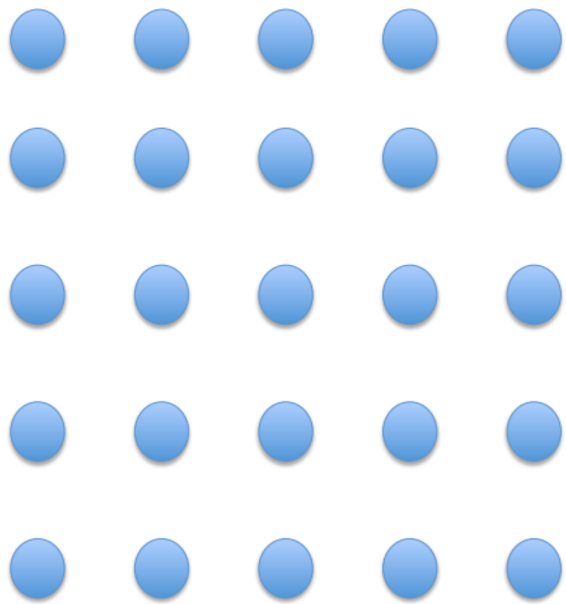
The Wright-Fisher model

Assumptions

- Finite population size N
- At each generation, each individual chooses its parent uniformly at random from the previous generation, independently of all other individuals (no selection, pure genetic drift)
- Every individual inherits the same allelic type as the parent's (no mutation)

