Endowed with an Extra Sense : Mathematics and Evolution

Todd Parsons

Laboratoire de Probabilités et Modèles Aléatoires - Université Pierre et Marie Curie
Center for Interdisciplinary Research in Biology - Collège de France

Talk to the Archimedeans
Cambridge, October 16th 2014

"I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics ; for men thus endowed seem to have an extra sense."
– Charles Darwin

1 The Modern Synthesis

2 The Molecular Revolution

3 The Present & Beyond

1 The Modern Synthesis

2 The Molecular Revolution

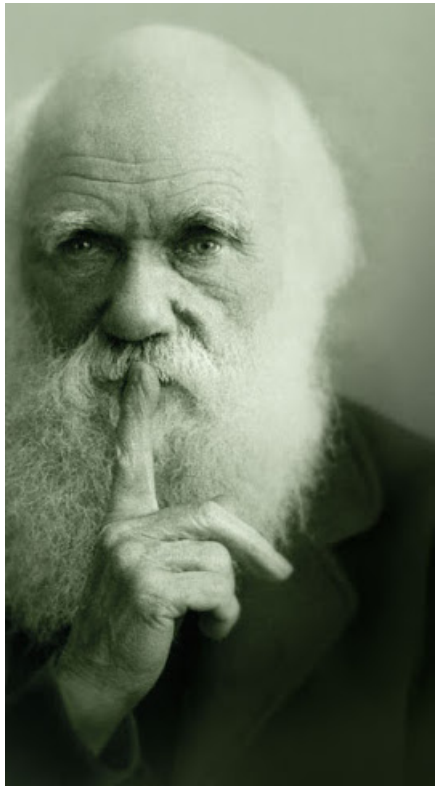
3 The Present & Beyond

"Darwin's" Theory of Evolution

- Populations contain genetic variation that arises by random mutation and recombination.
- Populations evolve by changes in gene frequency brought about by random genetic drift, gene flow, and especially natural selection.
- Most adaptive genetic variants have individually slight phenotypic effects so that phenotypic changes are gradual.
- Diversification comes about by speciation, which normally entails the gradual evolution of reproductive isolation among populations.
- These processes, continued for sufficiently long, give rise to changes of such great magnitude as to warrant the designation of higher taxonomic levels.

(from Futuyma, D.J. 2005. Evolution. Sinauer Associates, Sunderland, Massachusetts.)

Darwin's dirty secret. . .

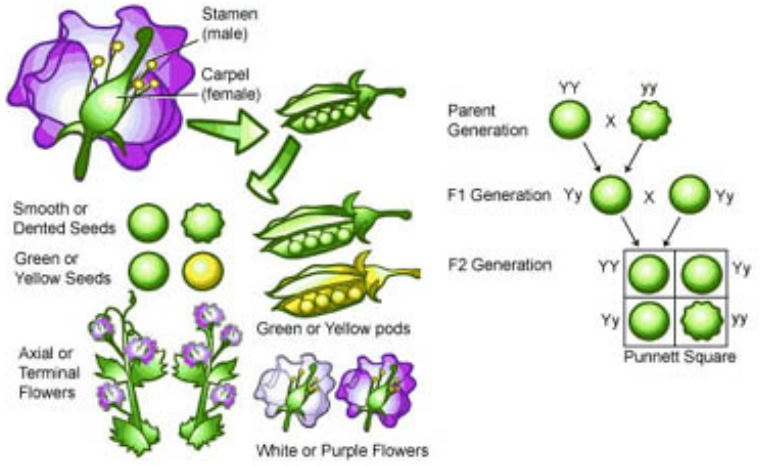


Charles Darwin
(1809–1882)

In fact, Darwin lacked a theory of heredity. He considered – but ultimately rejected – two ideas :

- Lamarckism : an organism can pass on characteristics *that it acquired during its lifetime* to its offspring.
- Blended inheritance : inherited traits were determined randomly, from a range bound by the homologous traits found in the parents.

"Mendelism killed Darwinism"

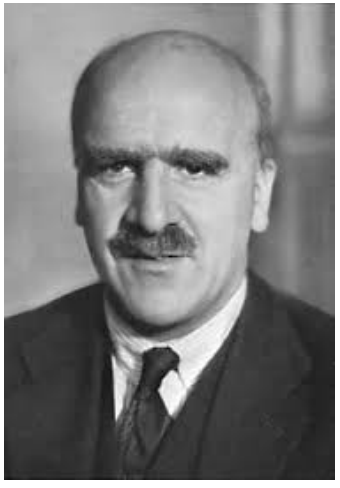


As hard as it is to imagine today, at the beginning of the 20th century, the rediscovery of Gregor Mendel's work on inheritance was seen as the final nail in the coffin for the theory of evolution :

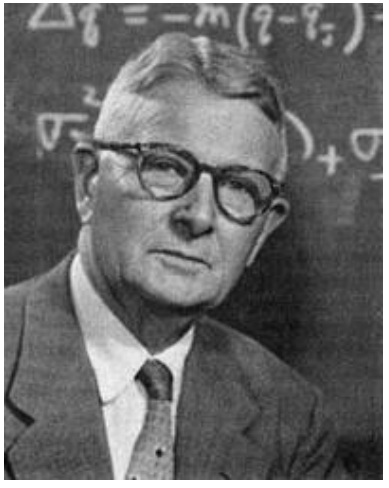
- blended inheritance would lead to uniform populations, and, moreover,
- Mendel's work suggested "mutationism" : inheritance is discrete, and that hereditary variants originate in rare mutations occurring in a single individual,
- and these would almost certainly perish, unless they conveyed substantial benefits – evolution would be a wait for "hopeful monsters"

Enter Mathematics. . .

R. A. Fisher
(1890–1962)



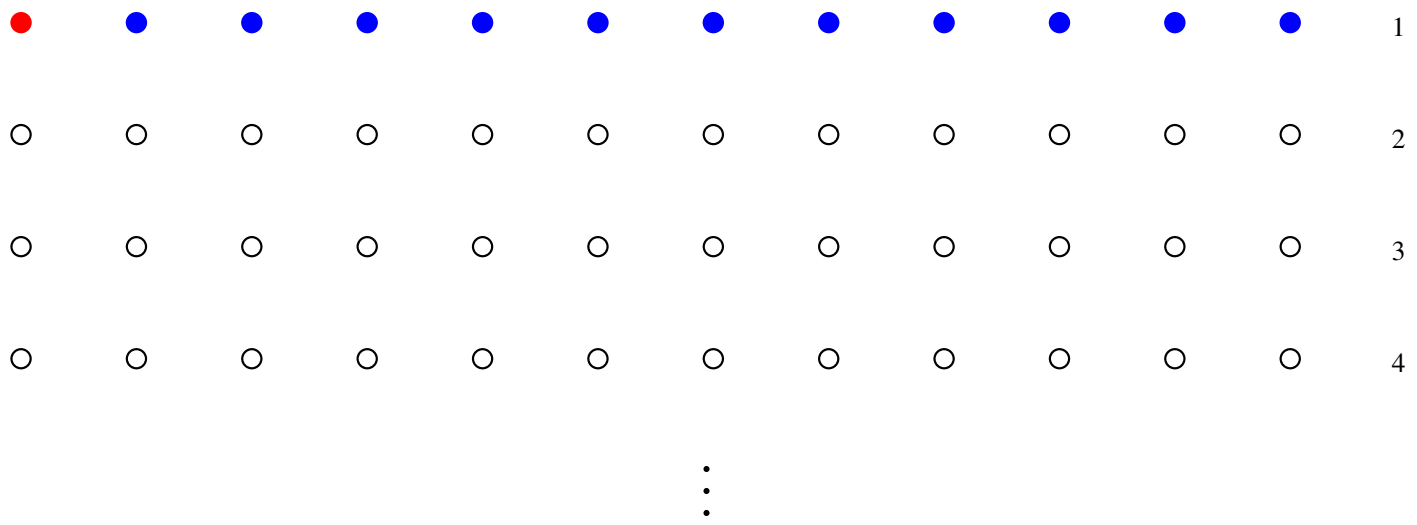
J. B. S. Haldane
(1892–1964)



Sewall Wright
(1889–1988)

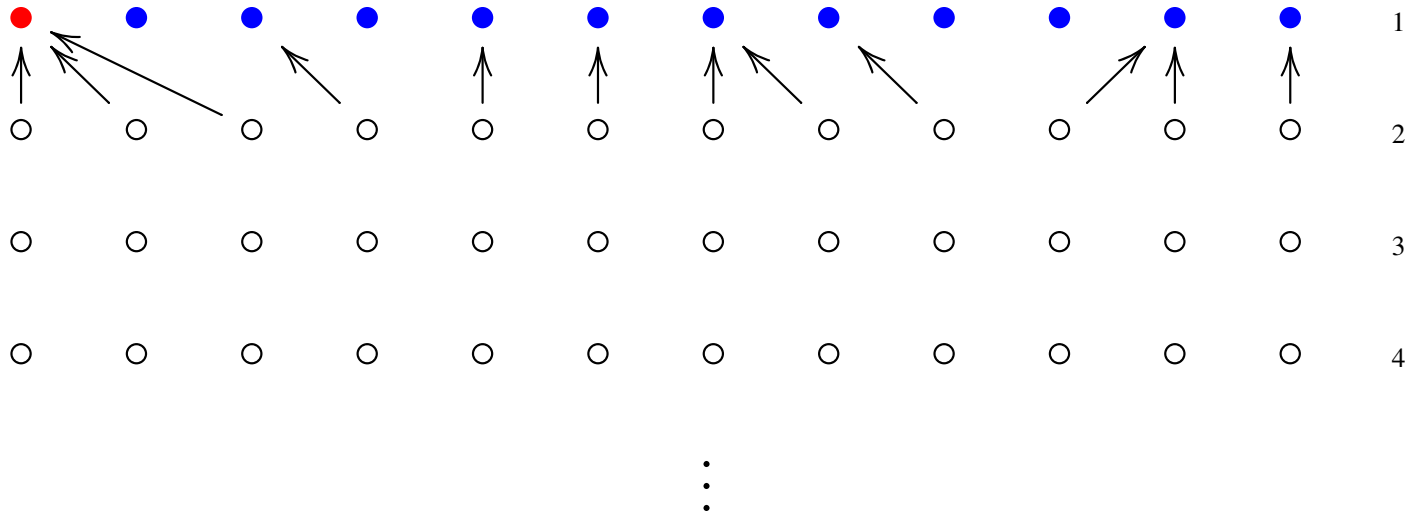
Darwin's theory of evolution nearly disappeared, until early in the 20th century, three young scholars realised that mathematics could be used to reconcile Darwin's theory with Mendel's discoveries.

The Wright-Fisher Model



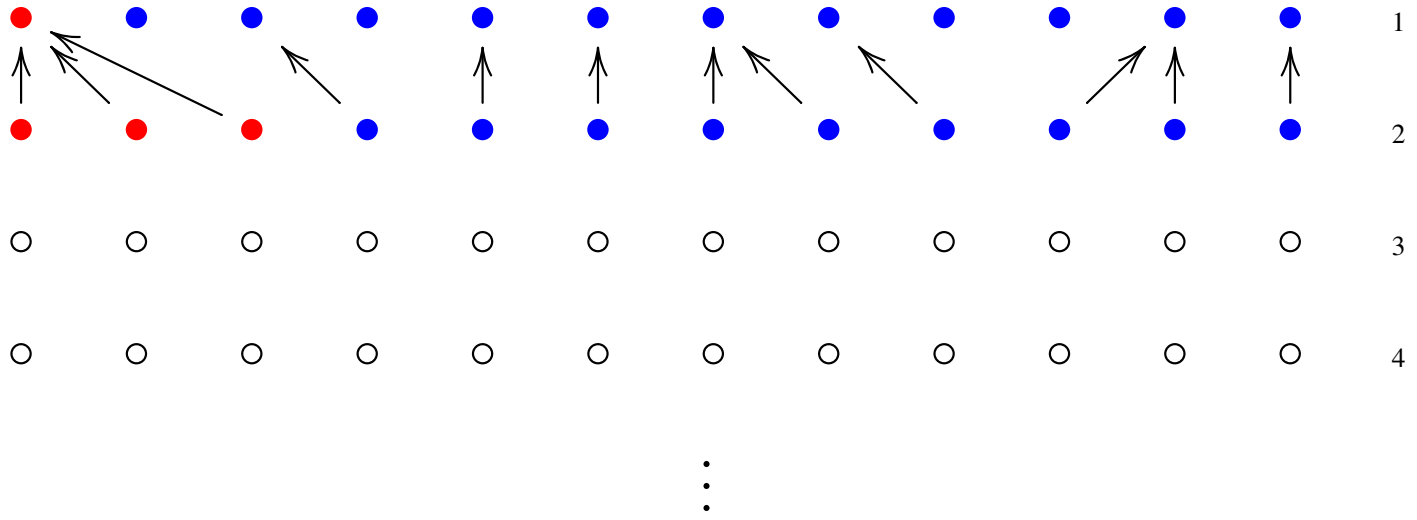
- Assume a haploid population with exactly N individuals with two possible types : a mutant and the ancestral “wild-type”
- Discrete time (generations).
- Each generation, each individual “looks back” to the previous and “picks” a parent.
- Offspring have the same type as their parent.

