

RANDOM SELECTION — A SIMPLE MODEL BASED ON LINEAR BIRTH AND DEATH PROCESSES†

■ P. SCHUSTER and K. SIGMUND‡

Institut für Theoretische Chemie und Strahlenchemie and

‡Institut für Mathematik,

Universität Wien,

A-1090 Wien, Austria

Linear birth and death processes are used to derive simple expressions for sequential extinction times and gene fixation probabilities in asexual populations.

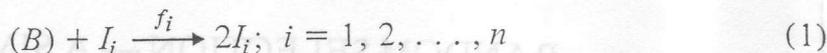
1. Introduction. The synthetic neo-Darwinian theory has been challenged by the development of the neutral theory. According to this theory the gene pool of species is also subject to evolutionary change in constant environment and in absence of selection. The basic idea is that a substantial percentage of those mutations not immediately eliminated is selectively neutral and may reach fixation in the population by recurrence and random drift. A recent mathematical discussion of the neutral theory clarifying the notions of average evolution rates and protein polymorphism is found in Ishii *et al.* (1982).

More and more data favouring the neutral theory have become available during the last two decades (Kimura, 1982). Nevertheless, the neutralist's view has not yet been accepted completely by the majority of biologists. Originally, molecular evolution was studied almost exclusively on diploid organisms. More recently, data have become available also for haploid plants (Yamazaki, 1982) and bacteria (Milkman, 1973), where overdominance is excluded as a variability-maintaining factor. Here, one finds the same phenomena as observed with diploid organisms and predicted by the neutral theory: protein polymorphism and regularities in nucleotide replacement.

In this paper 'fixation' of neutral mutant alleles in asexually replicating populations is described by a simple model system, which is identical with one used for *in vitro* RNA replication. The analysis makes use of well-known results of linear birth and death processes (Harris, 1963; Jagers, 1975) and derives simple expressions for expectation values of extinction times. The probability for fixation of newly formed mutants is similar to the classical expressions for random drift in population genetics (Crow and Kimura, 1970).

†This work was supported financially by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung, Project Nos. 3502 and 4506.

2. *Independent Birth and Death Processes.* An appropriate model system consists of $2n$ parallel reactions, n autocatalytic replication steps and n decay reactions:



All $2n$ reactions proceed under conditions of practical irreversibility. The n replicating elements are denoted by I_1, I_2, \dots, I_n . The raw material from which the elements are built, B , and the decomposition product, M , do not enter as variables into the model system: the amount of B is assumed to be constant or buffered and M is the end product of an irreversible reaction step. We describe the evolution of our system by a set of stochastic variables $X_1(t), X_2(t), \dots, X_n(t)$, where $X_i(t)$ is the number of replicating elements of type I_i at the time t . The corresponding probability densities are given by

$$P_{x_i}^{(i)}(t) = \text{Prob}\{X_i(t) = x_i\}; \quad i = 1, 2, \dots, n \quad (3)$$

and $x_i = 0, 1, 2, \dots$

Since we assumed independence, the joint probability density is simply given by the product of the individual probability densities:

$$P_{x_1, \dots, x_n}(t) = \text{Prob}\{X_1(t) = x_1, \dots, X_n(t) = x_n\} \\ = P_{x_1}^{(1)}(t) \cdot \dots \cdot P_{x_n}^{(n)}(t). \quad (4)$$

As is well known, the probabilities satisfy the master equations

$$\frac{d}{dt} P_{x_i}^{(i)} = (x_i - 1)f_i P_{x_i-1}^{(i)} + (x_i + 1)d_i P_{x_i+1}^{(i)} - x_i(f_i + d_i)P_{x_i}^{(i)}$$

corresponding to the linear birth and death process (see, e.g. Bartholomay, 1958; Jagers 1975).

In our model studies we assume $f_i = d_i$. This is the case where the deterministic approach based on conventional kinetic equations fails. Indeed, it yields the differential equations $\dot{x}_i = (f_i - d_i)x_i = 0$ for which x_i remains constant, while the stochastic model eventually leads with probability 1 to extinction, as random fluctuations are going to build up.

