Theoretical and Applied Genetics 45, 300–303 (1975)

Dynamics of Finite Populations

III. A Note on the Rate of Approach to Homozygosity in a Haploid Population whose Size is a Poisson Random Variable[1]

R. F. NASSAR and R. D. COOK

Department of Statistics, Kansas State University, Manhattan, Kansas (USA) and
School of Statistics, University of Minnesota, St. Paul, Minnesota (USA)

Summary. A time-heterogeneous stochastic process is used to describe the rate of approach to homozygosity, the expected time to fixation or loss and the ultimate probability of survival of a gene or type in a haploid population whose size is a Poisson random variable.

1. Intoduction

In a previous study, Cook and Nassar (1972a) and Nassar and Cook (1973) investigated the time to fixation or extinction of a gene and its probability of survival in a haploid population whose size was assumed to be a Poisson random variable with homogeneous parameter. In nature there are cases where a population size may be cyclic over time (Elton and Nicholson, 1952; Odum, 1959). Also situations might exist when a population grows in size in a new environment and then stabilizes due to limited resources. It is possible also that a stable population might encounter an environmental situation which causes it to decrease in size. Situations of this sort, where the average population size is variable over time, are best represented by a heterogeneous branching process where the average progeny number per individual can vary from generation to generation. There have been previous studies on the probability of survival of a gene using heterogeneous branching processes (Koutsky, 1962; Kojima and Kelleher, 1962; Pollak, 1966; and others). However, the limiting factor in these studies is the assumption of infinite population size.

The purpose of this note is to show how the same model in Nassar and Cook (1973) can be easily adapted to a heterogeneous case of varying average progeny number and to investigate the rate of approach to homozygosity.

2. Probability of Transition

Referring to Nassar and Cook (1973), we can express equation (3.2) in the form:

$$X_t^n(k, m) = 1 - exp\,[-k\alpha_n(t)] + exp\,[-k\alpha_n(t) - (m-k)\,\beta_n(t)] - exp\,[-(m-k)\,\beta_n(t)] \qquad (2.1)$$

where

$$\left.\begin{aligned} \alpha_1(t) &= \alpha(t) \\ \alpha_2(t) &= \alpha(t)\,(1 - e^{-\alpha_1(t+1)}) \\ \alpha_3(t) &= \alpha(t)\,(1 - e^{-\alpha_2(t+1)}) \\ &\vdots \\ \alpha_n(t) &= \alpha(t)\,(1 - exp\,[-\alpha_{n-1}(t+1)] \end{aligned}\right\} \qquad (2.2)$$

The same type of equation is true of $\beta_i(t)$, $i = 1, 2, \ldots, n$. Note here that α and β are allowed to vary from generation to generation. The sequences $\{\alpha_n(t)\}$ and $\{\beta_n(t)\}$ are decreasing sequences bounded below by zero. The ultimate probability of absorption is one $[\lim_{n\to\infty} X_t^n(k, m) = 0]$ if zero is the limit over n for both or either of $\alpha_n(t)$ and $\beta_n(t)$, otherwise it is less than 1. This is easily seen from inspection of (2.1). From (2.2), it is seen that

$$\lim_{n\to\infty} \alpha_n(t) = L_\alpha(t) = \alpha(t)\,(1 - e^{-L}\,\alpha^{(t+1)}) . \qquad (2.3)$$

Therefore, it is sufficient to consider the case $t = 0$.

Because of the recursive relationship between the limits and the arbitrary nature of the sequences, $\alpha_n(t)$ and $\beta_n(t)$, it is difficult to use (2.3) to determine necessary and sufficient conditions for absorption to occur with probability 1. However, under rather mild restrictive conditions it can be shown (see Cook and Nassar, 1972b) that

$$\lim_n \alpha_n(0) = 0$$

if and only if

$$\sum_{v=0}^{\infty} \prod_{j=0}^{v} \frac{1}{\alpha(j)} = \infty \qquad (2.4)$$

when (2.4) holds, it follows immediately that

$$\lim_n X_t^n(k, m) = 0 \text{ for all } t .$$

The same is true of $\beta(t)$.

