

Neutral models of allele frequency change

$$P_0 = \frac{i}{2N_p}$$

of copies of the allele in the population

P_0 ←

P frequency of an allele in generation 0

+ generations of evolution → P_t

of individuals in the population

↑

frequency of the allele in generation t

Neutral models of allele frequency change

$$P_0 = \frac{i}{2N_R}$$

of copies of the allele in the population
frequency of an allele in generation 0

of individuals in the population + generations of evolution → P_t

frequency of the allele in generation t

What should we expect for P_t , given P_0 ?

- To start, assume the allele has no impact on evolutionary fitness (i.e. that it is neutral)

Neutral models of allele frequency change

of copies of the allele in the population

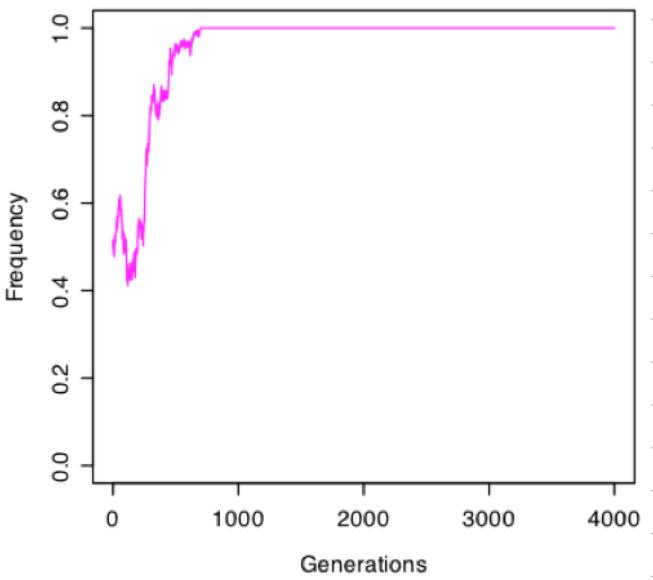
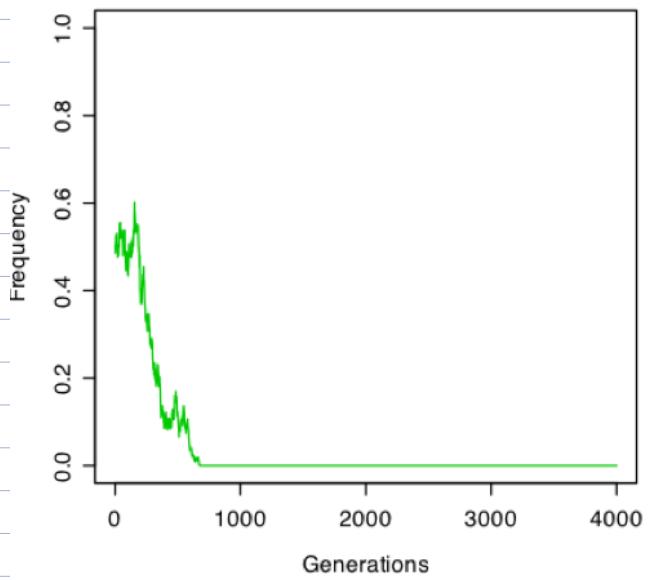
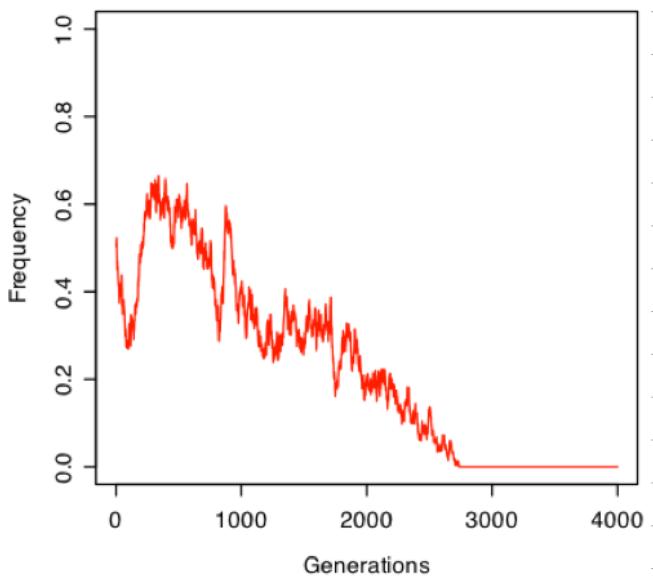
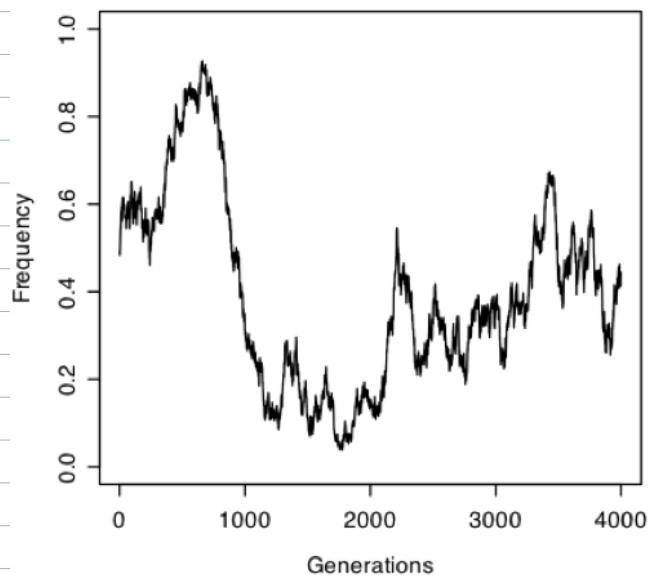
$$P_0 = \frac{i}{2N}$$

