Simple models of evolution with selection and genealogies

Bernard Derrida

ENS, Paris

E. Brunet 2011-2012

E. Brunet, A. H. Mueller, S. Munier 2006-2007

Seoul July 2012

Outline

Coalescent processes

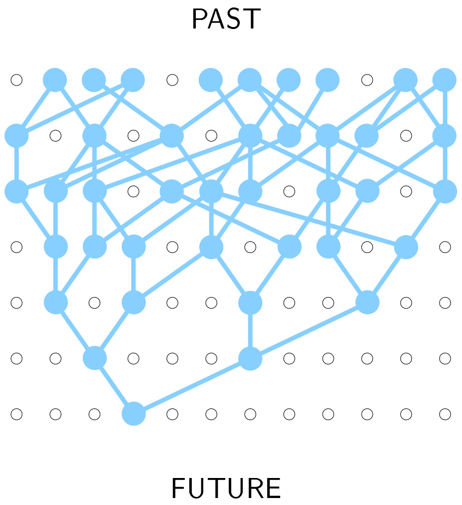
Models of evolution without or with selection

The exponential model

The generic case

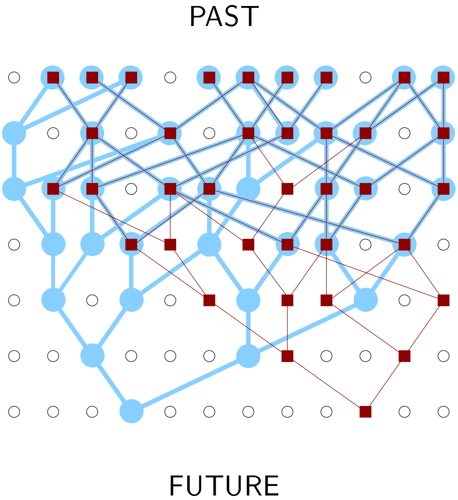
Genealogies with or without selection

Sexual Reproduction



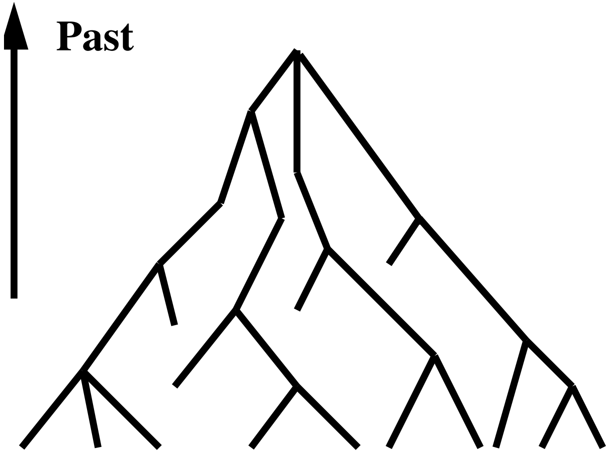
Genealogies with or without selection

Sexual Reproduction



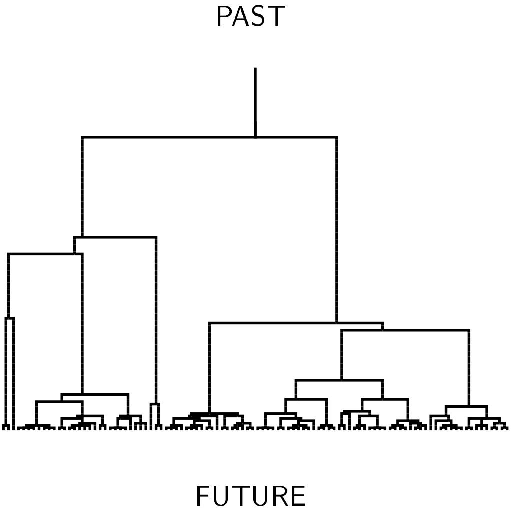
Genealogies with or without selection

Asexual Reproduction



Genealogies with or without selection

Asexual Reproduction



How to analyze the statistics of these random trees

At generation g, take k individuals $\alpha_1, \alpha_2, ... \alpha_k$ in the population

$$\boxed{T_k(\alpha_1, ... \alpha_k) = \text{age of their most recent common ancestor}}$$

