Some mathematical models from population genetics
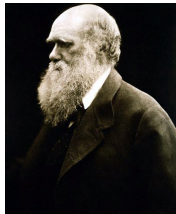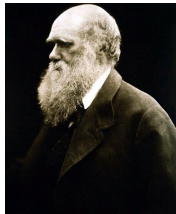# I. Classical models

Alison Etheridge
University of Oxford

RIMS, September 2023

The Modern Evolutionary Synthesis
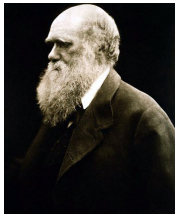
The Modern Evolutionary Synthesis

Darwin: Heritable traits that increase reproductive success will become more common in a population.

The Modern Evolutionary Synthesis

Darwin: Heritable traits that increase reproductive success will become more common in a population. Requires:

▶ Variation in population

▶ Offspring must be similar to parents

The Modern Evolutionary Synthesis

Darwin: Heritable traits that increase reproductive success will become more common in a population. Requires:

- ▶ Variation in population
- ▶ Offspring must be similar to parents

Mendel: Traits 'determined' by genes.

- ▶ Genes occur in different types (alleles)
- ▶ Offspring inherit genes from parents

The Modern Evolutionary Synthesis

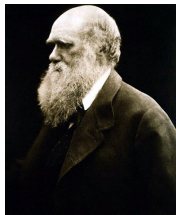Darwin: Heritable traits that increase reproductive success will become more common in a population. Requires:

▶ Variation in population
▶ Offspring must be similar to parents

Mendel: Traits 'determined' by genes.

▶ Genes occur in different types (alleles)
▶ Offspring inherit genes from parents

The Modern Evolutionary Synthesis

Darwin: Heritable traits that increase reproductive success will become more common in a population. Requires:

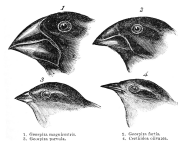- ▶ Variation in population
- ▶ Offspring must be similar to parents

Mendel: Traits 'determined' by genes.

- ▶ Genes occur in different types (alleles)
- ▶ Offspring inherit genes from parents

A fundamental question

At the beginning of the 20th Century, the modern evolutionary synthesis saw the theories of Darwin and Mendel united, but fundamental questions remained.

What is the relative importance of:

- ▶ natural selection;
- ▶ population structure (spatial and genetic);
- ▶ genetic drift (randomness due to reproduction in a finite population);
- ▶ . . .

A mathematical challenge

The pioneers could only observe genetic variation indirectly through phenotype.

www.alamy.com

Modern geneticists use differences in DNA sequences to infer 'relatedness' between sampled individuals. We require consistent

▶ forwards in time models for the evolution of the population,

▶ and backwards in time models for the relatedness between individuals in a sample.

Jonathan Marchini

Family trees

To trace the ancestry of a human, we record parents, grandparents, great-grandparents . . . .

Family trees

To trace the ancestry of a human, we record parents, grandparents, great-grandparents . . . .

After just 9 generations, Mark Wallinger's 'Y' has 512 leaves.

Some individuals must occur multiple times in a family tree.

photo courtesy of Magdalen College

The Ancestry of King Charles II of Spain (1661-1700)

Philip of Castile (1478-1506)

Joanna of Castile (1479-1555)

Isabella of Portugal (1503-39)

Charles V, Holy Roman Emperor (1500-58)

Ferdinand I, Holy Roman Emperor (1503-64)

Anna of Bohemia and Hungary (1503-47)

Isabella of Burgundy (1501-26)

Christian II of Denmark (1481-1559)

Philip II of Spain (1527-98)

Maria of Spain (1528-1603)

Maximilian II, Holy Roman Emperor (1527-76)

Charles II of Austria (1540-90)

Anne of Habsburg (1528-90)

Albert V, Duke of Bavaria (1528-79)

Christina of Denmark (1522-90)

Francis I, Duke of Lorraine (1517-45)

Anne of Austria (1549-80)

Maria Anna of Bavaria (1551-1608)

William V, Duke of Bavaria (1548-1626)

Renata of Lorraine (1544-1602)

Philip III of Spain (1578-1621)

Margarita of Austria (1584-1611)

Ferdinand II, Holy Roman Emperor (1578-1637)

Maria Anna of Bavaria (1574-1616)

Philip IV of Spain (1605-65)

Maria Anna of Spain (1606-46)

Ferdinand III, Holy Roman Emperor (1608-57)

Mariana of Austria (1634-96)

Charles II of Spain (1661-1700)

The diploid (neutral) Wright-Fisher model

Population of fixed size $N \gg 1$. (Hermaphrodite for simplicity)

- Evolves in discrete generations.
- Each offspring chooses two parents uniformly at random from previous generation.

