

Time Since Common Pedigree Ancestors with Two Progeny per Individual

R. B. Campbell

Department of Mathematics

University of Northern Iowa

Cedar Falls, IA 50614-0506

USA

e-mail: campbell@math.uni.edu

homepage: <http://www.math.uni.edu/~campbell>

phone: (319)273-2447

running head: Common Ancestry with Two Progeny

keywords: coalescent; fixation time; progeny distribution; population genetics;
pedigree

Abstract

Constraining individuals to two progeny (versus Poisson distribution) increases the time since a pedigree (non-genetic) common ancestor, but the time still increases logarithmically in the population size. This is confirmed by simulations for discrete generations and rigorously for expected time with a modification of the Moran model. Selfing increases the expected time since a common ancestor with both the Poisson progeny distribution and two progeny per individual. As selfing approaches one the time since a common ancestor asymptotically approaches infinity with two progeny per individual, but only twice the population size with the Poisson progeny distribution. Regular systems of inbreeding with two progeny per individual can either increase or decrease the time since a common ancestor as contrasted with random mating with two progeny per individual.

1 Introduction

Much of population genetics analysis is based on the Wright-Fisher model which assumes a binomial progeny distribution (which is often approximated by the Poisson distribution). This is true for the standard diffusion approximation (Crow and Kimura 1970) and the standard coalescent (Kingman 1982). Indeed, Wright (1931) recognized that having two progeny per individual halves the sampling variance from the Poisson distribution. Robinson and Bray (1965) contrasted two progeny per individual with the binomial distribution for the inbreeding coefficient and rate of allele extinction. Campbell (1995) contrasted two progeny per individual with the Poisson distribution for heterozygosity and the number of segregating alleles. This work contrasts two progeny per individual with the Poisson distribution for time since a common ancestor.

The time since a common ancestor has generally been interpreted in the genetic sense of the time since a common ancestor of all the genes at a specified locus. This is quite generally equal to the time until fixation of a new mutation in the population (Campbell 1999). An alternative pedigree sense (for, e.g., diploid human populations) is the time since a person who is an ancestor of everyone in the present population, whether or not there is common genetic material. This model was studied by Chang (1999). Although this may have limited biological relevance, it could be important for cultural inheritance, or pathogens which are transmitted from either parent. The primary focus of this paper is contrasting the results of Chang for the Poisson distribution with the case of two progeny per individual.

The structure of this paper is to first review the case of genetic ancestry as a basis for comparison. Then results will be presented for pedigree ancestry, and extended to the case of partial selfing. Finally, some results for regular systems

of inbreeding will be presented to illustrate the impact of mating structure.

2 Genetic ancestry

For a diploid population with N individuals, i.e., $2N$ gametes form each generation, the expected time until fixation of a new neutral mutation, which is the same as the expected time since a common ancestor of the genes at a locus, is $4N$ generations. This is for the Wright-Fisher model (binomial progeny distribution), and has been shown with both the diffusion approximation (Crow and Kimura 1970) and the coalescent (Kingman 1982). If the constraint of two progeny per individual replaces the binomial progeny distribution, the expected time until fixation (time since a common ancestor) doubles to $8N$ generations under both the diffusion approximation and the coalescent model as indicated in the following two paragraphs.

In the case of no selection the the diffusion approximation provides the fixation time

$$\int_{\frac{1}{2N}}^1 \frac{2x(1-x)}{V_{\delta x}} dx + (2N-1) \int_0^{\frac{1}{2N}} \frac{2x^2}{V_{\delta x}} dx \quad (1)$$

where $V_{\delta x}$ is the sampling variance at frequency x (Crow and Kimura 1970, p. 430). In the case of the binomial distribution, the sampling variance $V_{\delta x}$ is equal to $\frac{x(1-x)}{2N}$ where x is the relative frequency of an allele. With the restriction of two progeny per individual, only heterozygous individuals contribute to the variance, so the sampling variance is $2x(1-x) \times 2N \times \frac{1}{4} / (2N)^2 = \frac{x(1-x)}{4N}$. Hence the sampling variance is halved, which doubles the fixation time by (1).

This result follows from the coalescent because under the Poisson distribution the probability that two genes have the same parent gene (hence coalesce) is $\frac{1}{2N}$, but under the constraint of two progeny per individual the probability that two genes have the same parent gene is $\frac{.5}{2N-1} \doteq \frac{1}{4N}$ (this is because once

a gene is chosen, of the other $2N - 1$ genes in the progeny generation only one came from the same diploid parent, and only half the time will it be the same gene from that diploid parent). Thus the probability of coalescing is halved, and the coalescent time doubles. Kingman (1982) has shown this more generally as the effect of the variance on the time scale.

3 Pedigree ancestry

3.1 Preliminary comparisons

A concrete model for studying pedigree ancestry with two progeny per individual entails mating as every individual putting two gametes into a mating pool from which gametes are randomly paired. Most progeny will have one half-sib from each parent, but full sibs will occur if the two gametes from one parent are paired with the two gametes from a different parent, and selfing will occur in the event that the two gametes from one parent are paired. The first remark is that the constraint that every individual have exactly two progeny means that the ancestral process is the same as the descendant process. It can be viewed as every individual choosing two progeny, or every individual choosing two parents. (These choices are not independent, since only two individuals can have a given parent or progeny; indeed the independence assumed with the Poisson distribution is not quite satisfied with the binomial distribution, but the Poisson distribution is a good approximation to the binomial distribution. This symmetry with respect to time does not hold for the genetic process with two progeny per individual, a given copy of a gene may have 0, 1, or 2 descendants the next generation, but has exactly one parent gene.)