How to analyze the statistics of these random trees

At generation g, take k individuals $\alpha_1, \alpha_2, ...\alpha_k$ in the population

$$\boxed{T_k(\alpha_1, ...\alpha_k) = \text{age of their most recent common ancestor}}$$

Coalescence times

$$\boxed{T_k = \langle T_k(\alpha_1, ...\alpha_k) \rangle_{\alpha_1, ...\alpha_k}}$$

How to analyze the statistics of these random trees

At generation g, take k individuals $\alpha_1, \alpha_2, ...\alpha_k$ in the population

$$\boxed{T_k(\alpha_1, ...\alpha_k) = \text{age of their most recent common ancestor}}$$

Coalescence times

$$\boxed{T_k = \langle T_k(\alpha_1, ...\alpha_k) \rangle_{\alpha_1, ...\alpha_k}}$$

Size of the τ families (analogy with spin glasses)

Choose k individuals $\alpha_1, \alpha_2, ...\alpha_k$ at random at generation g

$$\boxed{Y_k(\tau) = \text{probability that } T_k(\alpha_1, ...\alpha_k) < \tau}$$

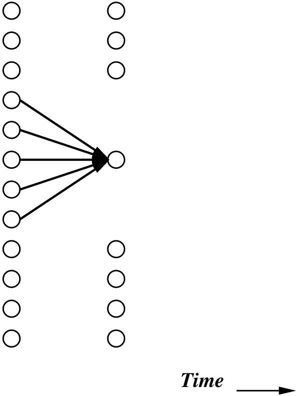
$Y_k(\tau)$ and T_k are non-self averaging quantities $\Rightarrow$

one measures or calculates averages of $Y_k(\tau)$ or T_k over generations

Coalescent processes

Take a large number N of points

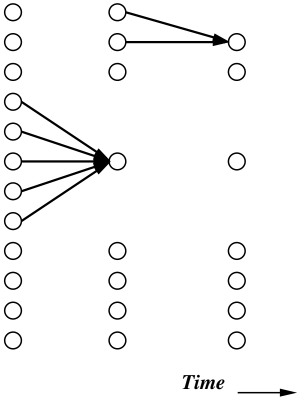
During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$



Coalescent processes

Take a large number N of points

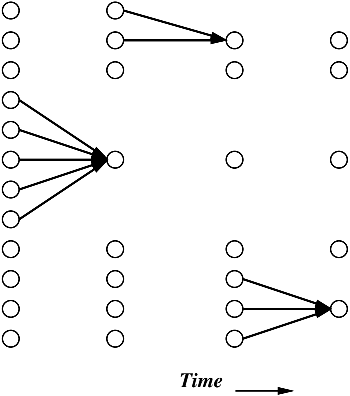
During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$



Coalescent processes

Take a large number N of points

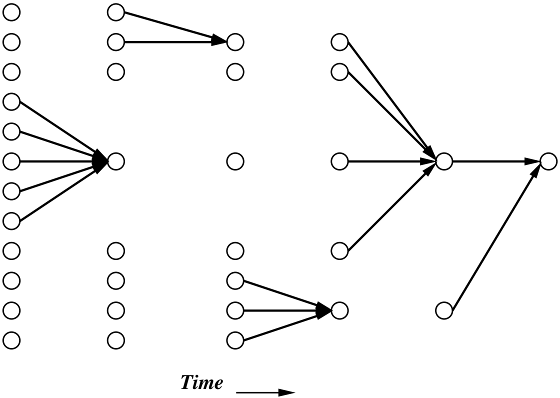
During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$



