Probability of Fixation in a Population of Variable Size

Pavlos Kazakopoulos
Physics Department,
Massachusetts Institute of Technology
(Dated: May 12, 2005)

We compute the probability of fixation of an allele in a population of haploids where the progeny distribution of each individual is Poisson. We show that for an increasing population there is a finite probability of fixation of an allele while for a deceasing or steady one any mutation is inevitabely absorbed. We ignore random mutations and selection forces. The method used is a two-dimensional stochastic model.

I. INTRODUCTION

Although the assumption of fixed population size simplifies the solution for the probability of fixation of an allele considerably, for most natural populations the size varies over time. These variations may or not average to zero over many generations, but in the short term some variation is definitely the rule. This makes it an interesting problem to study how the probability for the fixation of an allele behaves in this case.

In [1] Chia uses the diffusion equation to study the probability of fixation in population with cyclical change in size when the magnitude of the changes is much smaller than the population size. Ewans in [2] and Kimura in [3] studied the same problem using a combination of stochastic methods and the diffusion equation. Hill removed the small magnitude restriction by using a matrix series expansion in [4].

In this paper we study the probability of fixation for a population of haploids that is finite but of variable size. We assume each to reproduce according to a Poisson distribution, with common parameter μ. As a first approximation we ignore the forces of random mutation and selection, keeping random genetic drift as the only force. We see that other these assumptions the system is solvable as a two-dimensional stochastic model. The main result is that the probability of fixation depends critically on the value of μ. More specifically, for $0 < \mu \leq 1$ this probability is one, while for $\mu > 1$ it is non-zero but less than one. The plan of the paper is as follows: In Section II we set up the problem and derive the general formula for the evolution of the probability distribution of an allele. Then, in Section III we examine the probability of fixation for different values of the parameter μ and discuss the results. Section IV contains a discussion and suggests directions for further study. These results were first obtained[7] in [6].

II. THE MODEL

Consider a population of haploid individuals with two different alleles A_1 and A_2 of a given gene. We assume that neither of the two alleles conveys an evolutionary advantage and that the rate of mutations is small enough that they can be ignored. Each individual reproduces independently of the others and the probability distribution for producing offspring is the same for all individuals. We take this to be the Poisson distribution with parameter μ [5]:

$$P(n) = e^{-\mu}\frac{\mu^n}{n!} \tag{1}$$

where n is the number of offspring produced by a single individual. Then, since the variables are independent, the probability that at the ith reproduction the size of the population goes from N_i to N_{i+1} is:

$$P(N_i \to N_{i+1}) = e^{-N_i\mu}\frac{(N_i\mu)^{N_{i+1}}}{N_{i+1}!}. \tag{2}$$

Note that $\mu < 1$ means that the population is (on the average) decreasing while $\mu > 1$ gives an increasing population. An ever-inreasing population is not a realistic assumption since enviromental factors will eventually limit the growth, so the model can only be valid for a limited period of time. A more realistic approach should treat μ as a time (step) dependent variable but this would make the problem a lot harder.

We also assume that at the ith generation the number of individuals in the population carrying the allele A_1 is x_i and define the frequences $p_i \equiv \frac{x_i}{N_i}$ and $q_i \equiv 1 - p_i$. The frequency of A_1 in the next generation is governed by the binomial distribution:

$$P(x_i \to x_{i+1}) = \left(\begin{array}{c} N_{i+1} \\ x_{i+1} \end{array}\right) p_i^{x_{i+1}} q_i^{N_{i+1}-x_{i+1}}. \tag{3}$$

As mentioned before, this is the only force we will consider for changing the distribution of alleles.

With these assumptions, the joint distribution for going from the pair (N_i, x_i) to (N_{i+1}, x_{i+1}) is the product

$$\begin{aligned} &P(N_i; x_i \to N_{i+1}; x_{i+1}) \\ &\quad = e^{-N_i\mu}\frac{(N_i\mu)^{N_{i+1}}}{N_{i+1}!}\left(\begin{array}{c} N_{i+1} \\ x_{i+1} \end{array}\right) p_i^{x_{i+1}} q_i^{N_{i+1}-x_{i+1}}. \end{aligned} \tag{4}$$

This is the "master equation" for this particular stochastic model. To derive the probability of fixation we have to study the behavior of this over repeated iterations. We do this in the next section.