[1] Contribution Number 166, Department of Statistics, Statistical Laboratory, Kansas Agricultural Experiment Station, Manhattan, Kansas 66506.

Using (2.4) it is easily seen that if $\alpha(t)$ is constant (c_α) after some finite initial period of time, absorption occurs with probability one if and only if $0 < c_\alpha \leq 1$. This is the same condition as shown by Cook and Nassar (1972a). Note that (2.4) is satisfied if $lim\, \alpha(t) = c_\alpha$, $0 < c_\alpha < 1$. If $lim\, \alpha(t) = c_\alpha = 1$, then for (2.4) to be satisfied the convergence of $\alpha(t)$ must be relatively fast. An example would be $\alpha(t) = 1 + ce^{-kt}$. On the other hand $\alpha(t) = 1 + 2/t$ is an example of a sequence converging to 1 for which (2.4) is not satisfied.

As stated in the introduction, some populations might be cyclic in their size pattern over time. Assuming $\alpha(t) = \alpha(t + cn)$, $\beta(t) = \beta(t + cn)$, $n = 0, 1, 2, \ldots$, (2.1) can be written in the form

$$X_t^{cn}(k, m) = 1 - exp\,(-k\alpha_{cn}(t)) + exp\,(-k\alpha_{cn}(t) - (m - k)\,\beta_{cn}(t)) - exp\,(-(m - k)\,\beta_{cn}(t)) \quad (2.5)$$

where c is the cycle length and n is the cycle number. In expression (2.5)

$$\alpha_{cn}(t) = \alpha(t)\left[1 - exp\left[-\alpha(t+1)\left[1 - exp\left[-\alpha(t+2) \ldots [1 - exp[-\alpha(t+c-1)\cdot[1 - exp\,[-\alpha_{cn-1}(t)]]]]\right]\right]\right]\right]. \quad (2.6)$$

The sequence $\{\alpha_{cn}(t)\}$ is a decreasing sequence and has a limit. Taking the limit on both sides of (2.6) we get

$$L_{\alpha c}(t) = \alpha(t)\left[1 - exp\left[-\alpha(t+1)\left[1 - exp\left[-\alpha(t+2) \ldots [1 - exp\,[-\alpha(t+c-1)\cdot[1 - exp\,[-L_{\alpha c}(t)]]]]\right]\right]\right]\right]. \quad (2.7)$$

From (2.7),

$$L_{\alpha c}(t) = 0 \quad \text{if} \quad \prod_{i=0}^{c} \alpha(i) \leq 1\,.$$

The proof follows closely that of (2.4). Analogous expressions, of course, hold for $\beta_{cn}(t)$ and $L_{\beta c}(t)$.

Expression (2.4) is satisfied if

$$\prod_{i=0}^{c} \alpha(i) \leq 1\,.$$

Thus, for populations which are cyclic in their size patterns over time a necessary and sufficient condition for absorption to occur with probability 1 is that the product of the average progeny numbers per individual of at least one type in the population be less than or equal to 1.

Under this model expressions for the ultimate probability of fixation or loss and the expected time to absorption can be easily obtained by using the sequences defined by (2.2) and the expressions given in sections 4 and 5 of Nassar and Cook (1973). The procedure is to simply replace β_n and α_n in Nassar and Cook (1973) by $\beta_n(t)$ and $\alpha_n(t)$ respectively. These expressions will not be restated here. The extension to more than two types is also immediate and can be obtained from Section 6 in Nassar and Cook (1973) by replacing $\alpha_{k,n}$ and $\beta_{k,n}$ by $\alpha_{k,n}(t)$ and $\beta_{k,n}(t)$.

