

## WHAT IS FITNESS?

**Fitness** is *the propensity to increase in frequency*. Organisms have fitness, and so do genes—and in many cases, the two are tightly linked. That fitness is a *propensity* means that it is intrinsically probabilistic. Just as a fair coin—one with a propensity to land heads and tails in equal proportion—may sometimes generate long runs of heads, an organism with lower fitness may on occasion outcompete or out-reproduce higher-fitness organisms. By the same token, a genotype’s increases in frequency over time constitute evidence, rather than certainty, that it confers increased fitness.

## FITNESS AS GROWTH RATE

A simple model of a growing population helps to make these ideas concrete. Suppose we have a population of individuals which grow, reproduce and die in a synchronized fashion, like annual plants. Suppose each individual leaves a Poisson-distributed number of offspring with mean  $w$ .  $w$  is the absolute fitness.

If the population starts with  $n$  individuals, then after one generation the probability that there will be  $k$  individuals is  $\text{Poisson}(k; wn)$  which has mean  $wn$  and a standard deviation of  $\sqrt{wn}$ . If  $w = 1$  then on average the population remains of size  $n$  with an average fluctuation of  $\sqrt{n}$ .

In this sense, we can say that the absolute fitness  $w$  is the *expected* change in frequency of a genotype after one generation:

$$\langle n(1) \rangle = wn(0), \text{ so that } w = \left\langle \frac{n(1)}{n(0)} \right\rangle$$

Fitness is a propensity, not destiny, so we note that the *actual* change in frequency will rarely be  $w$ , even in this simple model. To see this, note that the expected number of individuals  $wn$  occur with probability  $\text{Poisson}(wn; wn)$ , which is zero if  $wn$  is not an integer, and  $p_{\text{mean}} = (wn)^{wn} e^{-wn} / (wn)!$  in the (rare) cases when it is. For example, if  $wn = 100$  individuals, then the probability of observing the expected value is less than 4%. Though this may seem a niggling detail, basing one’s evolutionary intuition on average outcomes is analogous to basing one’s clothing choices each morning on the average weather.

## OTHER WAYS OF WRITING FITNESS

Our model population will continue to grow on average as  $\langle n(t) \rangle = n(0)w^t$  with  $t$  marking generations. Commonly, we wish to measure in absolute time, with  $t$  marking seconds, hours or years. Then we generally prefer to use instantaneous growth rates  $r$  defined by

$$\frac{\partial n}{\partial t} = rn(t) \quad \text{with the average behavior} \quad \langle n(t) \rangle = n(0)e^{rt}.$$

In this case, the fitness  $w = e^r$ . Even more realistically, we note that growth is a balance of birth and death, and for a population with birth rate  $b$  and death rate  $d$ ,  $r = b - d$ . This formulation helps us distinguish the biologically very different situations leading to  $r = 0$ : having offspring that die ( $b = d > 0$ ), versus not having offspring at all ( $b = d = 0$ ).

### FITNESS AND RELATIVE FITNESS

Now suppose we have individuals of two types,  $a$  and  $b$ , reproducing in a mixed population with respective fitnesses  $w_a$  and  $w_b$  (and growth rates  $r_a$  and  $r_b$  with  $w = e^r$ ). Suppose that they start from equal numbers,  $n_a(0) = n_b(0)$ , so that the initial concentrations cancel. We know that they will grow such that the ratio of  $a$  to  $b$  will be, on average,

$$\frac{\langle n_a(t) \rangle}{\langle n_b(t) \rangle} = \frac{w_a^t}{w_b^t} = \left( \frac{w_a}{w_b} \right)^t = \left( \frac{e^{r_a}}{e^{r_b}} \right)^t = e^{(r_a - r_b)t}$$

We see that the relative growth depends only on the ratio of the fitnesses or the difference between the instantaneous growth rates. (For this reason,  $r$  is often called the **additive fitness**.)

We define the average proportion of individuals of type  $a$  as

$$p_a(t) := \frac{\langle n_a(t) \rangle}{\langle n_a(t) \rangle + \langle n_b(t) \rangle}$$

and  $p_b(t) = 1 - p_a(t)$ . So here we see that

$$\frac{p_a(t)}{p_b(t)} = \frac{w_a^t}{w_b^t} = e^{(r_a - r_b)t}.$$

### SELECTION COEFFICIENTS

Because of the appearance of fitness ratios (and additive fitness differences) above, we can define the **selection coefficient**

$$s_{ab} := \ln \frac{w_b}{w_a} = r_b - r_a$$

which we can also write as  $s = \Delta \ln w = \Delta r$ . In most cases,  $s$  is all we'll ever need to know about fitness to determine evolutionary outcomes. When the types  $a$  and  $b$  refer to mutant and wild-type forms of the same organism,  $s_{ba}$  is the fitness effect of the mutation.

