

A Logistic Branching Process Alternative to the Wright-Fisher Model

R. B. Campbell
Department of Mathematics
University of Northern Iowa
Cedar Falls IA 50614-0506
campbell@math.uni.edu
<http://www.math.uni.edu/~campbell>
(319) 273-2447

Running head: Logistic Branching Process

Keywords: Branching Process, Coalescent, Population Regulation

Introduction

Most of the theoretical work in population genetics is based on the Wright-Fisher Model (Ewens 1979), which has constant population size and discrete generations. However, most of its analysis has been done using a continuous approximation, in particular, the diffusion approximation of Kimura (1962) and the coalescent analysis of Kingman (1980a,b). The present work considers the extent to which a branching process, which has fluctuating population size but discrete generations, is a good approximation to the Wright-Fisher model. A logistic branching process is introduced in order to limit the fluctuation in population size.

Fisher (1922) introduced a branching process model with a Poisson progeny distribution, and Haldane (1927) used it to derive the now classical formula for the probability of fixation of a new advantageous mutation ($2s$). Most subsequent studies using branching processes have been for rare alleles where population size is unimportant, but it has also been used to model growing populations (O'Connell 1995). This study addresses the validity of the branching process approximation at all allele frequencies for constant population size.

First the expected time to fixation (which is the coalescent time, Campbell 1999) for a branching process is shown to be the same as obtained with the diffusion and coalescent models. Then the shape of the ancestral pedigree is discussed. Next the difference between the Wright-Fisher model and an unconstrained branching process is discussed. Then the logistic branching process is discussed. Finally, the effect of selection on these models is considered. The conclusion is that the logistic branching process (and even an unconstrained branching process) provides a reasonable approximation for the Wright-Fisher model. Whether a logistic branching process (or even an unconstrained branching process) is a more realistic model than the Wright-Fisher model is not discussed.

Expected Fixation Time

The branching process model starts with a population of N distinct individuals (i.e., a haploid population is assumed), each of which independently reproduces with a Poisson progeny distribution with parameter one. Generations are discrete, hence the population contains N synchronized branching processes. The population will go extinct with probability one, but the expected time until extinction is infinite (Karlin and Taylor 1975).

Fixation shall be defined as when exactly one of the original N branching processes remains. The probability of fixation is trivially $1/N$ by symmetry.

To calculate the expected time until fixation, let $p(t)$ be the probability that a branching process is extinct at time t , given that there was a single individual at time 1. Then if there were initially N individuals, the probability that exactly one predesignated lineage (i.e., descendants of a chosen

individual) remains at time t is $(1 - p(t))(p(t))^{N-1}$ by independence. Hence the expected time until fixation is given by $\int_1^\infty t(1 - p)(N - 1)p^{N-2}dp/dt dt$, which is intractable unless $p(t)$ is known. For the Poisson progeny distribution, it is known that $p(t) \doteq 1 - 2/t$ asymptotically (Holte 1974), which approximation is good for $t > 20$. Substituting $2/t$ for $1 - p$ in the integral cancels out the factor of t and leaves $\int_1^\infty 2(N - 1)p^{N-2}(dp/dt)dt = \int_0^1 2(N - 1)p^{N-2}dp = 2$; this must be multiplied by N because there are N original genes which could become fixed. Hence the expected time until fixation is $2N$ generations, as obtained with the diffusion and coalescent models.

Shape of the Pedigree

The question is: what is the ancestral pedigree for a population? That is, what remains of the history of the population after all the lineages which have terminated have been removed. The number of individuals each generation, called the reduced family tree, is known (O'Connell 1975), but we also wish to study the pedigree structure within that shape. The shape and structure can be inferred from three observations. First, almost all except for recent branchings in the ancestral pedigree are bifurcations. Second, the probability of a branching process surviving t generations is approximately $2/t$. Third, recent generations can be easily described from the generating function of the branching process. Put together, these provide many generations of slow growth, and a rapid increase in the number of individuals in recent generations, perhaps resembling the shape of a golf tee. The approximation $2/t$ for the relative size of the pedigree t generations ago provides, for example, that the pedigree size $\sqrt{2N}$ generations ago is approximately $N \times 2/\sqrt{2N} = \sqrt{2N}$. In particular, for a population of size 1000 (with expected time since a common ancestor equal to 2000), the expected size of the pedigree 45 generations ago is 45. If $N = 1\,000\,000$ (with expected time since a common ancestor equal to 2\,000\,000), the expected size of the pedigree 1414 generations ago is 1414. Furthermore, if $N = 1000$, 1/2 of the individuals in the pedigree lived, hence 1/2 of the allelic forms originated, in the most recent 35 generations; 90% of the individuals lived, hence 90% of the allelic forms originated in the most recent 983 generations. For $N = 1\,000\,000$, 1/2 of the individuals in the pedigree lived, hence 1/2 of the allelic forms originated, in the most recent 1861 generations; 90% of the individuals lived, hence 90% of the allelic forms originated in the most recent 495\,237 generations.

Similarity of Models

The branching structure within the ancestral pedigree is quite similar for the branching process and Wright-Fisher models. Essentially all of the branchings in the early ancestral pedigree are bifurcations, and similarity of the branching structure in recent generations follows from the fact that a branching process with Poisson progeny distribution conditioned on the final population size has a binomial progeny distribution.