Coalescent processes

Take a large number N of points

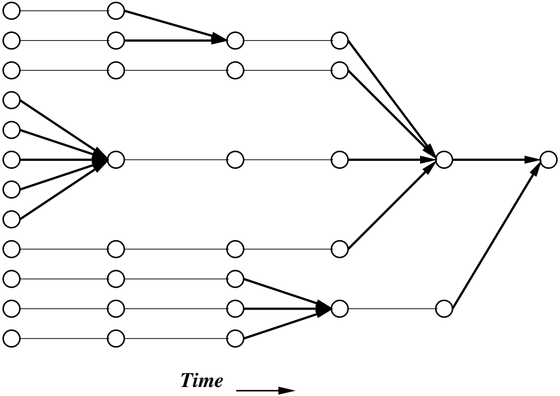
During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$



Coalescent processes

Take a large number N of points

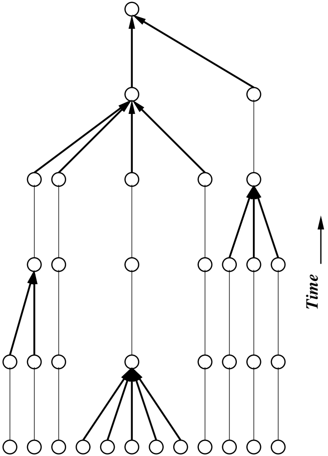
During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$



Coalescent processes

Take a large number N of points

During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$



Coalescent processes

Take a large number N of points

During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$

$$\langle T_2 \rangle = \frac{1}{q_2}, \qquad \frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{4q_2 - 3q_3}{3q_2 - 2q_3},$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{27q_2^2 - 56q_2q_3 + 28q_3^2 + 12q_2q_4 - 10q_3q_4}{(3q_2 - 2q_3)(6q_2 - 8q_3 + 3q_4)}$$

Coalescent processes

Take a large number N of points

During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$

$$\langle T_2 \rangle = \frac{1}{q_2}, \qquad \frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{4q_2 - 3q_3}{3q_2 - 2q_3},$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{27q_2^2 - 56q_2q_3 + 28q_3^2 + 12q_2q_4 - 10q_3q_4}{(3q_2 - 2q_3)(6q_2 - 8q_3 + 3q_4)}$$

Kingman: $q_k = 0$ for all $k > 2$

$$\langle T_2 \rangle = \frac{1}{q_2}, \qquad \frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{4}{3}, \qquad \frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{3}{2}.$$

Coalescent processes

Take a large number N of points

During $dt \ll 1$, each group of k points may coalesce into a single point with probability $q_k dt$

$$\langle T_2 \rangle = \frac{1}{q_2}, \qquad \frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{4q_2 - 3q_3}{3q_2 - 2q_3},$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{27q_2^2 - 56q_2q_3 + 28q_3^2 + 12q_2q_4 - 10q_3q_4}{(3q_2 - 2q_3)(6q_2 - 8q_3 + 3q_4)}$$

Kingman: $q_k = 0$ for all $k > 2$

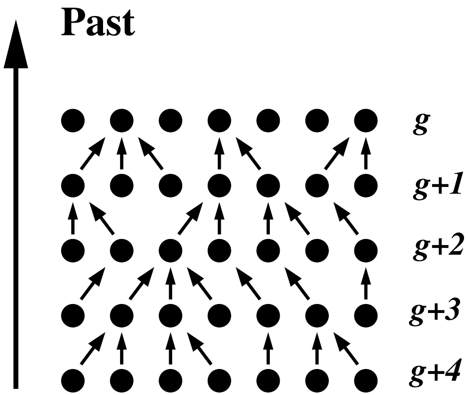
$$\langle T_2 \rangle = \frac{1}{q_2}, \qquad \frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{4}{3}, \qquad \frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{3}{2}.$$

Bolthausen-Sznitman: $q_k = q_2/(k-1)$ for all k

$$\langle T_2 \rangle = \frac{1}{q_2}, \qquad \frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{5}{4}, \qquad \frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{25}{18}.$$

