A Logistic Branching Process For Population Genetics

R. B. Campbell

Department of Mathematics, University of Northern Iowa, Cedar Falls, IA 50614-0506

Abstract

A logistic (regulated population size) branching process population genetic model is presented. It is a modification of both the Wright-Fisher and (unconstrained) branching process models, and shares several properties including the coalescent time and shape, and structure of the coalescent process with those models. An important feature of the model is that population size fluctuation and regulation are intrinsic to the model rather than externally imposed. A consequence of this model is that the fluctuation in population size enhances the prospects for fixation of a beneficial mutation with constant relative viability, which is contrary to a result for the Wright-Fisher model with fluctuating population size. Explanation of this result follows from distinguishing between expected and realized viabilities, in addition to the contrast between absolute and relative viabilities.

Key words: Branching Process, Coalescent, Population Regulation, Selection

Email address: campbell@math.uni.edu (R. B. Campbell).
URL: http://www.math.uni.edu/∼campbell (R. B. Campbell).

1 Introduction

Most of the theoretical work in population genetics is based on the Wright-Fisher Model (Ewens, 1979). In particular, the diffusion approximation of Kimura (1962) and the coalescent analysis of Kingman (1982a,b) are based on it. Hence it is the model for which fixation and coalescent times have been calculated. (Actually the diffusion and coalescent analyses are based on a continuous approximation to the discrete Wright-Fisher model. The analyses are valid for a large class of exchangeable models.)

The Wright-Fisher model entails a constant population size and binomial progeny distribution. The assumption of constant population size is supported by the fact that population size is regulated by external resources; yet population size does fluctuate, in part because individuals cannot anticipate what other individuals will do. Hence the Wright-Fisher model may not be the best model for natural populations.

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 ($2s$). That result has been generalized, including an extension to fluctuating population size by Otto & Whitlock (1997). The classic application of the model to the study of the probability of extinction of surnames is still being refined (Hull, 1998). Branching processes have also been used to study the coalescence of rare alleles (Rannala, 1997). Another context in which branching processes have been used is to model the spread of mutations in geographically structured populations (Crump & Gillespie, 1977; Sawyer, 1979). (The birth and death

process is the continuous analog to the branching process.)

However, a branching process entails great fluctuations in population size (of the order of $\sqrt{N}$ each generation, where N is the population size). Especially because this variation cumulates over generations, this may entail greater fluctuation than occurs in nature. The branching process has mainly been used to model rare alleles, because in that context it has little impact on the total population size and is a good approximation to the Wright-Fisher model. An exception to this generalization is recent work by O'Connell (1995) which uses a supercritical branching process to model population growth since the time of "Eve".

Despite the difference between fixed and varying population size, the Wright-Fisher and branching process models are quite similar. The Poisson distribution (with two types) conditioned on population size is binomial, hence the branching process is the same as the Wright-Fisher model conditioned on the varying population sizes resulting from the branching process. In the appendix we show for neutral alleles that the fixation probability, fixation time, coalescent shape, and coalescent structure are quite similar for the Wright-Fisher model and Poisson branching process, hence the branching process is a good approximation to the Wright-Fisher model, even when an allele is not rare. Similarity of coalescent structure is demonstrated by showing that the coalescent (ancestral pedigree) of the Wright-Fisher model can be generated by slight modification of the coalescent for an (unconstrained) branching process. The arguments in the appendix are extended to show that the logistic branching process (which is defined below) also shares those properties.

The body of this paper is concerned with the logistic branching process, which

is a size dependent branching model (Lipow, 1975). Such models have been studied in the context of growth of populations (Klebamer, 1993), and are readily interpreted for alternative alleles within a population. They incorporate features of both the Wright-Fisher model and the Poisson branching process. Biologically relevant features include that individuals each generation reproduce independently of each other (unlike the Wright-Fisher model); this necessarily entails fluctuation in population size. However, there is an equilibrium population size which is approached in expectation each generation (unlike the Poisson branching process). For the specific model considered here (logistic branching process) the equilibrium population size is the expected population size each generation; this facilitates the analysis. It is shown that the fixation probability and expected time until fixation of a neutral allele under the logistic branching process are the same as for the Wright-Fisher model and Poisson branching process, and it is shown in the appendix that the structure of the coalescent is similar. An analysis of fixation probabilities with fluctuating viabilities ensuing from population size variation contrasts results for the Wright-Fisher model, which contrast is explained by clarifying the definitions of viability.