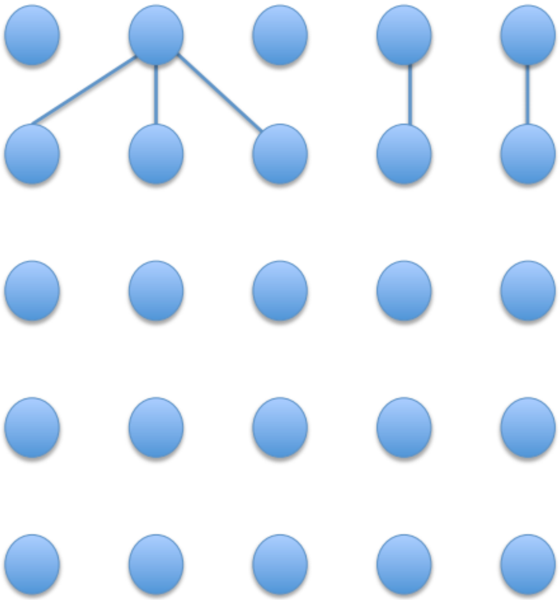
Wright-Fisher graph

0

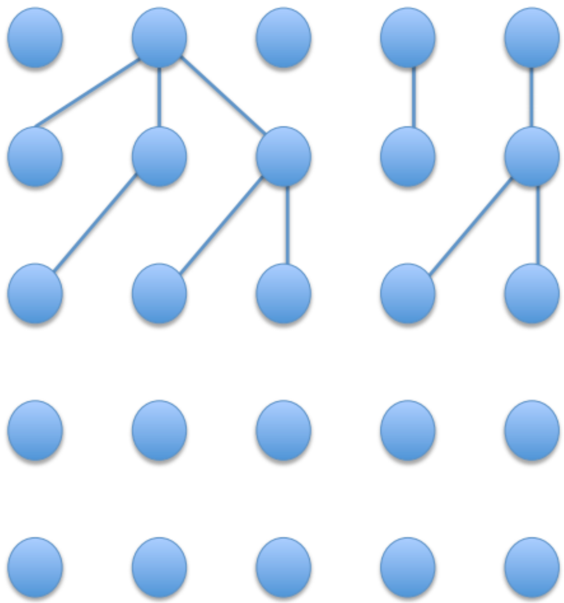


Wright-Fisher graph

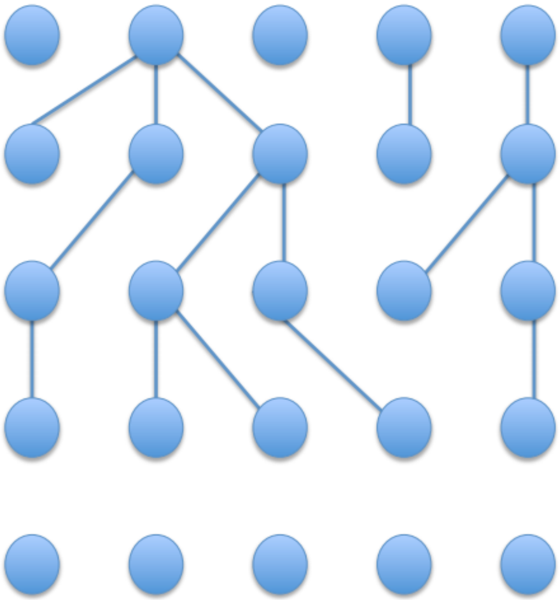
$N = 5$



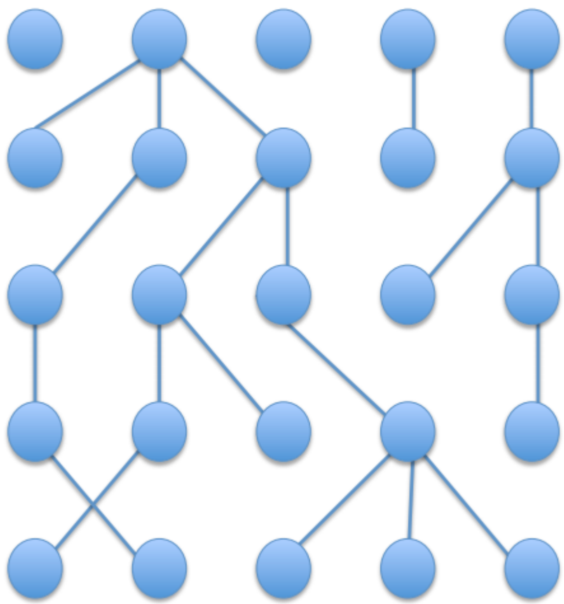
Wright-Fisher graph



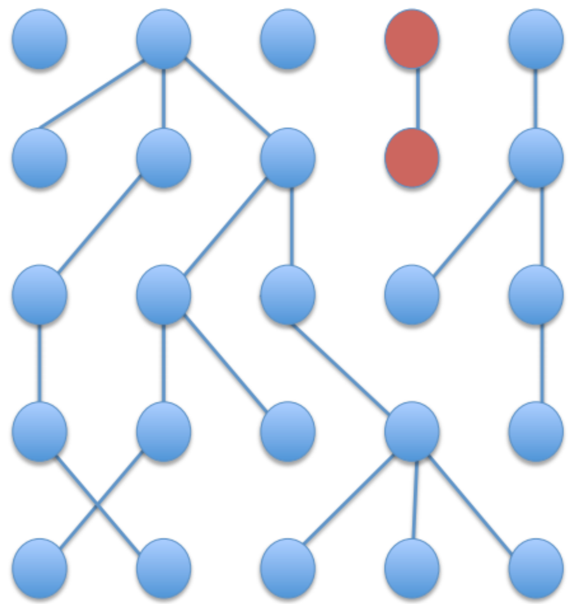
Wright-Fisher graph



Wright-Fisher graph



Wright-Fisher graph: allele frequencies



Y(0)=1/5

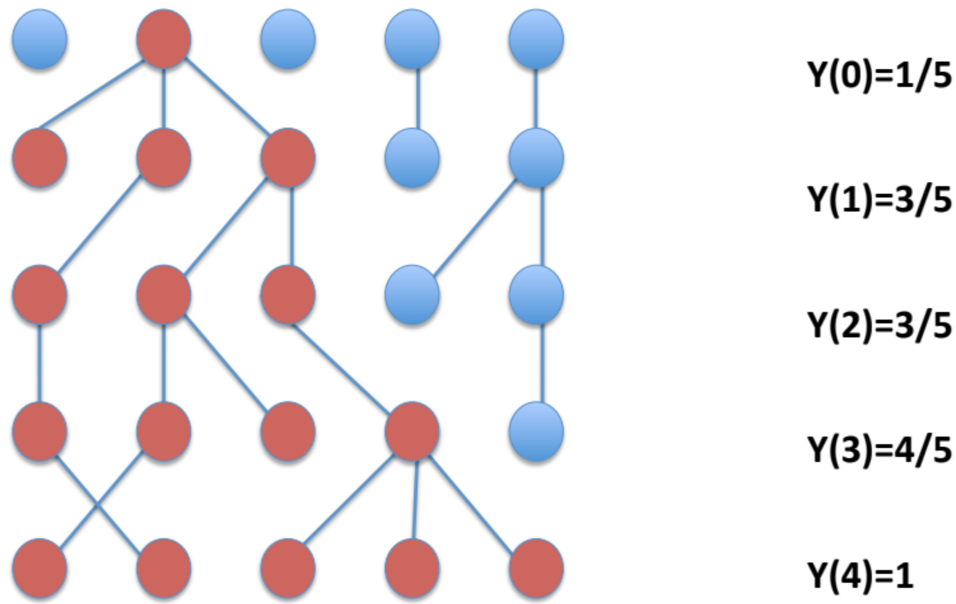
Y(1)=1/5

Y(2)=0

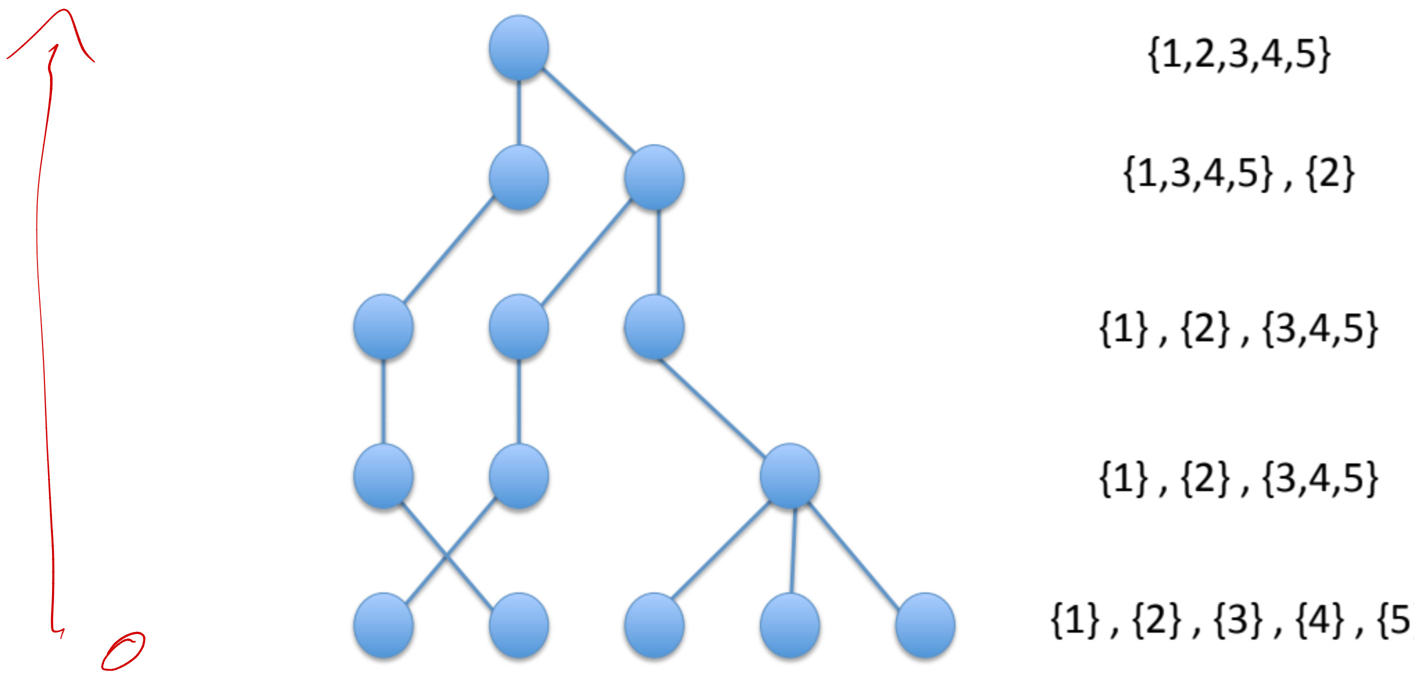
Y(3)=0

Y(4)=0

Wright-Fisher graph: allele frequencies



Wright-Fisher graph: genealogies



Questions (exercises):

Assume at generation 0 only one individual out of N=6 is of the “red” type

1. What is the probability that all the 6 individuals in generation 1 are of red type?

2. What is the probability that all in generation k are all of the red type?

3. Suppose we have observed that the history of red allele counts up to generation 5 is {1,3,2,5,2}. Given this information, what is the conditional probability that at generation 6 the number of red individuals is j, for any j=0,…,6 ?