3. Rate of Approach to Homozygosity

The theory of finite Markov chains enables one to say that the probability that both types of individuals are present in the nth generation is assymptotically of the order of

$$c(k)\,\lambda^n \quad n \to \infty \quad (3.1)$$

for some $0 < \lambda < 1$ where $c(k)$ is independent of n and depends only on the initial population configuration. The number λ is commonly called the rate of approach to homozygosity. It indicates that the extent of heterozygosity decreases essentially by a constant factor of λ per generation. This theory is applicable only to finite Markov chains and, unfortunately, an analogous theory for the case of an infinite chain is not available. That is, under this model, we are not assured that $X_0^n(k, m)$ (the probability that both types of individuals are present in the nth generation) is always of the order of (3.1). In what follows, when $X_0^n(k, m)$ is of the order $f(n)$, we shall refer to the rate of approach to homozygosity provided $f(n)$ is of the form (3.1); otherwise, we shall use $f(n)$ itself as a measure of the extent of heterozygosity.

Expanding each exponential function in (2.1) for $t = 0$, it is seen that the convergence of $X_0^n(k, m)$ is controlled by the dominant term, $\alpha_n(0)\,\beta_n(0)$. Also, it is seen that $\alpha_n(0)\,\beta_n(0)$ is of the order of

$$[A_nB_n]^{-1} \quad n \to \infty \quad (3.2)$$

where

$$B_n = \sum_{v=0}^{n} \prod_{j=0}^{v} \frac{1}{\beta(j)} \quad \text{and} \quad A_n = \sum_{v=0}^{n} \prod_{j=0}^{v} \frac{1}{\alpha(j)}\,.$$

Using (3.2) we now consider the rate of approach to homozygosity for three important special cases. Notice that if (3.2) is of the order $f(n)$ then so is $X_0^n(k, m)$.

a) $\lim_t \alpha(t) = c_\alpha < 1$ and $\lim_t \beta(t) = c_\beta < 1$ — Here it is seen that (3.2) and (therefore $X_0^n(k, m)$) is of the order $(c_\alpha c_\beta)^n$ as $n \to \infty$. Therefore, as in the case of a finite Markov chain, the extent of heterozygosity decreases essentially by a constant factor $(c_\alpha c_\beta)$ per generation.

b) $c_\alpha = c_\beta = 1$ — In this situation, it is difficult to present the rate of approach to homozygosity in a form that is more meaningful than (3.2). However, if we attach the conditions

$$0 < \prod_{t=0}^{\infty} \frac{1}{\alpha(t)} < \infty \quad \text{and} \quad 0 < \prod_{t=0}^{\infty} \frac{1}{\beta(t)} < \infty \quad (3.3)$$

then it can be shown that (3.2) is of the order $(1/n)^2$ as $n \to \infty$. It is felt that in most cases of practical interest (3.3) will be satisfied. For example, let

$\alpha(t) = 1 + xe^{-\lambda t}$ ($\lambda > 0$, $x > 0$). Here it is easily verified that (3.3) is satisfied and therefore $X_0^n(k, m)$ is of the order $(1/n)^2$ as $n \to \infty$. These forms were used by Kimura and Ohta (1969) to investigate the probability of fixation of a mutant gene in a finite population when selective advantage decreases with time.

c) $\alpha(t) = \alpha(t + cn)$, $\beta(t) = \beta(t + cn)$ — In this case, it is convenient to consider X_0^{cn} rather than X_0^n; that is, we will consider the rate of approach to homozygosity on a per cycle rather than a per generation basis. Let

$$P_\alpha = \prod_{t=1}^{c} \alpha(t) \quad \text{and} \quad P_\beta = \prod_{t=1}^{c} \beta(t) .$$

If $p_\alpha < 1$ and $p_\beta < 1$, it can be shown that (3.2) is of the order $(p_\alpha p_\beta)^n$ and therefore, the rate per cycle of approach to homozygosity is $p_\alpha p_\beta$. If $p_\alpha = p_\beta = 1$ then X_0^{cn} is of the order $(1/n)^2$.

