

# A Model for Random Genetic Drift in Finite Populations

PATRICK STALEY  
Mathematics Department  
Southwestern College  
900 Otay Lakes Road, Chula Vista California 91910  
UNITED STATES  
<http://www.mr-ideahamster.com>

*Abstract:* - An eigenvalue analysis of random genetic drift in populations of size two produces a geometrically declining heterozygosity with a rate different from that predicted by the Wright-Fisher model. This is posited as the reason that actual experiments do not match the Wright-Fisher model prediction.

*Key-Words:* - random genetic drift, finite populations, eigenvalue analysis, Wright-Fisher Model

## 1 Introduction

Standard mathematical models for random genetic drift in finite populations like the Wright-Fisher model do not effectively mimic results from actual experiments. The usual explanation for this discrepancy is that the experiments never exactly match the assumptions of simple random genetic drift. The thesis of this paper is that the standard mathematical model is what fails to match the assumptions of the random genetic drift experiment. Specifically it ignores the restricted choice of gametes—one egg and one sperm.

## 2 Assumptions for Random Genetic Drift in Finite Populations

From each generation  $N$  diploid individuals are chosen to be parents for the next generation. Reproduction is sexual with equal numbers of males and females. For simplicity we follow only one gene with two alleles. Both of the totally homozygous states are stable absorbing equilibria. This is true in both the standard model and in the experimental results. The discrepancy lies in the rate of loss of heterozygosity.

## 3 The Standard Model

The standard model has the  $N$  diploid individuals producing an infinite pool of gametes whose allele frequencies match those of the parent group, say  $p$  for the 'A' allele and  $q=1-p$  for 'a' allele. The offspring are then produced with frequencies  $p^2$  for genotype  $AA$ ,  $2pq$  for genotype  $Aa$ , and  $q^2$  for genotype  $aa$ .

From this group another  $N$  individuals are chosen at random to be parents for the next generation, etcetera. {See Chapter 7 of reference [3], which comes from [1] and [2]}

The problem with the Standard Model is that it fails to account for the assumption of sexual reproduction. That assumption leads us to two separate gamete pools—one for eggs and one for sperm.

### 3.1 An Example Illustrating the Problem with the Standard Model

Consider the case where  $N=2$  and  $p=3/4$ . The standard model predicts progeny genotype frequencies of  $p^2=9/16$  for  $AA$ ,  $2pq=3/8$  for  $Aa$ , and  $1/16$  for  $aa$ . Thus the random draw for the next parent group is  $(9/16)^2 = 81/256$  for  $AA AA$ ,  $2(9/16)(3/8) = 27/64$  for  $AA Aa$ ,  $(3/8)^2 = 9/64$  for  $Aa Aa$ ,  $2(9/16)(1/16) = 9/128$  for  $AA aa$ ,  $2(3/8)(1/16) = 3/64$  for  $Aa aa$ , and  $(1/16)^2 = 1/256$  for  $aa aa$ .

This might seem reasonable until we realize that  $N=2$  means one male and one female and  $p=3/4$  means exactly genotypes  $AA$  and  $Aa$ . For every progeny the gamete from the  $AA$  parent must be  $A$  and the gamete from the other is equally likely to be  $A$  or  $a$ . Thus the correct progeny genotype frequencies are  $1/2$  for  $AA$ ,  $1/2$  for  $Aa$ , and zero for  $aa$ . The correct frequencies for the next generation parent group are  $(1/2)^2=1/4$  for  $AA AA$ ,  $2(1/2)(1/2) = 1/2$  for  $AA Aa$ ,  $(1/2)(1/2) = 1/4$  for  $Aa Aa$ , zero for  $AA aa$ , zero for  $Aa aa$ , and zero for  $aa aa$ .

This is clearly different from the standard model predicted frequencies.