2 The Logistic Branching Process Model

The purpose of this model is to provide a compromise between the Wright-Fisher and Poisson branching process models. The logistic branching process model is the standard Poisson branching process (discrete generations, every individual reproduces independently with a Poisson progeny distribution with parameter λ), except that $\lambda = N_{eq}/N$ (where N_{eq} is the equilibrium popula-

tion size and N is $N(t) = N_t$, the population size at the time of reproduction) instead of the temporally constant value 1. This allows fluctuation in population size, but much less fluctuation than the Poisson branching process entails. Since there will be N Poisson distributions, 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 term "logistic" is used in its early sense of referring to ratios or fractions, and that the rate of increase falls as the population size increases; this model is not based on the logistic growth equation.)

The branching process analyses cannot be applied to this model, because the growth (size) of different branches are not independent in the logistic branching process (reproduction is independent within a generation, but the growth of one branch impacts the reproduction of all branches in the subsequent generations). However, the probability of fixation for neutral alleles must still be the reciprocal of the initial frequency by symmetry. The expected time until fixation for neutral alleles can be shown to be $2N_{eq}$ generations by the diffusion approximation as outlined below.

Within each generation, the reproduction of the genes obey independent Poisson distributions, hence the reproduction of each type of allele is a Poisson distribution (Feller, 1957). Furthermore, if one is considering only two different alleles, the number of alleles of each type, conditioned on the resultant population size, will have a binomial distribution (if each allele obeys a Poisson distribution (Feller, 1957)). Therefore the sampling variance will be $V_{\delta p} = \frac{p(1-p)}{N}$. In the neutral case, the integral for fixation time (Kimura & Ohta, 1969; Crow

& Kimura, 1970) becomes $\int_p^1 4N dx + \frac{1-p}{p} \int_0^{1-p} \frac{4Nx}{1-x} dx$. Indeed, $N = N(t)$ is a random variable, but because it is in the numerator and the expected value of an integral is the integral of the expected value, $E[N] = N_{eq}$ can be used to calculate the expected fixation time. This integral is approximately equal to $4N$ (i.e., $4N_{eq}$) for large N. Because we study the haploid model rather than the diploid model which Kimura analyzed, the integral is approximately equal to $2N$ in our context.

For the logistic branching process, the probability of extinction of the entire population is equal to one, as it is for the regular branching process. This can easily be seen because it returns to size N_{eq} (in expectation) each generation, and there is a positive probability of having no progeny. More specifically, the Poisson parameter $\lambda = N_{eq}/N$ provides that for each individual the probability of having no progeny is $e^{-N_{eq}/N}$, hence the probability of extinction for the entire population each generation is $(e^{-N_{eq}/N})^N = e^{-N_{eq}}$. The probability of extinction being equal to $e^{-N_{eq}}$ each generation provides that the expected time until extinction is $1/e^{-N_{eq}}$, by a geometric series.

However, for the regular branching process, although the probability of extinction is 1, the expected time until extinction is infinite. (This follows easily from the probability of extinction $1 - 1/(2t)$, and can be shown rigorously using the error bound ϵ/t in Holte (1974).)

The ancestral pedigree of the logistic branching process is shown in the appendix to be essentially the same as the ancestral pedigree of a Poisson branching process by the same argument which is employed to show that the ancestral pedigree of the Wright-Fisher process is essentially the same as the ancestral pedigree of a Poisson branching process. This is based on viewing reproduc-

tion under the logistic branching process as a modification of reproduction under the Poisson branching process with $\lambda = 1$. Similarity is measured as the extent to which one pedigree can be superimposed on another.

3 The Fate of an Advantageous Mutation

The consequences of the logistic branching process are of especial interest because they contrast a well known result for the Wright-Fisher model with fluctuating population size. Ewens (1967) has shown that fluctuating population size will increase the probability of extinction of a new beneficial mutation compared to a constant population size, which result was confirmed by Otto & Whitlock (1997) using the same model. This is expected since fluctuating population size should increase drift, hence the probability of extinction before the allele frequency has increased due to deterministic forces. However, the logistic branching process (which entails fluctuation in population size) provides a beneficial mutation with greater probability of fixation than a constant population size does. Constant relative viabilities are assumed in both analyses, and both models are essentially branching processes with discrete generations. The population size fluctuation can be the same for both models. We consider fitness as defined by the geometric mean of the viability, which is the appropriate mean for temporal variation in viability, as well as by the probability of fixation, which is the ultimate definition of fitness. Pairs of models can be constructed where one has a higher geometric viability, and the other a higher probability of fixation, but the logistic branching process provides both greater geometric mean viability and greater probability of fixation than Ewens' model. Ewens (1967) considers a population which is deterministically