A result of this is that no pedigree (as opposed to genetic) lineages dis-

appear. With the Poisson distribution, if one goes back far enough in time, approximately 80% of the individuals are ancestors of everyone in the current generation, with the remaining 20% having no descendants in the current generation (Chang 1999). But with two progeny per individual (or at least one progeny per individual), if you go back far enough in time, every individual is an ancestor of everyone in the current generation.

3.2 Analysis

The following calculations and simulations estimate the expected time until a specified individual becomes a pedigree ancestor of the entire population, which is the same as the expected time since the entire population was a pedigree ancestor of a specified individual. This will enable us to bound the time since the most recent pedigree ancestor of the entire population, and the time since the entire population was a pedigree ancestor of the entire present population, which are the quantities calculated by Chang (1999) (with the Poisson progeny distribution, 20% of the pedigree lineages go extinct, so Chang calculated the expected time since all individuals whose pedigree lineages do not go extinct were pedigree ancestors of the entire present population).

[FIGURE 1 NEAR HERE]

Figure 1 illustrates some of these concepts. The time until some individual becomes a pedigree ancestor of the entire population is three generations, which is achieved by the third and fifth individual. The time until a specified individual becomes a pedigree ancestor of the entire population is three, four, or five generations depending on which individual is specified (the expected time until a specified individual becomes a pedigree ancestor of the entire population is $\frac{2}{5} \times 3 + \frac{2}{5} \times 4 + \frac{1}{5} \times 5 = 3.8$). The time until every individual becomes a pedigree

ancestor of the entire population is five generations. Hence by “some” we mean minimum time and by “every” we mean maximum time.

The expected time until some individual becomes an ancestor of the entire population is greater than or equal to $\log_2 N$ (N is the population size) since every individual has two progeny, hence the number of descendants can only grow as 2^t . This lower bound is achieved with maximum avoidance of inbreeding as discussed below. The expected time until some individual becomes a pedigree ancestor of the entire population is less than or equal to the expected time until a specified individual becomes a pedigree ancestor of the entire population, because the former is obtained from minimums of the times the latter is based on. The expected time until every individual becomes an ancestor of the entire population is greater than or equal to the expected time until a specified individual becomes a pedigree ancestor of the entire population because the former is obtained from maximums of the times the latter is based on. The expected time until every individual becomes a pedigree ancestor of every individual in some future population is less than or equal to twice the expected time until a specified individual becomes a pedigree ancestor of the entire population, because every individual in a past population being a pedigree ancestor of a specified individual, and then that individual becoming a pedigree ancestor of an entire future population, entails every individual in the past population being a pedigree ancestor of every individual in the future population. Therefore $\log_2 N$ is less than or equal to the expected time until some individual becomes a pedigree ancestor of the entire population which is less than or equal to the expected time until every individual becomes a pedigree ancestor of the entire population which is less than or equal to twice the expected time until a specified individual becomes a pedigree ancestor of the entire population.

By symmetry as noted above (the process is the same if time is reversed), the expected time until every individual in the present population becomes a pedigree ancestor of the entire population some time in the future is equal to the expected time since everyone in the entire population some time in the past was a pedigree ancestor of everyone in the present population. The expected time since the most recent pedigree ancestor is greater than or equal to $\log_2 N$ because growth is bounded by 2^t . Hence $\log_2 N$ is less than or equal to the expected time since the most recent pedigree ancestor of the entire population which is less than or equal to the expected time since every individual was a pedigree ancestor of the entire population which is less than or equal to twice the expected time until a specified individual becomes a pedigree ancestor of the entire population. We have not been able to prove that the expected time until some individual becomes a pedigree ancestor of the entire population is equal to the expected time since the most recent pedigree ancestor of the population.

To calculate the expected time until a specified individual becomes a pedigree ancestor of the entire population, we assume a population of size N , hence $2N$ gametes form each generation. We count the number of descendants of an individual. As noted above, unlike with the Poisson progeny distribution, no individual can have its pedigree lineage go extinct. Let $k(t)$ be the number of individuals in generation t which are descended from a single specified individual in generation 0 (hence they will contribute $2k$ gametes to the next generation, and there will be $2(N - k)$ gametes from non descendants contributed to the next generation). Then the expected value of $k(t + 1)$ (time is going forward) is:

$$E[k(t + 1)] = N \times \left(\frac{2k}{2N} \times \frac{2k - 1}{2N - 1} + 2 \times \frac{2k}{2N} \times \frac{2(N - k)}{2N - 1} \right) \quad (2)$$

where k is $k(t)$, $\frac{2k}{2N} \times \frac{2k-1}{2N-1}$ is the probability an individual in generation $t + 1$ is formed from two gametes which came from descendants in generation t of the specified individual in generation 0, and $2 \times \frac{2k}{2N} \times \frac{2(N-k)}{2N-1}$ is the probability an individual is formed from one gamete from a descendant and one gamete from a non-descendant. This provides the the expected change in k (k is monotonely increasing)