III. PROBABILITY OF FIXATION

The transient states T for the system under consideration are those in which $0 < x_i < N_i$ and the absorbsion states A those with $x_i = 0$ or N_i. We want to compute the probability that a system starting from some point in T will ultimately end up in A. Since we have ignored mutations, once the population is in A it remains there forever.

We can define the probability that a system in T with initial conditions x_0, N_0 remains in T after i generations recursively as:

$$\mathcal{P}_i(x_0, N_0) = \sum_{(x_{i+1}, N_{i+1}) \in \mathrm{T}} P(N_i; x_i \to N_{i+1}; x_{i+1}) \mathcal{P}_{i-1}(x_0, N_0) \tag{5}$$

with

$$\mathcal{P}_0(x_0, N_0) = \sum_{(x_1, N_1) \in \mathrm{T}} P(N_0; x_0 \to N_1; x_1). \tag{6}$$

Using eq.(4) we can compute a quasi-explicit form for $\mathcal{P}_i$. The computation is similar to going from a canonical to a grand canonical ensemble in statistical mechanics. The result is:

$$\mathcal{P}_i = 1 - e^{-x_0 \mu_i} + e^{-N_0 \mu_i} - e^{-(N_0 - x_0)\mu_i} \tag{7}$$

where μ_i is defined recursively through

$$\mu_i = \mu(1 - e^{-\mu_{i-1}}) \quad , \quad \mu_0 = \mu. \tag{8}$$

All we need to do now is study the limiting value M of μ_i after many generations. This is the real root of the equation $M = \mu(1 - e^{-M})$. For $0 < \mu \leq 1$ the only real solution is $M = 0$. This makes $\mathcal{P}_i$ vanish for large i. For $\mu > 1$ there are two solutions, zero and $M' > 0$. It is not hard to show that the sequence of μ_i's actually converges to the positive solution in this case. For large μ we have $M \approx \mu$ and $M < \mu$ always. $\mathcal{P}_i$ approaches a finite positive value $\mathcal{P}$ in this case:

$$\mathcal{P} = 1 - e^{-x_0 M'} + e^{-N_0 M'} - e^{-(N_0 - x_0)M'}. \tag{9}$$

The interpretation is now straightforward. For $0 < \mu \leq 1$ one of the two alleles is ultimately absorbed, while for $\mu > 1$ (increasing population) there is a finite probability that both alleles will ultimately survive, given by (9). Note that larger values of μ are detrimental to the probability of fixation (or loss). This is also true for values of $\frac{N_0 - x_0}{N_0}$ closer to $\frac{1}{2}$. The next step in this computation would be to find the probability that fixation will happen after a given number of steps. This would allow one to calculate the expected number of generations before fixation (or loss) occurs. This has been done in [6]. The expected time to fixation is of course undefined for $\mu > 1$. For $\mu < 1$ it grows as μ approaches one from below, becoming infinite in this limit.

IV. DISCUSSION

We saw how to compute the probability of fixation for a haploid population of variable size using a stochastic model with two variables. The model uses a few assumes that the progeny distribution is Poisson and ignores random mutations and selections. These assumptions obviously limit the applicability, and it would be desirable to drop them. Another limiting factor is the constant value of μ which cannot be a realistic assumption over many generations. The problem does not seem to be solvable for a generic deterministic μ that depends on time, but maybe something can be done if μ is treated as a random variable. The result we obtained however agrees with the naive intuition. One would expect fixation to occcur in a decreasing population, and on the other hand an increasing population should be able to support both alleles.

[1] A. Chia "Random mating in a population of finite size," *J. Appl. Probability 2*,1968, 21-30.

[2] J. Ewens "The probability of survival of a new mutant in a fluctuating environment," *Heredity 22*,1967, 438-43.

[3] M. Kimura "On the probability of fixation of mutant genes in a population," *Genetics 47*,1962, 713-19.

[4] W. Hill " Probability of fixation of genes in populations of variable size," *Theor. Popul. Biol. 3*,1972, 27-40.

[5] S. Karlin and J. McGregor, "The role of the Poisson progeny distribution in population genetic models," *Math. Biosci.2*,1968, 11-17.

[6] R. Cook and R. Nassar "Dynamics of finite populations. I. The expected time to fixation or loss and the probability of fixation of an allele in a haploid population of variable size ," *Biometrics 2*,1972, 373-84.

[7] I derived all the results except eq.(8) independently. I became aware of [6] only yesterday and adopted some of its formalism and jargon because they seem to me more "professional" than what I was using.