cycling through a set of population sizes N_t. The Wright-Fisher model is used to generate subsequent generations of the population. The relative viability of the mutation is held constant at $1 + s$ each generation, hence the absolute viability is $(1 + s)N_{t+1}/N_t$ in the t^{th} generation in order to maintain constant relative viability using the Wright-Fisher model. Because the population sizes are cyclic, the geometric mean of the absolute viabilities is $1 + s$, hence fluctuating population size does not change the mean viability. The result that the probability of loss increases over that probability for a constant population size is obtained from the eigenvalues of a matrix representing one cycle of the population sizes. Some approximations are employed.

With the logistic branching process, a relative viability $1 + s$ is modelled by having the mutant alleles obey a Poisson progeny distribution with parameter $(1 + s)N_{eq}/N_t$. This will not impact the mean fitness of the population if the mutant is rare. Taking the geometric mean of $(1 + s)N_{eq}/N_t$ over time entails taking the geometric mean of the N_t in the denominator. N_{eq} is the arithmetic mean of the N_t, hence since the geometric mean is less than the artihmetic mean, the geometric mean absolute viability is larger than the constant relative viability $1 + s$.

It is not easy to obtain an exact, or approximate, analysis for the probability of fixation for the logistic branching process. Therefore random simulations were performed and the generating function analysis of Slatkin (1996) was employed. First, the generating function (Karlin & Taylor, 1975) for a Poisson branching process with parameter 1.02 was composed 1000-fold, which produced the probability of extinction 0.9610423. This did not entail any randomness, but provided the comparison of constant population size. It also showed that 1000 generations were sufficient to calculate the probability of

extinction, hence the possibility of extinction of the entire population was not important. Then 40 000 sequences of 1000 independent normally distributed population sizes with mean 10 000 and standard deviation 100 were generated. These were used to calculate 40 000 999-fold compositions of generating functions with parameter $(1.02)N(t+1)/N(t)$. Although this did not entail cyclic population size, extinction should be determined before a long cycle is repeated, hence it provides the same probability of extinction as Ewens' model; the probability of extinction was 0.9610446 (this average was weighted by initial population sizes). Another 40 000 sequences of 1000 independent normally distributed population sizes with mean 10 000 and standard deviation 100 were generated. These were used to calculate 40 000 1000-fold compositions of generating functions with parameter $(1.02)N_{eq}/N_t$ for the logistic branching process. The resulting probability of extinction was 0.9609555 (this average was weighted by initial population sizes). Since the standard deviation of extinction probabilities was 0.0019, the standard error was 0.00001, and Ewens' model had a nonsignificant increase in the probability of extinction, while the logistic branching process had a significant reduction ($p < .000000000000000001$)) in probability of extinction when compared to the constant population size.

An important difference between the models is that the absolute viability of the mutant allele depends on both the present and next population sizes in Ewens' model, but it only depends on the present population size in the logistic model. This provides Ewens' model with approximately double the variance of the absolute viabilities, and also provides a negative autocorrelation of absolute viabilities. A conceptual interpretation is that Ewen's model employs more of a realized absolute viability because it is defined based on the actual population growth, whereas the logistic branching model employs an expected

absolute viability, based on the anticipated growth of the population.

These results for probability of fixation of a selected allele assume that s is small, but not so small that selection is near-neutral ($s \ll 1$, but $4Ns > 1$). Like most analyses of probability of fixation for favored alleles, approximations valid when the mutation is rare are employed, because fixation of a favored allele is determined while the allele is rare.

4 Discussion

This investigation was originally motivated by the question: is the branching process a reasonable approximation to the Wright-Fisher model only when an allele is rare? The results in the appendix show that the Poisson branching process and Wright-Fisher model are quite similar, hence should approximate each other even when alleles are not rare. The logistic branching process was then introduced as a blend incorporating both the population size fluctuation of the branching process and the population size regulation of the Wright-Fisher model. Similarity to the branching process and Wright-Fisher model was not surprising, but the differences in fixation probabilities from the Wright-Fisher model were. These differences can be explained from several perspectives.

An essential difference of the logistic branching process from Ewens' and similar fluctuating population size models based on the Wright-Fisher model is that the population size fluctuation is intrinsic to the logistic branching process, but extrinsic to Ewens' model. Donnelly & Tavaré (1995) note that the coalescent structure remains the same for models with exogenously varying population size. The logistic branching process may serve as a prototype for

models with population size fluctuation intrinsic to the model.