$$E[\Delta k] = \frac{2k(N-k)}{2N-1} \quad (3)$$

which has the associated differential equation

$$\frac{dk}{dt} = \frac{2k(N-k)}{2N-1}. \quad (4)$$

This differential equation is readily solved for the time to increase from $k = 1$ to $k = N - 1$, $\frac{2N-1}{N} \ln(N - 1)$. Unfortunately, there are no nice error bounds for approximating the discrete process (3) with the continuous process (4), and there is further error from approximating a stochastic process with a deterministic process (Eq. (2) and (3) are not valid with expected values on the right hand side). (The fact that the solution to the differential equation blows up at $k = N$ is not a problem, it is readily verified that the discrete process Eq. (2) will increase from $N - 1$ to (more than) N in two generations, so stopping at $N - 1$ is not important.) Hence the utility of the estimate $\frac{2N-1}{N} \ln(N - 1)$ relies on simulations. Fifty simulations for each of the population sizes in Table 1 (500, 1000, 2000, 4000) had standard deviations less than .5 and mean values less than but within 6% of $2 \ln(N - 1)$. The result is also confirmed for a modification of the Moran model below, without error in calculating expected values, and small error for approximating the discrete process with a differential equation.

Because the expected time since the most recent pedigree ancestor of the

entire population and the expected time since every individual was a pedigree ancestor of the entire population are greater than $\log_2 N$, but less than twice the expected time until a specified individual becomes a pedigree ancestor of the entire population, which is approximately $4 \ln(N - 1)$, we have a rather narrow estimate of the expected time since the most recent pedigree ancestor and the expected time since the entire population was a pedigree ancestor ($\log_2 N = \frac{1}{\ln 2} \ln N \doteq 1.44 \ln N$, so our upper bound is approximately 2.8 times our lower bound).

To determine the discrepancy between the discrete process Eq. (2) (i.e., assuming $E[k(t)] = k(t)$ in order to iterate) and the continuous approximation Eq. (4), the time until N was attained determined by numerical iterations of Eq. (2) was contrasted to $2 \ln(N - 1)$, the (approximate) solution to the continuous approximation Eq. (4). For values of N ranging from 10 to 1 000 000, the discrepancy ranged from the continuous approximation 4.39 corresponding to the discrete value 6 for $N = 10$ to the continuous approximation 27.63 corresponding to the discrete value 25 for $N = 1\,000\,000$. Hence there is not a direction of inequality between the discrete value and the continuous approximation, but the approximation is reasonable for the values studied. For $N = 500, 1000, 2000,$ and 4000 (the population sizes used in Table 1), the integer value from discrete iteration is the same as the integer obtained by rounding up the value from the continuous approximation. (The time until N individuals is obtained by discrete iteration is not included in Table 1, the discrete iteration column in Table 1 terminates the iterations at $N - 1$.)

4 Overlapping generations

4.1 Modified Moran model

An alternative to the Wright-Fisher model is the Moran (1958) model which replaces individuals one by one rather than simultaneously each generation. In order to employ the constraint of two progeny per individual, individuals will be viewed as originally containing two gametes which they will ultimately contribute to future individuals, they lose one gamete each time they mate. (These are not the gametes that formed them, they are randomly formed from the genetic material in the individual; but this is not relevant for the question of pedigree ancestry.) Constant population size is interpreted as a total of $2N$ gametes inside the individuals, this will entail more than N individuals because individuals which have mated once will contain only one gamete (individuals die when they mate twice, hence give up both gametes). The pedigree structure can be viewed as $2N$ gamete lineages which randomly (selfing is allowed) meet (mate) and immediately separate two at a time (they immediately separate into two gametes, but those gametes stay together in an individual until that individual mates). This process is illustrated in Fig. 2, where the number of gametes in an individual is listed, and one mating occurs between each horizontal row. This graph also provides random meetings and separations of lineages going backward in time, so the expected time until a specified individual is a pedigree ancestor of the entire population is the same as the expected time since the entire population was pedigree ancestors of a specified individual.

[FIGURE 2 NEAR HERE]

Moran (1958) found that for the Poisson progeny distribution the (genetic) fixation time is half that for the Wright-Fisher model. This is readily confirmed

by calculating $V_{\delta x}$ and employing Eq. (1). With the Poisson progeny distribution under the Moran model $V_{\delta x} = \frac{x(1-x)}{N}$ (after rescaling for time; each mating under the Moran model corresponds to $\frac{1}{N}$ of a generation under the Wright-Fisher model, since a Wright-Fisher generation has N matings) so that the fixation time is $2N$ generations. In the case of two progeny per individual $V_{\delta x} = \frac{x(1-x)}{4N}$ (after rescaling for time), which provides that the fixation time is $8N$ generations, the same as with discrete generations.