A simple model of inheritance

The diploid (neutral) Wright-Fisher model

Population of fixed size $N \gg 1$. (Hermaphrodite for simplicity)

- ▶ Evolves in discrete generations.
- ▶ Each offspring chooses two parents uniformly at random from previous generation.

Slightly more biologically: During reproduction each individual produces an effectively infinite number of germ cells (same genotype as parent), which split into gametes (one copy of each gene), which combine to form a pool. Each offspring formed by fusing two randomly chosen gametes from the pool.

A simple model of inheritance

The diploid (neutral) Wright-Fisher model

Population of fixed size $N \gg 1$. (Hermaphrodite for simplicity)

- ▶ Evolves in discrete generations.
- ▶ Each offspring chooses two parents uniformly at random from previous generation.

Slightly more biologically: During reproduction each individual produces an effectively infinite number of germ cells (same genotype as parent), which split into gametes (one copy of each gene), which combine to form a pool. Each offspring formed by fusing two randomly chosen gametes from the pool.

In neutral model, all individuals make equal contribution to the pool of gametes.

A simple model of inheritance (diploid Wright-Fisher)
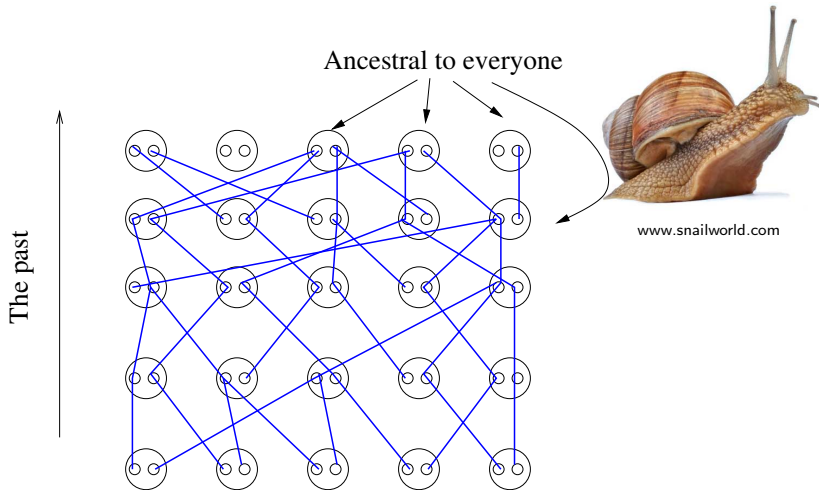
Ancestral to everyone

The past

www.snailworld.com

A simple model of inheritance (diploid Wright-Fisher)

Ancestral to everyone

The past

www.snailworld.com
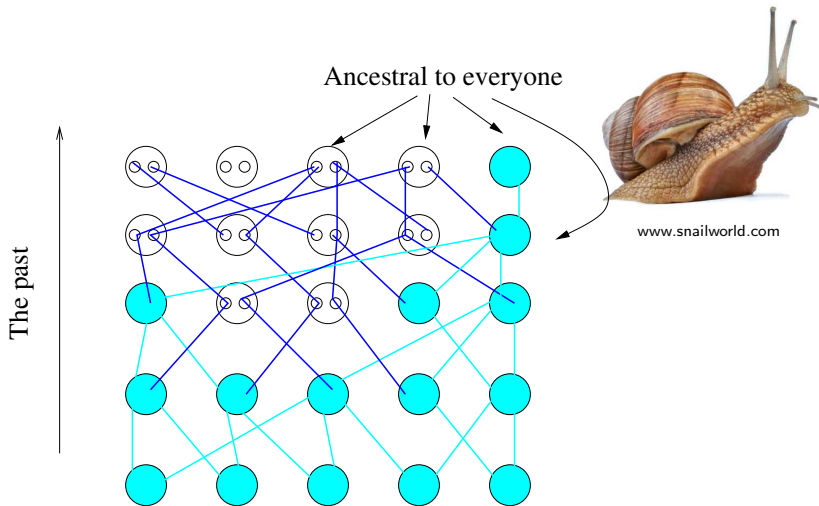
The (diploid) Wright-Fisher model revisited