For convenience, we'll often refer to  $s$  without subscripts. Still, selection coefficients are always a *difference* between one type and another, and we need to keep track of which direction that difference goes. We'll use  $s$  on its own to refer to a mutant's selective advantage over a wild-type organism. Subscripts provide specificity:  $s_{ba}$  can be read as "the fitness gained by mutating from  $b$  to  $a$ ." We can do this in reverse, and find  $s_{ab} = r_b - r_a = \ln \frac{w_b}{w_a}$ . This brings up some useful properties of the selection coefficient defined in this way, which you may verify:

$$\begin{aligned} s_{ab} &= -s_{ba} && \text{reversibility} \\ s_{ab} + s_{bc} &= s_{ac} && \text{additivity} \end{aligned}$$

A common alternative formulation of the selection coefficient lets  $w_a = 1$  and  $w_b = 1 + s'$ . (We use  $s'$  to avoid confusion with  $s$  above.) In general, this  $s'_{ab} := \frac{w_b}{w_a} - 1$  behaves less conveniently:

$$\begin{aligned} s'_{ab} &= \frac{-s'_{ba}}{1-s'_{ba}} && \text{non-reversibility} \\ s'_{ab} + s'_{bc} &\neq s'_{ac} && \text{non-additivity} \end{aligned}$$

Consequently, we will use  $s = \Delta \ln w$  whenever possible. Because

$$s' = e^s - 1 = s + s^2/2 + \dots,$$

in the common case where  $s$  is small such that  $s^2$  terms can be neglected,  $s' \rightarrow s$ .

Selection coefficients indicate much about the evolutionary process:  $s = 0$  signifies neutral evolution,  $s > 0$  a selective advantage or beneficial mutation, and  $s < 0$  a selective disadvantage or deleterious mutation.

If we only know  $s$ , what does that tell us about fitness? We know only that the mutant's fitness is larger than the wild type's fitness by a factor  $e^s$ . With  $a$  the mutant and  $b$  the wild type:

$$(1) \quad s_{ba} = \ln \frac{w_a}{w_b} \implies w_a = w_b e^{s_{ba}}$$

#### HOW SELECTION COEFFICIENTS BEHAVE

With  $s$  in hand, we return to our population of two initially equipopulous types, a mutant  $a$  and wild-type  $b$  for which we will write the average proportion at time  $t$  as

$$n_a := \langle n_a(t) \rangle$$

and similarly for  $b$ . Thus,

$$\frac{n_a}{n_b} = \frac{p_a(t)}{p_b(t)} = e^{(r_a - r_b)t} = e^{s_{ba}t}$$

The selection coefficient allows this simpler form, which says that the ratio of average proportions of  $a$  and  $b$  individuals changes exponentially with time, rising exponentially if  $s > 0$  (where the mutant has an advantage) and falling exponentially if  $s < 0$ . On a logarithmic scale, the ratio changes linearly with time with slope  $s$ . We can see this in Figure 1.

How does the proportion of  $a$  change with time, on average? Suppose the  $a$  allele confers a birth rate  $1 + r_a$  and the  $b$  allele  $1 + r_b$ . With a constant death rate of 1 for all types, the growth rate of  $a$  is  $r_a$  and of  $b$  is  $r_b$ . That is,

$$\begin{aligned} \frac{\partial n_a}{\partial t} &= n_a r_a \\ \frac{\partial n_b}{\partial t} &= n_b r_b \end{aligned}$$

By the chain rule, the derivative of the proportion

$$(2) \quad \frac{\partial}{\partial t} \left( \frac{n_a}{n_a + n_b} \right) = \frac{1}{(n_a + n_b)^2} \left[ \frac{\partial n_a}{\partial t} n_b - n_a \frac{\partial n_b}{\partial t} \right].$$

So we have

$$\begin{aligned}
 \frac{\partial}{\partial t} \left( \frac{n_a}{n_a + n_b} \right) &= \frac{1}{(n_a + n_b)^2} [n_a r_a n_b - n_a n_b r_b] \\
 &= [r_a - r_b] \frac{n_a}{n_a + n_b} \frac{n_b}{n_a + n_b} \\
 (3) \qquad \qquad \qquad &= sp(1 - p)
 \end{aligned}$$

where  $s = s_{ba} = r_a - r_b$ , the fitness advantage of the mutant  $a$  over the wild type  $b$ .