For the question of pedigree common ancestry, we shall use κ to designate the number of gametes descended from a specified individual. Each mating removes two gametes from the population, but creates a new individual which replaces those two gametes, so the total number of gametes in the population remains constant at $2N$. κ will not change if two individuals descended from the specified individual mate or if two individuals not descended from the specified individual mate, but will increase by 1 if an individual descended from the specified individual mates with an individual not descended from the specified individual (this is a perhaps arbitrary census time decision: when a mating occurs both gametes in the resultant individual are counted as descendants of the original individual if either gamete which formed that individual was a descendant). Thus random mating (weighted by the number of gametes and allowing selfing) produces the discrete time equation:

$$E[\kappa(t+1)] = \kappa + 2 \frac{\kappa(2N - \kappa)}{2N(2N - 1)} \quad (5)$$

where the time unit is per mating ($\frac{1}{N}$ of a discrete generation) and κ counts the number of gametes descended from the original individual. The quantity $2 \frac{\kappa(2N - \kappa)}{2N(2N - 1)}$ is the probability that the new individual was formed from a descendant ($\frac{\kappa}{2N}$) and a non-descendant ($\frac{2N - \kappa}{2N - 1}$) or a non-descendant ($\frac{2N - \kappa}{2N}$) and a descendant ($\frac{\kappa}{2N - 1}$), which matings increase the number of descendant gametes

by one since both the new gametes are descendants.

The associated differential equation

$$\frac{d\kappa}{dt} = \frac{2\kappa(2N - \kappa)}{2N(2N - 1)} \quad (6)$$

is the same as Eq. (4) after adjusting the time scale by N (because there are N matings in a generation) and substituting $2k = \kappa$. It has the same lack of error bounds.

4.2 Exact analysis

Although it is of interest that the discrete and overlapping generation models provide the same differential equation, that differential equation is an approximation, hence the solution is not precise (but it is buttressed by numerical simulations). The overlapping generation model allows one to explicitly calculate the expected time until a specified individual becomes a pedigree ancestor of the entire population, rather than rely on the differential equation from an heuristic argument and simulations. The crucial feature for the analysis of this model is that each reproductive event entails increasing the number of descendants by 0 or 1.

Theorem: With overlapping generations and two progeny per individual, the expected time until a specified individual becomes a pedigree ancestor of the entire population is $O(\ln N)$.

Proof: As noted above, the probability that the number of descendants of a specified individual increases (by one) when a new individual is formed is $\frac{2\kappa(2N - \kappa)}{2N(2N - 1)}$ where κ is the number of descendants, hence the expected time to increase from κ to $\kappa + 1$ is

$$\sum_{i=0}^{\infty} \left(1 - \frac{2\kappa(2N - \kappa)}{2N(2N - 1)}\right)^i = \frac{2N(2N - 1)}{2\kappa(2N - \kappa)}. \quad (7)$$

Summing this from $\kappa = 2$ (one individual has two gametes) to $2N - 1$ (so the additional descendant will make $2N$ descendants) gives

$$\frac{2N - 1}{2} \sum_{\kappa=2}^{2N-1} \frac{2N}{\kappa(2N - \kappa)} = \frac{2N - 1}{2} \sum_{\kappa=2}^{2N-1} \left(\frac{1}{\kappa} + \frac{1}{2N - \kappa} \right) \quad (8)$$

which is approximately equal to

$$\frac{2N - 1}{2} \int_2^{2N-1} \frac{1}{\kappa} + \frac{1}{2N - \kappa} d\kappa + N = \frac{2N - 1}{2} (\ln(2N-1) - \ln(2) - \ln(1) + \ln(2N-2)) + N. \quad (9)$$

(The summation is bounded by $\frac{2N-1}{2}(\ln(2N-1) - \ln(2) - \ln(1) + \ln(2N-2) - \frac{1}{2N-1} - 1) + N$ and $\frac{2N-1}{2}(\ln(2N-1) - \ln(2) - \ln(1) + \ln(2N-2) + \frac{1}{2} + \frac{1}{2N-2}) + N$ using integral bounds on Riemann sums. This provides that the difference between (8) and (9) is less than $\frac{3(2N-1)}{4} + \frac{4N-3}{2(2N-2)}$ on the mating time scale, which is less than $\frac{3}{2} + O(\frac{1}{N})$ on the generation time scale.) Dividing by N to change from the mating time scale to the generation time scale provides that the time until a specified individual becomes a pedigree ancestor of the entire population is approximately $2 \ln(2N)$ confirming the heuristic result from the differential equation.

As noted above, twice the expected time until a specified individual becomes a pedigree ancestor of the entire population is an upper bound on the the expected time since a common pedigree ancestor. A lower bound is found by noting that the number of descendants (κ) can only increase by one each mating, so $2N - 2$ matings or $\frac{2N-2}{N}$ generations is a lower bound on the time since a common pedigree ancestor, hence the expected time since a common pedigree ancestor is between 2 and $4 \ln N$ generations. A systematic mating system can make everybody $3N - 3$ matings ago a pedigree ancestor of everyone in the present population, so the time since everyone in the population was an ancestor of everyone in the present population is between 3 and $4 \ln N$ generations.