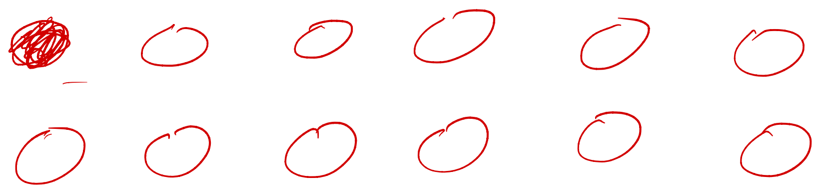
- 4. Suppose that, out of N individuals, a fraction (frequency) x is of red type at generation 0. What is the expected value of the frequency of reds at time k (k=1,2,…)?

Question 1

Assume at generation 0 only one individual out of N=6 is of the “red” type

What is the probability that all the 6 individuals in generation 1 are of red type?

$t = 0$



$$\mathbb{P}(\text{fix at time } 1) = \underbrace{\left(\frac{1}{6}\right)}_{P(\text{ind}_i \text{ chooses red parent})}^{6}$$

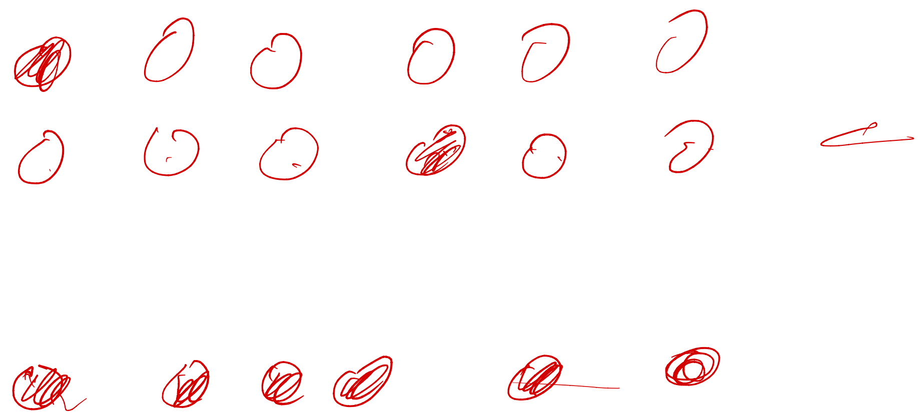
because of iid

Question 2

Assume at generation 0 only one individual out of N=6 is of the “red” type.

What is the probability that all in generation k are all of the red type?

Gen 0



Gen k

FOR K=2

$$\underbrace{\left[1 - P(\text{No red in gen. } 1)\right]}_{\mathbb{P}(\text{At least 1 red in gen. } 1)} \times \sum_{k=1}^{6} P(k \text{ reds in gen. } 1)\, P(6 \text{ reds in gen. } 2 \mid k \text{ reds in gen. } 1)$$

Question 3

Assume at generation 0 only one individual out of N=6 is of the “red” type.

Suppose we have observed that the history of red allele counts up to generation 5 is {1,3,2,5,2}. Given this information, what is the conditional probability that at generation 6 the number of red individuals is j, for any j=0,…,6 ?

Say N_k = # of red ind. in gen. k

So we have observed $\{N_0 = 1, N_1 = 1, N_2 = 3, N_3 = 2, N_4 = 5, N_5 = 2\} =: \underline{E}$

$$\mathbb{P}(N_6 = j \mid E) = ?$$

$$\binom{6}{j}\Big(\underbrace{\tfrac{1}{3}}_{P(\text{ind. } i \text{ chooses a red parent})}\Big)^{j}\Big(\tfrac{2}{3}\Big)^{6-j} \qquad j = 0, \dots, 6$$

Question 4

Suppose that, out of N individuals, a fraction (frequency) x is of red type at generation 0. What is the expected value of the frequency of reds at time k (k=1,2, …)?