In order to treat neutral mutations we study the degenerate case of maximum kinetic equivalence first: $f_1 = f_2 = \dots = f_n = f$. Despite the kinetic equivalence, the n replicating elements are distinguishable: we assume the existence of an appropriate analytical technique to separate them. In the case of polynucleotides, sequence analysis provides a tool to distinguish between neutral mutants. The main conclusions will also be valid in systems with distinct f values.

In Section 3 we discuss the sequence of extinction times and in Section 4 the probabilities for random selection of individual molecular species.

3. *Sequential Extinction Time.* Let us start with the (inessential) assumption that initially, each replicating element is present in a single copy. As is well known (see, e.g. Bartholomay, 1958), the probability that a given line is extinct at time t is

$$P_0(t) = \frac{ft}{1 + ft}. \quad (5)$$

Eventual extinction, then, is certain:

$$\lim_{t \rightarrow +\infty} P_0(t) = 1, \quad (6)$$

although the expectation value for the number of elements at time t is always equal to the initial number, i.e. 1.

Let us denote by T_k the time up to the extinction of $n - k$ types of replicators. This waiting time T_k is the random variable, which we characterize as 'sequential extinction time'. Thus n types of replicators are present between $T_n (\equiv 0)$ and T_{n-1} , $n - 1$ replicators between T_{n-1} and T_{n-2} etc. Between T_1 and T_0 only one kind of replicating element is still existing—the population is now uniform—and finally T_0 is the moment of extinction.

The probability distribution of T_k ,

$$H_k(t) = \text{Prob}\{T_k < t\}, \quad (7)$$

is calculated from the probability of extinction of a given replicator, $P_0(t)$. We start with the simplest case, $k = 0$. The probability H_0 is the probability that all replicators have vanished.

$$H_0 = P_{0,0,\dots,0} = P_0^{(1)} \cdot P_0^{(2)} \cdot \dots \cdot P_0^{(n)} = \left(\frac{ft}{1 + ft} \right)^n. \quad (8)$$

The event $T_1 < t$ can happen in several ways: either I_1 is present and all the other replicators have become extinct, or only I_2 is present, or only I_3 etc. $T_1 < t$, however, is also fulfilled if all replicators have died out:

$$H_1 = P_{x_1 \neq 0, 0, \dots, 0} + P_{0, x_2 \neq 0, \dots, 0} + \dots + P_{0, 0, \dots, x_n \neq 0} + H_0. \quad (9)$$

The probability that a given replicator is still present is obtained easily since existence and non-existence are complementary:

$$P_{x \neq 0} = 1 - P_0 = \frac{1}{1 + ft}. \quad (10)$$

Thus we find

$$P_{x_1 \neq 0, 0, \dots, 0} = P_{0, x_2 \neq 0, \dots, 0} = \dots = P_{0, 0, \dots, x_n \neq 0} = \frac{(ft)^{n-1}}{(1+ft)^n} \quad (11)$$

and

$$H_1 = (n+ft) \frac{(ft)^{n-1}}{(1+ft)^n}. \quad (12)$$

Similarly, we derive all higher probability distributions by the recursion formula

$$H_k = \binom{n}{k} \frac{(ft)^{n-k}}{(1+ft)^n} + H_{k-1}, \quad (13)$$

which finally yields in the formula

$$H_k = \sum_{j=0}^k \binom{n}{j} \frac{(ft)^{n-j}}{(1+ft)^n}. \quad (14)$$

Setting

$$A_j(t) = \binom{n}{j} \frac{(ft)^{n-j}}{(1+ft)^n}$$

one obtains for its derivative

$$A_j'(t) = \frac{f}{(1+ft)^{n+1}} \left[\binom{n}{j} (n-j) (ft)^{n-j-1} - \binom{n}{j} j (ft)^{n-j} \right].$$

One easily verifies that $H_k'(t) = A_k'(t) + A_{k-1}'(t) + \dots + A_0'(t)$ is a telescopic sum where all terms cancel except the first one. Thus the probability density for T_k is given by

$$H_k' = (n-k) \binom{n}{k} f^{n-k} \frac{t^{n-k-1}}{(1+ft)^{n+1}}. \quad (15)$$

In order to calculate the expectation value of T_k we have to evaluate the definite integral (see Gradshteyn and Ryzhik, 1965), $k \geq 1$:

$$\int_0^\infty \frac{t^{n-k}}{(1+ft)^{n+1}} dt = f^{-(n-k+1)} \sum_{i=0}^{n-k} \frac{(-1)^i}{k+i} \binom{n-k}{i} = \frac{f^{-(n-k+1)}}{k \binom{n}{k}}; \quad (16)$$

the analogous integral for $k=0$, however, diverges.