5 Partial selfing

For the Poisson progeny distribution, Wiuf and Hein (1999) found that the time since a common pedigree ancestor $\log_2 N$ became $\log_{2-s} N$ where a fraction s of the population had only one parent. This expression blows up as $s \rightarrow 1$, but the approximation is no longer valid as $s \rightarrow 1$. If the entire population is obligate selfers, the question of common pedigree ancestry is essentially the same as the question of common genetic ancestry for a haploid population of size N , hence the time until a common pedigree ancestor is approximately $2N$ generations.

The time until a specified individual becomes an ancestor of the entire population with partial selfing and two progeny per individual (one if it selfs) is modelled by randomly determining whether each individual will self (the probability an individual selfs, s , is specified), replicating the selfing individuals into the next generation, and then creating the rest of the next generation by randomly mating the non-selfing individuals (randomly pairing their two gametes); this allows chance selfing among the random mating individuals. That is the process that is used in the simulations in Table 1. However, in order to get an analytic formula, it is assumed that the fraction s of the descendants of the specified individual self and the fraction s of those not descended from the specified individual self (this makes the number of individuals who self, i.e., are removed from the mating pool before mating occurs, deterministic rather than random). This removes the variation in the rate of increase of the number of descendants, k , hence should reduce the time until the entire population is descended from a single individual. The formula for $E[k(t+1)]$ becomes analogous to (2)

$$E[k(t+1)] = sk + (1-s)N \times \left(\frac{2k(1-s)}{2N(1-s)} \times \frac{2k(1-s)-1}{2N(1-s)-1} + 2 \times \frac{2k(1-s)}{2N(1-s)} \times \frac{2(N-k)(1-s)}{2N(1-s)-1} \right) \quad (10)$$

where sk manifests that the selfed progeny of descendants of the specified in-

dividual are descendants of the specified individual, $(1 - s)N \times \left(\frac{2k(1-s)}{2N(1-s)}\right) \times \frac{2k(1-s)-1}{2N(1-s)-1}$ manifests that outcrossed progeny of two descendants of the specified individual are descendants of the specified individual, and $(1 - s)N \times 2 \times \frac{2k(1-s)}{2N(1-s)} \times \frac{2(N-k)(1-s)}{2N(1-s)-1}$ manifests that outcrossed progeny of one descendant of the specified individual and one non-descendant of the specified individual are descendants of the specified individual. Subtracting k provides that in the case of deterministic partial selfing

$$E[\Delta k] = \frac{2k(1-s)^2(N-k)}{2N(1-s)-1}, \quad (11)$$

which reduces to Eq. (3) in the case $s = 0$.

The associated differential equation provides the time until a specified individual becomes a pedigree ancestor of the entire population

$$\frac{2(1-s)N-1}{2(1-s)^2N} \times 2 \ln(N-1) \doteq \frac{2 \ln(N-1)}{1-s}. \quad (12)$$

This is not a rigorous approximation for the original model: selfing was changed from a random to a deterministic portion of the population; as with Eq. (3), Eq. (11) is not valid with $E[k]$ on the right hand side; and since Eq. (4) is a special case, there are no good error bounds in approximating the difference equation with a differential equation. It is a heuristic derivation to find a formula to compare with simulations. Its utility is the extent to which it is a concise description of the simulation results. It is generally within 10% of the simulation values in Table 1.

The ratio $\frac{1}{1-s}$ which multiplies the time until a specified individual becomes an ancestor of the entire population with two progeny per individual is greater than the ratio $\frac{1}{\log_2(2-s)}$ which multiplies the time since a common ancestor with the Poisson progeny distribution (Wiuf and Heine, 1999). Furthermore, although the validity of the approximation $\frac{\log_2 N}{\log_2(2-s)}$ ends before $s \rightarrow 1$ (the

time with $s = 1$ is approximately $2N$), with two progeny per individual (i.e., one progeny per individual in the case of selfing) the time until a specified individual becomes a pedigree ancestor of the entire population (hence the time since a common pedigree ancestor) indeed blows up as the formula $\frac{2 \ln(N-1)}{1-s}$ provides. The Poisson distribution with obligate selfing allows some lineages to replace others, hence common ancestors occur, but with two progeny per individual with obligate selfing, each individual has a lineage which persists until infinity without meeting other lineages.

If there is significant selfing (i.e., $s(1-s) > \frac{1}{2N}$), the value N will never be attained by iterating Eq. (10). Therefore iteration of Eq. (10) was terminated at $N - 1$ in Table 1. For large values of s , there are other problems with Eq. (10).

[TABLE 1 NEAR HERE]

6 Regular systems of inbreeding

6.1 Mating structures

[FIGURE 3 NEAR HERE]

Two regular systems of inbreeding which entail two progeny per individual have been studied by Kimura and Crow (1963). Half-sib mating (also called circular mating) consists of a linear (circular) array of individuals, where each mates with its neighbor on each side. The resultant progeny form a linear (circular) array of individuals, and each mates with its neighbor on each side. The individuals in the pedigree thus form a large quincuncial array, and the lines of descent look like chainlink fence. Maximum avoidance of inbreeding (2^n -fold n^{th} cousin mating in a population of $N = 2^{n+1}$ individuals) entails mating between pairs

of adjacent individuals in a linear (circular) array. Their progeny are then located in antipodal positions in the circular array of individuals in the next generation such that the sequence of the parental pairs is repeated twice in the sequence of their progeny. Pairs of adjacent individuals are formed, and the generations continue. Both of these mating structures are consistent with a bisexual population.