Because the result depends only on the difference of growth rates, it is typical to normalize these rates such that the mutant has birth rate  $1 + s$  and the wild type 1, which leaves the evolutionary dynamics unchanged.

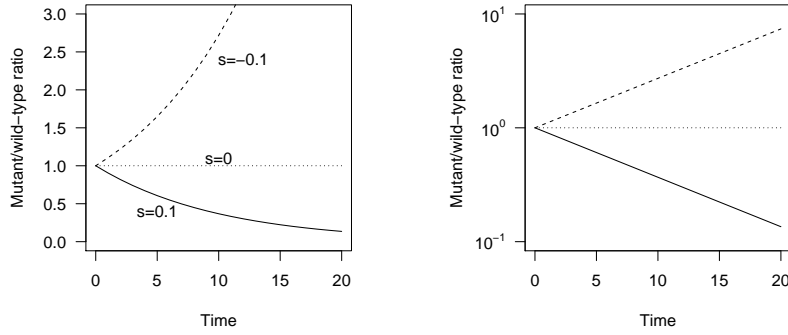


FIGURE 1. Average behavior of two types with a range of mutant selective differences  $s$ . On the right, the same results on a logarithmic  $y$ -axis show linear behavior.

#### DIPLOIDS AND DOMINANCE

In a diploid population, each individual's genotype includes two alleles for each gene. Whereas before we analyzed  $n_a := \langle n_a(t) \rangle$ , the average number of organisms with mutant allele  $a$  at time  $t$ , we now must consider the number of individuals whose genotypes include two copies of  $a$  ( $n_{aa}$ ) or of the wild-type allele  $b$  ( $n_{bb}$ ), termed **homozygotes**, and those carrying a copy of each ( $n_{ab}$  and  $n_{ba}$ ), the **heterozygotes**.

With two alleles, we need some way to quantify the contribution of each allele to the trait they encode. We introduce the **dominance**  $h$  of the mutant allele. Dominance quantifies the probability that the allele determines the trait. Heterozygotes carrying a mutant allele with dominance  $h = 1$  have the mutant trait, and such a mutant allele is **dominant**; heterozygotes carrying a **recessive** mutant alleles ( $h = 0$ ) have the wild-type trait. Intermediate values of  $h$  signal **incomplete dominance**.

**Model.** Suppose that the homozygous mutant, with genotype  $aa$ , has birth rate  $1 + s$ , the homozygous wild type has birth rate 1, and the heterozygotes have birth rate  $1 + sh$ .

All genotypes have death rate 1, so that the growth rates (birth minus death rates) are  $r_{aa} = s$ ,  $r_{ab} = r_{ba} = sh$ , and  $r_{bb} = 0$ . We assume **random mating** and

**free recombination**, meaning that all genotypes are randomized each generation, such that the expected number of  $aa$  alleles is the number of total alleles  $n_a + n_b$  multiplied by the probability of two  $a$  alleles ending up in the same individual if randomly inherited, and so on:

$$(4) \quad \begin{aligned} n_{aa} &= (n_a + n_b) \frac{n_a}{n_a + n_b} \frac{n_a}{n_a + n_b} = \frac{n_a^2}{n_a + n_b} \\ n_{ab} &= (n_a + n_b) \frac{n_a}{n_a + n_b} \frac{n_b}{n_a + n_b} = \frac{n_a n_b}{n_a + n_b} \dots \end{aligned}$$

We also define the allele fraction  $f_y^x$ , the fraction of genotype  $y$  that is allele  $x$ . For example,  $f_{aa}^a = 1$  and  $f_{ab}^b = \frac{1}{2}$ .