1. Label parents $\{1, \ldots, N\}$. Family sizes $(\nu_1, \ldots, \nu_N)$:

$$\mathbb{P}\big[(\nu_1, \ldots, \nu_N) = (k_1, \ldots, k_N)\big] = \frac{1}{N^{2N}} \binom{2N}{k_1, \ldots, k_N} \mathbf{1}_{\sum_{i=1}^{N} k_i = 2N}$$

Distribution ν_i is $\texttt{Binom}\big(2N, 1/N\big) \approx \texttt{Poiss}(2)$

The (diploid) Wright-Fisher model revisited

1. Label parents $\{1, \ldots, N\}$. Family sizes $(\nu_1, \ldots, \nu_N)$:

$$\mathbb{P}\big[(\nu_1, \ldots, \nu_N) = (k_1, \ldots, k_N)\big] = \frac{1}{N^{2N}} \binom{2N}{k_1, \ldots, k_N} \mathbf{1}_{\sum_{i=1}^{N} k_i = 2N}$$

Distribution ν_i is $\texttt{Binom}(2N, 1/N) \approx \texttt{Poiss}(2)$

2. Each individual has $\texttt{Poiss}(2)$ offspring and condition on total number offspring $N_1 = 2N$.

$$\mathbb{P}\big[(\nu_1, \ldots, \nu_N) = (k_1, \ldots, k_N)\big] = \frac{(2N)!}{(2N)^{2N} e^{-2N}} \prod_{i=1}^{N} \frac{2^{k_i} e^{-2}}{k_i!} \mathbf{1}_{N_1 = 2N}$$

$$= \frac{1}{N^{2N}} \frac{(2N)!}{k_1! \cdots k_N!} \mathbf{1}_{\sum_{i=1}^{N} k_i = 2N}$$

The (diploid) Wright-Fisher model revisited

1. Label parents $\{1, \ldots, N\}$. Family sizes $(\nu_1, \ldots, \nu_N)$:

$$\mathbb{P}\big[(\nu_1, \ldots, \nu_N) = (k_1, \ldots, k_N)\big] = \frac{1}{N^{2N}}\binom{2N}{k_1, \ldots, k_N}\mathbf{1}_{\sum_{i=1}^N k_i = 2N}$$

Distribution ν_i is $\texttt{Binom}\big(2N, 1/N\big) \approx \texttt{Poiss}(2)$

2. Each individual has $\texttt{Poiss}(2)$ offspring and condition on total number offspring $N_1 = 2N$.

$$\mathbb{P}\big[(\nu_1, \ldots, \nu_N) = (k_1, \ldots, k_N)\big] = \frac{(2N)!}{(2N)^{2N}e^{-2N}}\prod_{i=1}^N \frac{2^{k_i}e^{-2}}{k_i!}\mathbf{1}_{N_1 = 2N}$$

$$= \frac{1}{N^{2N}}\frac{(2N)!}{k_1! \cdots k_N!}\mathbf{1}_{\sum_{i=1}^N k_i = 2N}$$

If N large, marginal prob first $n \ll N$ individuals have family sizes $k_1, \ldots, k_N$ well approximated by independent Poisson probabilities $\rightsquigarrow$ branching process approximation

Ancestry under diploid Wright-Fisher

Lemma For large N, $\mathbb{P}[$randomly chosen individual from population at time 0 is ancestral to given individual in generation $t] \approx 0.8$ as $t \to \infty$.

Ancestral individual has `Binom`$(2N, 1/N) \approx$ `Poiss`(2) direct descendants.

$P(t) =$ prob individual $\notin$ pedigree.

$P(t+1) \approx \exp(-2 + 2P(t))$ (p.g.f. Poisson evaluated at $P(t)$)

Fixed point $p \approx 0.2$, so $1 - p \approx 0.8$.

(used branching process approximation)

More precise results (Chang, 1999)

With probability $\to 1$ as $N \to \infty$, $\log_2 N$ generations before the present $\exists$ individual in pedigree ancestral to everyone in present population.

More precise results (Chang, 1999)

With probability $\to 1$ as $N \to \infty$, $\log_2 N$ generations before the present $\exists$ individual in pedigree ancestral to everyone in present population.

With probability $\to 1$ as $N \to \infty$, $\approx 1.77 \log_2 N$ generations before the present everyone in population ancestral to everyone or to no-one in present population.

With probability $\to 1$ as $N \to \infty$, $\log_2 N$ generations before the present $\exists$ individual in pedigree ancestral to everyone in present population.