6.2 Genetic ancestry

I am not aware of estimates for the time since a common genetic ancestor under regular systems of inbreeding, but some remarks can be made. Avoidance of inbreeding will increase the sampling variance (the two alleles in an individual are less likely to be the same) while half-sib mating will decrease the sampling variance (individuals are more likely to have two genes which are identical by descent). The magnitude of the change in the sampling variance can be gauged by comparing the fixation times under simulations to the fixation times for the extreme possible sampling variances.

As noted above, the sampling variance for random mating with two progeny per individual is $\frac{x(1-x)}{4N}$. Only heterozygous individuals contribute to the sampling variance, each one contributes $.5/(2N)^2$. The greatest possible number of heterozygous individuals if one allele frequency is x is $N \times \min(2x, 2(1-x))$. This provides an upper bound $\frac{\min(2x, 2(1-x))}{8N}$ on the sampling variance possible with maximum avoidance of inbreeding. All individuals could be homozygous, but with half-sib mating there will need to be at least two interfaces between the two alleles, and those will need to entail a heterozygous individual at least half the time, hence with half-sib mating the sampling variance will average at least $.5/(2N)^2$ while two alleles are segregating.

Substituting these values into (1) gives the times until fixation. The sampling variance $\frac{x(1-x)}{4N}$ for random mating with two progeny per individual yields the fixation time approximately $8N$. The sampling variance $N \times \min(2x, 2(1-x)) \times \frac{1}{2}/(2N)^2$ yields the fixation time approximately $4N$. The sampling variance $\frac{1}{2}/(2N)^2$ yields the fixation time approximately $\frac{8}{3}N^2$.

Simulations for avoidance of inbreeding were performed for 8, 64, and 512 individuals (16, 128, and 1024 genes). The mean times until fixation were 54.2, 386.05, and 4753.1 generations, respectively, based on 20 simulations each. The standard deviations were $s = 29.67, 225.58, \text{ and } 2270.45$, respectively. Only two of these means are less than $8N$ which you would expect for random mating with two progeny per individual ($54 < 64, 386 < 512, 4753 > 4096$), but with the large standard deviations the results are consistent with with a slight reduction in expected fixation time below $8N$. However, they are not close to the limit $4N$ associated with maximum heterozygosity.

Simulations for half-sib (circular) mating for 10, 100, and 500 individuals (20, 200, and 1000 genes) produced mean fixation times of 103.2, 7368.85, and 180091.05 respectively, based on 20 simulations each. The standard deviations were 70.44, 5075.27, and 148555.75, respectively. The standard deviations were over half the value of the mean. The mean fixation times are well over $8N$ (80, 8000, and 4000) for random mating, but are not close to the limit values $\frac{8}{3}N^2$ (267, 26,667, and 666,667). But half-sib mating seems to have a greater effect on fixation time than maximum avoidance of inbreeding.

6.3 Pedigree ancestry

The time since a pedigree (non-genetic) ancestor is quite easily calculated for the regular systems of inbreeding considered here. For half-sib (circular pair) mat-

ing, the number of descendants (or ancestors) increases by one each generation, hence $N - 1$ generations are required to have N descendants (or ancestors). Before that time nobody is an ancestor of the entire population, but at that time everyone is an ancestor of the entire population. N is much less than the genetic fixation times determined by simulation above. But N is much greater than $2 \ln N$, the pedigree ancestry time with random mating and two progeny per individual.

Under maximum avoidance of inbreeding (2^n -fold n^{th} cousin mating in a population of $N = 2^{n+1}$ individuals) nobody is an ancestor of the entire population until the $n + 1^{\text{st}}$ generation ($\log_2(N)$), at which time everyone is an ancestor of the entire population. This gives a pedigree ancestry time of $\log_2 N$ which is much less than the genetic ancestry time from the above simulations. But $\log_2 N = \frac{\ln N}{\ln 2} \doteq 1.44 \ln 2$ is just a little less than $2 \ln N$, the pedigree ancestry time with random mating and two progeny per individual. Hence with pedigree ancestry time half-sib mating also has a greater effect than maximum avoidance of inbreeding.

7 Conclusion

The main result of this paper is the calculation of the expected time since a common pedigree ancestor when there are two progeny per individual. Heuristic calculations without error bounds are confirmed by simulations and exact calculations for the Moran model. The result is that the expected time since a common ancestor of the entire population and the expected time since everyone was a common ancestor of the entire population are between $\log_2 N$ and $4 \ln N$ generations (with two progeny per individual, everyone becomes an ancestor of the entire population, unlike with the Poisson progeny distribution under which

approximately 20% of individuals ultimately have no descendants). These values compare with Chang's (1999) results for the Poisson progeny distribution that the expected time since a common pedigree ancestor is $\log_2 N$ and the expected time since everyone whose lineage does not go extinct is a common ancestor is $1.77 \log_2 N$.

One difference from the Poisson distribution is that with two progeny per individual, no pedigree lineages disappear, and eventually everyone is a common pedigree ancestor of everyone in a subsequent generation. Another difference is that with the Poisson progeny distribution increased selfing only increases the expected time since a common pedigree ancestor to $2N$ generations when the probability of selfing becomes one, but with two progeny per individual the expected time since a common pedigree ancestor approaches infinity as the probability of selfing approaches one.