Given these parameters, how does the average frequency of the  $a$  allele change over time? The rate of change in the number of  $a$  alleles is the rate of change of each genotype, multiplied by the fraction of the genotype that is  $a$ , and likewise for  $b$ :

$$\begin{aligned} \frac{\partial n_a}{\partial t} &= f_{aa}^a n_{aa} r_{aa} + f_{ab}^a n_{ab} r_{ab} + f_{ba}^a n_{ba} r_{ba} + f_{bb}^a n_{bb} r_{bb} \\ \frac{\partial n_b}{\partial t} &= f_{aa}^b n_{aa} r_{aa} + f_{ab}^b n_{ab} r_{ab} + f_{ba}^b n_{ba} r_{ba} + f_{bb}^b n_{bb} r_{bb} \end{aligned}$$

which simplifies when we eliminate terms with allele fractions and growth rates that are zero, and note that  $n_{ab} = n_{ba}$ :

$$(5) \quad \begin{aligned} \frac{\partial n_a}{\partial t} &= n_{aa} r_{aa} + \frac{1}{2} n_{ab} r_{ab} + \frac{1}{2} n_{ba} r_{ba} \\ &= n_{aa} r_{aa} + n_{ab} r_{ab} \\ \frac{\partial n_b}{\partial t} &= \frac{1}{2} n_{ab} r_{ab} + \frac{1}{2} n_{ba} r_{ba} + n_{bb} r_{bb} \\ &= n_{ab} r_{ab} + n_{bb} r_{bb} \end{aligned}$$

Now, using Eq. 2, we can substitute in our expressions for the change in  $n_a$  and  $n_b$ :

$$\frac{\partial}{\partial t} \left( \frac{n_a}{n_a + n_b} \right) = \frac{1}{(n_a + n_b)^2} [n_b (n_{aa} r_{aa} + n_{ab} r_{ab}) - n_a (n_{ab} r_{ab} + n_{bb} r_{bb})]$$

and with the values for the growth rates from our model,

$$= \frac{1}{(n_a + n_b)^2} [n_b (n_{aa} s + n_{ab} sh) - n_a (n_{ab} sh + n_{bb} \cdot 0)]$$

and with the genotype numbers in terms of allele numbers, from Eq. 4,

$$\begin{aligned} &= \frac{1}{(n_a + n_b)^2} \left[ n_b \left( \frac{n_a^2}{n_a + n_b} s + \frac{n_a n_b}{n_a + n_b} sh \right) - n_a \left( \frac{n_a n_b}{n_a + n_b} sh \right) \right] \\ &= s \frac{n_a}{n_a + n_b} \frac{n_b}{n_a + n_b} \left[ \left( \frac{n_a}{n_a + n_b} + \frac{n_b}{n_a + n_b} \right) h - \left( \frac{n_a}{n_a + n_b} h \right) \right] \end{aligned}$$

and writing the average proportion of  $a$  as  $p$  and of  $b$  as  $1 - p$  we finally have

$$(6) \quad \frac{\partial p}{\partial t} = sp(1 - p) [h + (1 - 2h)p].$$

Eq. 6 is the rate of change of an advantageous mutant allele with dominance  $h$  and homozygous selective advantage  $s$ . Equivalently, we may think of the wild-type allele  $b$  as having dominance  $1 - h$  and selective advantage  $-s$ . Several important consequences derive from this result. When  $h = 1$ ,

$$(7) \quad \frac{\partial p}{\partial t} = sp(1 - p)^2$$

This is the case of a dominant advantageous allele  $a$  with proportion  $p$ , and of a recessive deleterious allele  $b$  with proportion  $1 - p$ ; compare this to Eq. 3. Note how the increase in frequency of  $a$  depends on the square of the proportion of  $b$ . When  $b$  is common,  $p$  rapidly increases, but as  $b$  becomes rare,  $a$ 's rate of change slows dramatically.

#### PROBABILITY OF FIXATION

What is the probability that a mutant allele becomes the most recent common ancestor at its locus? The details of the evolutionary process matter, most fixation formulae are approximate, and no fixation formula covers all situations. However, many situations are similar but for a few constants. For a mutant allele with selective advantage  $s$  currently at a proportion  $p$  in a population of  $N$  individuals, Table 1 shows the probability of fixation under various models.

A fixation formula can be difficult to understand just by looking at it. Let's look at how one behaves.

Figure 2 shows a few key features. Deleterious changes ( $s < 0$ ) have a very low probability of fixing, whereas beneficial mutations ( $s > 0$ ) have a much larger probability. A neutral allele fixes at the rate  $1/N$  haploids ( $1/2N$  in diploids), shown where the lines cross  $s = 0$ . Deleterious alleles have a substantially higher probability of fixing in smaller populations.