The Wright-Fisher Model



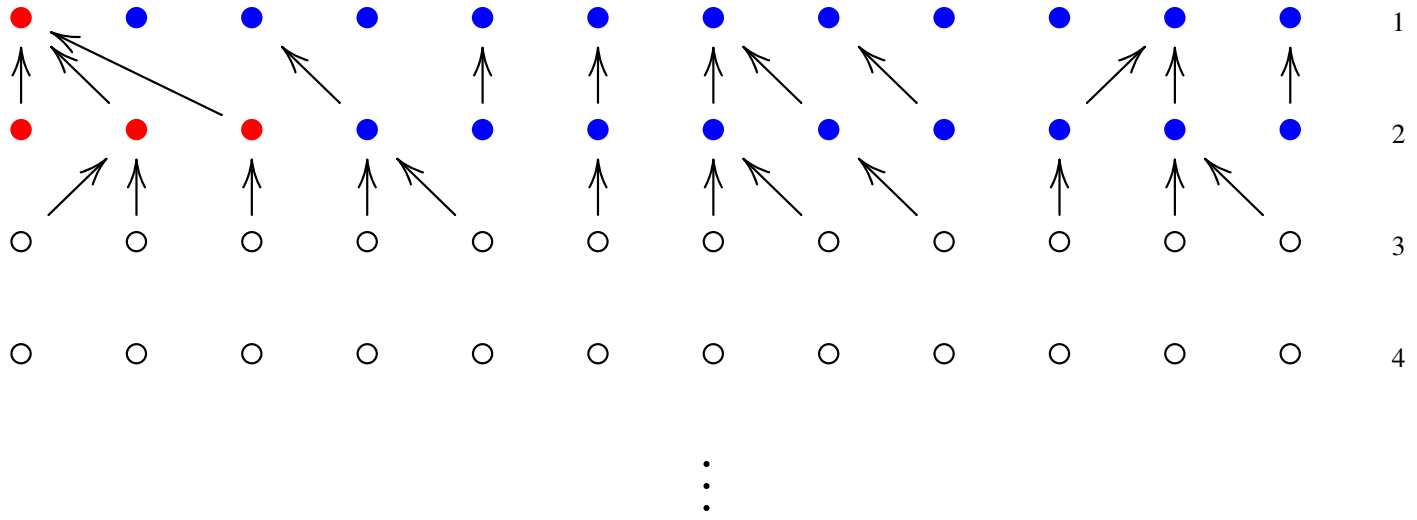
- Assume a haploid population with exactly N individuals with two possible types : a mutant and the ancestral "wild-type"
- Discrete time (generations).
- Each generation, each individual "looks back" to the previous and "picks" a parent.
- Offspring have the same type as their parent.

The Wright-Fisher Model



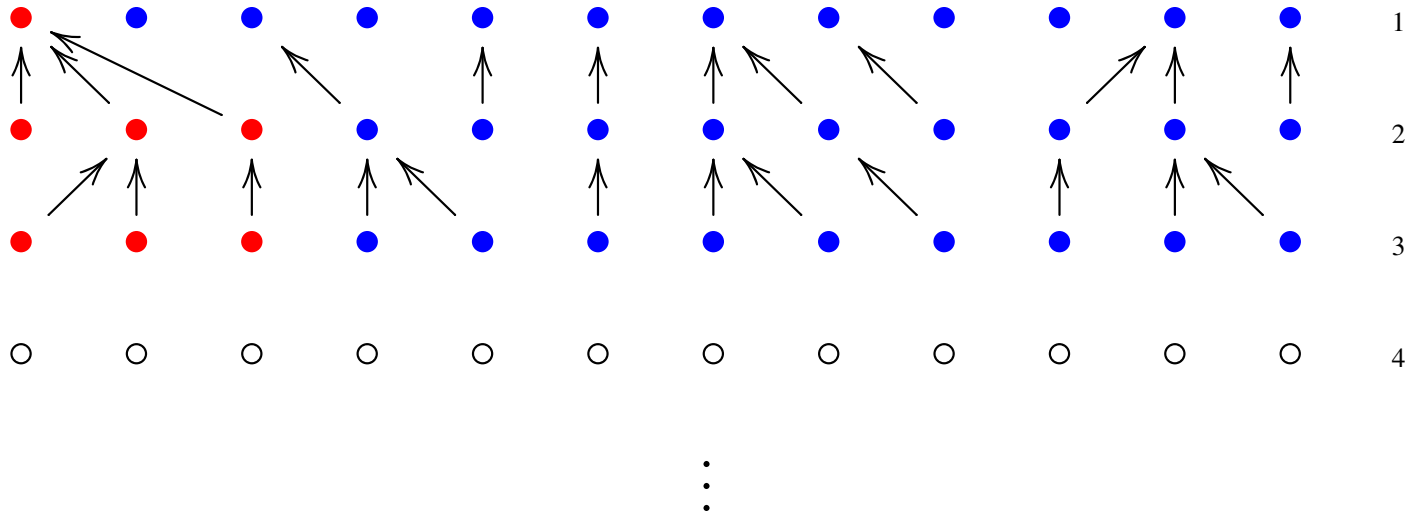
- Assume a haploid population with exactly N individuals with two possible types : a mutant and the ancestral "wild-type"
- Discrete time (generations).
- Each generation, each individual "looks back" to the previous and "picks" a parent.
- Offspring have the same type as their parent.

The Wright-Fisher Model



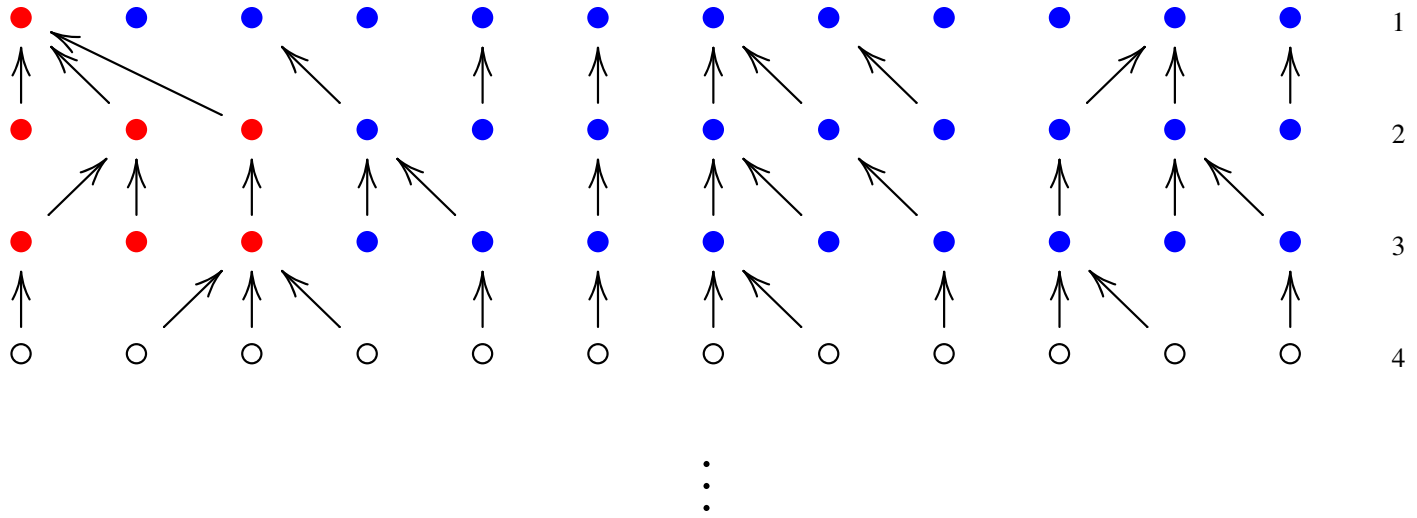
- Assume a haploid population with exactly N individuals with two possible types : a mutant and the ancestral "wild-type"
- Discrete time (generations).
- Each generation, each individual "looks back" to the previous and "picks" a parent.
- Offspring have the same type as their parent.

The Wright-Fisher Model



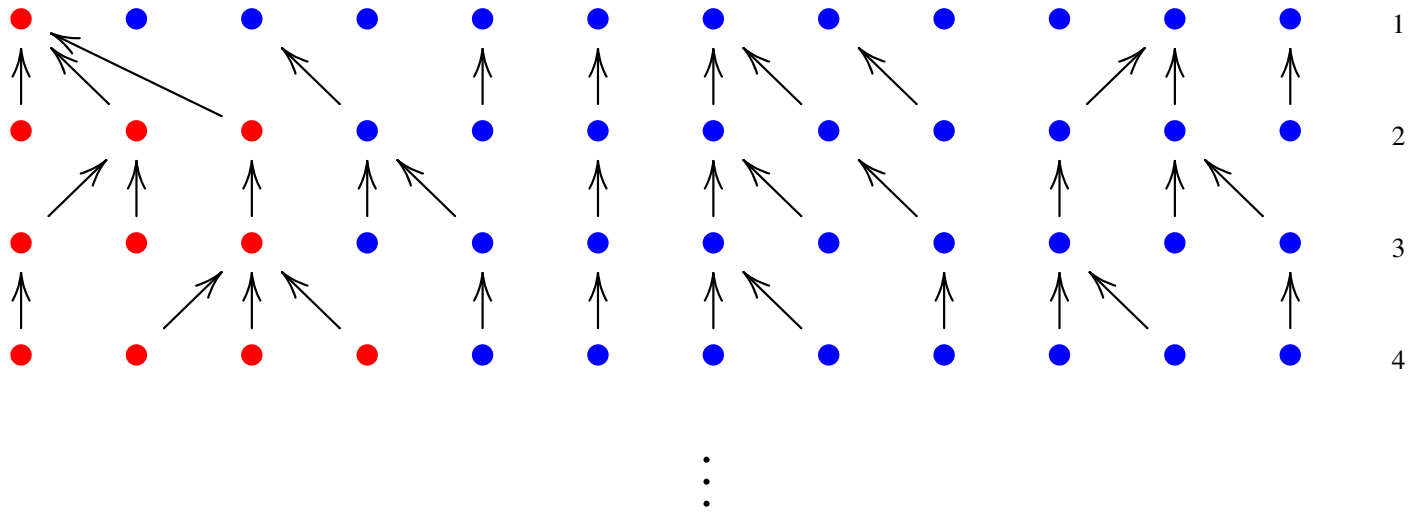
- Assume a haploid population with exactly N individuals with two possible types : a mutant and the ancestral "wild-type"
- Discrete time (generations).
- Each generation, each individual "looks back" to the previous and "picks" a parent.
- Offspring have the same type as their parent.

The Wright-Fisher Model



- Assume a haploid population with exactly N individuals with two possible types : a mutant and the ancestral "wild-type"
- Discrete time (generations).
- Each generation, each individual "looks back" to the previous and "picks" a parent.
- Offspring have the same type as their parent.

The Wright-Fisher Model



- Assume a haploid population with exactly N individuals with two possible types : a mutant and the ancestral "wild-type"
- Discrete time (generations).
- Each generation, each individual "looks back" to the previous and "picks" a parent.
- Offspring have the same type as their parent.

The Wright-Fisher Model : Mathematical Formulation

- Let X_n be the number of mutants in the n^{th} generation.
- Let's suppose that the mutant has an advantage : mutants will be picked as parents with probability proportional to $1+s$, whereas the wild-type is picked with probability proportional to 1.
- If $X_n = x_n$ mutant parents in the n^{th} generation, then the probability that a random individual in the $n+1^{\text{st}}$ generation picks a mutant parent is :

$$p_n = \frac{(1+s)x_n}{(1+s)x_n + 1(N-x_n)} = \frac{(1+s)x_n}{sx_n + N}.$$

- The probability that $X_{n+1} = x$ is then

$$\mathbb{P}(X_{n+1} = x | X_n = x_n) = \binom{N}{x} p_n^x (1-p_n)^{N-x},$$

  a Binomial(p_n) distributed random variable with mean Np_n and variance $Np_n(1-p_n)$.
- X_n is a Markov chain with these transition probabilities.

The Wright-Fisher Model : Mathematical Formulation

- Let X_n be the number of mutants in the n^{th} generation.
- Let's suppose that the mutant has an advantage : mutants will be picked as parents with probability proportional to $1+s$, whereas the wild-type is picked with probability proportional to 1.
- If $X_n = x_n$ mutant parents in the n^{th} generation, then the probability that a random individual in the $n+1^{\text{st}}$ generation picks a mutant parent is :

$$p_n = \frac{(1+s)x_n}{(1+s)x_n + 1(N-x_n)} = \frac{(1+s)x_n}{sx_n + N}.$$

- The probability that $X_{n+1} = x$ is then

$$\mathbb{P}(X_{n+1} = x | X_n = x_n) = \binom{N}{x} p_n^x (1-p_n)^{N-x},$$

  a Binomial(p_n) distributed random variable with mean Np_n and variance $Np_n(1-p_n)$.
- X_n is a Markov chain with these transition probabilities.

The Wright-Fisher Model : Mathematical Formulation

- Let X_n be the number of mutants in the n^{th} generation.
- Let's suppose that the mutant has an advantage : mutants will be picked as parents with probability proportional to $1+s$, whereas the wild-type is picked with probability proportional to 1.
- If $X_n = x_n$ mutant parents in the n^{th} generation, then the probability that a random individual in the $n+1^{\text{st}}$ generation picks a mutant parent is :

$$p_n = \frac{(1+s)x_n}{(1+s)x_n + 1(N-x_n)} = \frac{(1+s)x_n}{sx_n + N}.$$

- The probability that $X_{n+1} = x$ is then

$$\mathbb{P}(X_{n+1} = x | X_n = x_n) = \binom{N}{x} p_n^x (1-p_n)^{N-x},$$

a Binomial(p_n) distributed random variable with mean Np_n and variance $Np_n(1-p_n)$.

- X_n is a Markov chain with these transition probabilities.

The Wright-Fisher Model : Mathematical Formulation

- Let X_n be the number of mutants in the n^{th} generation.
- Let's suppose that the mutant has an advantage : mutants will be picked as parents with probability proportional to $1+s$, whereas the wild-type is picked with probability proportional to 1.
- If $X_n = x_n$ mutant parents in the n^{th} generation, then the probability that a random individual in the $n+1^{\text{st}}$ generation picks a mutant parent is :

$$p_n = \frac{(1+s)x_n}{(1+s)x_n + 1(N-x_n)} = \frac{(1+s)x_n}{sx_n + N}.$$

- The probability that $X_{n+1} = x$ is then

$$\mathbb{P}(X_{n+1} = x | X_n = x_n) = \binom{N}{x} p_n^x (1-p_n)^{N-x},$$

  a Binomial(p_n) distributed random variable with mean Np_n and variance $Np_n(1-p_n)$.
- X_n is a Markov chain with these transition probabilities.