Another perspective on the difference of the models is provided by the nature of the definition of viability. Both models have constant relative viability $1+s$, but the definition of relative viability is not quite the same in the two models. The absolute viabilities of the mutant allele, $(1 + s)N_{eq}/N_t$ for the logistic branching process and $(1 + s)N_{t+1}/N_t$ for Ewens' model, are both expected viabilities, with the realized viability being determined by the Poisson or binomial distribution, respectively. The absolute viability of the wild-type allele in the logistic branching process, N_{eq}/N_t, is also an expected viability with the realized viability depending on the actual number of progeny from the Poisson distribution, but the absolute viability of the wild-type allele in Ewens' model, N_{t+1}/N_t, is a realized viability since the subsequent population size N_{t+1} is predetermined, which mandates how many progeny the wild-type allele has (recall that we are assuming that the muatant allele is rare, hence essentially all the alleles are wild-type). Hence constant relative viability is defined as a ratio of two expected viabilities in the logistic branching process, but the ratio of an expected to a realized viability in Ewens' model. Mathematically, the presence of N_{t+1} rather than its arithmetic mean N_{eq} in the numerator of the Poisson parameter provides that there will be greater extinction of mutant alleles under Ewens' model than under the logistic branching process, by the convexity of the probability of no progeny $e^{-\lambda}$.

The discrepancy between the the probabilities of fixation of a favored mutation between the two models can also be explained from properties previously mentioned. Since the geometric mean viability of the mutant allele is greater in the logistic model than in Ewens' model, its probability for fixation should also be greater. (Actually the geometric mean viability of the wild-type allele,

hence population, is greater than one in the logistic model, but that is because expected rather than realized viabilities are used. If realized viabilities were used, the geometric mean of the wild-type viability in the logistic branching process would be one.) Another result was that the variance of the viability of the mutant allele in Ewens' model should be approximately twice the variance for the logistic branching process. The greater variance should afford greater prospects for extinction in Ewens' model. The variance in viability will also affect the variance of the offspring number, which Gillespie (1975) has shown is related to fitness.

It is worthwhile to have a model which questions the association between fluctuation in population size and increased prospects for extinction of beneficial mutations. It is also useful to consider the many tacit assumptions which are made when the words viability or fitness are used. But this model has utility beyond the above results which it has motivated. It is most important as a prototype, building fluctuation in population size into the model rather than having it externally imposed. Indeed the equilibrium population size is just a generalization of the carrying capacity of the logistic growth model, but it introduces it into a new context.

A Fixation Probability

The probability of fixation of a selectively neutral mutation is $1/N$ for a population of N individuals ($1/(2N)$ for a diploid population) by symmetry, hence this result merely merits mention. (Indeed, it must be shown that one of the alleles will become fixed.) However, an alternative proof for the branching process lays a foundation for the derivation of the expected time until fixation,

hence is presented here.

The branching process model starts with a population of N distinct individuals (i.e., a haploid population is assumed; for comparison with diploid results replace N with $2N$), 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 & Taylor, 1975).

Fixation shall be defined as when exactly one of the original N branching processes remains (or alternatively, all but one have become extinct, the second largest extinction time among the N original individuals). This is the same definition as is employed with the Wright-Fisher model, that only one of the original alleles has descendants in the population. However, unlike the Wright-Fisher model, it is possible that the last two (or more) remaining branching processes will go extinct the same generation, hence such fixation will not occur; but that will have very small probability, and we may ignore it. Notation (derivatives and integrals) for a continuous process is used below, but remains valid when interpreted for a discrete process.

To calculate the probability of fixation of an original individual, let $p(t)$ be the probability that a branching process (not specifically the branching process ensuing from the designated original individual, but by independence $p(t)$ is the same for all the original individuals) is extinct at time t, given that there was a single individual at time 1. Then if there were initially N individuals, the probability for a given initial individual that its lineage (i.e., descendants of that individual) remains and the other lineages are extinct at

time t is $(1-p(t))(p(t))^{N-1}$ by independence. (Since this expression is for a designated original individual, the probability that descendants of exactly one of the original individuals remains (with the other lineages extinct) is $N(1-p(t))(p(t))^{N-1}$ because the events of different individuals remaining are mutually exclusive.)