General form	New haploid mutation	Biological conditions
$\frac{1 - e^{-4Nsp}}{1 - e^{-4Ns}}$	$\frac{1 - e^{-2s}}{1 - e^{-4Ns}}$	Randomly mating diploids
$\frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}$	$\frac{1 - e^{-2s}}{1 - e^{-2Ns}}$	Haploids, and the Wright-Fisher process
$\frac{1 - e^{-Nsp}}{1 - e^{-Ns}}$	$\frac{1 - e^{-s}}{1 - e^{-Ns}}$	The Moran process (derived below)

TABLE 1. Probabilities of fixation in various models.

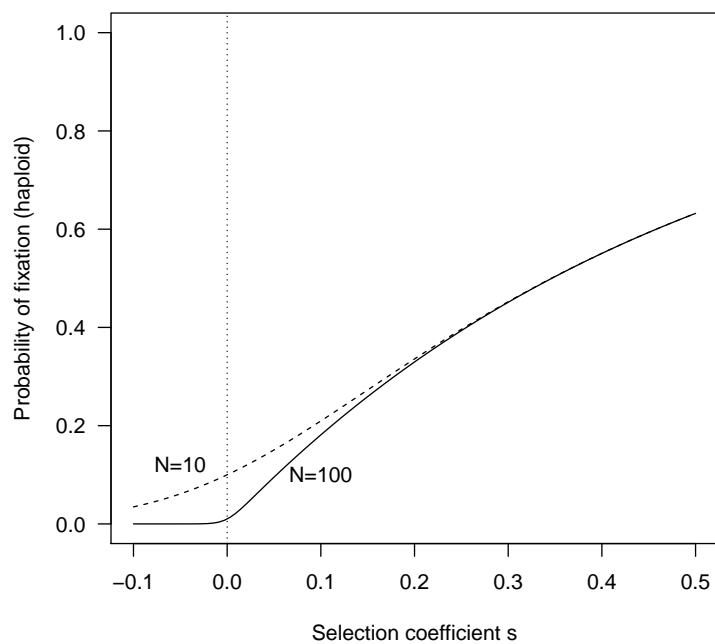


FIGURE 2. The probability of fixation of a new allele ( $p = 1/N$ ) in a haploid population.

## DERIVING THE PROBABILITY OF FIXATION

For a classic model of evolution, the Moran process, the probability of fixation can be determined without approximation. The result has the same form as other (approximate but more useful)  $p_{\text{fix}}$  formulae, making it worth exploring.

In the Moran process, a population of  $N$  individuals are born, reproduce, and die. At each time step, one individual is chosen to reproduce by binary fission, and one individual is chosen to die, preserving  $N$ . The probability that an individual reproduces is proportional to its fitness, while all individuals have an equal probability of death. A generation in this process is roughly  $N$  time steps, long enough for most organisms to reproduce once.

We will tackle the probability that a mutant allele conferring fitness advantage  $s$ , currently present in  $k$  out of the  $N$  organisms, becomes fixed, that is, grows until it is present in all  $N$  organisms. And we'll start with an odd pair of tricks. First, we note that there are three options for fixation starting from  $k$  copies (which we term  $p_k^{\text{fix}}$ : in the next time step, gain a copy, and then fix from  $k+1$  copies with probability  $p_{k+1}^{\text{fix}}$ ; lose a copy and fix from  $k-1$  copies with probability  $p_{k-1}^{\text{fix}}$ ; or neither gain nor lose a copy, and attempt to fix again with probability  $p_k^{\text{fix}}$ . Let the probability of gaining a copy (going "up") be  $\Pr(k \rightarrow k+1) = u_k$  and the probability of losing a copy be  $\Pr(k \rightarrow k-1) = d_k$ . Then,

$$p_k^{\text{fix}} = u_k p_{k+1}^{\text{fix}} + d_k p_{k-1}^{\text{fix}} + (1 - u_k - d_k) p_k^{\text{fix}}$$

so we collect terms in  $p_k^{\text{fix}}$  and get

$$p_k^{\text{fix}} = \frac{u_k}{u_k + d_k} p_{k+1}^{\text{fix}} + \frac{d_k}{u_k + d_k} p_{k-1}^{\text{fix}}$$