The Wright-Fisher Model : Mathematical Formulation

- Let X_n be the number of mutants in the n^{th} generation.
- Let's suppose that the mutant has an advantage : mutants will be picked as parents with probability proportional to $1+s$, whereas the wild-type is picked with probability proportional to 1.
- If $X_n = x_n$ mutant parents in the n^{th} generation, then the probability that a random individual in the $n+1^{\text{st}}$ generation picks a mutant parent is :

$$p_n = \frac{(1+s)x_n}{(1+s)x_n + 1(N-x_n)} = \frac{(1+s)x_n}{sx_n + N}.$$

- The probability that $X_{n+1} = x$ is then

$$\mathbb{P}(X_{n+1} = x | X_n = x_n) = \binom{N}{x} p_n^x (1-p_n)^{N-x},$$

  a Binomial(p_n) distributed random variable with mean Np_n and variance $Np_n(1-p_n)$.
- X_n is a Markov chain with these transition probabilities.

The Wright-Fisher Model : Diffusion Approximation

- Although easy to simulate, the Wright-Fisher model is hard to analyse.
- Inspired by the heat equation, in 1922, Fisher had the insight to approximate his model via a continuous equation : let

$$f(x,y,t) = \mathbb{P}\left(X_{\lfloor Nt \rfloor} = \lfloor Nx \rfloor \middle| X_0 = \lfloor Ny \rfloor\right),$$

  so x and y are now the *frequency* of the mutant type at the first and current generations.
- Then, if N is very big, and s is very small, say $s = \frac{\alpha}{N}$, then

$$\frac{\partial f}{\partial t}(x,y,t) \approx \alpha y(1-y)\frac{\partial f}{\partial y}(x,y,t) + \frac{1}{2}y(1-y)\frac{\partial^2 f}{\partial y^2}(x,y,t).$$

- This is one of the early examples of a diffusion equation.

The Wright-Fisher Model : Diffusion Approximation

- Although easy to simulate, the Wright-Fisher model is hard to analyse.
- Inspired by the heat equation, in 1922, Fisher had the insight to approximate his model via a continuous equation : let

$$f(x,y,t) = \mathbb{P}\left(X_{\lfloor Nt \rfloor} = \lfloor Nx \rfloor \middle| X_0 = \lfloor Ny \rfloor\right),$$

so x and y are now the *frequency* of the mutant type at the first and current generations.

- Then, if N is very big, and s is very small, say $s = \frac{\alpha}{N}$, then

$$\frac{\partial f}{\partial t}(x,y,t) \approx \alpha y(1-y)\frac{\partial f}{\partial y}(x,y,t) + \frac{1}{2}y(1-y)\frac{\partial^2 f}{\partial y^2}(x,y,t).$$

- This is one of the early examples of a diffusion equation.

The Wright-Fisher Model : Diffusion Approximation

- Although easy to simulate, the Wright-Fisher model is hard to analyse.
- Inspired by the heat equation, in 1922, Fisher had the insight to approximate his model via a continuous equation : let

$$f(x,y,t) = \mathbb{P}\left(X_{\lfloor Nt \rfloor} = \lfloor Nx \rfloor \middle| X_0 = \lfloor Ny \rfloor\right),$$

  so x and y are now the *frequency* of the mutant type at the first and current generations.
- Then, if N is very big, and s is very small, say $s = \frac{\alpha}{N}$, then

$$\frac{\partial f}{\partial t}(x,y,t) \approx \alpha y(1-y)\frac{\partial f}{\partial y}(x,y,t) + \frac{1}{2}y(1-y)\frac{\partial^2 f}{\partial y^2}(x,y,t).$$

- This is one of the early examples of a diffusion equation.

The Wright-Fisher Model : Diffusion Approximation

- Although easy to simulate, the Wright-Fisher model is hard to analyse.
- Inspired by the heat equation, in 1922, Fisher had the insight to approximate his model via a continuous equation : let

$$f(x,y,t) = \mathbb{P}\left(X_{\lfloor Nt \rfloor} = \lfloor Nx \rfloor \middle| X_0 = \lfloor Ny \rfloor\right),$$

  so x and y are now the *frequency* of the mutant type at the first and current generations.
- Then, if N is very big, and s is very small, say $s = \frac{\alpha}{N}$, then

$$\frac{\partial f}{\partial t}(x,y,t) \approx \alpha y(1-y)\frac{\partial f}{\partial y}(x,y,t) + \frac{1}{2}y(1-y)\frac{\partial^2 f}{\partial y^2}(x,y,t).$$

- This is one of the early examples of a diffusion equation.

Fixation Probability

- With the diffusion, we can approach one of the questions that interested Fisher : what is the probability that the mutant *fixes i.e.*, that eventually all N individuals in the population have the mutation.
- The probability this happens before time t, starting from frequency y, is $F(y,t) = f(1,y,t)$, which satisfies the same diffusion equation.
- Thus, the probability the mutant fixes is $F(y) = \lim_{t\to\infty} F(t,y)$ which satisfies

$$\alpha y(1-y)\frac{\partial F}{\partial y}(y) + \frac{1}{2}y(1-y)\frac{\partial^2 F}{\partial y^2}(y) = 0.$$

- This has solution $F(y) = \frac{1-e^{2\alpha y}}{1-e^{2\alpha}}$.
- In particular, starting from 1 mutant (*i.e.*, frequency $\frac{1}{N}$), we see the fixation probability is

$$\frac{1-e^{2\frac{\alpha}{N}}}{1-e^{2\alpha}} = \frac{1-e^{s}}{1-e^{2Ns}} \approx 2s.$$

- Proceeding similarly, the expected time to fixation is $\frac{1}{s}\ln(Ns)$.

Fixation Probability

- With the diffusion, we can approach one of the questions that interested Fisher : what is the probability that the mutant *fixes i.e.*, that eventually all N individuals in the population have the mutation.
- The probability this happens before time t, starting from frequency y, is $F(y,t) = f(1,y,t)$, which satisfies the same diffusion equation.
- Thus, the probability the mutant fixes is $F(y) = \lim_{t\to\infty} F(t,y)$ which satisfies

$$\alpha y(1-y)\frac{\partial F}{\partial y}(y) + \frac{1}{2}y(1-y)\frac{\partial^2 F}{\partial y^2}(y) = 0.$$

- This has solution $F(y) = \frac{1-e^{2\alpha y}}{1-e^{2\alpha}}$.
- In particular, starting from 1 mutant (*i.e.*, frequency $\frac{1}{N}$), we see the fixation probability is

$$\frac{1-e^{2\frac{\alpha}{N}}}{1-e^{2\alpha}} = \frac{1-e^{s}}{1-e^{2Ns}} \approx 2s.$$

- Proceeding similarly, the expected time to fixation is $\frac{1}{s}\ln(Ns)$.

Fixation Probability

- With the diffusion, we can approach one of the questions that interested Fisher : what is the probability that the mutant *fixes i.e.*, that eventually all N individuals in the population have the mutation.
- The probability this happens before time t, starting from frequency y, is $F(y,t) = f(1,y,t)$, which satisfies the same diffusion equation.
- Thus, the probability the mutant fixes is $F(y) = \lim_{t\to\infty} F(t,y)$ which satisfies

$$\alpha y(1-y)\frac{\partial F}{\partial y}(y) + \frac{1}{2}y(1-y)\frac{\partial^2 F}{\partial y^2}(y) = 0.$$

- This has solution $F(y) = \frac{1-e^{2\alpha y}}{1-e^{2\alpha}}$.
- In particular, starting from 1 mutant (*i.e.*, frequency $\frac{1}{N}$), we see the fixation probability is

$$\frac{1-e^{2\frac{\alpha}{N}}}{1-e^{2\alpha}} = \frac{1-e^{s}}{1-e^{2Ns}} \approx 2s.$$

- Proceeding similarly, the expected time to fixation is $\frac{1}{s}\ln(Ns)$.

Fixation Probability

- With the diffusion, we can approach one of the questions that interested Fisher : what is the probability that the mutant *fixes i.e.*, that eventually all N individuals in the population have the mutation.
- The probability this happens before time t, starting from frequency y, is $F(y,t) = f(1,y,t)$, which satisfies the same diffusion equation.
- Thus, the probability the mutant fixes is $F(y) = \lim_{t\to\infty} F(t,y)$ which satisfies

$$\alpha y(1-y)\frac{\partial F}{\partial y}(y) + \frac{1}{2}y(1-y)\frac{\partial^2 F}{\partial y^2}(y) = 0.$$

- This has solution $F(y) = \frac{1-e^{2\alpha y}}{1-e^{2\alpha}}$.
- In particular, starting from 1 mutant (*i.e.*, frequency $\frac{1}{N}$), we see the fixation probability is

$$\frac{1-e^{2\frac{\alpha}{N}}}{1-e^{2\alpha}} = \frac{1-e^{s}}{1-e^{2Ns}} \approx 2s.$$

- Proceeding similarly, the expected time to fixation is $\frac{1}{s}\ln(Ns)$.

Fixation Probability

- With the diffusion, we can approach one of the questions that interested Fisher : what is the probability that the mutant *fixes i.e.*, that eventually all N individuals in the population have the mutation.
- The probability this happens before time t, starting from frequency y, is $F(y,t) = f(1,y,t)$, which satisfies the same diffusion equation.
- Thus, the probability the mutant fixes is $F(y) = \lim_{t\to\infty} F(t,y)$ which satisfies

$$\alpha y(1-y)\frac{\partial F}{\partial y}(y) + \frac{1}{2}y(1-y)\frac{\partial^2 F}{\partial y^2}(y) = 0.$$

- This has solution $F(y) = \frac{1-e^{2\alpha y}}{1-e^{2\alpha}}$.
- In particular, starting from 1 mutant (*i.e.*, frequency $\frac{1}{N}$), we see the fixation probability is

$$\frac{1-e^{2\frac{\alpha}{N}}}{1-e^{2\alpha}} = \frac{1-e^{s}}{1-e^{2Ns}} \approx 2s.$$

- Proceeding similarly, the expected time to fixation is $\frac{1}{s}\ln(Ns)$.

Fixation Probability

- With the diffusion, we can approach one of the questions that interested Fisher : what is the probability that the mutant *fixes i.e.*, that eventually all N individuals in the population have the mutation.
- The probability this happens before time t, starting from frequency y, is $F(y,t) = f(1,y,t)$, which satisfies the same diffusion equation.
- Thus, the probability the mutant fixes is $F(y) = \lim_{t\to\infty} F(t,y)$ which satisfies

$$\alpha y(1-y)\frac{\partial F}{\partial y}(y) + \frac{1}{2}y(1-y)\frac{\partial^2 F}{\partial y^2}(y) = 0.$$

- This has solution $F(y) = \frac{1-e^{2\alpha y}}{1-e^{2\alpha}}$.
- In particular, starting from 1 mutant (*i.e.*, frequency $\frac{1}{N}$), we see the fixation probability is

$$\frac{1-e^{2\frac{\alpha}{N}}}{1-e^{2\alpha}} = \frac{1-e^{s}}{1-e^{2Ns}} \approx 2s.$$

- Proceeding similarly, the expected time to fixation is $\frac{1}{s}\ln(Ns)$.

Mathematics saves Darwinism

- A bit of mathematics shows us mutations have fixation probability approximately proportional to their benefit, $2s$.
- Even if s is very small, the probability that none of M mutations goes to fixation is $(1-2s)^M$ which quickly becomes very small (*e.g.*, the probability that none of 60 mutations that give a 5% benefit is less than 5%.
- Moreover, those mutations spread rapidly throughout the whole population : even in a population of $N = 10^{10}$ individuals, $lnN \approx 23$.
- The upshot : mutations of small benefit can accumulate over reasonable timescales, producing the gradual change that Darwin observed.
- The theory that grew out of Fisher, Haldane and Wright's contribution is what we now call the *Modern (Evolutionary) Synthesis*.

Mathematics saves Darwinism

- A bit of mathematics shows us mutations have fixation probability approximately proportional to their benefit, $2s$.
- Even if s is very small, the probability that none of M mutations goes to fixation is $(1-2s)^M$ which quickly becomes very small (*e.g.*, the probability that none of 60 mutations that give a 5% benefit is less than 5%.
- Moreover, those mutations spread rapidly throughout the whole population : even in a population of $N = 10^{10}$ individuals, $lnN \approx 23$.
- The upshot : mutations of small benefit can accumulate over reasonable timescales, producing the gradual change that Darwin observed.
- The theory that grew out of Fisher, Haldane and Wright's contribution is what we now call the *Modern (Evolutionary) Synthesis*.

Mathematics saves Darwinism

- A bit of mathematics shows us mutations have fixation probability approximately proportional to their benefit, $2s$.
- Even if s is very small, the probability that none of M mutations goes to fixation is $(1-2s)^M$ which quickly becomes very small (*e.g.*, the probability that none of 60 mutations that give a 5% benefit is less than 5%.
- Moreover, those mutations spread rapidly throughout the whole population : even in a population of $N = 10^{10}$ individuals, $lnN \approx 23$.
- The upshot : mutations of small benefit can accumulate over reasonable timescales, producing the gradual change that Darwin observed.
- The theory that grew out of Fisher, Haldane and Wright's contribution is what we now call the *Modern (Evolutionary) Synthesis*.