A mathematically aesthetic feature of pedigree ancestry with two progeny per individual is that the time reversal of the process is the same as the original process: every individual has two progeny and two parents.

The progeny distribution of humans is not binomial, but rather negative binomial with the variance 1.5 to 3 times the mean (Cavalli-Sforza and Bodmer 1971, p.311). This suggests that two progeny per individual which halves the variance from the Poisson distribution is a worse approximation than the Poisson distribution which has the variance equal to the mean. But if progeny in large sibships are less fecund, a lower variance provides a better model. Further, increased use of birth control and government policies regulating the number of children may reduce the variance. Even if two progeny per individual is not an appropriate model for humans, this work still illustrates the importance of progeny distribution for time since a common pedigree ancestor.

Regular systems of inbreeding are of interest in this context because they provide extreme cases of what is possible under random mating with two progeny per individual. Maximum avoidance of inbreeding provides the lower bound on the time since a common pedigree ancestor $\log_2 N$, which is the same order as the expected time since a common pedigree ancestor under random mating $O(\ln N)$. Half-sib mating requires $N - 1$ generations until a common pedigree ancestor, which is larger than $O(\ln N)$. Maximum avoidance of inbreeding and circular (half-sib) mating both have the property that the generation where a common pedigree ancestor first occurs is the generation when everyone is a common pedigree ancestor, i.e., the time since a common pedigree ancestor is the same as the time since everyone is a common pedigree ancestor.

References

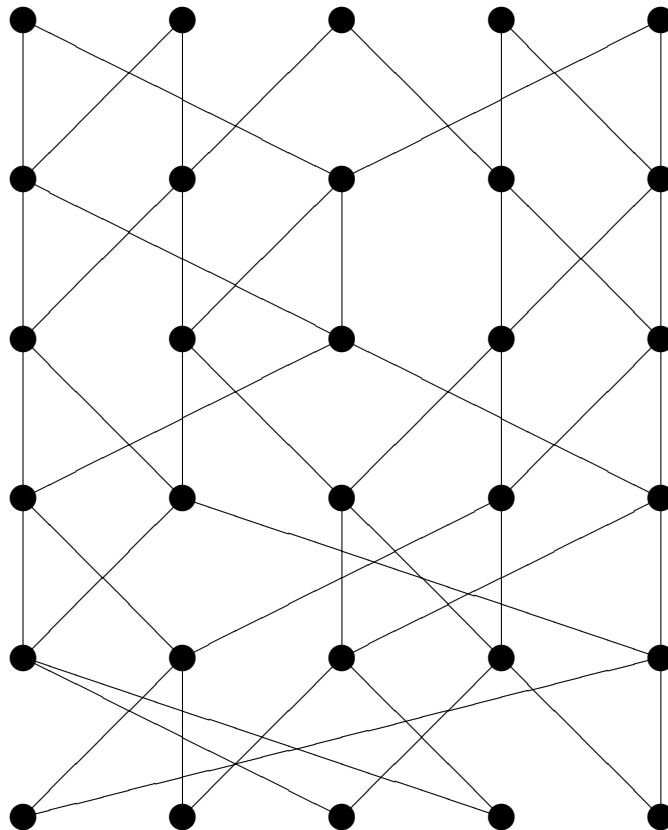
- CAMPBELL, R. B. (1995). The effect of mating structure and progeny distribution on heterozygosity versus the number of alleles as measures of variation. *J. Theoret. Biol.* **175**, 503–509.
- CAMPBELL, R. B. (1999). The coalescent in the presence of background fertility selection. *Theoret. Popn Biol.* **55**, (1999), 260–269.
- CAVALLI-SFORZA, L. L. AND BODMER, W. F. (1971). *The Genetics of Human Populations*. W. H. Freeman, San Francisco.
- CHANG, J. T. (1999). Recent common ancestors of all present-day individuals. *Adv. Appl. Prob.* **31**, 1002–1026.
- CROW, J. F. AND KIMURA, M. (1970). *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- KIMURA, M. AND CROW, J. F. (1963). On the maximum avoidance of inbreeding. *Genet. Res. (Camb.)*. **4**, 399–416.
- KINGMAN, J. F. C. (1982). On the genealogy of large populations. *J. Appl. Prob.* **19A**, 27–43.
- MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.* **154**, 60–71.
- ROBINSON, P. AND BRAY, D. F. (1965). Expected effects on the inbreeding coefficient and rate of gene loss of four methods of reproducing finite diploid populations. *Biometrics*. **21**, 447–458.
- WIUF, C. AND HEIN, J. (1999). Discussion: Recent common ancestors of all present-day individuals. *Adv. Appl. Prob.* **31**, 1029–1030.
- WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics*. **16**, 97–159.