The expectation values for the sequential extinction times T_k are thus

$$E\{T_k\} = \int_0^\infty H'_k \cdot t dt = \frac{n-k}{k} \cdot \frac{1}{f}; \quad n \geq k \geq 1 \quad (17)$$

and $E\{T_0\} = \infty$. In the figure we illustrate this series for a concrete numerical example.

Similarly, we compute the variance of the sequential extinction times T_k :

$$D^2\{T_k\} = \frac{n(n-k)}{k^2(k-1)} \cdot \frac{1}{f^2}; \quad n \geq k \geq 2. \quad (18)$$

Note that the variance diverges for $k=0$ and $k=1$.

In the general case, the rate parameters f_i are distinct and the initial particle numbers x_1^0, \dots, x_n^0 are not all equal to 1. The expressions for $E\{T_k\}$ become considerably more complicated, but the main conclusion is unaffected: $E\{T_1\}$ is finite and $E\{T_0\}$ is not.

Indeed, the derivative of $(1 + f_i t)^{-1}$ is just the opposite of that of $f_i t(1 + f_i t)^{-1}$, namely $-f_i(1 + f_i t)^{-2}$. Now if $x_1^0 = \dots = x_n^0 = 1$, then

$$H_0 = \prod_i \left(\frac{f_i t}{1 + f_i t} \right)$$

and its derivative H'_0 is a rational function with leading term t^{-2} . Hence $H'_0 t$ is not integrable in $[0, +\infty)$. On the other hand, H_1 is the sum of H_0 and expressions of the type

$$\left(\frac{1}{1 + f_i t} \right) \prod_{j \neq i} \left(\frac{f_j t}{1 + f_j t} \right).$$

It is easy to see that H'_0 cancels with the sum of the expressions

$$\left(\frac{1}{1 + f_i t} \right)' \prod_{j \neq i} \left(\frac{f_j t}{1 + f_j t} \right).$$

H'_1 then consists of expressions of the type

$$\left(\frac{1}{1 + f_i t} \right) \left(\frac{f_k t}{1 + f_k t} \right)' \prod_{j \neq i, k} \left(\frac{f_j t}{1 + f_j t} \right),$$

which are rational functions with leading term t^{-3} . Hence $H'_1 t$ is integrable in $[0, +\infty)$. If the initial concentrations x_i^0 are not all equal to 1, then the expression for H_0 is like the one given before, except that the terms $(f_i t)(1 + f_i t)^{-1}$ are repeated x_i^0 times in the product. If we denote by T'_1 the time until there are only copies of one individual left, the preceding argument

shows that T_1' is integrable and T_0 is not. Finally, the time T_1 until there are only replicating elements of one type left is certainly not larger than T_1' . Hence T_1 is integrable.

After T_1 , one type of replicating element has been selected. It will eventually die out, but since the random interval $[T_1, T_0]$ during which the ensemble is homogeneous has a length whose expectation value is infinite, we may interpret this as 'fixation' of the corresponding type, in analogy to the diploid case.

