

## 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

campbell@math.uni.edu

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

June 2008

(full manuscript at <http://www.math.uni.edu/~campbell/twoprograms.pdf>)

### Background

This work studies the time since a common pedigree ancestor (a pedigree ancestor is an ancestor which appears in a pedigree whether or not it passed genetic material on to its descendants) rather than genetic ancestry (the standard identity by descent), and two progeny per individual rather than the Poisson progeny distribution. The standard result for the Wright-Fisher model is that the time since a common (genetic) ancestor is  $4N$  generations. Chang (1999) found that for the Poisson progeny distribution the time since a common pedigree ancestor is  $O(\ln N)$ . For both the diffusion model and the coalescent model for genetic ancestry, it is known that two progeny per individual (versus the Poisson progeny distribution) doubles the expected time since a common ancestor to  $8N$  generations.

### Results

Results for time since a pedigree (non-genetic) common ancestor include that constraining individuals to two progeny (versus Poisson distribution) increases the time, but the time remains  $O(\ln N)$ . Selfing increases the expected time since a common pedigree ancestor with both the Poisson progeny distribution and two progeny per individual, but that time asymptotically approaches infinity as selfing approaches 1 with two progeny per individual which is contrasted with twice the population size with the Poisson distribution. Regular systems of inbreeding with two progeny per individual can either increase or decrease the time since a common pedigree ancestor as contrasted with random mating with two progeny per individual; maximum avoidance of inbreeding provides the time  $\log_2 N$  since common ancestry, which is the lower bound for all mating systems with two progeny per individual.

### Mathematical Aesthetics

The model for pedigree ancestry with two progeny per individual is mathematically aesthetic in that it can be viewed as either individuals randomly choosing two parents, or individuals randomly choosing two progeny (the ancestral process is the same as the descendant process), hence it is the same model if time is reversed. A result of this is that no lineages disappear, everyone will eventually become an ancestor of the entire population. This symmetry with respect to time does not hold with the Poisson progeny distribution for pedigree ancestry (where ancestry can be traced back to infinity, but approximately 20% of the individuals in a generation ultimately have no descendants). Neither does it hold for genetic ancestry with Poisson progeny distribution or two progeny per individual (where ancestry can be traced back to infinity, but all but one individual in a generation ultimately have no descendants).

### The Solution

The difference equation for the expected increase in the number of descendants each generation with two progeny per individual provides the differential equation approximation

$$E[\Delta k] = \frac{2k(N-k)}{2N-1} \doteq \frac{dk}{dt}$$

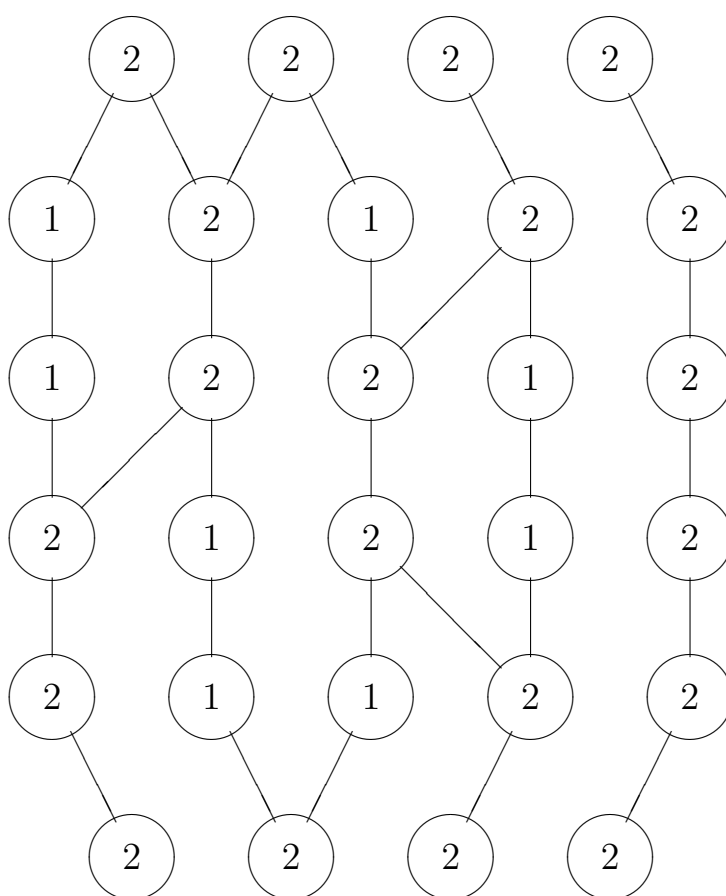
which has the solution  $2\ln(N-1)$  for the time since a common ancestor, which is  $O(\ln(N))$ , but greater than in the case of the Poisson distribution. However, this result must be considered as heuristic, since there are errors from approximating a stochastic process with a deterministic process, and from approximating a discrete process with a continuous process, none of which have nice bounds. But simulations confirm that the result is valid. Furthermore, a modification of the Moran model (which replaces one individual at a time versus discrete generations) produces the same differential equation as above, but with exact calculations of expected times and small error from approximating a discrete process with a continuous process.

### Moran Model

Replacing one individual at a time within the constraint of two progeny per individual is obtained by assuming every individual initially contains two reproductive potentials, one of which is lost each time it contributes to a progeny. Parents are chosen by randomly choosing among the reproductive potentials. Reproductive potentials have access to the entire genetic composition of an individual, they are not restricted to half the genome. When both reproductive potentials have been used an individual is removed from the population. This allows the actual number of individuals to fluctuate, but the number of reproductive potentials in the population remains constant at  $2N$ . This is illustrated in the following schematic.

(see reverse)

Figure 1: Moran model with two progeny per individual



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

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 a discrete deterministic iteration 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.