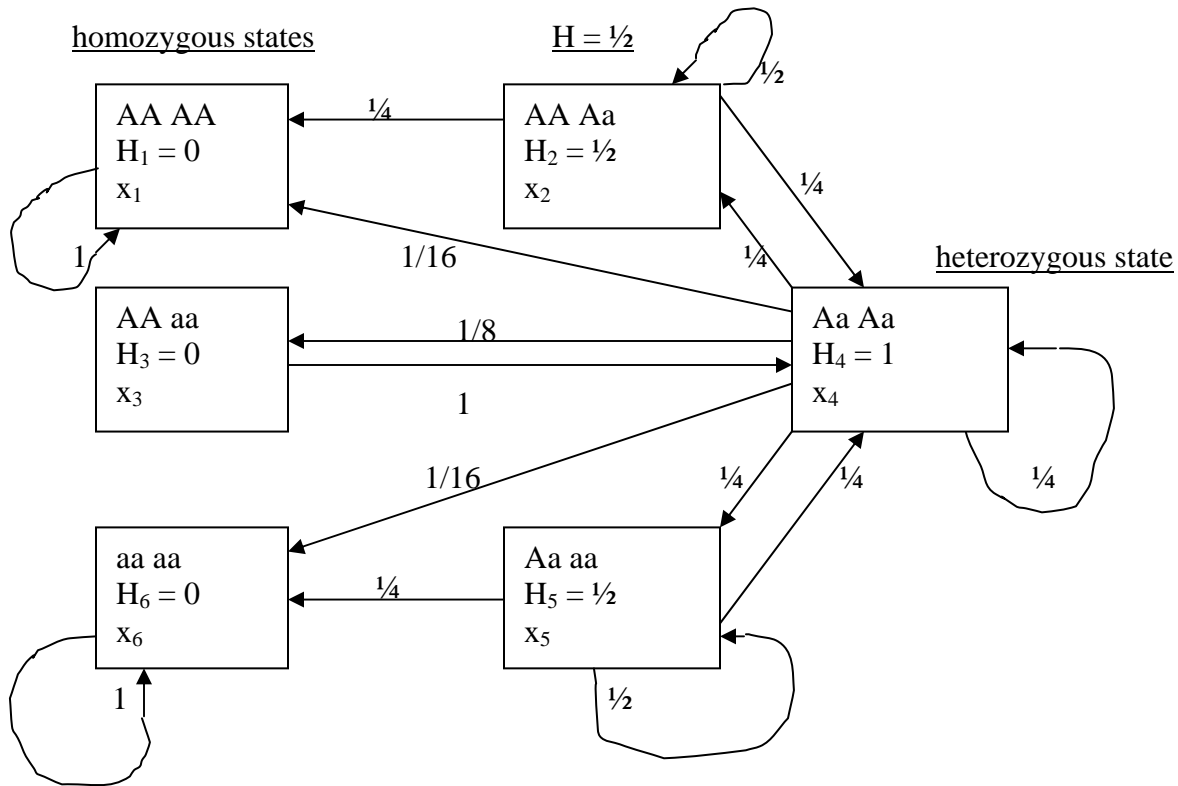
### 3.2 What to Do

It seems obvious to the author that correcting the model for two separate gamete pools is the first step to resolving differences between the mathematical model and the experimental results.

### 4 Corrected Model for N=2

For N=2 we have the six states: AA AA, AA Aa, AA aa, Aa Aa, Aa aa, and aa aa. Let state i have frequency  $x_i$  and heterozygosity  $H_i$ . The state diagram showing transition probabilities is given below.

Fig. 1 State Diagram of Transition Probabilities.



The state vector at generation k is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}_k \rightarrow = x_k$$

And from the state diagram above we get the stochastic transition matrix:

$$\Phi = \begin{pmatrix} 1 & 1/4 & 0 & 1/16 & 0 & 0 \\ 0 & 1/2 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/8 & 0 & 0 \\ 0 & 1/4 & 1 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 & 1/2 & 0 \\ 0 & 0 & 0 & 1/16 & 1/4 & 1 \end{pmatrix}$$

The transition to the next generation is

$$\vec{x}_k = \Phi \vec{x}_{k-1} = \Phi^k \vec{x}_0$$

Let the heterozygosity vector be

$$\vec{H} = \begin{pmatrix} H_{state\ 1} \\ \vdots \\ H_{state\ 6} \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \\ 0 \\ 1 \\ 1/2 \\ 0 \end{pmatrix}$$

Then the heterozygosity at generation k is

$$\vec{H} \cdot \vec{x}_k = \vec{H}^T \vec{x}_k = (1/2)x_{2,k} + x_{4,k} + (1/2)x_{5,k}$$

#### 4.1 Preservation of Allele Frequencies

We can easily verify that allele frequency is preserved. The vector for the 'A' allele

frequency by state is  $\vec{p} = \begin{pmatrix} 1 \\ 3/4 \\ 1/2 \\ 1/2 \\ 1/4 \\ 0 \end{pmatrix}$ . Likewise

the vector for the 'a' allele is  $\vec{q} = \begin{pmatrix} 0 \\ 1/4 \\ 1/2 \\ 1/2 \\ 3/4 \\ 1 \end{pmatrix}$ .