It should be noted that the rate of approach to homozygosity can be easily determined for any combination of the above cases. For example, if $c_\alpha < 1$ and $c_\beta = 1$ then X_0^n is of the order c_α^n/n.

4. Discussion

To demonstrate the effect a variable number of progeny from generation to generation would have on the dynamics of a mutant gene, we evaluated (for certain $\alpha(t)$ and $\beta(t)$ values) the expected time to fixation, given that fixation occurs, and the ultimate probability of fixation. The results were compared with the homogeneous case. The ultimate probability of fixation is given as

$$D_t(k, m) = \sum_{j=1}^{\infty} D_t^j(k, m)$$

and the expected time to fixation, given that the gene is ultimately fixed, as

$$E_t(k, m) = \sum_{j=1}^{\infty} j\, D_t^j(k, m)/D_t(k, m) .$$

A selectively neutral gene in a population of initial size 100 which is non-cyclic ($\alpha = \beta = 1$) has an ultimate probability of fixation μ_x of .00994 and an expected time to fixation t_x of 193. If we choose the average number of progeny for the gene to be cyclic every two generations with $\alpha(0) = 2$, $\alpha(1) = .5$ and $\beta(0) = \beta(1) = 1$, the ultimate probability of fixation is found to be .0113 and the time to fixation to be 195.2. Thus very little increase in μ_x and t_x has been achieved inspite of the drastic fluctuation in the selection intensity of the gene. However, if the average number of progeny is cyclic for the population as a whole, ($\alpha(0) = \beta(0) = 2$ and $\alpha(1) = \beta(1) = 1/2$) then μ_x is found to be .00994 and t_x 259.3. Hence, a noticeable increase in t_x results when the population oscillates as a unit. In another example where a cycle occurs every 6 generations with $\beta(t) = \alpha(t) = (4, \frac{1}{2}, \ldots, \frac{1}{2}, 4)$; $t = 0, 1, \ldots, 5$ the time to fixation was decreased from that of the homogeneous case ($t_x = 134.3$). Also a small decrease in the ultimate probability of fixation ($\mu_x = .0098$) was observed. If the mutant gene is cyclic, but not the rest of the population ($\beta(t) = 1$), the results show that t_x is not affected ($t_x = 194.8$), but μ_x is decreased ($\mu_x = .0068$). These two examples illustrate that a mutant gene, that may be neutral on the average over a cycle, can in effect become advantageous or disadvantageous depending on its pattern of oscillation. If the gene occurs at the point in time when its average number of progeny is decreasing before it increases again in the cycle, then the gene is going to be on the average selected against and vice versa. As our second example points out, even a gene that occurs at a point in time where an expansion is taking place can still be selected against if contraction is for a relatively long number of generations thereafter. The above argument is general in that it applies also to genes that are not selectively neutral. The time to fixation of a mutant gene is seen to increase or decrease (from the case of constant fitness) depending on the pattern of oscillation in the population. The effect of cycling on the time of fixation is most pronounced when the population oscillates as a whole. Oscillation in the fitness of the mutant gene alone has little effect on its time to fixation.

Fossil records indicate that a large percentage (i.e., probably about 99%) of once existent populations are now extinct. Extinction of natural populations would, therefore, seem to be the rule rather than the exception. Also, it is generally concluded that extinction usually occurs after the population has been fixed in one pure type. For fixation to occur with probability one, under this model, we must have $\lim_t \alpha(t) \leq 1$ for at least $r - 1$ of the types in the population. Furthermore, it can be shown that, if $\lim_t \alpha(t) = 1$ for all types then the expected time to extinction is infinite. Thus, it would seem that in most populations we would expect $\lim_t \alpha(t) < 1$; for otherwise there would probably be a much smaller percentage of once existent populations now extinct.