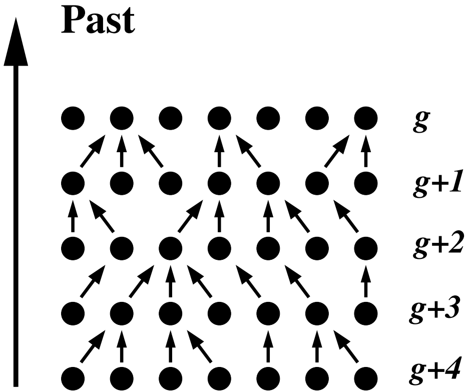
Wright-Fisher model (1930-1931)

- One parent model (asexual reproduction)
- Population of fixed size N
- Each individual has its parent chosen at random in the previous generation (neutrality)



Wright-Fisher model (1930-1931)

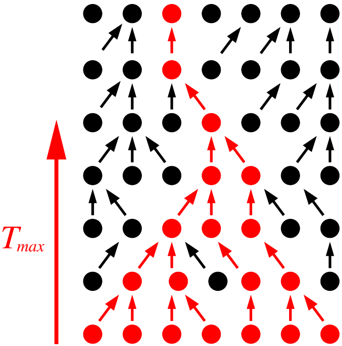
- One parent model (asexual reproduction)
- Population of fixed size N
- Each individual has its parent chosen at random in the previous generation (neutrality)



$$q_2 = \frac{1}{N} \quad ; \quad q_3 = \frac{1}{N^2} \quad \dots \quad q_k = \frac{1}{N^{k-1}}$$

Coalescence times:

Age of the most recent common ancestor T_{max}

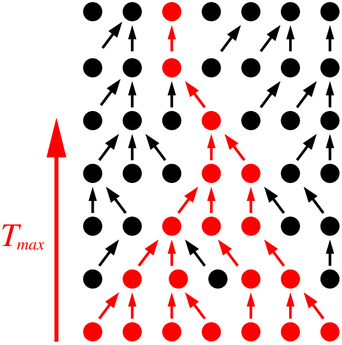


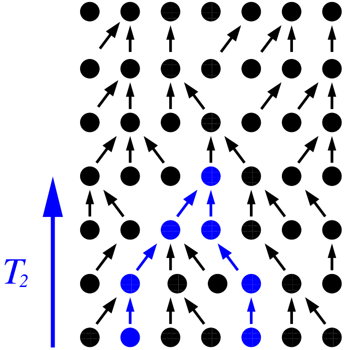
$$T_{max} \sim N$$

T_{max} is a non self-averaging quantity

Coalescence times:

Ages of the most recent common ancestors T_{max} and T_2





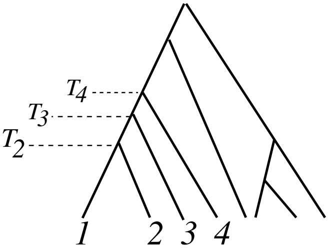
$$T_{max} \sim T_2 \sim N$$

T_{max} and T_2 are non self-averaging quantities

Coalescence times:

Age T_p Kingman theory

T_p = age of the most recent common ancestor of p individuals chosen at random

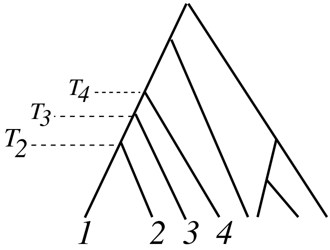


$$\langle T_p \rangle \simeq \frac{2(p-1)}{p} N$$

Coalescence times:

Age T_p Kingman theory

T_p = age of the most recent common ancestor of p individuals chosen at random

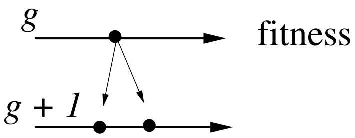


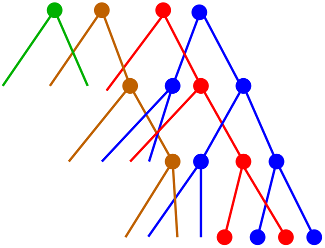
$$\langle T_p \rangle \simeq \frac{2(p-1)}{p} N$$