P_0 frequency of an allele in generation 0

+ generations of evolution $\rightarrow P_t$

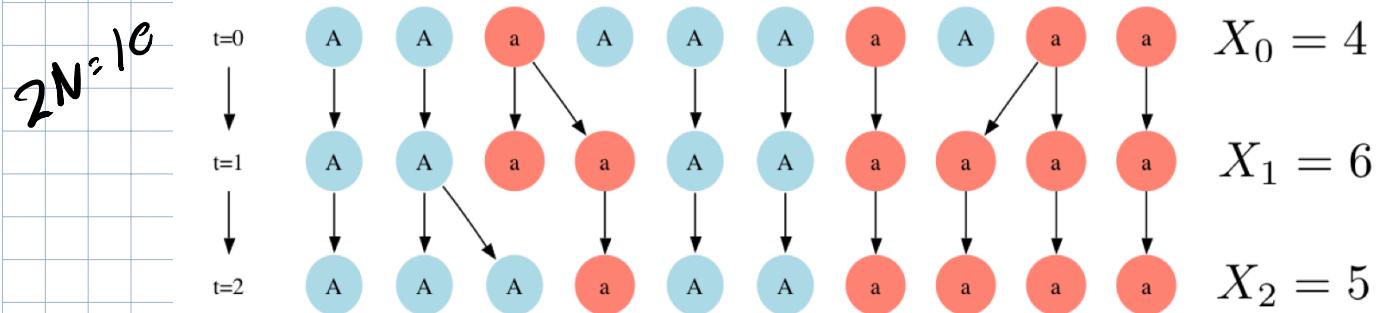
of individuals in the population

frequency of the allele in generation t



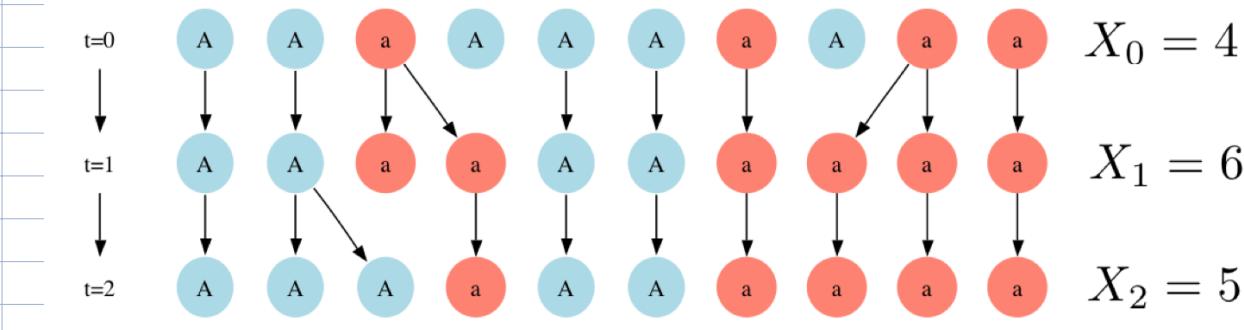
The "Wright-Fisher model"

Schematic of two generations of drift



The "Wright-Fisher model"

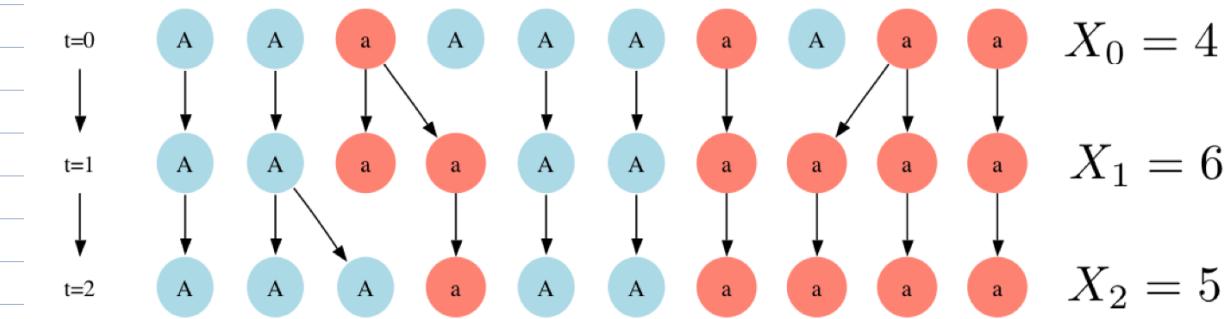
Schematic of two generations of drift



Binomial sampling as a model of genetic drift

The "Wright-Fisher model"