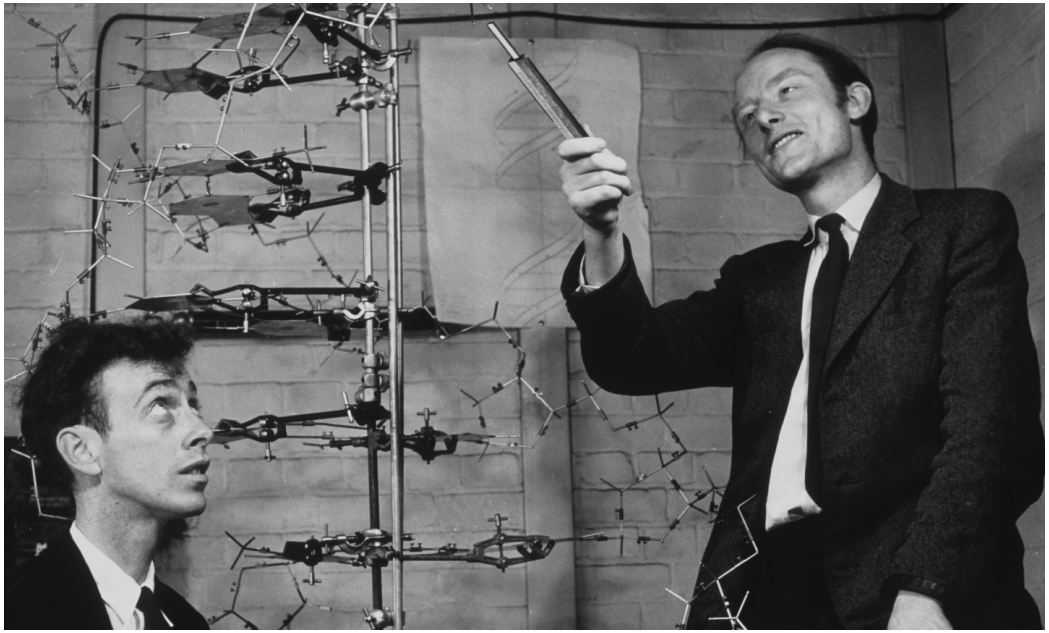
Mathematics saves Darwinism

- A bit of mathematics shows us mutations have fixation probability approximately proportional to their benefit, $2s$.
- Even if s is very small, the probability that none of M mutations goes to fixation is $(1-2s)^M$ which quickly becomes very small (*e.g.*, the probability that none of 60 mutations that give a 5% benefit is less than 5%.
- Moreover, those mutations spread rapidly throughout the whole population : even in a population of $N = 10^{10}$ individuals, $lnN \approx 23$.
- The upshot : mutations of small benefit can accumulate over reasonable timescales, producing the gradual change that Darwin observed.
- The theory that grew out of Fisher, Haldane and Wright's contribution is what we now call the *Modern (Evolutionary) Synthesis*.

Mathematics saves Darwinism

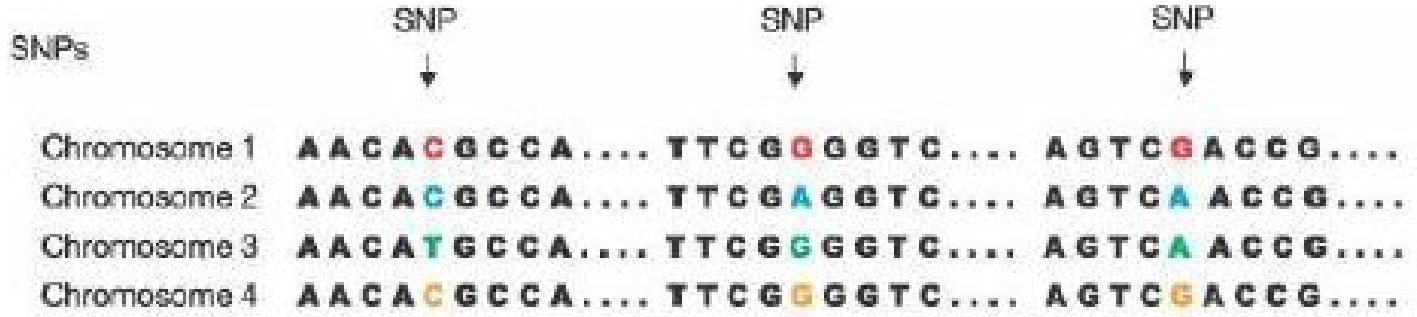
- A bit of mathematics shows us mutations have fixation probability approximately proportional to their benefit, $2s$.
- Even if s is very small, the probability that none of M mutations goes to fixation is $(1-2s)^M$ which quickly becomes very small (*e.g.*, the probability that none of 60 mutations that give a 5% benefit is less than 5%.
- Moreover, those mutations spread rapidly throughout the whole population : even in a population of $N = 10^{10}$ individuals, $lnN \approx 23$.
- The upshot : mutations of small benefit can accumulate over reasonable timescales, producing the gradual change that Darwin observed.
- The theory that grew out of Fisher, Haldane and Wright's contribution is what we now call the *Modern (Evolutionary) Synthesis*.

Molecular (r)Evolution



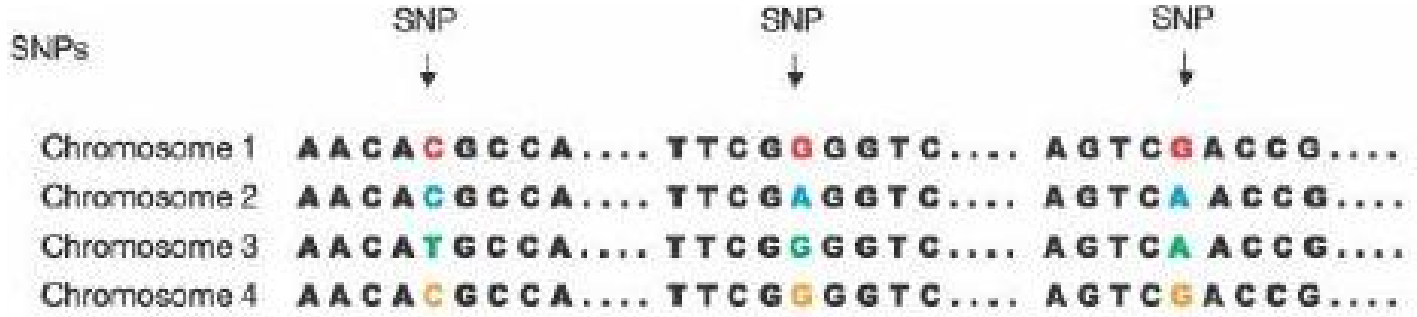
Fast forward a few generations. . . In 1953, Cambridge physicists Francis Crick and James D. Watson published their discovery structure of the DNA molecule, which in subsequent decades allowed first the development of gel electrophoresis and then DNA sequencing.

Confronting (Mathematical) Models with (Molecular) Data



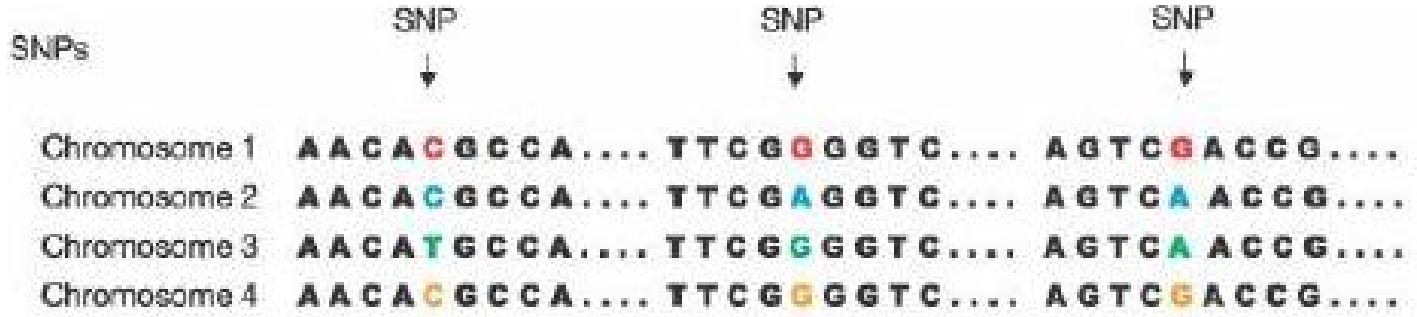
- We could now count the mutations that distinguished individuals. This presented a new mathematical challenge : given sequenced DNA from the present, what could we say about its evolutionary history ?
- We learned that many mutations had no effect : DNA is a string of triples of nucleotides, adenine (A), cytosine (C), guanine (G) and thymine (T), and each triple encodes an amino acid, but changing a single nucleotide (esp. the third) often leaves the amino acid unchanged. Is there evidence for the selection ($s > 0$) in molecular evolution ?
- Attempting to answer such questions eventually led to the infinite sites model, which inspired Kingman's coalescent.

Confronting (Mathematical) Models with (Molecular) Data



- We could now count the mutations that distinguished individuals. This presented a new mathematical challenge : given sequenced DNA from the present, what could we say about its evolutionary history ?
- We learned that many mutations had no effect : DNA is a string of triples of nucleotides, adenine (A), cytosine (C), guanine (G) and thymine (T), and each triple encodes an amino acid, but changing a single nucleotide (esp. the third) often leaves the amino acid unchanged. Is there evidence for the selection ($s > 0$) in molecular evolution ?
- Attempting to answer such questions eventually led to the infinite sites model, which inspired Kingman's coalescent.

Confronting (Mathematical) Models with (Molecular) Data



- We could now count the mutations that distinguished individuals. This presented a new mathematical challenge : given sequenced DNA from the present, what could we say about its evolutionary history ?
- We learned that many mutations had no effect : DNA is a string of triples of nucleotides, adenine (A), cytosine (C), guanine (G) and thymine (T), and each triple encodes an amino acid, but changing a single nucleotide (esp. the third) often leaves the amino acid unchanged. Is there evidence for the selection ($s > 0$) in molecular evolution ?
- Attempting to answer such questions eventually led to the infinite sites model, which inspired Kingman's coalescent.

Infinite Sites model

James Crow (1916 – 2012) & Motoo Kimura (1924–1994)

- Crow & Kimura argued that because genomes are very long (*e.g.*, the human genome has length 3,200,000,000bp) the probability of observing any mutation twice was very low (effectively zero)
- They used this assumption to make many (hard) calculations to estimate the number of observed mutations in a sample, and thereby infer the mutation rate.
- We'll use the benefit of hindsight, and use a more elegant approach. . .

Infinite Sites model

James Crow (1916 – 2012) & Motoo Kimura (1924–1994)

- Crow & Kimura argued that because genomes are very long (*e.g.*, the human genome has length 3,200,000,000bp) the probability of observing any mutation twice was very low (effectively zero)
- They used this assumption to make many (hard) calculations to estimate the number of observed mutations in a sample, and thereby infer the mutation rate.
- We'll use the benefit of hindsight, and use a more elegant approach. . .

Infinite Sites model

James Crow (1916 – 2012) & Motoo Kimura (1924–1994)

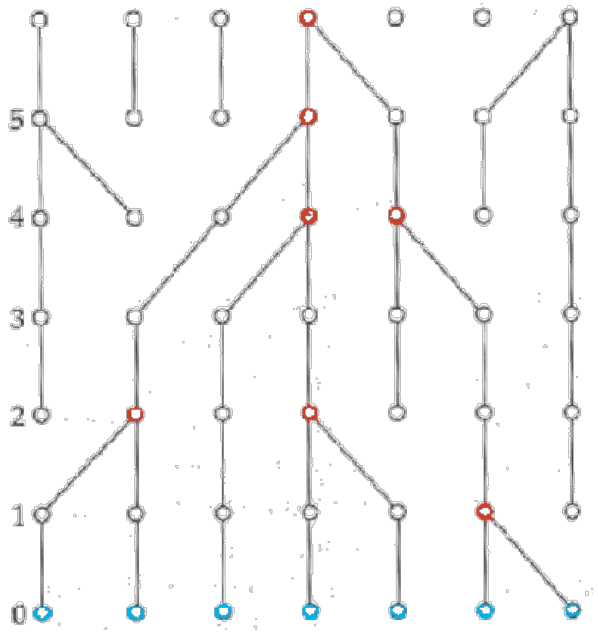
- Crow & Kimura argued that because genomes are very long (*e.g.*, the human genome has length 3,200,000,000bp) the probability of observing any mutation twice was very low (effectively zero)
- They used this assumption to make many (hard) calculations to estimate the number of observed mutations in a sample, and thereby infer the mutation rate.
- We'll use the benefit of hindsight, and use a more elegant approach. . .

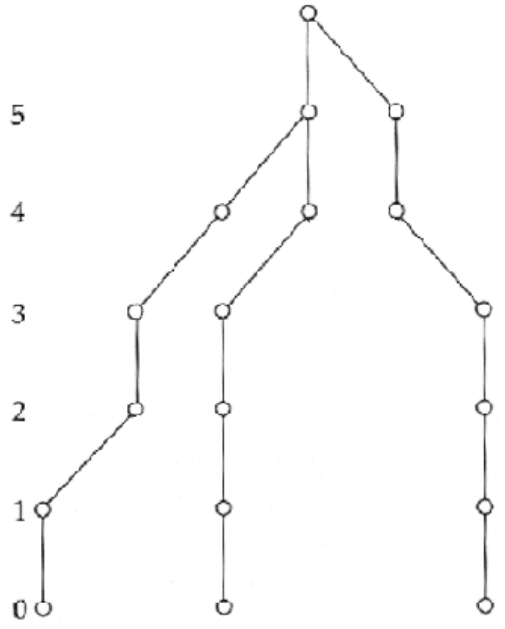
Kingman's Coalescent

- Kingman turned thinking about evolution on its head, by thinking about ancestral lineages backward in time, rather than individuals forward in time.
- Tracing the ancestors of a sample backwards in time gives us a random genealogical tree.
- Some simple maths allows us to estimate the length of the branches of this tree, and thus determine mutation rates.