The probability that a given lineage becomes fixed during a time interval dt is $(N-1)(1-p(t))(p(t))^{N-2}\, dp/dt\, dt$ where $(1-p(t))$ is the probability that the given lineage is not extinct, there are $N-1$ lineages which may go extinct with probability $dp/dt\, dt$, and $(p(t))^{N-2}$ is the probability that the other $N-2$ lineages are already extinct. The probability that a given individual becomes fixed is therefore given by $\int_1^\infty (1-p)(N-1)p^{N-2}dp/dt\, dt = \int_0^1 (1-p)(N-1)p^{N-2}dp = 1/N$, which, as noted above, is what symmetry requires if fixation eventually occurs since there are N original individuals. The reader may have noted that the integrand is only the second half of the product rule for the derivative $d((1-p(t))(p(t))^{N-1})/dt$ of the probability of one type being present; the other half $(-(p(t))^{N-1}dp/dt)$ reflects the given lineage, hence the entire population, going extinct.

B Expected Time until Fixation

The expected time until fixation for the Wright-Fisher model is a classic result from the diffusion approximation (Crow & Kimura, 1970), and has also been calculated as the coalescent time (Kingman, 1982a,b), which is equal to the fixation time (Campbell, 1999); the expected time until fixation is approximately $2N$ generations. For the branching process model, employing the above notation, the expected time until fixation is given by

$N \int_1^\infty t(1-p)(N-1)p^{N-2}dp/dt\ dt$, which is intractable unless $p(t)$ is known. For the Poisson progeny distribution, it is known that $p(t) \doteq 1 - 2/t$ asymptotically (e.g. Holte, 1974; O'Connell, 1995), and numerical evaluation of the composition of the generating function e^{s-1} evaluated at $s = 0$ confirms that this approximation is good for $t > 20$. The integration for $t < 20$ will be small (assuming N is reasonably large) because p is small there (i.e., $p < .91$,hence $Nt(1-p)(N-1)p^{N-2} < 1$ if $N > 200$ and $t < 20$, which implies $N \int_1^{20} t(1-p)(N-1)p^{N-2}dp/dt\ dt < 1$). Substituting $2/t$ for$1-p$ in the integral cancels out the factor of t and leaves $N \int_1^\infty 2(N-1)p^{N-2}(dp/dt)dt = N \int_0^1 2(N-1)p^{N-2}dp = 2N$. (Using the $P(20) = 0.91$ for the lower bound of integration provides the same result to six decimal places for $N > 200$.) Hence the time until fixation is approximately $2N$ generations. (This time would be $4N$ for a diploid model.) This approximation remains valid if $p(t)$ is interpreted as a discrete function changing only at the integers, in this case $p(t+1) - p(t)$ is approximately $2/t^2$.

For this neutral branching process model (Poisson progeny distribution) the population size will fluctuate by about $\sqrt{N}$ (where N is $N(t)$, not the original population size) each generation, because the variance for each individual is 1, and the variance of the sum is the sum of the variances. But the time until fixation is based upon the initial population size.

C The Shape of the Ancestral Pedigree

The ultimate 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. This ancestral pedigree is the coalescent

in its most general sense, and these terms shall be used interchangeably. One aspect of this question is: what is the size of the ancestral population? This process, N_t, is called the reduced branching process or reduced family tree (O'Connell, 1995). [This should not be confused with the skeleton, which is individuals who will have descendants in all future generations. The skeleton does not exist for critical branching processes, but for critical branching processes conditioned on non-extinction the skeleton has one member per generation (O'Connell, 1993). We refer to the skeleton as the ancestral lineage.] Although a critical Poisson branching process (conditioned on non-extinction) grows linearly with t, the reduced family tree will be convex because of the many branches which will have died off. There is much information on the shape of the reduced family tree, although its explicit shape may not have been obtained. Of course, there are large variances to any results which may be obtained.

Kingman's coalescent studies are based on the Wright-Fisher model. The approximation $dk/dt = k(k-1)/2N$ of Kingman (1982a,b) (a negative sign has been omitted to change the direction of time) where k is the number of ancestors of the present population t generations after the most recent common ancestor cannot be solved explicitly because of difficulties specifying the initial conditions. However, setting $k(2N) = N$ provides $k = (1 - (1-(1/N))e^{(t/2N-1)})^{-1}$ which suggests the nature of the growth. The cruder approximation $dk/dt = k^2/2N$ with the same initial condition yields $k = (1 + 1/N - t/2N)^{-1}$ which may be easier to comprehend.