Schematic of two generations of drift

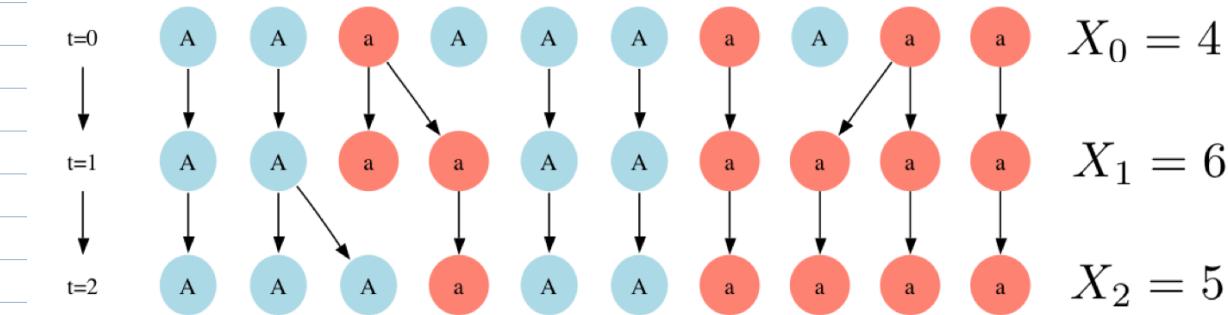


Binomial sampling as a model of genetic drift

- Assume that each copy of a gene present in generation $t+1$ is inherited at random from a gene copy present in generation t .

The "Wright-Fisher model"

Schematic of two generations of drift

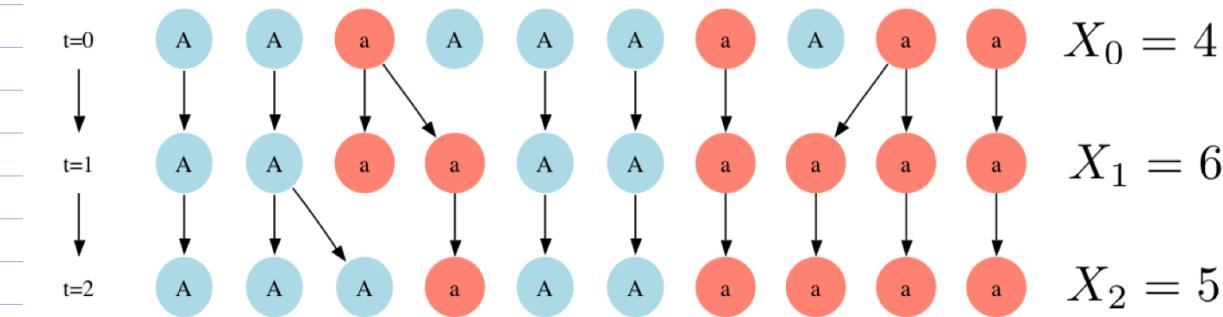


Binomial sampling as a model of genetic drift

- Assume that each copy of a gene present in generation $t+1$ is inherited at random from a gene copy present in generation t .
- if there are i copies of the a allele out of $2N$ chromosomes, then the probability that any given gene copy in generation $t+1$ will be the a type is $P = \frac{i}{2N}$

The "Wright-Fisher model"

Schematic of two generations of drift



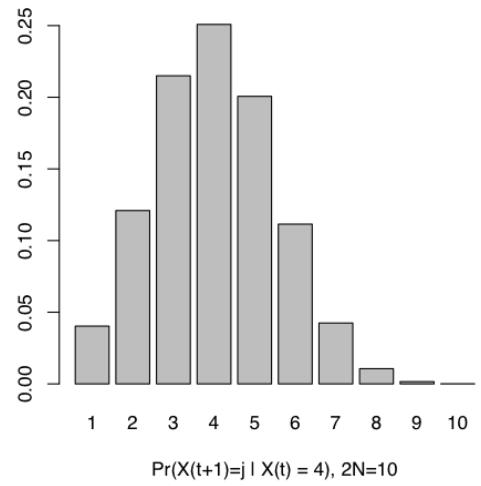
Binomial sampling as a model of genetic drift