J. F. C. Kingman
(1939–)

Kingman's Coalescent





- Suppose we have a sample of n individuals from a population of size N that has been evolving according to the Wright-Fisher model. How long do we have to wait before two of those individuals have a common ancestor ?
- If no individual has a selective advantage ($s = 0$), each parent is chosen with equal probability, $\frac{1}{N}$.
- The probability that two individuals choose the same parent is thus $\frac{1}{N}$.
- No pair in our sample have the same parent with probability $1 - \frac{1}{N}\binom{n}{2}$.
- Now, let $t > 0$. The probability that after $\lfloor Nt \rfloor$ generations, no pair of individuals in our sample has chosen the same parent is

$$\left(1 - \frac{1}{N}\binom{n}{2}\right)^{\lfloor Nt \rfloor} \underset{\infty}{\overset{N}{\longrightarrow}} e^{-\binom{n}{2}t}.$$

- The probability that three or more individuals choose the same parent, or two pairs choose the same two parents, is $\leq \frac{C}{N^2}$; the probability this happens in $\lfloor Nt \rfloor$ generations is $\leq \frac{C\lfloor Nt \rfloor}{N^2} \to 0$.
- This describes a continuous time Markov chain : if $n(t)$ is the number of ancestors of the sample at time t in the past, $n(t)$ transitions from n to $n-1$ with rate $\binom{n}{2}$.

- Suppose we have a sample of n individuals from a population of size N that has been evolving according to the Wright-Fisher model. How long do we have to wait before two of those individuals have a common ancestor ?
- If no individual has a selective advantage ($s = 0$), each parent is chosen with equal probability, $\frac{1}{N}$.
- The probability that two individuals choose the same parent is thus $\frac{1}{N}$.
- No pair in our sample have the same parent with probability $1 - \frac{1}{N}\binom{n}{2}$.
- Now, let $t > 0$. The probability that after $\lfloor Nt \rfloor$ generations, no pair of individuals in our sample has chosen the same parent is

$$\left(1 - \frac{1}{N}\binom{n}{2}\right)^{\lfloor Nt \rfloor} \underset{\infty}{\overset{N}{\longrightarrow}} e^{-\binom{n}{2}t}.$$

- The probability that three or more individuals choose the same parent, or two pairs choose the same two parents, is $\leq \frac{C}{N^2}$; the probability this happens in $\lfloor Nt \rfloor$ generations is $\leq \frac{C\lfloor Nt \rfloor}{N^2} \to 0$.
- This describes a continuous time Markov chain : if $n(t)$ is the number of ancestors of the sample at time t in the past, $n(t)$ transitions from n to $n-1$ with rate $\binom{n}{2}$.

- Suppose we have a sample of n individuals from a population of size N that has been evolving according to the Wright-Fisher model. How long do we have to wait before two of those individuals have a common ancestor ?
- If no individual has a selective advantage ($s = 0$), each parent is chosen with equal probability, $\frac{1}{N}$.
- The probability that two individuals choose the same parent is thus $\frac{1}{N}$.
- No pair in our sample have the same parent with probability $1 - \frac{1}{N}\binom{n}{2}$.
- Now, let $t > 0$. The probability that after $\lfloor Nt \rfloor$ generations, no pair of individuals in our sample has chosen the same parent is

$$\left(1 - \frac{1}{N}\binom{n}{2}\right)^{\lfloor Nt \rfloor} \underset{\infty}{\overset{N}{\longrightarrow}} e^{-\binom{n}{2}t}.$$

- The probability that three or more individuals choose the same parent, or two pairs choose the same two parents, is $\leq \frac{C}{N^2}$; the probability this happens in $\lfloor Nt \rfloor$ generations is $\leq \frac{C\lfloor Nt \rfloor}{N^2} \to 0$.
- This describes a continuous time Markov chain : if $n(t)$ is the number of ancestors of the sample at time t in the past, $n(t)$ transitions from n to $n-1$ with rate $\binom{n}{2}$.

- Suppose we have a sample of n individuals from a population of size N that has been evolving according to the Wright-Fisher model. How long do we have to wait before two of those individuals have a common ancestor ?
- If no individual has a selective advantage ($s = 0$), each parent is chosen with equal probability, $\frac{1}{N}$.
- The probability that two individuals choose the same parent is thus $\frac{1}{N}$.
- No pair in our sample have the same parent with probability $1 - \frac{1}{N}\binom{n}{2}$.
- Now, let $t > 0$. The probability that after $\lfloor Nt \rfloor$ generations, no pair of individuals in our sample has chosen the same parent is

$$\left(1 - \frac{1}{N}\binom{n}{2}\right)^{\lfloor Nt \rfloor} \underset{\infty}{\longrightarrow} e^{-\binom{n}{2}t}.$$

- The probability that three or more individuals choose the same parent, or two pairs choose the same two parents, is $\leq \frac{C}{N^2}$; the probability this happens in $\lfloor Nt \rfloor$ generations is $\leq \frac{C\lfloor Nt \rfloor}{N^2} \to 0$.
- This describes a continuous time Markov chain : if $n(t)$ is the number of ancestors of the sample at time t in the past, $n(t)$ transitions from n to $n-1$ with rate $\binom{n}{2}$.

- Suppose we have a sample of n individuals from a population of size N that has been evolving according to the Wright-Fisher model. How long do we have to wait before two of those individuals have a common ancestor ?
- If no individual has a selective advantage ($s = 0$), each parent is chosen with equal probability, $\frac{1}{N}$.
- The probability that two individuals choose the same parent is thus $\frac{1}{N}$.
- No pair in our sample have the same parent with probability $1 - \frac{1}{N}\binom{n}{2}$.
- Now, let $t > 0$. The probability that after $\lfloor Nt \rfloor$ generations, no pair of individuals in our sample has chosen the same parent is

$$\left(1 - \frac{1}{N}\binom{n}{2}\right)^{\lfloor Nt \rfloor} \xrightarrow[\infty]{N} e^{-\binom{n}{2}t}.$$

- The probability that three or more individuals choose the same parent, or two pairs choose the same two parents, is $\leq \frac{C}{N^2}$; the probability this happens in $\lfloor Nt \rfloor$ generations is $\leq \frac{C\lfloor Nt \rfloor}{N^2} \to 0$.
- This describes a continuous time Markov chain : if $n(t)$ is the number of ancestors of the sample at time t in the past, $n(t)$ transitions from n to $n-1$ with rate $\binom{n}{2}$.

- Suppose we have a sample of n individuals from a population of size N that has been evolving according to the Wright-Fisher model. How long do we have to wait before two of those individuals have a common ancestor ?
- If no individual has a selective advantage ($s = 0$), each parent is chosen with equal probability, $\frac{1}{N}$.
- The probability that two individuals choose the same parent is thus $\frac{1}{N}$.
- No pair in our sample have the same parent with probability $1 - \frac{1}{N}\binom{n}{2}$.
- Now, let $t > 0$. The probability that after $\lfloor Nt \rfloor$ generations, no pair of individuals in our sample has chosen the same parent is

$$\left(1 - \frac{1}{N}\binom{n}{2}\right)^{\lfloor Nt \rfloor} \underset{\infty}{\overset{N}{\longrightarrow}} e^{-\binom{n}{2}t}.$$

- The probability that three or more individuals choose the same parent, or two pairs choose the same two parents, is $\leq \frac{C}{N^2}$; the probability this happens in $\lfloor Nt \rfloor$ generations is $\leq \frac{C\lfloor Nt \rfloor}{N^2} \to 0$.
- This describes a continuous time Markov chain : if $n(t)$ is the number of ancestors of the sample at time t in the past, $n(t)$ transitions from n to $n-1$ with rate $\binom{n}{2}$.

- Suppose we have a sample of n individuals from a population of size N that has been evolving according to the Wright-Fisher model. How long do we have to wait before two of those individuals have a common ancestor ?
- If no individual has a selective advantage ($s = 0$), each parent is chosen with equal probability, $\frac{1}{N}$.
- The probability that two individuals choose the same parent is thus $\frac{1}{N}$.
- No pair in our sample have the same parent with probability $1 - \frac{1}{N}\binom{n}{2}$.
- Now, let $t > 0$. The probability that after $\lfloor Nt \rfloor$ generations, no pair of individuals in our sample has chosen the same parent is

$$\left(1 - \frac{1}{N}\binom{n}{2}\right)^{\lfloor Nt \rfloor} \underset{\infty}{\overset{N}{\longrightarrow}} e^{-\binom{n}{2}t}.$$

- The probability that three or more individuals choose the same parent, or two pairs choose the same two parents, is $\leq \frac{C}{N^2}$; the probability this happens in $\lfloor Nt \rfloor$ generations is $\leq \frac{C\lfloor Nt \rfloor}{N^2} \to 0$.
- This describes a continuous time Markov chain : if $n(t)$ is the number of ancestors of the sample at time t in the past, $n(t)$ transitions from n to $n-1$ with rate $\binom{n}{2}$.

The Most Recent Common Ancestor

How long do we have to wait for a common ancestor to everyone ?

- Starting from n individuals, let T_n be the amount of time (measured in units of N) until two have a common ancestor.
- We saw that that $\mathbb{P}(T_n > t) = e^{-\binom{n}{2}t}$.
- Then, using integration by parts,

$$\mathbb{E}[T_n] = \int_0^t t\mathbb{P}(t+dt \geq T_n > t) = \int_0^t t\frac{d}{dt}\mathbb{P}(T_n > t)$$
$$= \int_0^t \mathbb{P}(T_n > t)\, dt = \int_0^t e^{-\binom{n}{2}t}\, dt = \frac{1}{\binom{n}{2}}.$$

- After T_n we have $n-1$ individuals.
- Thus the expected time until a single ancestor is

$$\mathbb{E}\left[\sum_{i=2}^n T_i\right] = \sum_{i=2}^n \mathbb{E}[T_i] = \sum_{i=2}^n \frac{1}{\binom{i}{2}} = 2\sum_{i=2}^n \frac{1}{i-1} - \frac{1}{i} = 2\left(1 - \frac{1}{n}\right).$$

i.e., approximately $2N$ generations.

The Most Recent Common Ancestor

How long do we have to wait for a common ancestor to everyone ?

- Starting from n individuals, let T_n be the amount of time (measured in units of N) until two have a common ancestor.
- We saw that that $\mathbb{P}(T_n > t) = e^{-\binom{n}{2}t}$.
- Then, using integration by parts,

$$\mathbb{E}[T_n] = \int_0^t t\mathbb{P}(t+dt \geq T_n > t) = \int_0^t t\frac{d}{dt}\mathbb{P}(T_n > t)$$
$$= \int_0^t \mathbb{P}(T_n > t)\,dt = \int_0^t e^{-\binom{n}{2}t}\,dt = \frac{1}{\binom{n}{2}}.$$

- After T_n we have $n-1$ individuals.
- Thus the expected time until a single ancestor is

$$\mathbb{E}\left[\sum_{i=2}^{n} T_i\right] = \sum_{i=2}^{n}\mathbb{E}[T_i] = \sum_{i=2}^{n}\frac{1}{\binom{i}{2}} = 2\sum_{i=2}^{n}\frac{1}{i-1} - \frac{1}{i} = 2\left(1-\frac{1}{n}\right).$$

i.e., approximately $2N$ generations.

The Most Recent Common Ancestor

How long do we have to wait for a common ancestor to everyone ?

- Starting from n individuals, let T_n be the amount of time (measured in units of N) until two have a common ancestor.
- We saw that that $\mathbb{P}(T_n > t) = e^{-\binom{n}{2}t}$.
- Then, using integration by parts,

$$\mathbb{E}[T_n] = \int_0^t t\mathbb{P}(t+dt \geq T_n > t) = \int_0^t t\frac{d}{dt}\mathbb{P}(T_n > t)$$
$$= \int_0^t \mathbb{P}(T_n > t)\,dt = \int_0^t e^{-\binom{n}{2}t}\,dt = \frac{1}{\binom{n}{2}}.$$

- After T_n we have $n-1$ individuals.
- Thus the expected time until a single ancestor is

$$\mathbb{E}\left[\sum_{i=2}^{n} T_i\right] = \sum_{i=2}^{n}\mathbb{E}[T_i] = \sum_{i=2}^{n}\frac{1}{\binom{i}{2}} = 2\sum_{i=2}^{n}\frac{1}{i-1} - \frac{1}{i} = 2\left(1-\frac{1}{n}\right).$$

i.e., approximately $2N$ generations.

The Most Recent Common Ancestor

How long do we have to wait for a common ancestor to everyone ?

- Starting from n individuals, let T_n be the amount of time (measured in units of N) until two have a common ancestor.
- We saw that that $\mathbb{P}(T_n > t) = e^{-\binom{n}{2}t}$.
- Then, using integration by parts,

$$\mathbb{E}[T_n] = \int_0^t t\mathbb{P}(t+dt \geq T_n > t) = \int_0^t t\frac{d}{dt}\mathbb{P}(T_n > t)$$
$$= \int_0^t \mathbb{P}(T_n > t)\, dt = \int_0^t e^{-\binom{n}{2}t}\, dt = \frac{1}{\binom{n}{2}}.$$

- After T_n we have $n-1$ individuals.
- Thus the expected time until a single ancestor is

$$\mathbb{E}\left[\sum_{i=2}^{n} T_i\right] = \sum_{i=2}^{n} \mathbb{E}[T_i] = \sum_{i=2}^{n} \frac{1}{\binom{i}{2}} = 2\sum_{i=2}^{n} \frac{1}{i-1} - \frac{1}{i} = 2\left(1 - \frac{1}{n}\right).$$

i.e., approximately $2N$ generations.

The Most Recent Common Ancestor

How long do we have to wait for a common ancestor to everyone ?