O'Connell (1994) presents a pure birth process approximation to the reduced branching process (i.e., for the Poisson branching process). With rescaling and writing as a derivative this becomes $dk/dt = \frac{k}{2N}/(1-\frac{t}{2N})$. The initial condition

$k(2N) = N$ cannot be used for this equation, but using $k(0) = 1$ yields the solution $k = (1 - t/2N)^{-1}$, which is similar to $k = (1 + 1/N - t/2N)^{-1}$ above.

The asymptotic approximation that the probability that a critical Poisson branching process has not become extinct by time t is $2/t$ (e.g. Holte, 1974; O'Connell, 1995) can be used to find the shape of the reduced family tree. This provides (assuming constant population size N) that the reduced family tree $2N - t$ generations before the present will have size $N(2/(2N - t))$ which is equal to $k = (1-t/2N)^{-1}$ obtained from the pure birth process approximation.

The three approximations $k = (1 - (1 - (1/N))e^{(t/2N-1)})^{-1}$, $k = (1 + 1/N - t/2N)^{-1}$, and $k = (1 - t/2N)^{-1}$ give a consistent picture of the shape of the reduced family tree (i.e., shape of the coalescent). For $N = 1000$, all the approximations have less than 10 individuals for the first 1791 generations, less than 100 individuals for the first 1979 generations, and the former two approximations have 1000 individuals in generation 2000, while the latter approximation has infinite size in generation 2000. Hence the shape of the reduced family tree resembles a long stemmed rose with very little breadth until the top.

This shape of the reduced family tree confirms why most mutations should be rare: most mutations in the ancestral pedigree will have occurred in recent generations, because most individuals in the ancestral pedigree lived in recent generations. Calculations which confirm this also allow us to address the extent to which the coalescent structure of the Wright-Fisher and Poisson branching process coalescents coincide. Specifically, we quantify the distribution of how many generations ago individuals in the coalescent lived. We refine the above approximations by combining the generating function for the Poisson

process for the most recent 20 generations with the asymptotic approximation $2/t$ which is reasonable for earlier generations in order to get numerical information on the shape of the reduced family tree.

The generating function e^{s-1} for the Poisson distribution provides that $1/e$ of the previous generation did not have any progeny, hence only $1 - 1/e$ of the previous generation contributed to the current generation, i.e., is part of the pedigree. Going back another generation, composition of the generating function provides that only $1 - (1/e)^{(1-1/e)}$ of the individuals two generations ago have descendants in the current generation. (Composition of the generating function evaluated at zero gives the probability of extinction, *ergo* the composition of the generating function is subtracted from one. Because a branching process may fluctuate in size with time, this result is stated in terms of the population size t generations ago.) Only going back two generations, the size of the ancestral pool is about 47% of the current population size. This is the probability of having descendants after two generations.

If one is using the Wright-Fisher model with constant population size and binomial (multinomial) progeny distribution, the same result obtains as a good approximation. If the population size is N, the probability that an individual in the previous generation does not have a descendant in the present generation is $((N-1)/N)^N$ which is equal to $(1/(1+1/(N-1)))^N$ which is approximately $1/e$ for large N. The extension to more generations follows *mutatis mutandis*.

Using the approximation $2/t$ for the relative size of the pedigree t generations ago for $t > 20$ 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 ancestral 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 ancestral pedigree 1414 generations ago is 1414.

In addition to the size of the reduced family tree t generations ago, the cumulative occurrence of individuals can be calculated by integrating $2/t$ for the $2N$ generations since the most recent common ancestor occurred (substituting explicit calculations with the generating function for the most recent 20 generations). In particular, this answers the question of the distribution of when mutations in the population occurred, since the number of individuals in the ancestral pedigree as a function of the number of generations should give the relative frequency of when mutations occurred. The total number of individuals in the ancestral pedigree since the most recent common ancestor is approximately $2N(\ln(2N) - .5)$. Specific calculations assuming $N = 1000$ (with the expected time since a common ancestor of 2000 generations) provide that 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 $N = 1\,000\,000$ (with the expected time since a common ancestor of $2\,000\,000$ generations), 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.

We are not addressing questions such as the age of a mutant dependent on its frequency, which have been considered elsewhere (Watterson & Guess, 1977), merely the distribution of ages of mutants. The shape of the ancestral pedigree (i.e., the reduced family tree) provides the distribution of ages of the mutant

alleles, but the structure of the ancestral pedigree and further mathematical analysis is necessary to relate frequency and age from this approach.