One way to generate a Poisson distribution is as the number of events which occur in a Poisson process in a specified period of time. From this perspective, the Wright-Fisher model with constant population size can be obtained from the branching process with Poisson progeny distribution by increasing or decreasing the time that the Poisson process runs so that the total resultant number of progeny is N (i.e., the total for all individuals, hence all branching processes since each individual has an independent branching process). Thus the Wright-Fisher model is obtained by adding branches to or removing branches from the unconstrained branching process each generation. If a population has size N , on average approximately $.8\sqrt{N}$ of the branches will have to be added or removed from a branching process each generation to obtain the Wright-Fisher model (.8 is the average of $|z|$ with the standard normal distribution). For example, for $N = 1\,000\,000$, going back 25 generations, $(1 - .8/1000)^{12.5} = .99$ and approximately 99% of the branches in an unconstrained branching process will be in the Wright-Fisher model. Similarly, approximately 1% of the branches in the Wright-Fisher pedigree will not be from the unconstrained branching process. The ancestral pedigree 25 generations ago is less than 8% of the present population size. Going back 200 generations over 92% of the branches will be the same; the ancestral pedigree 200 generations ago is less than 1% of the present population size.

The Logistic Model

The fluctuation in population size in a branching process model will be reduced if the Poisson distribution governing the number of progeny has parameter N_{eq}/N instead of 1, where N_{eq} is the equilibrium population size and N is $N(t)$, the population size at the time of reproduction. Since there will be N branching processes, each with expected value N_{eq}/N , the expected population size will be N_{eq} each generation. Furthermore, since each individual reproduces independently, the variance of the population size will be N_{eq} each generation (hence the standard deviation $\sqrt{N_{eq}}$).

The probability of fixation must be the reciprocal of the initial frequency by symmetry. The expected time until fixation can be shown to be $2N_{eq}$ generations by the diffusion approximation (the expected value of N can be substituted for N in the integrand of the integral for fixation time (Kimura and Ohta 1969, Crow and Kimura 1970)). Thus this model gives the same fixation probability and fixation time as the Wright-Fisher model. The fixation probability is determined by the original population size $N(0)$, but the fixation time is determined by the expected population size (N_{eq}).

For the logistic branching process, the probability of extinction is equal to one, however, the expected time until extinction is $1/e^{-N_{eq}}$, rather than infinite.

It can be shown that the logistic branching process is essentially identical to an unconstrained branching process in the same manner in which the Wright-Fisher model was demonstrated essentially identical to an unconstrained branching process, with the same results.

Selection

The classic result that the probability of fixation of an allele with relative fitness $1 + s$ is approximately $2s$ (Haldane 1927) is based on absolute viabilities in an unconstrained branching process, but is a reasonable approximation for the Wright-Fisher model.

For the logistic branching model, $2s$ is of course the probability of fixation if $1 + s$ is an absolute fitness. However, $1 + s$ can also be modelled as a relative fitness with the logistic branching process. In this case, the analysis of Ewens (1967) for a fluctuating population size provides that the probability of fixation is $2s\tilde{N}/N(0)$, where $N(0)$ is the initial population size and \tilde{N} is the harmonic mean of the population sizes over time. This provides that the average probability of fixation is $2s\tilde{N}/\bar{N}$, where the arithmetic mean appears in the denominator. Hence there will be a slight reduction in the probability of fixation as contrasted to the Wright-Fisher model since the harmonic mean is less than the arithmetic mean. However, if immigration is the source of new “mutations”, the probability of fixation is just $2s$.

This is not intended as a definitive statement on selection. Rather, the branching process models allow great flexibility in modelling selection, and it is anticipated that new insights will be provided by modelling selection with branching process models.

Literature Cited

- Campbell, R. B.** 1999. The coalescent time in the presence of background fertility selection. *Theor. Popul. Biol.* 55:260-269.
- Crow, J. F. and M. Kimura.** 1970. *An Introduction to Population Genetics Theory*. Harper and Row. New York.
- Ewens, W. J.** 1967. The probability of survival of a new mutant in a fluctuating environment. *Heredity* 22:438-443.
- Ewens, W. J.** 1979. *Mathematical Population Genetics*. *Biomathematics* vol. 9. Springer-Verlag. New York.
- 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. V. Selection and mutation. *Proc. Camb. Phil. Soc.* 23:838-844.
- Holte, J. M.** 1974. Extinction probability for a critical general branching process. *Stoch. Proc.*

Appl. 2:303-309.

Karlin, S. and H. M. Taylor. 1975. A First Course in Stochastic Processes (second edition). Academic Press. New York.

Kimura, M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713-719.

Kimura, M. and T. Ohta. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61:763-771.

Kingman, J. F. C. 1982a. The coalescent. *Stoch. Proc. Applicat.* 13:235-248.

Kingman, J. F. C. 1982b. On the Genealogy of large populations. *J. Appl. Probability* 19A::27-43.

O'Connell, N. 1995. The genealogy of branching processes and the age of our most recent common ancestor. *Adv. Appl. Prob.* 27:418-442.