- Assume that each copy of a gene present in generation $t+1$ is inherited at random from a gene copy present in generation t .
- if there are i copies of the A allele out of $2N$ chromosomes, then the probability that any given gene copy in generation $t+1$ will be the A type is $P = \frac{i}{2N}$
- Assuming each allele is inherited randomly in this way the probability that there will be j copies of the A allele follows a binomial distribution

$$P(X_{t+1}=j | X_t=i) = \binom{2N}{j} p^j (1-p)^{2N-j}$$

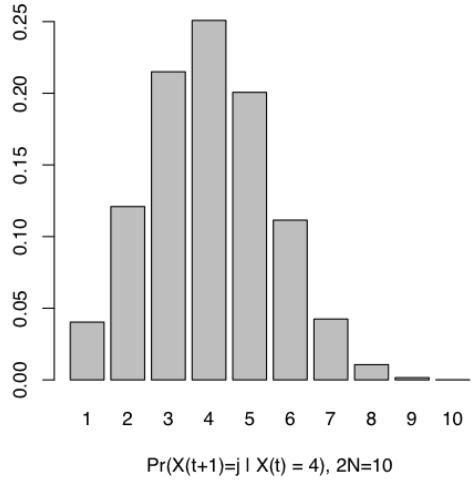
$$P(X_{t+1}=j \mid X_t=i) = \binom{2N}{j} p^j (1-p)^{2N-j}$$