D The Structure of the Ancestral Pedigree

In addition to the overall shape of the ancestral pedigree, the branching structure within the ancestral pedigree is also quite similar for the branching process and Wright-Fisher models. This is demonstrated by showing that essentially all of the branchings in the early ancestral pedigree are bifurcations, and that the branching structure in recent generations is similar. The former is true because for each individual in the early pedigree all but at most two of the descendant branches die off, the latter because the branching process with Poisson distribution conditioned on final population size has the multinomial distribution. Ancestral pedigree and coalescent shall be used interchangeably to refer to all ancestors of the present population (reduced family tree refers to only the number of individuals in each generation).

In the continuous approximations to the ancestral pedigree (Kingman, 1982a,b; O'Connell, 1994) all branching events are bifurcations. This is clearly not the case for discrete generations with the Poisson progeny distribution, but the following shows that at most two lineages remain after sufficient time. The probability that an individual t generations ago had two or more progeny with descendants in the present generation can be calculated assuming the Poisson progeny distribution (with $\lambda = 1$). This is approximately equal to $\sum_{n=2}^{\infty} \frac{1}{en!}\binom{n}{2}(\frac{2}{t})^2$. Because each pair within a sibship is counted, but the events of different pairs surviving are not mutually exclusive, this is an approximation. However, the concurrence of such events is sufficiently rare that this

is not important. Similarly, the probability that an individual had three or more progeny with descendants in the present generation is approximately $\sum_{n=3}^{\infty} \frac{1}{en!}\binom{n}{3}(\frac{2}{t})^3$. Rewriting these summations as $\sum_{n=k}^{\infty} \frac{1}{e(n-k)!k!}(\frac{2}{t})^k$ (with $k = 2, 3$), it is immediate the summation with $k = 3$ is $\frac{2}{3t}$ times the summation with $k = 2$. Hence conditioning on a branching event in the ancestral pedigree t generations ago (note that we are referring to a branching event that remains in the pedigree after lineages have died off, not to a branching event in the ancestral population), the probability that it was a bifurcation is approximately $1 - \frac{2}{3t}$.

Combining this result with the previous result that the pedigree size $\sqrt{2N}$ generations ago is approximately $N \times 2/\sqrt{2N} = \sqrt{2N}$ provides that the probability that all branchings in the ancestral pedigree more than $\sqrt{2N}$ generations ago are bifurcations is greater than $(1 - \frac{2}{3\sqrt{2N}})^{\sqrt{2N}}$ (The $\sqrt{2N}$ in the denominator is a bound on the time, and the $\sqrt{2N}$ in the exponent is a bound on the number of branchings). For the population sizes $N = 1000$ and $N = 1\,000\,000$ considered before, the probability is greater than 50% that all the branchings in the ancestral pedigree prior to 44 or 1414 generations ago, respectively, were bifurcations.

Note that this is not claiming that there was a bifurcation in the ancestral pedigree every generation, in most generations the ancestral pedigree did not branch. Neither is it claiming that members of the ancestral pedigree only had one or two progeny, it is claiming that only one or two of their progeny had lineages which extend to the present.

Although essentially all of the early branchings in the ancestral pedigree are bifurcations, many of the recent branchings in the ancestral pedigree will not be

bifurcations. However, the recent branching structures for the Poisson branching process and Wright-Fisher model are quite similar. We show that we can superimpose most of the recent ancestral pedigree of the Wright-Fisher model upon a corresponding ancestral pedigree of the branching process model. This can be done because a branching process with Poisson progeny distribution conditioned on the final population size has a multinomial (binomial if there are two types) 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 Poisson distributions since each individual has an independent Poisson distribution). This entails adding branches to or removing branches from the branching process. The resultant progeny distribution is the Wright-Fisher model, since it is the Poisson distribution conditioned on final population size. 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, then the standard deviation of the population size resulting from a Poisson progeny distribution with parameter $\lambda = 1$ is $\sqrt{N}$, hence on average approximately $.8\sqrt{N}$ of the branches will have to be added or removed from a branching process to obtain the Wright-Fisher model each generation (.8 is the mean of $|z|$ with the standard normal distribution). Indeed N will change over time for a branching process, but cumulating these added and deleted branches across generations accounts for this. Half the genera-

tions should entail adding lineages, and half removing lineages. For example, for $N = 1\,000\,000$, going back 25 generations, $(1 - .8/1000)^{12.5} = .99$ and approximately 99% of the branches in a Poisson branching process will remain when it is modified to the Wright-Fisher model. Similarly, approximately 1% of the branches in the Wright-Fisher pedigree will not be in the Poisson branching process it was based upon. Exchangeability of branches (an extra branch one generation may persist when an original branch is eliminated the next generation) will reduce the difference between the unconstrained branching process and Wright-Fisher pedigree.