- Starting from n individuals, let T_n be the amount of time (measured in units of N) until two have a common ancestor.
- We saw that that $\mathbb{P}(T_n > t) = e^{-\binom{n}{2}t}$.
- Then, using integration by parts,

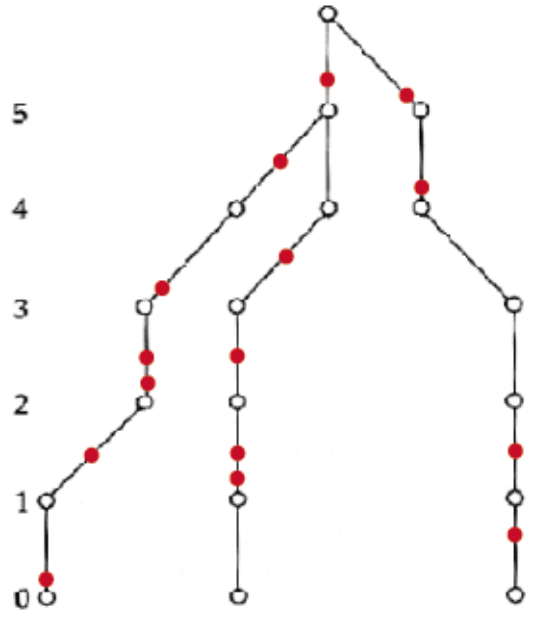
$$\mathbb{E}[T_n] = \int_0^t t\mathbb{P}(t+dt \geq T_n > t) = \int_0^t t\frac{d}{dt}\mathbb{P}(T_n > t)$$
$$= \int_0^t \mathbb{P}(T_n > t)\, dt = \int_0^t e^{-\binom{n}{2}t}\, dt = \frac{1}{\binom{n}{2}}.$$

- After T_n we have $n-1$ individuals.
- Thus the expected time until a single ancestor is

$$\mathbb{E}\left[\sum_{i=2}^{n} T_i\right] = \sum_{i=2}^{n} \mathbb{E}[T_i] = \sum_{i=2}^{n} \frac{1}{\binom{i}{2}} = 2\sum_{i=2}^{n} \frac{1}{i-1} - \frac{1}{i} = 2\left(1 - \frac{1}{n}\right).$$

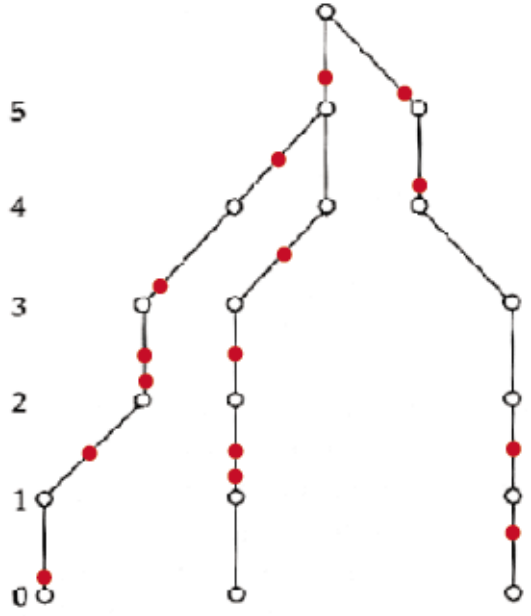
i.e., approximately $2N$ generations.

Adding Mutations



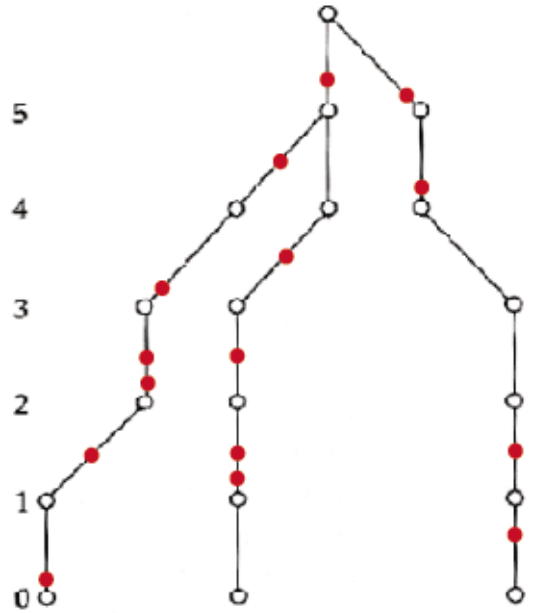
- We know that mutations are rare, so if μ is the probability an individual has a mutation in a generation, we might write $\mu = \frac{\theta}{N}$.
- If we only consider neutral mutations, they leave the genealogy unchanged.
- Then, for the coalescent process, the probability a mutation occurs in an interval of length dt is $\lfloor N\,dt \rfloor \mu \approx \theta\,dt$. *i.e.*, mutations happen at rate θ.

Adding Mutations



- We know that mutations are rare, so if μ is the probability an individual has a mutation in a generation, we might write $\mu = \frac{\theta}{N}$.
- If we only consider neutral mutations, they leave the genealogy unchanged.
- Then, for the coalescent process, the probability a mutation occurs in an interval of length dt is $\lfloor N\,dt \rfloor \mu \approx \theta\,dt$. *i.e.*, mutations happen at rate θ.

Adding Mutations



- We know that mutations are rare, so if μ is the probability an individual has a mutation in a generation, we might write $\mu = \frac{\theta}{N}$.
- If we only consider neutral mutations, they leave the genealogy unchanged.
- Then, for the coalescent process, the probability a mutation occurs in an interval of length dt is $\lfloor N\,dt \rfloor \mu \approx \theta\,dt$. *i.e.*, mutations happen at rate θ.

Inferring Mutation Rates

- Suppose we have a sample of n individuals, and a count of the number of pairwise differences between them, say S_n.
- The expected number of mutation each individual has acquired since their last coalescence (conditional on T_n) is θT_n. Thus, the n lines have on average $n\theta T_n$ differences.
- Recalling that $\mathbb{E}[T_n] = \binom{n}{2}$, going back to the common ancestor, the expected number of differences is

$$\mathbb{E}\left[\sum_{i=2}^{n} n\theta T_n\right] = \sum_{i=2}^{n} \frac{\theta}{n-1} = \theta H(n-1),$$

  where $H(n-1) = \sum_{i=1}^{n-1} \frac{1}{i} \sim \ln(n-1)$.
- Thus, for large samples, $\hat{\theta} = \frac{S_n}{\ln(n-1)}$ gives a (good) estimate of the mutation rate.

Inferring Mutation Rates

- Suppose we have a sample of n individuals, and a count of the number of pairwise differences between them, say S_n.
- The expected number of mutation each individual has acquired since their last coalescence (conditional on T_n) is θT_n. Thus, the n lines have on average $n\theta T_n$ differences.
- Recalling that $\mathbb{E}[T_n] = \binom{n}{2}$, going back to the common ancestor, the expected number of differences is

$$\mathbb{E}\left[\sum_{i=2}^{n} n\theta T_n\right] = \sum_{i=2}^{n} \frac{\theta}{n-1} = \theta H(n-1),$$

  where $H(n-1) = \sum_{i=1}^{n-1} \frac{1}{i} \sim \ln(n-1)$.
- Thus, for large samples, $\hat{\theta} = \frac{S_n}{\ln(n-1)}$ gives a (good) estimate of the mutation rate.

Inferring Mutation Rates

- Suppose we have a sample of n individuals, and a count of the number of pairwise differences between them, say S_n.
- The expected number of mutation each individual has acquired since their last coalescence (conditional on T_n) is θT_n. Thus, the n lines have on average $n\theta T_n$ differences.
- Recalling that $\mathbb{E}[T_n] = \binom{n}{2}$, going back to the common ancestor, the expected number of differences is

$$\mathbb{E}\left[\sum_{i=2}^{n} n\theta T_n\right] = \sum_{i=2}^{n} \frac{\theta}{n-1} = \theta H(n-1),$$

where $H(n-1) = \sum_{i=1}^{n-1} \frac{1}{i} \sim \ln(n-1)$.

- Thus, for large samples, $\hat{\theta} = \frac{S_n}{\ln(n-1)}$ gives a (good) estimate of the mutation rate.

Inferring Mutation Rates

- Suppose we have a sample of n individuals, and a count of the number of pairwise differences between them, say S_n.
- The expected number of mutation each individual has acquired since their last coalescence (conditional on T_n) is θT_n. Thus, the n lines have on average $n\theta T_n$ differences.
- Recalling that $\mathbb{E}[T_n] = \binom{n}{2}$, going back to the common ancestor, the expected number of differences is

$$\mathbb{E}\left[\sum_{i=2}^{n} n\theta T_n\right] = \sum_{i=2}^{n} \frac{\theta}{n-1} = \theta H(n-1),$$

  where $H(n-1) = \sum_{i=1}^{n-1} \frac{1}{i} \sim \ln(n-1)$.
- Thus, for large samples, $\hat{\theta} = \frac{S_n}{\ln(n-1)}$ gives a (good) estimate of the mutation rate.

Reconstructing the Past

- Estimators like those obtained from Kingman's coalescent, applied to mutations we know to be neutral (like those that leave the amino acid unchanged) have allowed us to infer the mutation rates for many organisms :
  - Unicellular eukaryotes and bacteria have roughly 0.003 mutations per genome per generation.
  - Viruses have much higher mutation rates : DNA viruses have mutation rates between 10^{-6} to 10^{-8} mutations per site per generation, and RNA viruses have mutation rates between 10^{-3} to 10^{-5} per base per generation.
  - Human mitochondrial DNA has been estimated to have mutation rates of 2.7×10^{-5} per base per 20 year generation.
- Looking for portions of DNA that have pairwise differences substantially higher or lower than predicted via the neutral rate allow us to discover sites where selection is occurring – higher than neutral suggest pressure to innovate, lower than the neutral suggests pressure to stay the same. With maths, we can infer what is the important DNA, even when we don't know how or why it's important !

Reconstructing the Past

- Estimators like those obtained from Kingman's coalescent, applied to mutations we know to be neutral (like those that leave the amino acid unchanged) have allowed us to infer the mutation rates for many organisms :
  - Unicellular eukaryotes and bacteria have roughly 0.003 mutations per genome per generation.
  - Viruses have much higher mutation rates : DNA viruses have mutation rates between 10^{-6} to 10^{-8} mutations per site per generation, and RNA viruses have mutation rates between 10^{-3} to 10^{-5} per base per generation.
  - Human mitochondrial DNA has been estimated to have mutation rates of 2.7×10^{-5} per base per 20 year generation.
- Looking for portions of DNA that have pairwise differences substantially higher or lower than predicted via the neutral rate allow us to discover sites where selection is occurring – higher than neutral suggest pressure to innovate, lower than the neutral suggests pressure to stay the same. With maths, we can infer what is the important DNA, even when we don't know how or why it's important !

Reconstructing the Past

- Estimators like those obtained from Kingman's coalescent, applied to mutations we know to be neutral (like those that leave the amino acid unchanged) have allowed us to infer the mutation rates for many organisms :
  - Unicellular eukaryotes and bacteria have roughly 0.003 mutations per genome per generation.
  - Viruses have much higher mutation rates : DNA viruses have mutation rates between 10^{-6} to 10^{-8} mutations per site per generation, and RNA viruses have mutation rates between 10^{-3} to 10^{-5} per base per generation.
  - Human mitochondrial DNA has been estimated to have mutation rates of 2.7×10^{-5} per base per 20 year generation.
- Looking for portions of DNA that have pairwise differences substantially higher or lower than predicted via the neutral rate allow us to discover sites where selection is occurring – higher than neutral suggest pressure to innovate, lower than the neutral suggests pressure to stay the same. With maths, we can infer what is the important DNA, even when we don't know how or why it's important !

Reconstructing the Past

- Estimators like those obtained from Kingman's coalescent, applied to mutations we know to be neutral (like those that leave the amino acid unchanged) have allowed us to infer the mutation rates for many organisms :
  - Unicellular eukaryotes and bacteria have roughly 0.003 mutations per genome per generation.
  - Viruses have much higher mutation rates : DNA viruses have mutation rates between 10^{-6} to 10^{-8} mutations per site per generation, and RNA viruses have mutation rates between 10^{-3} to 10^{-5} per base per generation.
  - Human mitochondrial DNA has been estimated to have mutation rates of 2.7×10^{-5} per base per 20 year generation.
- Looking for portions of DNA that have pairwise differences substantially higher or lower than predicted via the neutral rate allow us to discover sites where selection is occurring – higher than neutral suggest pressure to innovate, lower than the neutral suggests pressure to stay the same. With maths, we can infer what is the important DNA, even when we don't know how or why it's important !

Reconstructing the Past

- Estimators like those obtained from Kingman's coalescent, applied to mutations we know to be neutral (like those that leave the amino acid unchanged) have allowed us to infer the mutation rates for many organisms :
  - Unicellular eukaryotes and bacteria have roughly 0.003 mutations per genome per generation.
  - Viruses have much higher mutation rates : DNA viruses have mutation rates between 10^{-6} to 10^{-8} mutations per site per generation, and RNA viruses have mutation rates between 10^{-3} to 10^{-5} per base per generation.
  - Human mitochondrial DNA has been estimated to have mutation rates of 2.7×10^{-5} per base per 20 year generation.
- Looking for portions of DNA that have pairwise differences substantially higher or lower than predicted via the neutral rate allow us to discover sites where selection is occurring – higher than neutral suggest pressure to innovate, lower than the neutral suggests pressure to stay the same. With maths, we can infer what is the important DNA, even when we don't know how or why it's important !

1 The Modern Synthesis

2 The Molecular Revolution

3 The Present & Beyond

- We've looked at two of many examples of mathematical thinking being important in the development of our understanding of evolution.
- In many ways, we're still at the beginning of the story – physics and mathematics have had a relationship going back centuries, whereas math and evolution have less than 100 years of interacting.
- In biology, the disciplines of population genetics and bioinformatics have grown out of extending the work of Fisher, Haldane, Wright, Kingman, *etc.* .
- They also inspired a growing body of work in probability theory. Let's look at a few examples !