In the two-type case, the probability that the process is transient in generation n was found to be of the order

$$(c_\alpha c_\beta)^n \text{ if } \lim_t \alpha(t) = c_\alpha < 1 \text{ and } \lim_t \beta(t) = c_\beta < 1;$$

$$(c_\alpha)^n/n \text{ if } c_\beta = 1;\ (c_\beta)^n/n \text{ if } c_\alpha = 1 \text{ and } (1/n)^2 \text{ if } c_\alpha = c_\beta = 1 .$$

In view of preceding comments, it would seem that the extent of heterozygosity decreases esentially by a constant factor per generation in most populations and that the much slower rate of approach to homozygosity would be found only in a small percentage of populations. Analogous remarks can of course be made about populations that are cyclic in nature.

When the sequence of average progeny is cyclic the rate, per generation, of approach to homozygosity is $\sqrt[c]{p_\alpha p_\beta}$ ($p_\alpha < 1$, $p_\beta < 1$). It is of interest to note that if the average progeny number per individual for a non-cyclic population is chosen to be constant and equal to the mean of the progeny numbers within a cycle of a cyclic population, then the cyclic population will approach homozygosity at a rate faster than that of the noncyclic population. This follows by the familiar relation between arithmetic and geometric means.

The results of this model are more general than they might seem. We have found through simulation (Cook and Nassar 1972) that setting an upper limit on population growth (the upper limit was taken to be six times the initial population size m) did not alter the results of the Poisson model. Nassar and Cook (1973) examined also by simulation the effects that a deviation from a Poisson progeny distribution would have on the time to fixation and on the probability of fixation of a mutant gene. They found in the case of a negative binomial progeny distribution with a variance equal to twice the mean, that the time to fixation and the probability of fixation agreed well with those of the Poisson case. In all likelihood, our results are good approximations to real situations when independence in reproduction among individuals is in effect and when the variance of the progeny distribution is not much larger than the mean. The assumption of independence is probably met in uncrowded populations or in those where resources are not a limiting factor. Also in a population of small initial size where a limitation in food and/or space is not likely to be of immediate consequence. We do not know how good our results would be in approximating situations where selection through competition is predominant. Hence, it is desirable to further examine our model in the light of this possibility.

Literature

Cook, R. D., Nassar, R. F.: 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 **28**, 373–384 (1972a).

Cook, R. D., Nassar, R. F.: Probability of ultimate survival of a newly occurred inversion in natural populations. Theoretical and Applied Genetics **42**, 368–370 (1972b).

Elton, C., Nicholson, M.: The ten-year cycle in numbers of the lynx in Canada. J. Animal Ecology **11**, 215 to 244 (1952).

Kimura, M., Ohta, T.: The average number of generations until extinction of an individual mutant gene in a finite population. Genetics **63**, 701–709 (1969).

Kojima, K., Kelleher, T. M.: Survival of Mutant Genes. The Am. Naturalist **XCVI**, 329–345 (1962).

Koutsky, Z.: Periodische Verzweigungsfolgen. Transaction of the Third Prague Conference, 429–439 (1962).

Nassar, R. F., Cook, R. D.: Dynamics of Finite Populations II. A time-homogeneous stochastic process describing the ultimate probability of and the expected time of fixation or loss of an allele or type in a population of variable size. Theoretical and Applied Genetics **43**, 255–260 (1973).

Pollak, E.: Some effects of fluctuating offspring distribution on the survival of a gene. Biometrika **53**, 391 to 396 (1966).

Odum, E. P.: Fundamentals of Ecology. Philadelphia: Saunders 1959.

Received July 2, 1974

Communicated by W. Seyffert

R. F. Nassar
Department of Statistics
Kansas State University
Manhattan, Kansas 66506 (USA)

R. D. Cook
School of Statistics
University of Minnesota
St. Paul, Minnesota 55105 (USA)