$$i = 4, 2N = 10, p = 0.4$$



$$P(X_{t+1}=j \mid X_t=i) = \binom{2N}{j} p^j (1-p)^{2N-j}$$

$$i = 4, 2N = 10, p = 0.4$$

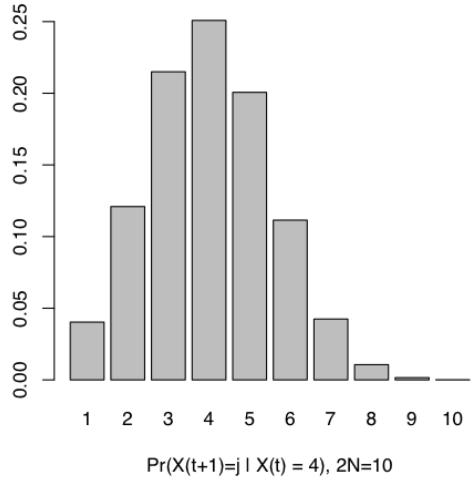


$$\mathbb{E}[P_{t+1} \mid P_t] = P_t$$

$$\text{Var}[P_{t+1} \mid P_t] = \frac{P_t(1-P_t)}{2N}$$

$$P(X_{t+1}=j \mid X_t=i) = \binom{2N}{j} p^j (1-p)^{2N-j}$$

$$i = 4, 2N = 10, p = 0.4$$



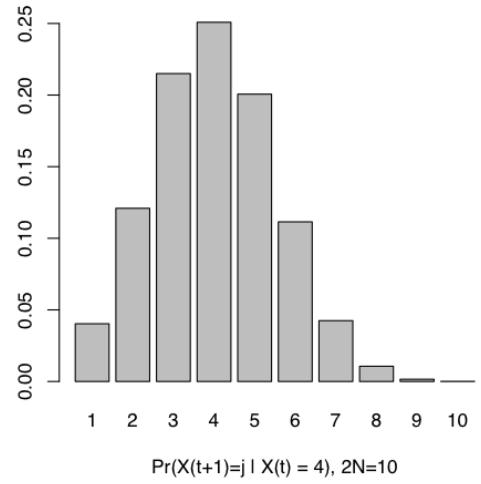
$$\mathbb{E}[P_{t+1} \mid P_t] = P_t$$

$$\text{Var}[P_{t+1} \mid P_t] = \frac{P_t(1-P_t)}{2N}$$

$$P(X_{t+1}=j \mid X_t=i) = \binom{2N}{j} p^j (1-p)^{2N-j}$$

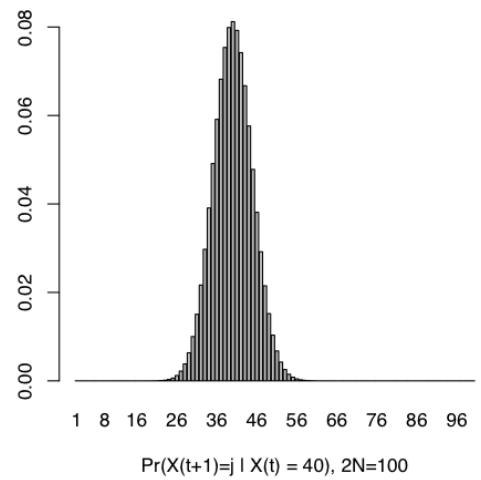
$$i = 4, 2N = 10, p = 0.4$$

$$\begin{aligned} N &= 5 \\ 2N &= 10 \end{aligned}$$

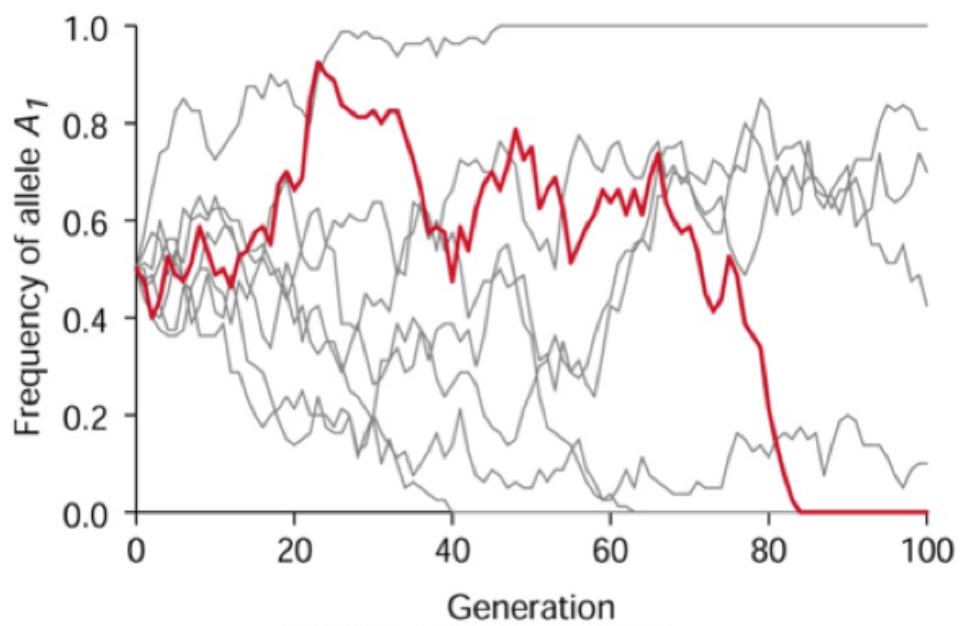


$$i = 40, 2N = 100, p = 0.4$$

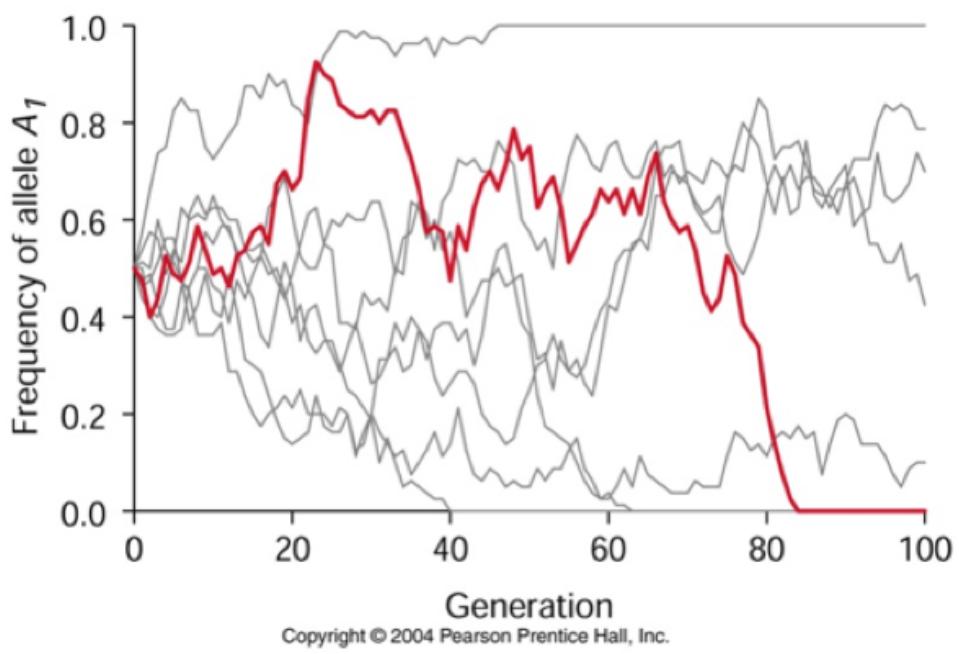
$$\begin{aligned} N &= 50 \\ 2N &= 100 \end{aligned}$$