Kingman theory is Universal

MODELS OF EVOLUTION WITH SELECTION

- Population of size N
- Each individual has 2 offspring at the next generation
- The fitness is transmitted up to some small change due to mutations
- The N right-most individuals are selected

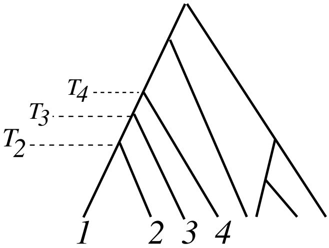




QUESTIONS

For a population of fixed size N

- Speed of evolution
- Ages of the most recent common ancestors



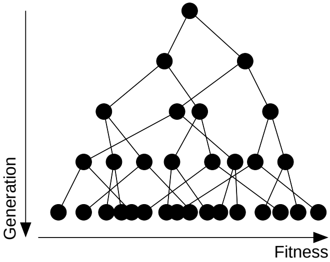
- Shape of the genealogical trees

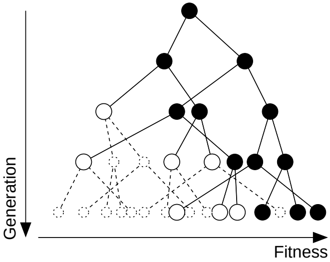
Example: branching random walk with selection

At each generation, only N individuals survive after selection

Example: branching random walk with selection

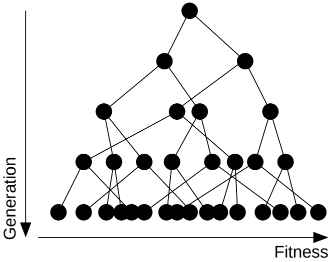
At each generation, only N individuals survive after selection





Example: branching random walk with selection

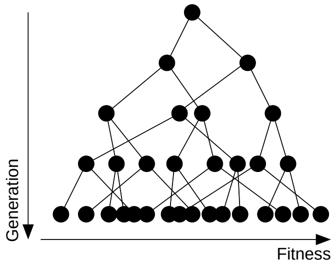
At each generation, only N individuals survive after selection



Genealogy given by the Bolthausen Sznitman

Example: branching random walk in the neutral case

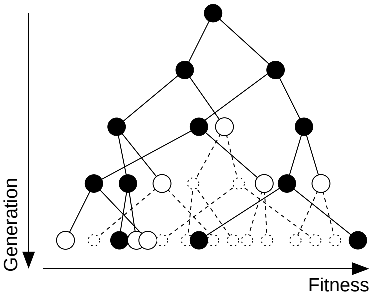
At each generation, N individuals survive, chosen at random

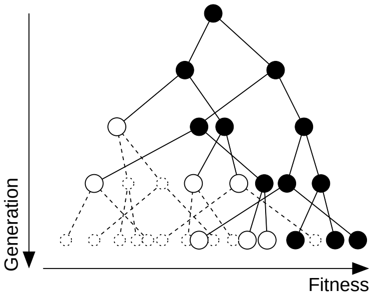


Genealogy given by the Kingman coalescent

Example: branching random walk

$N = 3$





NEUTRAL

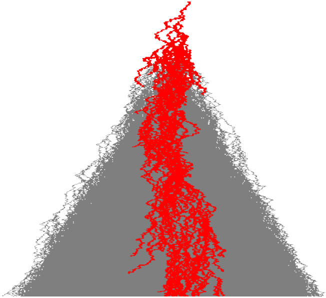
Kingman

SELECTION

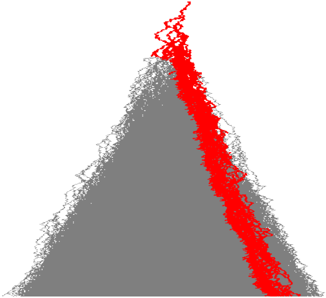
Bolthausen Sznitman

Neutral versus Selection

Size of the population $N = 25$



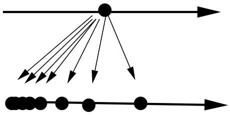
chosen at random



the fittest

Exponential model

- Population of size N
- An individual at position x has an offspring in $(x+y, x+y+dy)$ with probability $e^{-y}dy$ (Poisson process).