With probability $\to 1$ as $N \to \infty$, $\approx 1.77 \log_2 N$ generations before the present everyone in population ancestral to everyone or to no-one in present population.

But at a particular genetic locus, an individual can have at most two genetic ancestors.
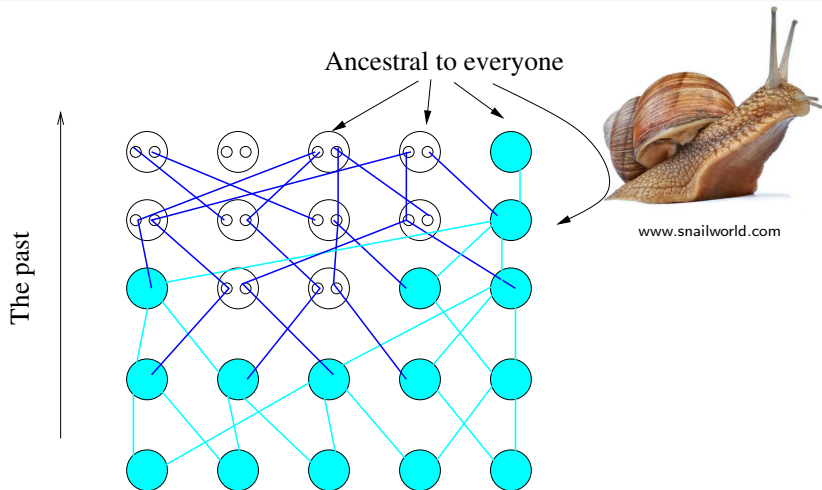
More precise results (Chang, 1999)

With probability $\to 1$ as $N \to \infty$, $\log_2 N$ generations before the present $\exists$ individual in pedigree ancestral to everyone in present population.

With probability $\to 1$ as $N \to \infty$, $\approx 1.77 \log_2 N$ generations before the present everyone in population ancestral to everyone or to no-one in present population.
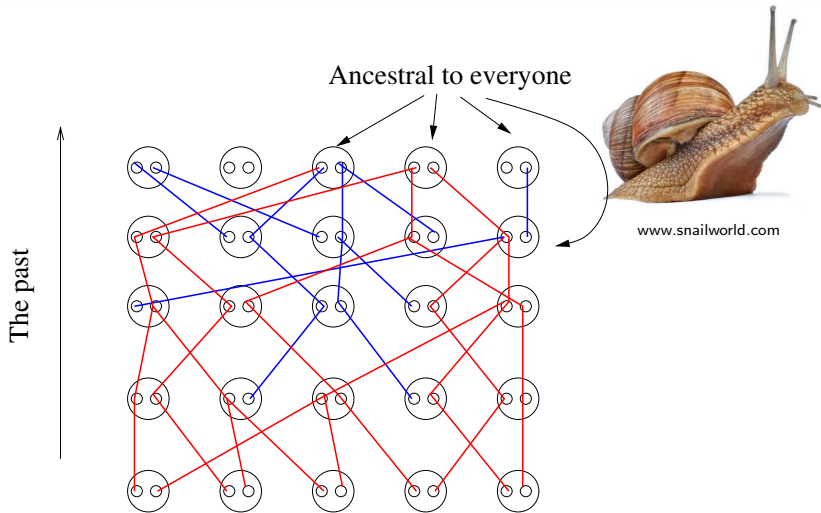
But at a particular genetic locus, an individual can have at most two genetic ancestors.

Under the diploid Wright-Fisher model, ancestry at a single genetic locus is not adequately captured by the pedigree

A simple model of inheritance

Ancestral to everyone

The past

www.snailworld.com

A simple model of inheritance

Ancestral to everyone

The past

www.snailworld.com

The simplest imaginable model of inheritance

The haploid Wright-Fisher model: each offspring (gene) chooses a *single* parent, uniformly at random   *to ease notation, ~~2~~N genes*

The simplest imaginable model of inheritance

The haploid Wright-Fisher model: each offspring (gene) chooses a *single* parent, uniformly at random   *to ease notation, ~~2~~N genes*

Kingman 1982
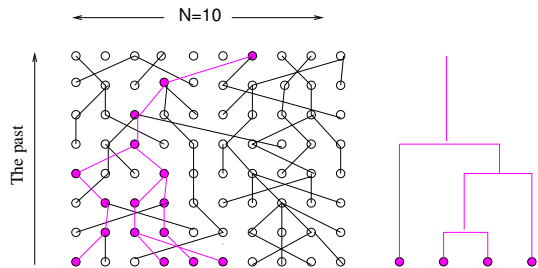
The simplest imaginable model of inheritance