Then the allele frequencies for a given state vector  $\mathbf{x}$  are  $\vec{p} \cdot \mathbf{x} = p$  and  $\vec{q} \cdot \mathbf{x} = q$ .

Notice that  $\vec{p} \Phi = \vec{p}$  and  $\vec{q} \Phi = \vec{q}$  so that the allele frequency at generation k will equal the initial allele frequency:

$$\vec{p} \cdot \vec{x}_k = \vec{p} \cdot \Phi^k \vec{x}_0 = \vec{p} \cdot \vec{x}_0$$

#### 4.2 Behavior of the Corrected Model for N=2

The behavior of this model can be understood from an eigenvalue analysis. The eigenvectors for the matrix  $\Phi$  are those state vectors for which  $\Phi$  will change their magnitude but leave them in the same direction. The magnitude multiple change is called the eigenvalue. Specifically if  $\Phi$  has eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_6$  with corresponding eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_6$  then  $\Phi \mathbf{v}_i = \lambda_i \mathbf{v}_i$ . Arrange the eigenvalues in decreasing order of magnitude and express the initial state in terms of the eigenvectors and we have

$$\vec{x}_0 = a_1 \vec{v}_1 + a_2 \vec{v}_2 + a_3 \vec{v}_3 + \dots + a_6 \vec{v}_6$$

$$\vec{x}_k = \Phi^k \vec{x}_0 = a_1 \Phi^k \vec{v}_1 + a_2 \Phi^k \vec{v}_2 + a_3 \Phi^k \vec{v}_3 + \dots + a_6 \Phi^k \vec{v}_6$$

$$\vec{x}_k = \Phi^k \vec{x}_0 = a_1 \lambda_1^k \vec{v}_1 + a_2 \lambda_2^k \vec{v}_2 + a_3 \lambda_3^k \vec{v}_3 + \dots + a_6 \lambda_6^k \vec{v}_6$$

For this  $\Phi$  --  $\lambda_1 = \lambda_2 = 1$  ( $\mathbf{v}_1$  and  $\mathbf{v}_2$  are the pure homozygous states AA AA and aa aa),

$$\lambda_3 = \frac{1 + \sqrt{5}}{4} \approx 0.809, \quad \lambda_4 = 0.5,$$

$$\lambda_5 = \frac{1 - \sqrt{5}}{4} \approx -0.309, \quad \text{and } \lambda_6 = 0.25.$$

Observe that  $\lambda_4^k, \lambda_5^k$ , and  $\lambda_6^k$  are going to zero faster than  $\lambda_3^k$ . This means that if  $a_3 \neq 0$  and k is larger than ten then the first three eigenvectors will be the dominant components:

$$\vec{x}_k = \Phi^k \vec{x}_0 \approx a_1 \vec{v}_1 + a_2 \vec{v}_2 + a_3 0.809^k \vec{v}_3$$

To predict the heterozygosity keep in mind that  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are pure homozygous, i.e.  $\mathbf{H}^T \mathbf{v}_1 = \mathbf{H}^T \mathbf{v}_2 = 0$ . So as k gets larger the heterozygosity is

$$\vec{H}^T \vec{x}_k \approx a_1 \vec{H}^T \vec{v}_1 + a_2 \vec{H}^T \vec{v}_2 + a_3 0.809^k \vec{H}^T \vec{v}_3 = a_3 0.809^k \vec{H}^T \vec{v}_3$$

[In the N=2 case this is true even sooner since it turns out that  $\vec{H}^T \vec{v}_4 = \vec{H}^T \vec{v}_6 = 0$ .]