- We've looked at two of many examples of mathematical thinking being important in the development of our understanding of evolution.
- In many ways, we're still at the beginning of the story – physics and mathematics have had a relationship going back centuries, whereas math and evolution have less than 100 years of interacting.
- In biology, the disciplines of population genetics and bioinformatics have grown out of extending the work of Fisher, Haldane, Wright, Kingman, *etc.* .
- They also inspired a growing body of work in probability theory. Let's look at a few examples !

- We've looked at two of many examples of mathematical thinking being important in the development of our understanding of evolution.
- In many ways, we're still at the beginning of the story – physics and mathematics have had a relationship going back centuries, whereas math and evolution have less than 100 years of interacting.
- In biology, the disciplines of population genetics and bioinformatics have grown out of extending the work of Fisher, Haldane, Wright, Kingman, *etc.* .
- They also inspired a growing body of work in probability theory. Let's look at a few examples !

- We've looked at two of many examples of mathematical thinking being important in the development of our understanding of evolution.
- In many ways, we're still at the beginning of the story – physics and mathematics have had a relationship going back centuries, whereas math and evolution have less than 100 years of interacting.
- In biology, the disciplines of population genetics and bioinformatics have grown out of extending the work of Fisher, Haldane, Wright, Kingman, *etc.* .
- They also inspired a growing body of work in probability theory. Let's look at a few examples !

Λ- and Ξ-Coalescents

- In the neutral ($s = 0$) Wright-Fisher model, we assumed that the vector of offspring numbers for parent was a Multinomial $\left(N, \frac{1}{N}\right)$ variable.
- We could equally well have considered more general exchangeable random vectors : $(\nu_1, \ldots, \nu_N)$ is exchangeable if

$$\mathbb{P}\left((\nu_1, \ldots, \nu_N) = (n_1, \ldots, n_N)\right) = \mathbb{P}\left((\nu_{\sigma(1)}, \ldots, \nu_{\sigma(N)}) = (n_1, \ldots, n_N)\right)$$

  for any permutation σ of $\{1, \ldots, N\}$.
- Unlike the Kingman coalescent, when we look at more general offspring distributions, we can have multiple mergers (Λ-coalescents) and simultaneous mergers (Ξ-coalescents)
- Understanding the properties of these coalescents is a topic of ongoing research.

Λ- and Ξ-Coalescents

- In the neutral ($s = 0$) Wright-Fisher model, we assumed that the vector of offspring numbers for parent was a Multinomial $\left(N, \frac{1}{N}\right)$ variable.
- We could equally well have considered more general exchangeable random vectors : $(\nu_1, \ldots, \nu_N)$ is exchangeable if

$$\mathbb{P}\left((\nu_1, \ldots, \nu_N) = (n_1, \ldots, n_N)\right) = \mathbb{P}\left((\nu_{\sigma(1)}, \ldots, \nu_{\sigma(N)}) = (n_1, \ldots, n_N)\right)$$

  for any permutation σ of $\{1, \ldots, N\}$.
- Unlike the Kingman coalescent, when we look at more general offspring distributions, we can have multiple mergers (Λ-coalescents) and simultaneous mergers (Ξ-coalescents)
- Understanding the properties of these coalescents is a topic of ongoing research.

Λ- and Ξ-Coalescents

- In the neutral ($s = 0$) Wright-Fisher model, we assumed that the vector of offspring numbers for parent was a Multinomial $\left(N, \frac{1}{N}\right)$ variable.
- We could equally well have considered more general exchangeable random vectors : $(\nu_1, \ldots, \nu_N)$ is exchangeable if

$$\mathbb{P}\left((\nu_1, \ldots, \nu_N) = (n_1, \ldots, n_N)\right) = \mathbb{P}\left((\nu_{\sigma(1)}, \ldots, \nu_{\sigma(N)}) = (n_1, \ldots, n_N)\right)$$

  for any permutation σ of $\{1, \ldots, N\}$.
- Unlike the Kingman coalescent, when we look at more general offspring distributions, we can have multiple mergers (Λ-coalescents) and simultaneous mergers (Ξ-coalescents)
- Understanding the properties of these coalescents is a topic of ongoing research.

Λ- and Ξ-Coalescents

- In the neutral ($s = 0$) Wright-Fisher model, we assumed that the vector of offspring numbers for parent was a Multinomial $\left(N, \frac{1}{N}\right)$ variable.
- We could equally well have considered more general exchangeable random vectors : $(\nu_1, \ldots, \nu_N)$ is exchangeable if

$$\mathbb{P}\left((\nu_1, \ldots, \nu_N) = (n_1, \ldots, n_N)\right) = \mathbb{P}\left((\nu_{\sigma(1)}, \ldots, \nu_{\sigma(N)}) = (n_1, \ldots, n_N)\right)$$

  for any permutation σ of $\{1, \ldots, N\}$.
- Unlike the Kingman coalescent, when we look at more general offspring distributions, we can have multiple mergers (Λ-coalescents) and simultaneous mergers (Ξ-coalescents)
- Understanding the properties of these coalescents is a topic of ongoing research.

The Ancestral Selection Graph

- Kingman's coalescent (and Λ- and Ξ-coalescents) all assume that all types are essentially equivalent.
- What about the $s \neq 0$ case ?
  - This is more complicated because now some parents are more likely than others, but just looking at the sequence data, we don't know *a priori* which ones.
- One approach is the *ancestral selection graph*, which introduces a complicated graph "potential ancestors", which, by tracing and forward in time, can be resolved into a genealogy under selection.
- This, however, is only adequate for two types. To find a theory with infinite alleles and selection is an important open problem !

The Ancestral Selection Graph

- Kingman's coalescent (and Λ- and Ξ-coalescents) all assume that all types are essentially equivalent.
- What about the $s \neq 0$ case ?
  - This is more complicated because now some parents are more likely than others, but just looking at the sequence data, we don't know *a priori* which ones.
- One approach is the *ancestral selection graph*, which introduces a complicated graph "potential ancestors", which, by tracing and forward in time, can be resolved into a genealogy under selection.
- This, however, is only adequate for two types. To find a theory with infinite alleles and selection is an important open problem !

The Ancestral Selection Graph

- Kingman's coalescent (and Λ- and Ξ-coalescents) all assume that all types are essentially equivalent.
- What about the $s \neq 0$ case ?
  - This is more complicated because now some parents are more likely than others, but just looking at the sequence data, we don't know *a priori* which ones.
- One approach is the *ancestral selection graph*, which introduces a complicated graph "potential ancestors", which, by tracing and forward in time, can be resolved into a genealogy under selection.
- This, however, is only adequate for two types. To find a theory with infinite alleles and selection is an important open problem !

The Ancestral Selection Graph

- Kingman's coalescent (and Λ- and Ξ-coalescents) all assume that all types are essentially equivalent.
- What about the $s \neq 0$ case ?
  - This is more complicated because now some parents are more likely than others, but just looking at the sequence data, we don't know *a priori* which ones.
- One approach is the *ancestral selection graph*, which introduces a complicated graph "potential ancestors", which, by tracing and forward in time, can be resolved into a genealogy under selection.
- This, however, is only adequate for two types. To find a theory with infinite alleles and selection is an important open problem !

The Ancestral Selection Graph

- Kingman's coalescent (and Λ- and Ξ-coalescents) all assume that all types are essentially equivalent.
- What about the $s \neq 0$ case ?
  - This is more complicated because now some parents are more likely than others, but just looking at the sequence data, we don't know *a priori* which ones.
- One approach is the *ancestral selection graph*, which introduces a complicated graph "potential ancestors", which, by tracing and forward in time, can be resolved into a genealogy under selection.
- This, however, is only adequate for two types. To find a theory with infinite alleles and selection is an important open problem !

Space. . . The Final(?) Frontier

- In our models, offspring picked their parent at random from all possible parents.
- In the real world, spatial separation often makes this impossible !
- Starting in the 1950's, people began to consider spatial population models, first on lattices, and then in continuous space.
- The Fleming-Viot model arose from considering the continuum limit of lattice models.
- Mathematical problems plagued most of these efforts : the limiting equations are ill-posed, or blow up in finite time, or making sense of a genealogy becomes impossible.
- The spatial Λ-Fleming Viot model is a recent, important step forward.
- Incorporating selection is again an important topic of current research.

Space. . . The Final(?) Frontier

- In our models, offspring picked their parent at random from all possible parents.
- In the real world, spatial separation often makes this impossible !
- Starting in the 1950's, people began to consider spatial population models, first on lattices, and then in continuous space.
- The Fleming-Viot model arose from considering the continuum limit of lattice models.
- Mathematical problems plagued most of these efforts : the limiting equations are ill-posed, or blow up in finite time, or making sense of a genealogy becomes impossible.
- The spatial Λ-Fleming Viot model is a recent, important step forward.
- Incorporating selection is again an important topic of current research.

Space. . . The Final(?) Frontier

- In our models, offspring picked their parent at random from all possible parents.
- In the real world, spatial separation often makes this impossible !
- Starting in the 1950's, people began to consider spatial population models, first on lattices, and then in continuous space.
- The Fleming-Viot model arose from considering the continuum limit of lattice models.
- Mathematical problems plagued most of these efforts : the limiting equations are ill-posed, or blow up in finite time, or making sense of a genealogy becomes impossible.
- The spatial Λ-Fleming Viot model is a recent, important step forward.
- Incorporating selection is again an important topic of current research.

Space. . . The Final(?) Frontier

- In our models, offspring picked their parent at random from all possible parents.
- In the real world, spatial separation often makes this impossible !
- Starting in the 1950's, people began to consider spatial population models, first on lattices, and then in continuous space.
- The Fleming-Viot model arose from considering the continuum limit of lattice models.
- Mathematical problems plagued most of these efforts : the limiting equations are ill-posed, or blow up in finite time, or making sense of a genealogy becomes impossible.
- The spatial Λ-Fleming Viot model is a recent, important step forward.
- Incorporating selection is again an important topic of current research.

Space. . . The Final(?) Frontier

- In our models, offspring picked their parent at random from all possible parents.
- In the real world, spatial separation often makes this impossible !
- Starting in the 1950's, people began to consider spatial population models, first on lattices, and then in continuous space.
- The Fleming-Viot model arose from considering the continuum limit of lattice models.
- Mathematical problems plagued most of these efforts : the limiting equations are ill-posed, or blow up in finite time, or making sense of a genealogy becomes impossible.
- The spatial Λ-Fleming Viot model is a recent, important step forward.
- Incorporating selection is again an important topic of current research.

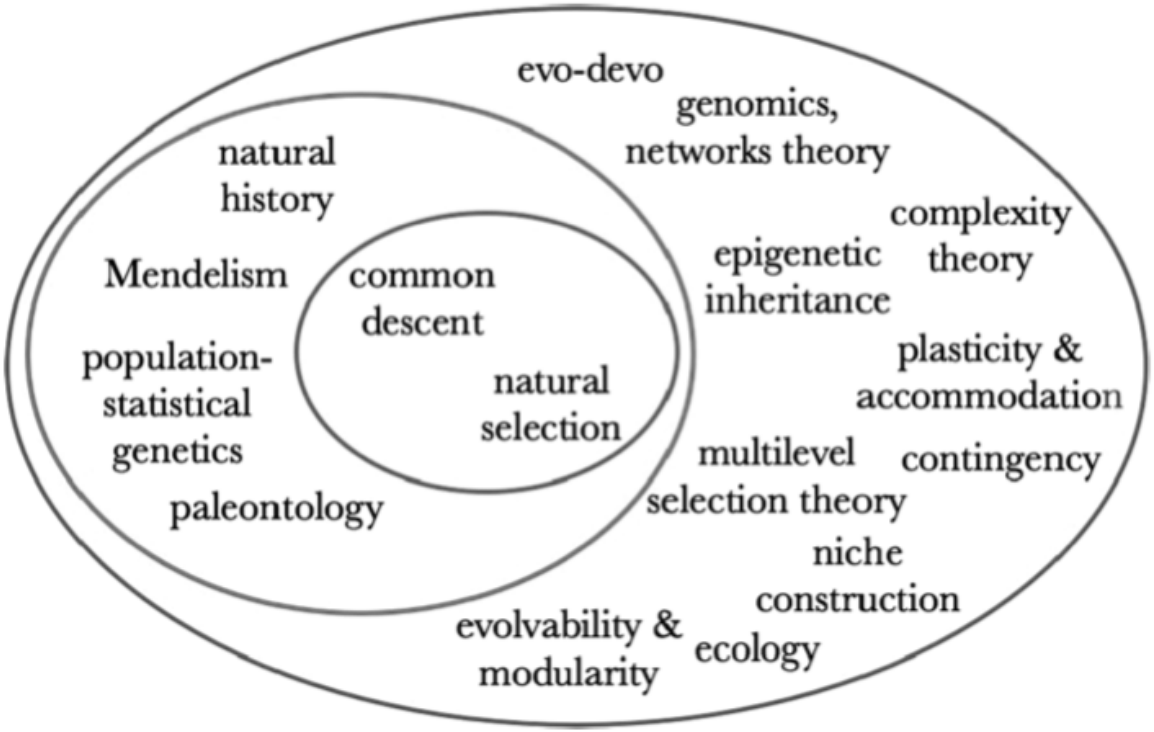
Space. . . The Final(?) Frontier

- In our models, offspring picked their parent at random from all possible parents.
- In the real world, spatial separation often makes this impossible !
- Starting in the 1950's, people began to consider spatial population models, first on lattices, and then in continuous space.
- The Fleming-Viot model arose from considering the continuum limit of lattice models.
- Mathematical problems plagued most of these efforts : the limiting equations are ill-posed, or blow up in finite time, or making sense of a genealogy becomes impossible.
- The spatial Λ-Fleming Viot model is a recent, important step forward.
- Incorporating selection is again an important topic of current research.