- The N right-most individuals are selected

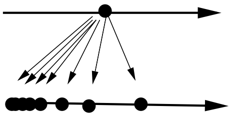
Brunet D. Mueller Munier 2006-2007

$$\frac{\langle X_t \rangle}{t} \simeq \log \log N + ...$$

$$\frac{\langle X_t^2 \rangle - \langle X_t \rangle^2}{t} \sim \frac{1}{\log N}$$

Exponential model

- Population of size N
- An individual at position x has an offspring in $(x+y, x+y+dy)$ with probability $e^{-y}dy$ (Poisson process).



- The N right-most individuals are selected

Brunet D. Mueller Munier 2006-2007

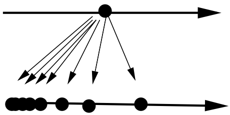
$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} \to \frac{5}{4} \neq \frac{4}{3}$$

$$\langle T_2 \rangle \simeq \log N \neq N$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} \to \frac{25}{18} \neq \frac{3}{2}$$

Exponential model

- Population of size N
- An individual at position x has an offspring in $(x+y, x+y+dy)$ with probability $e^{-y}dy$ (Poisson process).



- The N right-most individuals are selected

Brunet D. Mueller Munier 2006-2007

$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} \to \frac{5}{4} \neq \frac{4}{3}$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} \to \frac{25}{18} \neq \frac{3}{2}$$

$$\langle T_2 \rangle \simeq \log N \neq N$$

Bolthausen Sznitman $\neq$ Kingman

Statistics of the trees

	spin–glass	neutral
	$\frac{3}{4}$	1
	$\frac{1}{4}$	0

spin-glass $\equiv$ mean-field spin glasses

Parisi 79-80

Mézard-Parisi-Sourlas-Toulouse-Virasoro 84

… Bolthausen-Sznitman 98

neutral $\equiv$ Wright-Fisher model

	spin–glass	neutral
	$\frac{1}{3}$	$\frac{2}{3}$
	$\frac{1}{6}$	$\frac{1}{3}$
	$\frac{1}{6}$	0
	$\frac{2}{9}$	0
	$\frac{1}{9}$	0

Conditionning on the speed in the exponential model

Brunet D. 2011

X_t position of the population at time t

Weight the events by $e^{-\beta X_t}$

Conditionning on the speed in the exponential model

Brunet D. 2011

X_t position of the population at time t

Weight the events by $e^{-\beta X_t}$

Then

$$\left.\frac{\langle T_3\rangle}{\langle T_2\rangle}\right|_\beta = \frac{5+4\beta}{4+3\beta}$$

$$\left.\frac{\langle T_4\rangle}{\langle T_2\rangle}\right|_\beta = \frac{100+204\beta+133\beta^2+27\beta^3}{72+142\beta+90\beta^2+18\beta^3}$$

Conditionning on the speed in the exponential model

Brunet D. 2011

X_t position of the population at time t

Weight the events by $e^{-\beta X_t}$

Then

$$\left.\frac{\langle T_3\rangle}{\langle T_2\rangle}\right|_\beta = \frac{5+4\beta}{4+3\beta}$$

$$\left.\frac{\langle T_4\rangle}{\langle T_2\rangle}\right|_\beta = \frac{100+204\beta+133\beta^2+27\beta^3}{72+142\beta+90\beta^2+18\beta^3}$$

Genealogies with $\beta \neq 0$ $\neq$ Spin Glass Trees with $n_{\text{replicas}} \neq 0$

Heart of the calculation

q_p rate at which p branches coalesce into 1.

