be made in particular cases. For example, Haldane (1927) has shown, in the case  $k = 1$ , that if  $\mu_1$  is only slightly greater than 1, so that the new mutant is only slightly more fit than the wild type, then

$$1 - s \approx 2[\mu_1 - 1]/\sigma^2, \tag{2.8}$$

where  $\sigma^2$  is the variance of the offspring distribution  $f_1(s)$ .

In the present more general case we may obtain a similar result when each  $\mu_i$  is only slightly greater than 1. Then both  $s$  and  $s_j$  ( $j = 1, \dots, k$ ) will be close to 1, so that expanding  $\sum p_i f_i(s)$  in equation (2.6) about  $s = 1$  we get, to a close approximation,

$$s = 1 + (s - 1) \sum p_i \mu_i + \frac{1}{2} (s - 1)^2 \sum p_i \sigma_i^2$$

so that 
$$1 - s = 2[\bar{\mu} - 1]/\bar{\sigma}^2 \tag{2.9}$$

where  $\bar{\mu} = \sum p_i \mu_i$ ,  $\bar{\sigma}^2 = \sum p_i \sigma_i^2$ . This generalises Haldane's result (2.8). Individual  $s_i$  values may be obtained by a similar expansion of (2.7) to get

$$1 - s_i = (1 - s) \mu_i - \frac{1}{2} (1 - s)^2 \sigma_i^2, \quad (i = 1, \dots, k). \tag{2.10}$$

### 3. POISSON OFFSPRING DISTRIBUTION

If we make the common assumption that the offspring distributions  $f_i(s)$  are Poisson we can obtain further results. Here (2.5) becomes

$$s_i = \exp [\mu_i (p_1 s_1 + \dots + p_k s_k - 1)], \quad (i = 1, \dots, k). \tag{3.1}$$

Considering rather survival probabilities  $\pi_i = 1 - s_i$  we get

$$\begin{aligned} 1 - \pi_i &= \exp -[\mu_i (p_1 \pi_1 + \dots + p_k \pi_k)] \\ &= \exp (-\mu_i \pi), \quad (i = 1, \dots, k). \end{aligned} \tag{3.2}$$

Here  $\pi = \sum p_i \pi_i$ . Equation (2.6) becomes

$$1 - \pi = \sum p_i \exp (-\mu_i \pi). \tag{3.3}$$

It is interesting to note that the value of  $\pi$  obtained from (3.3) bears an interesting relationship to that obtained if we do not carry out the above multiple-type theory, but use as an approximation previous theory with the fitness of  $B$  being  $\sum p_i \mu_i$ . The survival probability  $\pi^*$  calculated from the latter method would solve

$$1 - \pi^* = \exp (-\pi^* \sum p_i \mu_i). \tag{3.4}$$

For any given  $p_i$  and  $\mu_i$  values for which  $\sum p_i \mu_i > 1$ , let the solution of (3.4) be  $x$ . This will be greater than or equal to the solution of (3.3) if the function

$$1 - \pi - \sum p_i \exp (-\mu_i \pi),$$

evaluated at  $\pi = x$ , is negative or zero. But this will be the case if and only if

$$\exp (-x \sum p_i \mu_i) \leq \sum p_i \exp (-x \mu_i). \tag{3.5}$$

But this inequality is always true, since the left-hand side is a weighted geometric mean of the positive quantities  $\exp (-x \mu_i)$  while the right-hand

side is the corresponding arithmetic mean. The latter will always exceed the former except in the limiting case when the  $\mu_i$  are equal, when they coincide. We conclude that the approximation using (3.4) will always overestimate the true mean survival probability when the  $\mu_i$  are not all equal.

A second conclusion derived from (3.3) is that  $\pi$  is not a linear function, and not necessarily an increasing function, of  $\Sigma p_i \mu_i$ . Thus it should be possible to find two cases, in one of which the value of  $\Sigma p_i \mu_i$  exceeds the corresponding value in the second, but for which the value of  $\pi$  is less than that for the second. We provide the following example:

Case A.  $\mu_1 = \mu_2 = 1.5$ ,  $p_1, p_2$  arbitrary.

Here  $\Sigma p_i \mu_i = 1.5$ ,  $\pi = 0.58281$ .

Case B.  $\mu_1 = 1$ ,  $\mu_2 = 2$ ,  $p_1 = 0.429$ ,  $p_2 = 0.571$ .

Here  $\Sigma p_i \mu_i = 1.571$ ,

exceeding the corresponding value for case A, while

$$\pi_1 = 0.44122, \pi = 0.687765, \pi = 0.582,$$

which is less than the corresponding value for case A.

While this result shows that  $\Sigma p_i \mu_i$  is not the only relevant quantity for determining survival probabilities one suspects that in the vast majority of cases increasing  $\Sigma p_i \mu_i$  increases  $\pi$ .

#### 4. REFERENCES

- BODMER, W. F., AND PARSONS, P. A. 1960. The initial progress of new genes with various genetic systems. *Heredity*, 15, 283-289.
- FISHER, R. A. 1922. On the dominance ratio. *Proc. Roy. Soc. Edinb.*, 42, 321-341.
- HALDANE, J. B. S. 1927. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proc. Camb. Phil. Soc.*, 23, 838-844.
- HARRIS, T. E. 1963. *The theory of branching processes*. Springer Verlag.