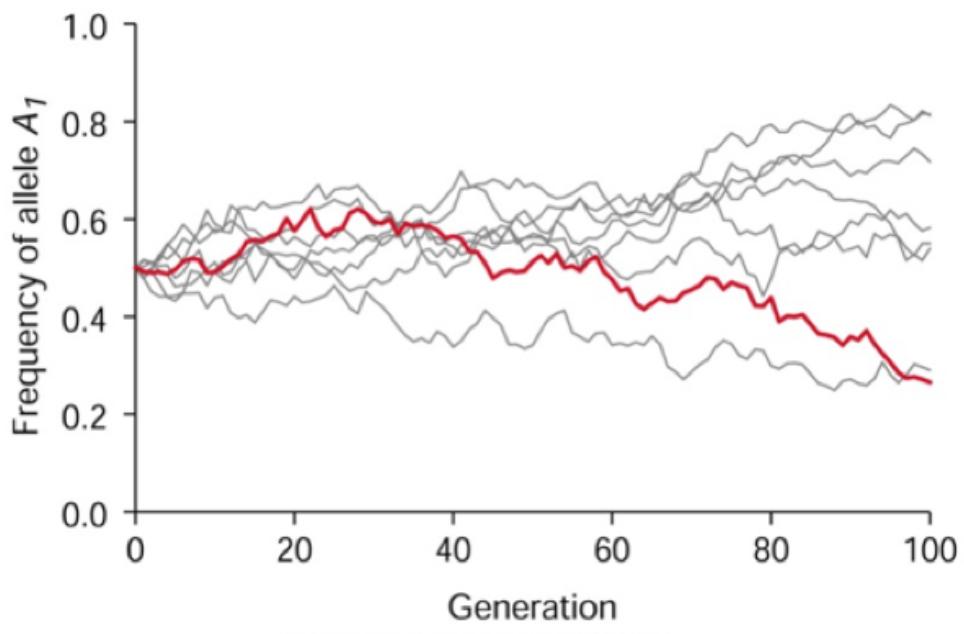
(b) Population size = 40



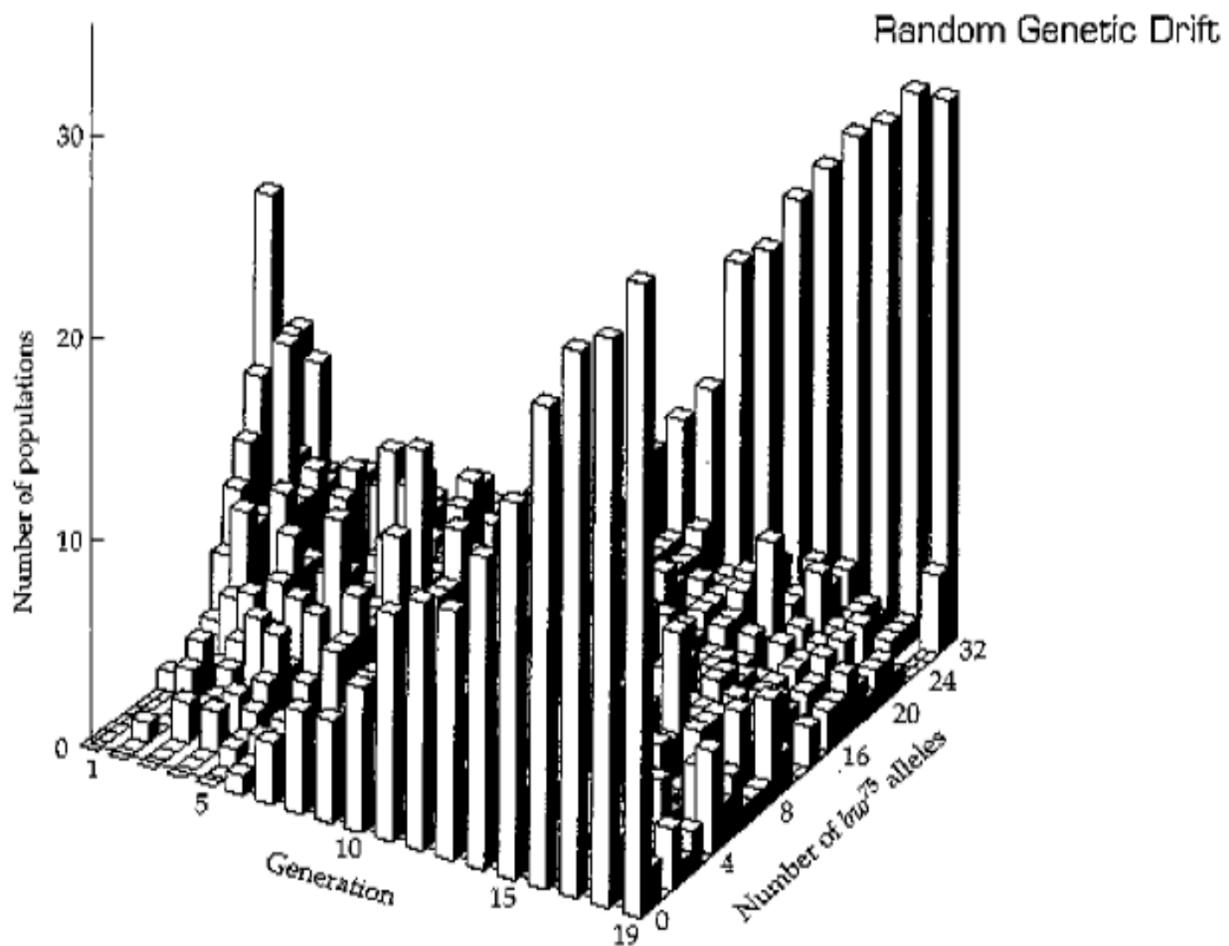
(b) Population size = 40



(c) Population size = 400

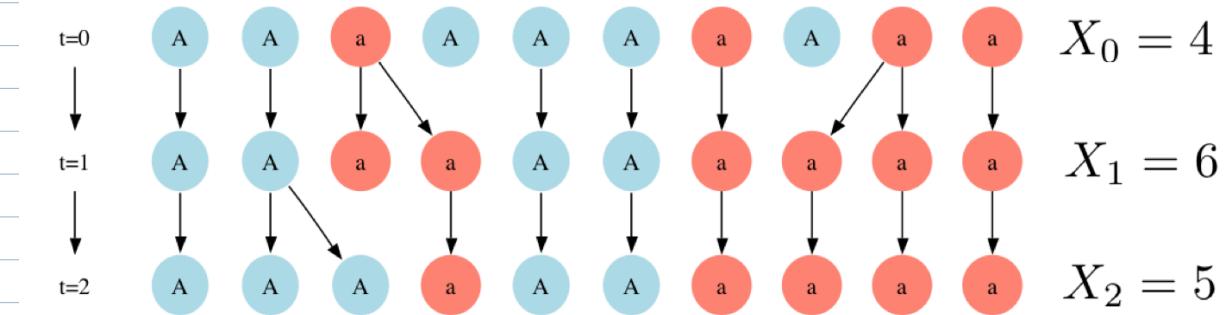


Buri 1956 genetic drift experiment
in *Prosophila melanaster*



fixation probability
of an allele at
frequency $p_0 \rightarrow p_0$

The Wright-Fisher model is a Markov Chain



The probability of being in any particular state at time $t+1$ depends only on the state of the system at time t .

$$P(X_2 | X_1, X_0) = P(X_2 | X_1)$$

The Wright-Fisher model

is mathematically intractable

The Wright-Fisher model

is mathematically tractable

$$P(X_1=j | X_0=i) = \binom{2N}{j} p_0^j (1-p_0)^{2N-j}$$

$$p_0 = \frac{i}{2N}$$

The Wright-Fisher model

is mathematically tractable

$$P(X_1=j | X_0=i) = \binom{2N}{j} p_o^j (1-p_o)^{2N-j}$$

$$p_o = \frac{i}{2N}$$

$$P(X_2=k | X_0=i) = \sum_{j=0}^{2N} P(X_2=k | X_1=j) P(X_1=j | X_0=i)$$

The Wright-Fisher model

is mathematically tractable

$$P(X_1=j | X_0=i) = \binom{2N}{j} p_0^j (1-p_0)^{2N-j}$$

$$p_0 = \frac{i}{2N}$$

$$P(X_2=k | X_0=i) = \sum_{j=0}^{2N} P(X_2=k | X_1=j) P(X_1=j | X_0=i)$$