### 4.3 Conclusion for N=2 Model

Thus the heterozygosity declines geometrically with the rate factor 0.809. For the standard model  $\lambda_3 = 1 - 1/2N$  or .75 for N=2. The heterozygosity behavior in the corrected model corresponds to  $N \approx 2.6$  in the standard model. This is a significant difference that can affect the interpretation of experimental results. In particular if a

species of termites is in actuality totally monogamous with mates always chosen from siblings, i.e. the N=2 case discussed here, then an analysis using the standard Wright-Fisher model (which expects a heterozygosity loss rate of 0.75) would underestimate the actual inbreeding [4].

## 5 Appendix

This appendix contains the specific components for the eigenvectors ( $v_i$ 's), the coefficients for the heterozygous starting state ( $a_i$ 's for  $Aa Aa$ ), and results for the first 10 generations starting from the heterozygous state.

Table 1 Eigenvalues/Eigenvectors with Heterozygosity for N=2

	$\lambda_1=1$	$\lambda_2=1$	$\lambda_3=(1+\sqrt{5})/4$	$\lambda_4=1/2$	$\lambda_5=(1-\sqrt{5})/4$	$\lambda_6=1/4$
	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$
AA AA	1	0	-11-5 $\sqrt{5}$	1	-11+5 $\sqrt{5}$	1
AA Aa	0	0	4+4 $\sqrt{5}$	-2	4-4 $\sqrt{5}$	-4
AA aa	0	0	-2+2 $\sqrt{5}$	0	-2-2 $\sqrt{5}$	2
Aa Aa	0	0	16	0	16	4
Aa aa	0	0	4+4 $\sqrt{5}$	2	4-4 $\sqrt{5}$	-4
aa aa	0	1	-11-5 $\sqrt{5}$	-1	-11+5 $\sqrt{5}$	1
$\vec{H}^T v_i$	0	0	20+4 $\sqrt{5}$	0	20-4 $\sqrt{5}$	0

Initial state  $Aa Aa =$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{40} * (20v_1 + 20v_2 + v_3 + 0v_4 + v_5 + 2v_6)$$

so the heterozygosity at generation k starting from state  $Aa Aa$  should be  $(1/40)[(\lambda_3)^k(20+4\sqrt{5})+(\lambda_5)^k(20-4\sqrt{5})] \approx 0.7236(0.809^k)+0.2764(-0.309)^k$ . Which is what we see in the Table 2 below showing the results for the first 10 generations.

Table 2 State Frequency by Generation Starting from *Aa Aa*

t =	0	1	2	3	4	5	6	7	8	9
AA AA	0	.0625	.1406	.2070	.2627	.3079	.3445	.3742	.3982	.4177
AA Aa	0	.25	.1875	.1719	.1367	.1123	.0906	.0734	.0594	.0480
AA aa	0	.125	.03125	.0391	.0254	.0220	.0172	.0141	.0113	.0092
AaAa	1	.25	.3125	.2031	.1758	.1377	.1125	.0906	.0734	.0594
Aa aa	0	.25	.1875	.1719	.1367	.1123	.0906	.0734	.0594	.0480
aa aa	0	.0625	.1406	.2070	.2627	.3079	.3445	.3742	.3982	.4177
$H_t$	1	.5	.5	.375	.3125	.25	.2031	.1641	.1328	.1074
$H_t/H_{t-1}$		.5	1	.75	.8333	.8	.8124	.8080	.8093	.8087

Notice that  $H_t/H_{t-1}$  will tend to  $\lambda_3$ . In fact for larger N this is an appropriate method for approximating  $\lambda_3$ , the geometric heterozygosity loss rate.

*References:*

[1] Fisher, R.A., *The Genetical Theory of Natural Selection, Second Edition*, Dover Press, New York, 1958.

[2] Wright, S. 1931. Evolution in Mendelian populations. *Genetics* Vol.16, 1931, pp97-159.

[3] Hartl, D. and Clark, A. *Principles of Population Genetics, Third Edition*, Sinauer Associates, Inc., 1997.

[4] Thompson, G. and Herbert, P. *Population genetic structure of the Neotropical termite *Nasutitermes nigriceps* (Isoptera: Termitidae)* *Heredity*, Vol 80 Num 1, Jan 1998, pp. 48-55.

[5]Larson, E. and Edwards, B. *Elementary Linear Algebra, Fourth Edition*, Houghton Mifflin, 2000