For our second trick, let's guess that  $p_k^{\text{fix}}$  takes a certain form,  $p_k^{\text{fix}} = S^k$ . The reason for this guess is that we can see a lot of  $k$  subscripts that would be turned into  $k$  exponents and therefore cancel each other if our guess were true. With that guess,

$$S^k = \frac{u_k}{u_k + d_k} S^{k+1} + \frac{d_k}{u_k + d_k} S^{k-1}$$

and doing the hoped-for cancellation, dividing by  $S^{k-1}$ , we obtain

$$S = \frac{u_k}{u_k + d_k} S^2 + \frac{d_k}{u_k + d_k}$$

which can be solved for  $S$  using the quadratic formula, yielding two solutions:  $S_1 = d_k/u_k$  and  $S_2 = 1$ . (Only a tiny hint here, to notice that the constants multiplying the  $S$  terms add to 1.) Now, linear combinations of these two solutions are also solutions, so in general,

$$p_k^{\text{fix}} = a_1 S_1^k + a_2 S_2^k = a_1 S_1^k + a_2.$$

To find the values of the  $a$  constants, we need to know the boundary conditions. And these are easy: the probability of fixation when  $k=0$  is zero (no fixation from death), and when  $k=N$  is one (already fixed!). Using these conditions, we have

$$\begin{aligned} p_0^{\text{fix}} = 0 &= a_1 + a_2 \\ a_1 &= -a_2 \end{aligned}$$



and

$$\begin{aligned} p_N^{\text{fix}} = 1 &= a_1 \left( \frac{d_k}{u_k} \right)^N + a_2 \\ 1 &= a_2 \left( 1 - \left( \frac{d_k}{u_k} \right)^N \right) \\ a_2 &= \frac{1}{1 - \left( \frac{d_k}{u_k} \right)^N} \end{aligned}$$

so that

$$(8) \quad p_k^{\text{fix}} = \frac{1 - \left( \frac{d_k}{u_k} \right)^k}{1 - \left( \frac{d_k}{u_k} \right)^N}.$$

What about these  $u_k$  and  $d_k$  terms? To deal with them, we circle back to thinking about the evolutionary process. What is  $u_k$ , the probability that the allele in question rises in frequency from  $k$  to  $k+1$  copies? We only know that the mutant allele has selective advantage  $s$ , and we know that individuals are chosen to reproduce proportional to their fitness. The fitness of the mutant is  $e^s$  times greater than the wild-type's by Eq. 1, so mutants will be chosen for reproduction  $e^s$  times more often than the wild type, all else equal. All else is not equal, because the mutant and wild type will occur at different frequencies, specifically  $k$  mutant and  $N-k$  wild-type individuals if we consider  $p_k^{\text{fix}}$ . So we need only think about the specific events that lead to  $k \rightarrow k+1$ —an increase in the number of mutants—and  $k \rightarrow k-1$ , a decrease.

Now  $k \rightarrow k+1$  requires that a mutant individual be chosen to reproduce (which has probability  $\frac{k}{N}e^s$ ) and a wild type be chosen to die (with probability  $\frac{N-k}{N}$ ) such that

$$\Pr(k \rightarrow k+1) = \frac{N-k}{N} \frac{k}{N} e^s.$$

By contrast,  $k \rightarrow k-1$  requires that a wild-type individual be chosen to reproduce, with probability  $\frac{N-k}{N}$ , and a mutant be chosen to die with probability  $\frac{k}{N}$ , such that

$$\Pr(k \rightarrow k-1) = \frac{k}{N} \frac{N-k}{N}.$$

The shockingly simple result is that

$$(9) \quad \frac{d_k}{u_k} = \frac{\Pr(k \rightarrow k-1)}{\Pr(k \rightarrow k+1)} = \frac{\frac{k}{N} \frac{N-k}{N}}{\frac{N-k}{N} \frac{k}{N} e^s} = e^{-s}.$$

And with Eqs. 8 and 9, we can complete the probability of fixation:

$$(10) \quad p_k^{\text{fix}} = \frac{1 - \left( \frac{d_k}{u_k} \right)^k}{1 - \left( \frac{d_k}{u_k} \right)^N} = \frac{1 - e^{-ks}}{1 - e^{-Ns}}$$

That's the exact probability that the mutant allele goes to fixation, given that it's present in  $k$  individuals. The case of  $k=1$  (a new mutant arising) is obvious.

Usually, the mutant frequency is expressed as a proportion,  $p = k/N$ , and in this case

$$\frac{1 - e^{-Nsp}}{1 - e^{-Ns}}$$

which is as close to the classical fixation formula as we'll get. This works for any  $s$ , large or small. The difference between this formula and Kimura's fixation formula for a haploid mutant allele involves only a factor of 2, as we saw in Table 1.