N random variables y_i

$$P(y) = e^{-y} \quad \text{for} \quad y > 0$$

and then

$$q_p = N \frac{\langle e^{py_1} Z^{\beta - p} \rangle}{\langle Z^\beta \rangle}$$

where

$$Z = \sum_{i=1}^{N} e^{y_i}$$

Coalescence rates

q_p rate at which p branches coalesce into 1.

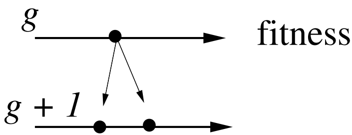
Coalescence rates

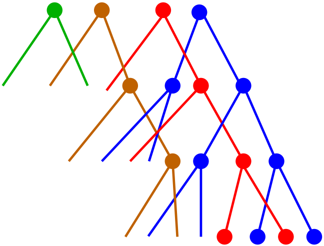
q_p rate at which p branches coalesce into 1.

	q_p/q_2 for $p > 2$
neutral	0
selection	$\frac{1}{p-1}$
selection + weight $e^{-\beta X_t}$	$\frac{(p-2)!\ \Gamma(\beta+2)}{\Gamma(\beta+p)}$

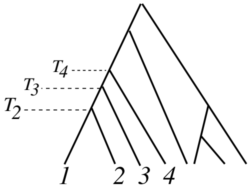
MODELS OF EVOLUTION WITH SELECTION

- Population of size N
- Each individual has 2 offspring at the next generation
- The fitness is transmitted up to some small change due to mutations
- The N right-most individuals are selected





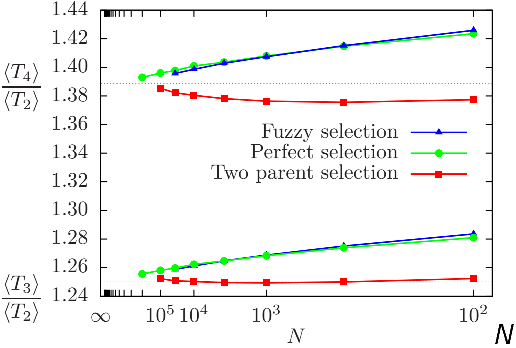
Simulations $N \to 10^5$ in the generic model



T_p = age of the most common ancestor of p individuals chosen at random

$$T_p \sim \log^3 N$$

Ratios $\frac{\langle T_p \rangle}{\langle T_2 \rangle}$



$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} \to \frac{5}{4} \neq \frac{4}{3}$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} \to \frac{25}{18} \neq \frac{3}{2}$$

selection $\neq$ neutral

Genealogies with $\beta \neq 0$ $\neq$ Spin Glass Trees with n replicas $\neq 0$

In the replica theory (Parisi)

$$\langle Y_k \rangle_\mu = \frac{(1-\mu)(2-\mu)\cdots(k-1-\mu)}{(1-n)(2-n)\cdots(k-1-n)}$$

Genealogies with $\beta \neq 0$ $\neq$ Spin Glass Trees with n replicas $\neq 0$

In the replica theory (Parisi)

$$\langle Y_k \rangle_\mu = \frac{(1-\mu)(2-\mu)\cdots(k-1-\mu)}{(1-n)(2-n)\cdots(k-1-n)}$$

In the coalescence process

$Y_k(\tau)$= probability that $T_k(\alpha_1, ...\alpha_k) < \tau$

$$\langle Y_2(\tau) \rangle = 1 - e^{-q_2 \tau} \quad ; \quad \langle Y_3(\tau) \rangle = 1 - \frac{3}{2} e^{-q_2\tau} + \frac{1}{2} e^{-(3q_2 - 2q_3)\tau}$$

$$\langle Y_4(\tau) \rangle = 1 - \frac{9q_2 - 14q_3 + 5q_4}{5q_2 - 8q_3 + 3q_4} e^{-q_2\tau} + e^{-(3q_2-2q_3)\tau} - \frac{q_2 - 2q_3 + q_4}{5q_2 - 8q_3 + 3q_4} e^{-(6q_2 - 8q}$$