The haploid Wright-Fisher model: each offspring (gene) chooses a *single* parent, uniformly at random   *to ease notation, $\not{2}N$ genes*

Kingman 1982

$\mathbb{P}[2 \text{ lineages coalesce in previous generation}] \approx \frac{1}{N}$

Time in units of N generations, $N \to \infty$, $\rightsquigarrow$ time to coalescence pair of lineages $\sim \text{Exp}(1)$
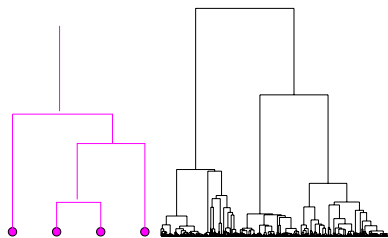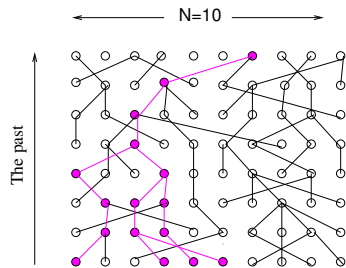
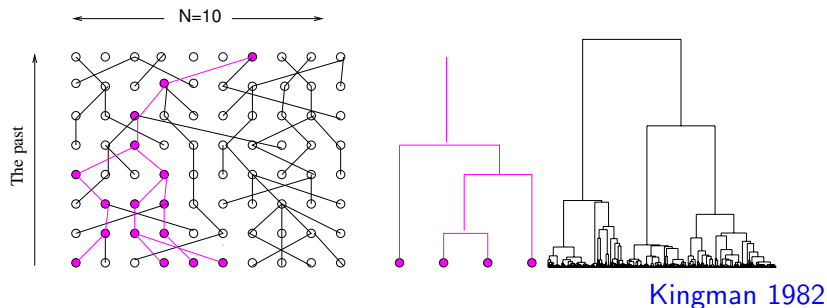# The simplest imaginable model of inheritance

The haploid Wright-Fisher model: each offspring (gene) chooses a *single* parent, uniformly at random   *to ease notation, $\cancel{2}N$ genes*
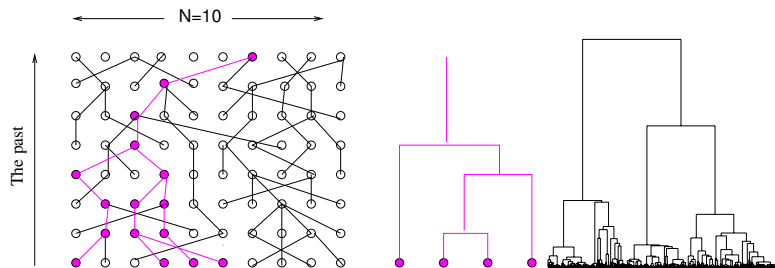
Kingman 1982

$\mathbb{P}[2 \text{ lineages coalesce in previous generation}] \approx \frac{1}{N}$

The most recent common ancestor in the pedigree was $\approx \log_2 N$ generations in the past. The most recent common *genetic* ancestor was $\approx 2N$ generations ago.

Sample size k

If currently j ancestral lineages:

- Probability two pairs of lineages merge into separate parents

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}$$

- Probability three-merger

$$\approx \binom{j}{3}\frac{1}{N^2}$$

Probability one of these events before single pairwise merger

$$\approx \frac{N}{\binom{j}{2}}\binom{j}{2}\binom{j-2}{2}\frac{1}{N^2} \approx \frac{j^2}{2N}$$

Sample size k

If currently j ancestral lineages:

- Probability two pairs of lineages merge into separate parents

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}$$

- Probability three-merger

$$\approx \binom{j}{3}\frac{1}{N^2}$$

Probability such an event somewhere in the genealogical tree

$$\approx \sum_{j=1}^{k} \frac{j^2}{N} \approx \frac{k^3}{3N}$$

Sample size k

If currently j ancestral lineages:

- Probability two pairs of lineages merge into separate parents

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}$$

- Probability three-merger

$$\approx \binom{j}{3}\frac{1}{N^2}$$

Probability such an event somewhere in the genealogical tree

$$\approx \sum_{j=1}^{k}\frac{j^2}{N} \approx \frac{k^3}{3N}$$

Sample size $k \ll N^{1/3}$, pair of lineages coalesces rate $\approx \frac{1}{N}\binom{k}{2}$