$$P(X_T=k | X_0=i) =$$

$$\sum_{\ell=0}^{2N} P(X_T=\ell | X_{T-1}=\ell) \sum_{h=0}^{2N} P(X_{T-1}=h | X_{T-2}=h) \dots \sum_{j=0}^{2N} P(X_2=j | X_1=j) P(X_1=j | X_0=i)$$

The Wright-Fisher model

is mathematically tractable

$$P(X_1=j | X_0=i) = \binom{2N}{j} p_o^j (1-p_o)^{2N-j}$$

$$p_o = \frac{i}{2N}$$

$$P(X_2=k | X_0=i) = \sum_{j=0}^{2N} P(X_2=k | X_1=j) P(X_1=j | X_0=i)$$

$$P(X_T=k | X_0=i) =$$

$$\sum_{\ell=0}^{2N} P(X_T=\ell | X_{T-1}=\ell) \sum_{h=0}^{2N} P(X_{T-1}=h | X_{T-2}=h) \dots \sum_{j=0}^{2N} P(X_2=j | X_1=j) P(X_1=j | X_0=i)$$

The Diffusion Approximation

Kolmogorov forward equation

$$\frac{\partial \phi(p, x; t)}{\partial t} = -\frac{\partial M(p)\phi(p, x; t)}{\partial p} + \frac{1}{2} \frac{\partial^2 V(p)\phi(p, x; t)}{\partial p^2}$$

where:

- $M(p)$ is the mean change in the allele frequency per generation (i.e. 0 for models with genetic drift, $sp(1-p)$ for additive selection)
- $V(p)$ is the variance in the change in allele frequency change per generation, i.e. $\frac{p(1-p)}{2N}$
- Boundary conditions model allele loss/fixation and creation via mutation.