Genealogies with $\beta \neq 0$ $\neq$ Spin Glass Trees with n replicas $\neq 0$

In the replica theory (Parisi)

$$\langle Y_k \rangle_\mu = \frac{(1-\mu)(2-\mu)\cdots(k-1-\mu)}{(1-n)(2-n)\cdots(k-1-n)}$$

In the coalescence process

$Y_k(\tau)$= probability that $T_k(\alpha_1, ...\alpha_k) < \tau$

$$\langle Y_2(\tau) \rangle = 1 - e^{-q_2\tau} \quad ; \quad \langle Y_3(\tau) \rangle = 1 - \frac{3}{2} e^{-q_2\tau} + \frac{1}{2} e^{-(3q_2-2q_3)\tau}$$

$$\langle Y_4(\tau) \rangle = 1 - \frac{9q_2 - 14q_3 + 5q_4}{5q_2 - 8q_3 + 3q_4} e^{-q_2\tau} + e^{-(3q_2-2q_3)\tau} - \frac{q_2 - 2q_3 + q_4}{5q_2 - 8q_3 + 3q_4} e^{-(6q_2-8q}$$

$$\langle T_k \rangle = \int_0^\infty (1 - \langle Y_k(\tau)) \rangle d\tau$$

pause

$$\langle Y_2 \rangle_\mu = \langle Y_2(\tau) \rangle \quad \Rightarrow \quad \mu = n + (1-n)e^{-q_2\tau}$$

Genealogies with $\beta \neq 0$ $\neq$ Spin Glass Trees with n replicas $\neq 0$

In the replica theory (Parisi)

$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{5 - 3n}{4 - 2n}$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{50 - 49n + 11n^2}{36 - 30n + 6n^2}$$

In the evolution with selection

$$\left.\frac{\langle T_3 \rangle}{\langle T_2 \rangle}\right|_\beta = \frac{5 + 4\beta}{4 + 3\beta}$$

$$\left.\frac{\langle T_4 \rangle}{\langle T_2 \rangle}\right|_\beta = \frac{100 + 204\beta + 133\beta^2 + 27\beta^3}{72 + 142\beta + 90\beta^2 + 18\beta^3}$$

Conclusion

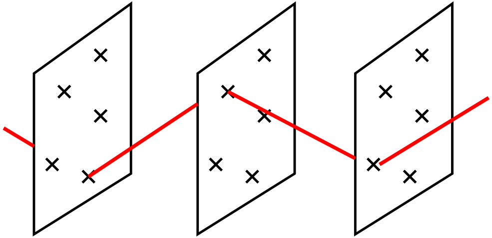
Selection $\Rightarrow$ Bolthausen-Sznitman coalescent
(Exponential and Generic)

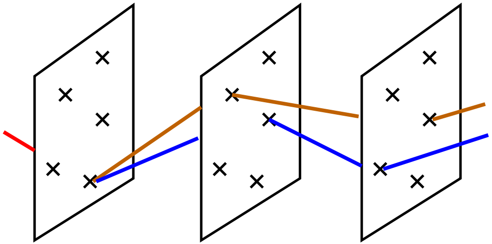
Conditionning on the speed interpolates between
Kingman and Bolthausen-Sznitman

$\beta \neq n$

Steady state measure for large N

Replica for the directed polymer problem





E. Brunet, B. Derrida
How genealogies are affected by the speed of evolution
Philosophical Magazine 92, 255-271 (2012)

E. Brunet, B. Derrida, A. H. Mueller, S. Munier
Noisy traveling waves: Effect of selection on genealogies
Europhysics Letters, 76, 1-7 (2006)
Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization Physical Review E 76, 041104 (2007)

P. Maillard
Branching Brownian motion with selection of the N right-most particles: An approximate model arXiv:1112.0266

J. Berestycki, N. Berestycki, J. Schweinsberg
The genealogy of branching Brownian motion with absorption
arXiv:1001.2337

L.P. Arguin, A. Bovier, N. Kistler *The genealogy of extremal particles of Branching Brownian Motion* arXiv:1008.4386