Melfi & Viswanath (2018)

Sample size k

If currently j ancestral lineages:

- Probability two pairs of lineages merge into separate parents

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}$$

- Probability three-merger

$$\approx \binom{j}{3}\frac{1}{N^2}$$

Probability such an event somewhere in the genealogical tree

$$\approx \sum_{j=1}^{k}\frac{j^2}{N} \approx \frac{k^3}{3N}$$

Sample size $k \ll N^{1/3}$, pair of lineages coalesces rate $\approx \frac{1}{N}\binom{k}{2}$

Melfi & Viswanath (2018)   but, Wakeley & Takahashi (2003)

Some formal definitions

- Ξ_k = eqivalence relations on $[k] = \{1, \ldots, k\}$
- A k-coalescent is a continuous time Markov chain on ξ_k with transition rates $q_{\xi,\eta} = 1$ if η obtained by coalescing two equiv classes of ξ, 0 otherwise
- The Kingman coalescent on equiv classes of $\mathbb{N}$ is such that $\forall k$, restriction to $[k]$ is a k-coalescent.

Label individuals in sample $\{1, \ldots, k\}$

$$
\begin{aligned}
\text{block coalescent} &\longleftrightarrow \text{ancestral lineage} \\
\text{elements of block} &\longleftrightarrow \text{descendants of that ancestor}
\end{aligned}
$$

Sampling consistency: the restriction of the $(k + l)$-coalescent to $[k]$ is the k-coalescent

The Kingman coalescent

Tracing back in time, if there are currently k ancestral lineages, the next event will occur after an exponentially distributed time with parameter $\binom{k}{2}$, when a pair of lineages (chosen at random) will coalesce.

For a vast array of models in which
- ▶ population size large and constant;
- ▶ all individuals are equally fit;
- ▶ there is no spatial structure;

measuring time in units of N_e generations, the genealogy of a sample is well approximated by the Kingman coalescent.

Overlapping generations: the Moran Model
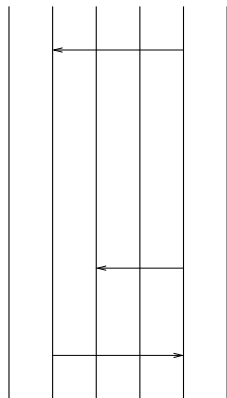
Time for Moran model

- ▶ Events determined by Poisson Process intensity $\binom{N}{2}dt$;
- ▶ Pair chosen at random;
- ▶ One reproduces, the other dies.

Genealogies under the Moran model
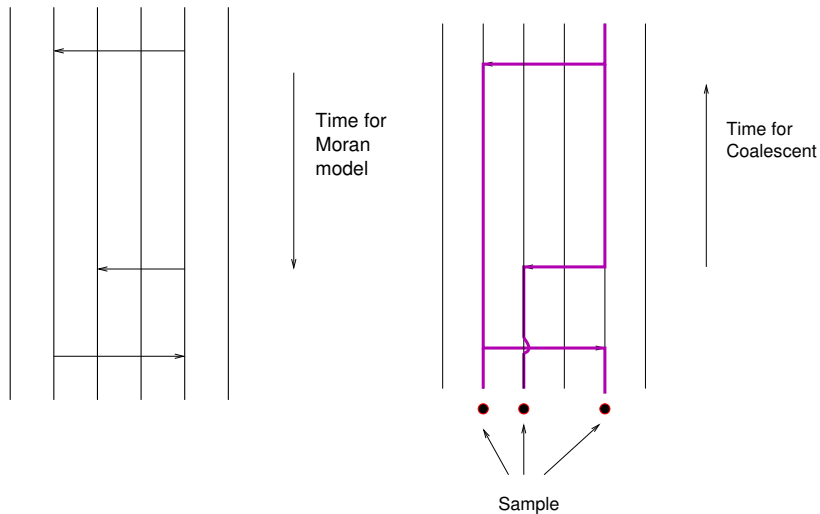
Time for Moran model

Time for Coalescent

Sample

Forwards in time: scaling the Wright-Fisher model

Population of fixed size N evolves in discrete generations.

▶ Each individual chooses parent uniformly at random from the previous generation;

▶ Offspring inherit the type of their parent.

Forwards in time: scaling the Wright-Fisher model

Population of fixed size N evolves in discrete generations.

- ▶ Each individual chooses parent uniformly at random from the previous generation;
- ▶ Offspring inherit the type of their parent.