The ancestral pedigree 25 generations ago is less than 8% of the present population size. Hence for either the Wright-Fisher model or Poisson branching process the early ancestral pedigree grew almost entirely by bifurcation, and the most recent 25 generations, entailing over 90% of the growth of the ancestral pedigree, coincide for 99% of their branches. For the most recent 200 generations, which manifest 99% of the growth of the ancestral pedigree, over 92% of the branches coincide.

An analogous argument shows that the recent ancestral pedigree of a logistic branching process is quite similar to the ancestral pedigree of an unconstrained branching process. The former can be obtained from the latter by adding and removing branches. Taking the derivative of the probability mass function for the Poisson distribution with respect to the parameter λ, $\frac{d}{d\lambda}\frac{e^{-\lambda}\lambda^n}{n!} = \frac{-e^{-\lambda}\lambda^n}{n!} + \frac{e^{-\lambda}n\lambda^{n-1}}{n!}$, shows that increasing λ entails adding individuals to existing (perhaps empty) sibships compared to $\lambda = 1$. (The negative summand represents the loss of sibships of size n when another individual is added to them, the positive summand represents the increase of sibships of size n when individuals are added to sibships of size $n - 1$.) The increase in

the expected number of progeny when λ is increased is the result of additional progeny in individual sibships. If λ is decreased, the reduction in the expected number of progeny is achieved by removing individuals from individual sibships. The random variable λ equals N_{eq}/N where N is a random variable with mean N_{eq} and standard deviation $\sqrt{N_{eq}}$, which means that on average approximately $.8\sqrt{N_{eq}}/N_{eq}$ of the branches will have to be added or removed from a $\lambda = 1$ pedigree (i.e., $.8\sqrt{N_{eq}}$ branches each generation). This is the same number of additions/removals as in the comparison between the Poisson branching process and Wright-Fisher model, hence the same similarity ensues. (By transitivity, at most twice that number of modifications would achieve the logistic branching process from the Wright-Fisher model).

The random variable N_{eq}/N which governs the logistic branching process has an expected value greater than 1, but when weighted by population size (which is the number of individuals which reproduce) the expected value is 1.

References

Campbell, R. B. 1999. The coalescent time in the presence of background fertility selection. Theor. Popul. Biol. 55,260-269.

Crow, J. F. & Kimura, M. 1970. An Introduction to Population Genetics Theory. Harper and Row. New York.

Crump, K. S. & Gillespie, J. H. 1977. Geographical distribution of a neutral allele considered as a branching process. Theor. Popul. Biol. 12,10-20.

Donnelly, P. & Tavaré, S. 1995. Coalescents and genealogical structure under neutrality. Annu. Rev. Genet. 29,401-421.

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.

Feller, W. 1957. An Introduction to Probability Theory and its Applications, Volume I. John Wiley & Sons. New York.

Fisher, R. A. 1922. On the dominance ratio. Proc. Roy. Soc. Edinb. 42,321-341.

Gillespie, J. H. 1975. Natural selection forwithin-generation variance in offspring number II. Discrete haploid models. Genetics 81,403-413.

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.

Hull, D. M. 1998. A reconsideration of Galton's problem (using a two-sex population). Theor. Popul. Biol. 54,105-116.

Karlin, S. & Taylor, H. M. 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. & Ohta, T. 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.

Klebamer, F. C. 1993. Population-dependent branching processes with a threshold. Stoch. Proc. Appl. 46,115-127.

Lipow, C. 1975. A branching model with population size dependence. Adv. Appl. Prob. 7,495-510.

O'Connell, N. 1993. Yule process approximation for the skeleton of a branching process. J. Appl. Prob. 30,725-729.

O'Connell, N. 1994. Branching and inference in population genetics. Progress in Population Genetics and Human Evolution 87,97-106.

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

Otto, S. P. & Whitlock, M. C. 1997. The probability of fixation in populations of changing size. Genetics 146,723-733.

Rannala, B. 1997. On the genealogy of a rare allele.Theor. Popul. Biol. 52,216-223.

Sawyer, S. 1979. A limit theorem for patch size in a selectively-neutral migration model. J. Appl. Prob. 16,482-495.

Slatkin, M. 1996. In defense of founder-flush theories of speciation. American Naturalist 147,493-505.

Watterson, G. A. & Guess, H. A. 1977. Is the most frequent allele the oldest? Theor. Popul. Biol. 11,141-160.