It works quite well!

Diffusion

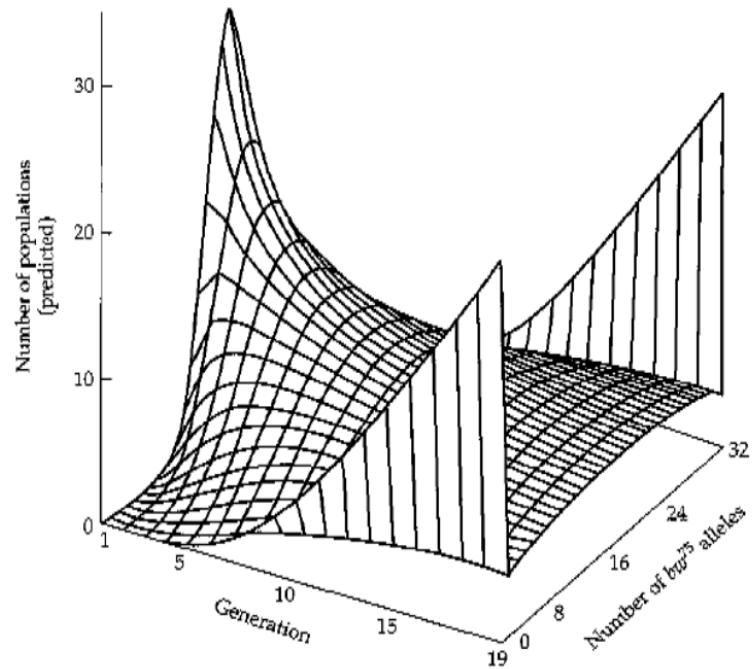


FIGURE 3.7 Kimura's (1955) solution to the diffusion equation for the particular case of $N = 16$. This is the three-dimensional view of Figure 3.6, and represents the diffusion approximation to the exact solution obtained from the Wright-Fisher model in Figure 3.5.

Buri

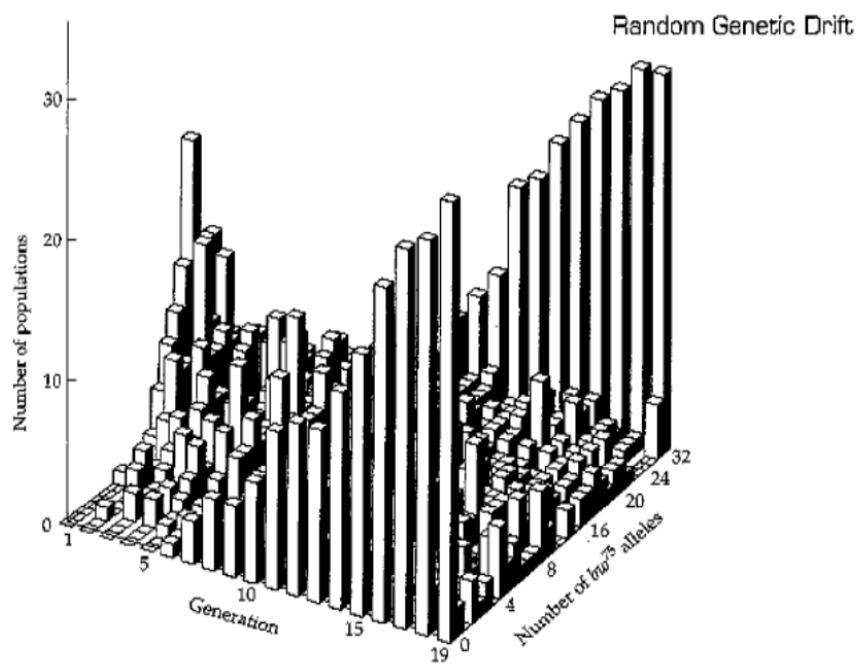


FIGURE 3.4 Random genetic drift in 107 actual populations of *Drosophila melanogaster*. Each of the initial 107 populations consisted of 16 bw^{75}/bw heterozygotes ($N = 16$; bw = brown eyes). From among the progeny in each generation, eight males and eight females were chosen at random to be the parents of the next generation. The horizontal axis of each curve gives the number of bw^{75} alleles in the population, and the vertical axis gives the corresponding number of populations. (Data from Buri 1956.)

Forward in time models

- the Wright-Fisher model is a **Markov chain** that describes how allele frequencies change from one generation to the next
- It is not the only forward in time model
 - e.g. the Moran model is a continuous time Markov chain model of allele frequency change
- The diffusion approximates both of these models quite well as long as N is large
- The diffusion equations can be used to calculate a lot about the long term behavior of the WF or Moran model
 - transition probabilities
 - can be adapted to handle selection as well
 - fixation probabilities
 - possible (but not easy) to handle changing population sizes
 - can use diffusion based methods together with data on allele frequencies at neutral sites to infer the history of population size changes