'Alleles' a, A.   Proportion p of a alleles among parents.

$\Delta p =$ increment

- ▶ Number of a-offspring $\texttt{Bin}(N, p)$.
- ▶ $\mathbb{E}[\Delta p] = 0$ (neutral); $\mathbb{E}[(\Delta p)^2] = \frac{1}{N} p(1 - p)$.

Forwards in time: scaling the Wright-Fisher model

Population of fixed size N evolves in discrete generations.

- ▶ Each individual chooses parent uniformly at random from the previous generation;
- ▶ Offspring inherit the type of their parent.

'Alleles' a, A.   Proportion p of a alleles among parents.

$\Delta p =$ increment

- ▶ Number of a-offspring $\texttt{Bin}(N, p)$.
- ▶ $\mathbb{E}[\Delta p] = 0$ (neutral); $\mathbb{E}[(\Delta p)^2] = \frac{1}{N} p(1 - p)$.

$\rightsquigarrow$ changes in p over timescales $\mathcal{O}(N)$ generations.

Drift (large population limit)

Time in units of N generations, $\quad h = \frac{1}{N}, \quad\quad N \to \infty$

Drift (large population limit)

Time in units of N generations, $h = \frac{1}{N}$, $N \to \infty$

Forwards in time,
$\Delta p = p_{t+h} - p_t$,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = hp(1 - p)$
- $\mathbb{E}[(\Delta p)^4] = O(h^2)$

$dp_t = \sqrt{p_t(1 - p_t)}dW_t$

Drift (large population limit)

Time in units of N generations, $h = \frac{1}{N}$, $N \to \infty$

Forwards in time,
$\Delta p = p_{t+h} - p_t$,

Backwards in time

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = hp(1-p)$
- $\mathbb{E}[(\Delta p)^4] = O(h^2)$

$dp_t = \sqrt{p_t(1-p_t)}dW_t$

Coalescence rate $\binom{k}{2}$.

Drift (large population limit)

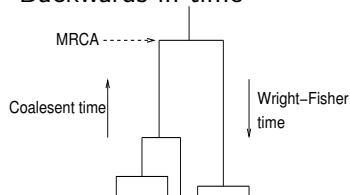Time in units of N generations, $\quad h = \frac{1}{N}$, $\qquad N \to \infty$

Forwards in time,
$\Delta p = p_{t+h} - p_t$,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = hp(1-p)$
- $\mathbb{E}[(\Delta p)^4] = O(h^2)$

$dp_t = \sqrt{p_t(1-p_t)}dW_t$

$dp_\tau = \sqrt{\frac{1}{N_e}p_\tau(1-p_\tau)}dW_\tau$

Backwards in time

MRCA

Coalesent time

Wright–Fisher time

Coalescence rate $\binom{k}{2}$.

Coalescence rate $\dfrac{1}{N_e}\dbinom{k}{2}$

Drift (large population limit)

Time in units of N generations, $\quad h = \frac{1}{N}$, $\qquad N \to \infty$

Forwards in time,
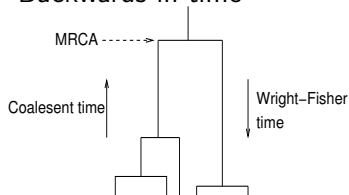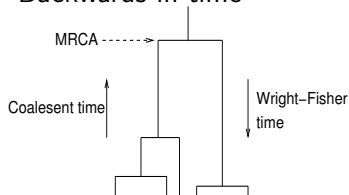$\Delta p = p_{t+h} - p_t$,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = hp(1-p)$
- $\mathbb{E}[(\Delta p)^4] = O(h^2)$

$dp_t = \sqrt{p_t(1-p_t)}dW_t$

$dp_\tau = \sqrt{\dfrac{1}{N_e} p_\tau(1-p_\tau)}dW_\tau$

Backwards in time

MRCA ·······▶

Coalesent time

Wright–Fisher time

Coalescence rate $\binom{k}{2}$.

Coalescence rate $\dfrac{1}{N_e}\binom{k}{2}$

Kingman coalescent = genealogy *random* sample

Buri's experiment

Gene that affects eye colour (but not fitness). Two alleles, a, A.

~ 100 populations, 8 males, 8 females. Each started with proportion $p = 0.5$ type a.

Buri's experiment

Gene that affects eye colour (but not fitness). Two alleles, a, A.