Space. . . The Final(?) Frontier

- In our models, offspring picked their parent at random from all possible parents.
- In the real world, spatial separation often makes this impossible !
- Starting in the 1950's, people began to consider spatial population models, first on lattices, and then in continuous space.
- The Fleming-Viot model arose from considering the continuum limit of lattice models.
- Mathematical problems plagued most of these efforts : the limiting equations are ill-posed, or blow up in finite time, or making sense of a genealogy becomes impossible.
- The spatial Λ-Fleming Viot model is a recent, important step forward.
- Incorporating selection is again an important topic of current research.

An Extended Synthesis for Evolutionary Biology ?



Pigliucci, *Ann. N.Y. Acad. Sci.* 1168 : 218–228 (2009).

Finally, much like Fisher, Haldane and Wright synthesised the (then) new theory of Mendelian inheritance with Darwin's theory of evolution to create population genetics, there's a need today for mathematicians to help devise theories to unite population genetics with new discoveries in biology.

Thank you for the invitation !

Any questions ?

Fixation Probability

- Not easily, but we can find a good approximation.
- Let's look at the number of offspring of a *single* mutant. As before, it's binomial distributed, with the probability of x offspring equal to

$$\binom{N}{x}\left(\frac{(1+s)}{sx_n+N}\right)^x\left(1-\frac{(1+s)}{sx_n+N}\right)^{N-x} \sim e^{-(1+s)}\frac{(1+s)^x}{x!}$$

  for large N.
- The number of mutants is approximated by a branching process with a Poisson$(1+s)$ offspring law.
- Let p_e be the probability that the mutants go extinct (in the branching process, starting from a single mutant). Then,

$$p_e = \sum_{x=0}^{\infty} e^{-(1+s)}\frac{(1+s)^x}{x!}p_e^x \approx 1-(1+s)(1-p_e)+\frac{1}{2}(1+s)^2(1-p_e)^2,$$

  which can be solved to give $p_e = 1-2s$.
- The mutant persists (and eventually takes over) with probability $2s$.

Fixation Probability

- Not easily, but we can find a good approximation.
- Let's look at the number of offspring of a *single* mutant. As before, it's binomial distributed, with the probability of x offspring equal to

$$\binom{N}{x}\left(\frac{(1+s)}{sx_n+N}\right)^x\left(1-\frac{(1+s)}{sx_n+N}\right)^{N-x}\sim e^{-(1+s)}\frac{(1+s)^x}{x!}$$

  for large N.
- The number of mutants is approximated by a branching process with a Poisson$(1+s)$ offspring law.
- Let p_e be the probability that the mutants go extinct (in the branching process, starting from a single mutant). Then,

$$p_e=\sum_{x=0}^{\infty}e^{-(1+s)}\frac{(1+s)^x}{x!}p_e^x\approx 1-(1+s)(1-p_e)+\frac{1}{2}(1+s)^2(1-p_e)^2,$$

  which can be solved to give $p_e=1-2s$.
- The mutant persists (and eventually takes over) with probability $2s$.

Fixation Probability

- Not easily, but we can find a good approximation.
- Let's look at the number of offspring of a *single* mutant. As before, it's binomial distributed, with the probability of x offspring equal to

$$\binom{N}{x}\left(\frac{(1+s)}{sx_n+N}\right)^x\left(1-\frac{(1+s)}{sx_n+N}\right)^{N-x}\sim e^{-(1+s)}\frac{(1+s)^x}{x!}$$

  for large N.
- The number of mutants is approximated by a branching process with a Poisson$(1+s)$ offspring law.
- Let p_e be the probability that the mutants go extinct (in the branching process, starting from a single mutant). Then,

$$p_e=\sum_{x=0}^{\infty}e^{-(1+s)}\frac{(1+s)^x}{x!}p_e^x\approx 1-(1+s)(1-p_e)+\frac{1}{2}(1+s)^2(1-p_e)^2,$$

  which can be solved to give $p_e=1-2s$.
- The mutant persists (and eventually takes over) with probability $2s$.

Fixation Probability

- Not easily, but we can find a good approximation.
- Let's look at the number of offspring of a *single* mutant. As before, it's binomial distributed, with the probability of x offspring equal to

$$\binom{N}{x}\left(\frac{(1+s)}{sx_n+N}\right)^x\left(1-\frac{(1+s)}{sx_n+N}\right)^{N-x}\sim e^{-(1+s)}\frac{(1+s)^x}{x!}$$

  for large N.
- The number of mutants is approximated by a branching process with a Poisson$(1+s)$ offspring law.
- Let p_e be the probability that the mutants go extinct (in the branching process, starting from a single mutant). Then,

$$p_e=\sum_{x=0}^{\infty}e^{-(1+s)}\frac{(1+s)^x}{x!}p_e^x\approx 1-(1+s)(1-p_e)+\frac{1}{2}(1+s)^2(1-p_e)^2,$$

  which can be solved to give $p_e=1-2s$.
- The mutant persists (and eventually takes over) with probability $2s$.

Fixation Probability

- Not easily, but we can find a good approximation.
- Let's look at the number of offspring of a *single* mutant. As before, it's binomial distributed, with the probability of x offspring equal to

$$\binom{N}{x}\left(\frac{(1+s)}{sx_n+N}\right)^x\left(1-\frac{(1+s)}{sx_n+N}\right)^{N-x}\sim e^{-(1+s)}\frac{(1+s)^x}{x!}$$

  for large N.
- The number of mutants is approximated by a branching process with a Poisson$(1+s)$ offspring law.
- Let p_e be the probability that the mutants go extinct (in the branching process, starting from a single mutant). Then,

$$p_e=\sum_{x=0}^{\infty}e^{-(1+s)}\frac{(1+s)^x}{x!}p_e^x\approx 1-(1+s)(1-p_e)+\frac{1}{2}(1+s)^2(1-p_e)^2,$$

  which can be solved to give $p_e=1-2s$.
- The mutant persists (and eventually takes over) with probability $2s$.

Fixation Time (branching process approximation)

- Each mutant in the n^{th} generation produces a Poisson$(1+s)$ number of offspring, say ξ_n^i, so $X_{n+1} = \sum_{i=1}^{X_n} \xi_n^i$.
- Conditional on X_n, this has expectation

$$\mathbb{E}[X_{n+1}|X_n] = \sum_{i=1}^{X_n} \mathbb{E}\left[\xi_n^i | X_n\right] = \sum_{i=1}^{X_n} (1+s) = (1+s)X_n$$

- Thus, $\mathbb{E}[X_{n+1}] = (1+s)\mathbb{E}[X_n]$, so $\mathbb{E}[X_n] = (1+s)^n X_0 = (1+s)^n$.
- Conditioning on non-extinction is equivalent to dividing by $2s$:

$$\begin{aligned}\mathbb{E}[X_n] = \mathbb{E}[X_n|\text{non-extincton}]\,\mathbb{P}(\text{non-extincton}) \\ + \mathbb{E}[X_n|\text{extincton}]\,\mathbb{P}(\text{extincton}) \approx 2s\mathbb{E}[X_n|\text{non-extincton}]\end{aligned}$$

- Let n_N be the first time (conditioned on non-extinction) that $X_n = N$:

$$N = X_{n_N} \approx \mathbb{E}[X_n|\text{non-extincton}] \approx \frac{1}{2s}\mathbb{E}[X_{n_N}] = \frac{1}{2s}(1+s)^{n_N},$$

so $n_N \approx \frac{1}{\ln(1+s)} \ln(2Ns) \approx \frac{1}{s} \ln(2Ns)$.

Fixation Time (branching process approximation)

- Each mutant in the n^{th} generation produces a Poisson$(1+s)$ number of offspring, say ξ_n^i, so $X_{n+1} = \sum_{i=1}^{X_n} \xi_n^i$.
- Conditional on X_n, this has expectation

$$\mathbb{E}[X_{n+1}|X_n] = \sum_{i=1}^{X_n} \mathbb{E}\left[\xi_n^i \middle| X_n\right] = \sum_{i=1}^{X_n} (1+s) = (1+s)X_n$$

- Thus, $\mathbb{E}[X_{n+1}] = (1+s)\mathbb{E}[X_n]$, so $\mathbb{E}[X_n] = (1+s)^n X_0 = (1+s)^n$.
- Conditioning on non-extinction is equivalent to dividing by $2s$:

$$\begin{aligned}\mathbb{E}[X_n] = \mathbb{E}[X_n|\text{non-extincton}]\,\mathbb{P}(\text{non-extincton}) \\ + \mathbb{E}[X_n|\text{extincton}]\,\mathbb{P}(\text{extincton}) \approx 2s\mathbb{E}[X_n|\text{non-extincton}]\end{aligned}$$

- Let n_N be the first time (conditioned on non-extinction) that $X_n = N$:

$$N = X_{n_N} \approx \mathbb{E}[X_n|\text{non-extincton}] \approx \frac{1}{2s}\mathbb{E}[X_{n_N}] = \frac{1}{2s}(1+s)^{n_N},$$

so $n_N \approx \frac{1}{\ln(1+s)} \ln(2Ns) \approx \frac{1}{s} \ln(2Ns)$.

Fixation Time (branching process approximation)

- Each mutant in the n^{th} generation produces a Poisson$(1+s)$ number of offspring, say ξ_n^i, so $X_{n+1} = \sum_{i=1}^{X_n} \xi_n^i$.
- Conditional on X_n, this has expectation

$$\mathbb{E}[X_{n+1}|X_n] = \sum_{i=1}^{X_n} \mathbb{E}\left[\xi_n^i \middle| X_n\right] = \sum_{i=1}^{X_n} (1+s) = (1+s)X_n$$

- Thus, $\mathbb{E}[X_{n+1}] = (1+s)\mathbb{E}[X_n]$, so $\mathbb{E}[X_n] = (1+s)^n X_0 = (1+s)^n$.
- Conditioning on non-extinction is equivalent to dividing by $2s$:

$$\mathbb{E}[X_n] = \mathbb{E}[X_n|\text{non-extincton}]\,\mathbb{P}(\text{non-extincton}) + \mathbb{E}[X_n|\text{extincton}]\,\mathbb{P}(\text{extincton}) \approx 2s\mathbb{E}[X_n|\text{non-extincton}]$$

- Let n_N be the first time (conditioned on non-extinction) that $X_n = N$:

$$N = X_{n_N} \approx \mathbb{E}[X_n|\text{non-extincton}] \approx \frac{1}{2s}\mathbb{E}[X_{n_N}] = \frac{1}{2s}(1+s)^{n_N},$$

so $n_N \approx \frac{1}{\ln(1+s)} \ln(2Ns) \approx \frac{1}{s} \ln(2Ns)$.

Fixation Time (branching process approximation)

- Each mutant in the n^{th} generation produces a Poisson$(1+s)$ number of offspring, say ξ_n^i, so $X_{n+1} = \sum_{i=1}^{X_n} \xi_n^i$.
- Conditional on X_n, this has expectation

$$\mathbb{E}[X_{n+1}|X_n] = \sum_{i=1}^{X_n} \mathbb{E}\left[\xi_n^i \middle| X_n\right] = \sum_{i=1}^{X_n} (1+s) = (1+s)X_n$$

- Thus, $\mathbb{E}[X_{n+1}] = (1+s)\mathbb{E}[X_n]$, so $\mathbb{E}[X_n] = (1+s)^n X_0 = (1+s)^n$.
- Conditioning on non-extinction is equivalent to dividing by $2s$:

$$\begin{aligned}\mathbb{E}[X_n] = \mathbb{E}[X_n|\text{non-extincton}]\,\mathbb{P}(\text{non-extincton}) \\ + \mathbb{E}[X_n|\text{extincton}]\,\mathbb{P}(\text{extincton}) \approx 2s\mathbb{E}[X_n|\text{non-extincton}]\end{aligned}$$

- Let n_N be the first time (conditioned on non-extinction) that $X_n = N$:

$$N = X_{n_N} \approx \mathbb{E}[X_n|\text{non-extincton}] \approx \frac{1}{2s}\mathbb{E}[X_{n_N}] = \frac{1}{2s}(1+s)^{n_N},$$

so $n_N \approx \frac{1}{\ln(1+s)} \ln(2Ns) \approx \frac{1}{s} \ln(2Ns)$.

Fixation Time (branching process approximation)

- Each mutant in the n^{th} generation produces a Poisson$(1+s)$ number of offspring, say ξ_n^i, so $X_{n+1} = \sum_{i=1}^{X_n} \xi_n^i$.
- Conditional on X_n, this has expectation

$$\mathbb{E}[X_{n+1}|X_n] = \sum_{i=1}^{X_n} \mathbb{E}\left[\xi_n^i \middle| X_n\right] = \sum_{i=1}^{X_n} (1+s) = (1+s)X_n$$

- Thus, $\mathbb{E}[X_{n+1}] = (1+s)\mathbb{E}[X_n]$, so $\mathbb{E}[X_n] = (1+s)^n X_0 = (1+s)^n$.
- Conditioning on non-extinction is equivalent to dividing by $2s$:

$$\begin{aligned}\mathbb{E}[X_n] = \mathbb{E}[X_n|\text{non-extincton}]\,\mathbb{P}(\text{non-extincton}) \\ + \mathbb{E}[X_n|\text{extincton}]\,\mathbb{P}(\text{extincton}) \approx 2s\mathbb{E}[X_n|\text{non-extincton}]\end{aligned}$$

- Let n_N be the first time (conditioned on non-extinction) that $X_n = N$:

$$N = X_{n_N} \approx \mathbb{E}[X_n|\text{non-extincton}] \approx \frac{1}{2s}\mathbb{E}[X_{n_N}] = \frac{1}{2s}(1+s)^{n_N},$$

so $n_N \approx \frac{1}{\ln(1+s)} \ln(2Ns) \approx \frac{1}{s} \ln(2Ns)$.