Number of reds at 0 is $N \cdot x$

Start with $k=2$. Actually start first $k=1$

Call the frequency of reds at time k $X_k^{(N)} := \frac{N_k}{N}$

$k = 0, 1, 2, \ldots$

$$\mathbb{E}[X_k \mid X_0 = x] \qquad x = \{0, \tfrac{1}{N}, \ldots, \tfrac{N-1}{N}, 1\}$$

For $k=1$

Question 4

Suppose that, out of N individuals, a fraction (frequency) x is of red type at generation 0. What is the expected value of the frequency of reds at time k (k=1,2, …)?

$$\mathbb{E}[X_1 \mid X_0 = x]$$

know $\quad NX_1 = N_1 \sim \mathrm{Bin}(N, x)$

N fixed

$$\mathbb{E}[X_1 \mid X_0 = x] = \mathbb{E}\left[\frac{N \cdot X_1}{N} \mid X_0 = x\right] = \mathbb{E}\left[\frac{N_1}{N} \mid X_0 = x\right]$$

$$= \frac{1}{N}\,\mathbb{E}[N_1 \mid X_0 = x] = \frac{1}{N} \times Nx = x.$$

Bin (N, x)

Question 4

Suppose that, out of N individuals, a fraction (frequency) x is of red type at generation 0. What is the expected value of the frequency of reds at time k (k=1,2, …)?

$$\mathbb{E}[X_2 \mid X_0 = x] = \mathbb{E}\big[\underbrace{\mathbb{E}[X_2 \mid X_1, X_0]} \mid X_0 = x\big]$$

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

$$\sum_{j=0}^{N} \frac{j}{N} \underbrace{\mathbb{P}\left(X_2 = \frac{j}{N} \mid X_0 = x\right)}$$

$$\sum_{i=0}^{N} \mathbb{P}\left(X_2 = \frac{j}{N}, X_1 = \frac{i}{N} \mid X_0 = x\right)$$

TOT. LAW

$$= \sum_{i=0}^{N} \mathbb{P}\left(X_2 = \frac{j}{N} \mid X_1 = \frac{i}{N}, X_0 = x\right)$$

DEF. OF COND. DIST

$$\mathbb{P}\left(X_1 = \frac{i}{N}, X_0 = x\right)$$

$$E[X_2 \mid X_0 = x] = \sum_{j=0}^{N} \frac{j}{N} \underbrace{P\left(X_2 = \frac{j}{N} \,\middle|\, X_0 = x\right)}_{*} \qquad (\cancel{A})$$

$$* \; P\left(X_2 = \frac{j}{N} \,\middle|\, X_0 = x\right) = \sum_{i=0}^{N} P\left(X_2 = \frac{j}{N}, X_1 = \frac{i}{N} \,\middle|\, X_0 = x\right)$$

$$(*) = \sum_{i=0}^{N} \underbrace{P\left(X_2 = \frac{j}{N} \,\middle|\, X_1 = \frac{i}{N}, X_0 = x\right)} P\left(X_1 = \frac{i}{N} \,\middle|\, X_0 = x\right)$$

$$P\left(X_2 = \frac{j}{N} \,\middle|\, X_1 = \frac{i}{N}, X_0 = x\right) = P\left(X_2 = \frac{j}{N} \,\middle|\, X_1 = \frac{i}{N}\right)$$

$$= \sum_{i=0}^{N} \binom{N}{j} \left(\frac{i}{N}\right)^{j} \left(\frac{N-i}{N}\right)^{N-j} \binom{N}{i} x^{i} (1-x)^{N-i}$$

Plug this into RHS of $(\cancel{A})$

It turns out that

$$\mathbb{E}[X_2 \mid X_0 = x] = x \quad !$$

IN FACT

$$\mathbb{E}[X_2 \mid X_1, X_0 = x] = \mathbb{E}[X_2 \mid X_1] = X_1$$

$$\mathbb{E}[X_2 \mid X_0 = x] = \mathbb{E}\big[\, \mathbb{E}[X_2 \mid X_1, X_0 = x] \mid X_0 = x \,\big]$$

$$= \mathbb{E}[X_1 \mid X_0 = x] = x \quad !!$$

The Wright-Fisher model

- Suppose the gene has two alleles, *A* and *a*.
- Let X_k denote the number of type *A* alleles in the population in generation *k*, so that X_k takes values between 0 and N. Then it is easy to see that:

$$(X_k | X_{k-1} = i) \sim Binomial(N, i/N).$$

- (X_k) is a ***Markov chain*** on {0, …, N}.
- In the absence of mutation, if the Markov chain hits 0 or *N* then it stays there; these are ***absorbing states***. The *A* allele either dies out or reaches fixation.

Digression:

A primer on Markov Chains