~ 100 populations, 8 males, 8 females. Each started with proportion $p = 0.5$ type a.

$$dp_t = \sqrt{\tfrac{1}{N_e} p_t(1-p_t)} \, dW_t,$$

$$\frac{d}{dt} \mathbb{E}[p_t(1-p_t)] = -\frac{1}{N_e} \mathbb{E}[p_t(1-p_t)]$$

Variance across populations

$$V_t = p_0(1-p_0)\left(1 - e^{-t/N_e}\right)$$

Buri's experiment

Gene that affects eye colour (but not fitness). Two alleles, a, A.

~ 100 populations, 8 males, 8 females. Each started with proportion $p = 0.5$ type a.

Eventually, each population will be entirely one type (with equal probabilities).

Under Wright-Fisher model, variance in p across populations increases from 0 to $1/4$ over time.
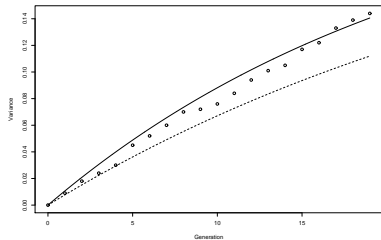
Buri's experiment

Gene that affects eye colour (but not fitness). Two alleles, a, A.

~ 100 populations, 8 males, 8 females. Each started with proportion $p = 0.5$ type a.

Eventually, each population will be entirely one type (with equal probabilities).

Under Wright-Fisher model, variance in p across populations increases from 0 to $1/4$ over time.

# Adding mutation (heritable change of genetic material)

Reconstruct tree using *differences* in DNA sequences in sample.

Extend Wright-Fisher model: constant probability, μ, per individual per generation of mutation.

Time until first mutation along single ancestral lineage $\sim \mathtt{Geom}(\mu)$

Assume $\theta := 2N_e\mu = \mathcal{O}(1)$, (coalescence/mutation on same timescale)
$\rightsquigarrow$ Poisson process along branches.
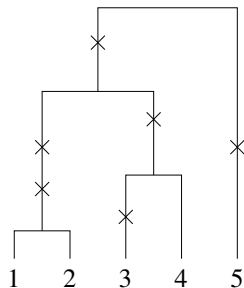
Note individuals 1 and 2 must have same type

Patterns in data reflect evolution over timescales dictated by the neutral mutation rate

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \le u \le t. \qquad (*)$$

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u),\underline{n}(t-u)\right)\right] = 0, \quad 0 \leq u \leq t. \qquad (*)$$

$$dp_t = \sqrt{p_t(1-p_t)}dW_t, \quad n_t \mapsto n_t - 1 \text{ rate } \binom{n_t}{2}, \quad f(p,n) = p^n$$

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u),\underline{n}(t-u)\right)\right]=0, \quad 0\leq u\leq t. \qquad (*)$$

$$dp_t=\sqrt{p_t(1-p_t)}dW_t, \quad n_t\mapsto n_t-1 \text{ rate } \binom{n_t}{2}, \quad f(p,n)=p^n$$

$$dp_u^{n(t-u)}=n(t-u)p_u^{n(t-u)-1}\sqrt{p_u(1-p_u)}dW_u$$
$$+\binom{n(t-u)}{2}p_u^{n(t-u)-2}p_u(1-p_u)du$$
$$-\binom{n(t-u)}{2}\left(p_u^{n(t-u)-1}-p_u^{n(t-u)}\right)$$

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u),\underline{n}(t-u)\right)\right] = 0, \quad 0 \leq u \leq t. \qquad (*)$$

$dp_t = \sqrt{p_t(1-p_t)}dW_t, \quad n_t \mapsto n_t - 1 \text{ rate } \binom{n_t}{2}, \quad f(p,n) = p^n$

Sampling probabilities:

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}]$$

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \le u \le t. \qquad (*)$$

$$dp_t = \sqrt{p_t(1-p_t)}dW_t, \quad n_t \mapsto n_t - 1 \text{ rate } \binom{n_t}{2}, \quad f(p,n) = p^n$$

Sampling probabilities:

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}]$$

Weaker than saying genealogy given by Kingman coalescent

What we have so far

In time units of N_e generations,

- ▶ (Forwards time) The Wright-Fisher diffusion

$$dp_t = \sqrt{p_t(1 - p_t)}dW_t;$$

- ▶ (Backwards time) The Kingman coalescent

$$n_t \mapsto n_t - 1 \text{ at rate } \binom{n_t}{2};$$

- ▶ Sampling probabilities

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}].$$

Stronger result holds. Kingman coalescent really describes genealogy of random sample from (neutral) population.