Table 1: Generations until an individual becomes a common ancestor with two progeny per individual.

s	N	mean	min	max	standard deviation	$\frac{2\ln(N-1)}{1-s}$	discrete iteration
0	500	12.16	12	13	0.37	12.43	12
0	1000	13.28	13	14	0.45	13.81	13
0	2000	14.38	14	15	0.49	15.20	14
0	4000	15.64	15	16	0.48	16.58	16
0.2	500	15.92	14	19	1.14	15.53	15
0.2	1000	17.58	16	22	1.16	17.27	17
0.2	2000	19.22	17	22	1.09	19.00	19
0.2	4000	20.84	19	24	1.15	20.73	20
0.4	500	22.16	20	30	2.05	20.71	21
0.4	1000	24.52	20	30	2.14	23.02	23
0.4	2000	26.24	23	31	1.78	25.33	25
0.4	4000	28.40	25	35	1.84	27.65	28
0.6	500	33.94	28	45	3.73	31.06	31
0.6	1000	37.36	32	53	3.83	34.53	35
0.6	2000	39.92	33	54	4.00	38.00	38
0.6	4000	43.08	38	54	3.97	41.47	42
0.8	500	67.48	55	91	9.01	62.13	62
0.8	1000	73.44	62	93	7.43	69.07	69
0.8	2000	84.98	68	107	8.64	76.00	76
0.8	4000	88.80	73	111	9.34	82.94	83

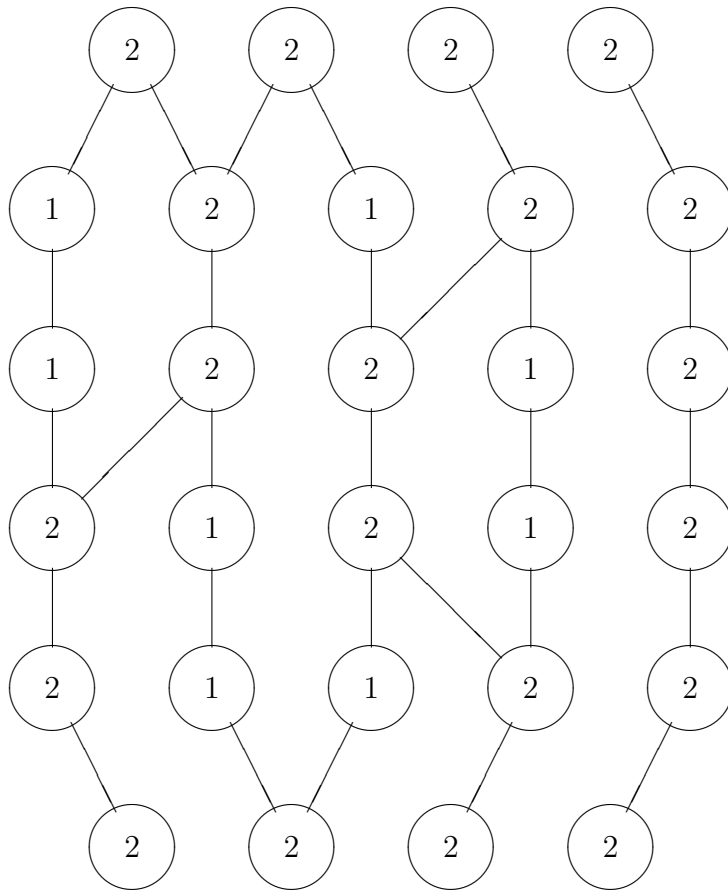
The population size is N , the proportion which selfs is s , and the discrete iteration column gives the number of generations (using Eq. (10)) until there are $N - 1$ (rather than N) descendants. The mean, minimum, maximum, and standard deviation are based on 50 simulations for each set of parameter values.

Figure 1: Times until pedigree ancestry of the entire population



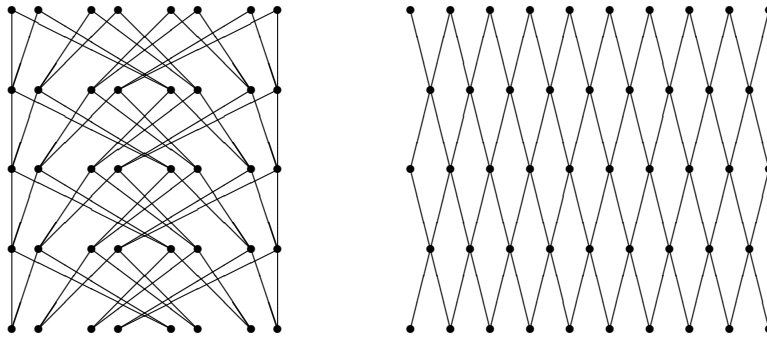
Time advances going down the page. The time until some individual is a pedigree ancestor of the entire population is 3 (achieved by the third and fifth individual), the time until a specified individual is a pedigree ancestor of the entire population is 3, 4, or 5 depending on which individual is specified, and the time until all individuals are pedigree ancestors of the entire population is 5.

Figure 2: Moran model with two progeny per individual



Each new individual (individuals with two lines entering them from above) removes a gamete from each parent (if a parent had only one gamete, it does not persist).

Figure 3: Maximum avoidance of inbreeding and half-sib mating



Generations proceed down the page with lines indicating parentage. For half-sib mating (the right schematic), either the the rightmost individuals are the same as the left most individuals (thereby forming a cylinder) or the figure extends infinitely in the horizontal direction. (After Kimura and Crow (1963).)