4. *Probabilities for Random Selection.* If all f_i are equal and the ensemble consists initially of one copy of each type, the probability that a given type is selected as the 'last survivor' is, of course, $1/n$. If the f_i are distinct, the probability that at time t copies of type I_i survive while all others have vanished is

$$\left(\frac{1}{1+f_i t}\right) \prod_{j \neq i} \left(\frac{f_j t}{1+f_j t}\right) = \frac{1}{f_i t} \prod_{j=1}^n \left(\frac{f_j t}{1+f_j t}\right). \quad (19)$$

The probability that type I_i is the exclusive survivor at time t , under the condition that fixation has occurred, is

$$\frac{1}{f_i t} \left\{ \sum_j \frac{1}{f_j t} \right\}^{-1}, \quad (20)$$

which is independent of t . Hence the probability for the eventual fixation of I_i is just the same expression, namely

$$\frac{f_i^{-1}}{f_1^{-1} + \dots + f_n^{-1}}. \quad (21)$$

Now if the initial ensemble consists of x_j^0 copies of I_j (for $j = 1, \dots, n$), the probability for the fixation of I_i is

$$\frac{x_i^0 f_i^{-1}}{\sum_{j=1}^n x_j^0 f_j^{-1}}. \quad (22)$$

In particular, if a mutation occurs in a homogeneous ensemble of x copies, and if this mutation is neutral in the sense that it does not affect the rate f , then the probability of its fixation is $1/x$. This is similar to the classical result, where the probability of fixation of a neutral mutation in a population of x diploid individuals is $1/2x$ (see, e.g. Kimura, 1982, p. 20).

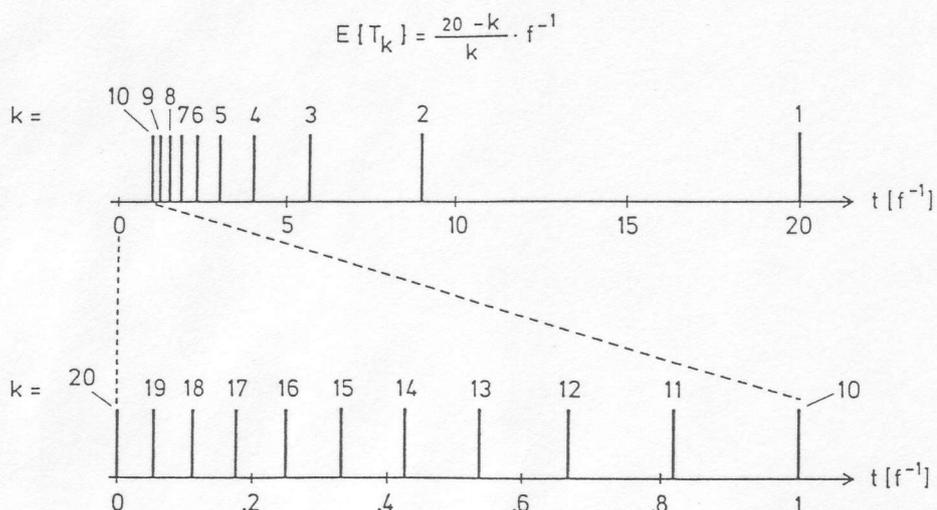


Figure 1. The distribution of sequential extinction times, $E\{T_k\}$, for $n = 20$.

LITERATURE

- Bartholomay, A. F. 1958. "On the Linear Birth and Death Processes of Biology as Markoff Chains." *Bull. math. Biophys.* 20, 97-118.
- Crow, J. F. and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. New York: Harper & Row.
- Gradshteyn, I. S. and I. Ryzhik. 1965. *Table of Integrals, Series and Products*. New York: Academic Press, p. 292.
- Harris, P. E. 1963. *The Theory of Branching Processes*. Berlin: Springer.
- Ishii, K., H. Matsuda and N. Ogita. 1982. "A Mathematical Model of Biological Evolution." *J. math. Biol.* 14, 327-353.
- Jagers, P. 1975. *Branching Processes with Biological Application*. London: Wiley.
- Kimura, M. 1982. *Molecular Evolution, Protein Polymorphism and the Neutral Theory*. Berlin: Springer.
- Milkman, R. 1973. "Electrophoretic Variation in *E. coli* from Natural Sources." *Science* 182, 1024-1026.
- Yamazaki, T. 1982. "Genic Variability in Natural Populations of the Haploid Plant, *Conocephalum conicum*. In *Molecular Evolution, Protein Polymorphism and the Neutral Theory*, Ed. M. Kimura, pp. 123-134. Berlin: Springer.

RECEIVED 11-2-82

REVISED 1-27-83