Some mathematical models from population genetics

Alison Etheridge
University of Oxford

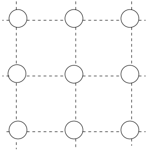
with thanks to numerous collaborators, especially Nick Barton, IST Austria

CMAP5, May 2023

Spatial models so far: subdivided populations

Kimura's stepping stone model

Special case $N_e(i) \equiv N_e$



$$dp_i = \sum_j m_{ji}(p_j - p_i)dt + \sqrt{\frac{1}{N_e} p_i(1-p_i)} dW_i$$

$\{W_i\}_{i \in I}$ independent Brownian motions

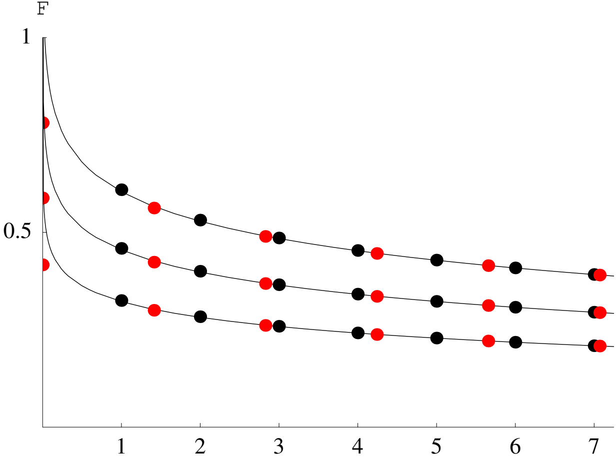
System of W-F diffusions coupled through migration

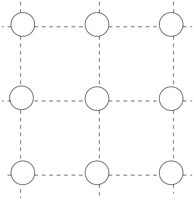
The structured coalescent $\underline{n}$:

- $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ at rate $n_i m_{ji}$
- $n_i \mapsto n_i - 1$ at rate $\frac{1}{2N_e} n_i (n_i - 1)$

Isolation by distance

Malécot-Wright approximation for the stepping stone model





$F = \mathbb{P}[\text{identity}]$

The unreasonable effectiveness of the Kingman coalescent

Common to use Kingman coalescent even for natural populations
Replace census population size by an effective population size

Sampling uniformly from the torus $\mathbb{T}(L) \subset \mathbb{Z}^2$
$T =$ time to MRCA two individuals sampled uniformly from $\mathbb{T}(L)$

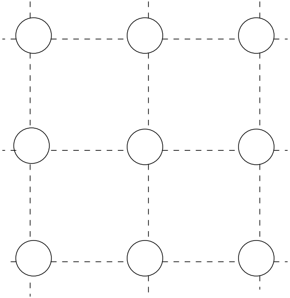
- $T_0 =$ time to first come into same deme
- $t_0 =$ time to coalesce started from same deme

$T = T_0 + t_0$
$t_0 \ll T_0$

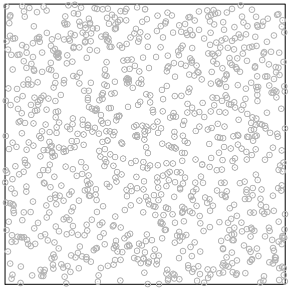
On timescale $L^2 \log L$ genealogy uniform sample from $\mathbb{T}(L) \rightarrow$ Kingman coalescent as $L \rightarrow \infty$ Zähle, Cox, Durrett (2005)

Census population size grows with L^2 so this does not explain the timescale seen in real populations

An obvious challenge

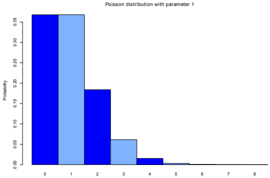


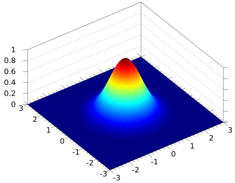
Modelling a spatial continuum: the Wright-Malécot model



$N = 1000$

- Individuals are scattered across a two-dimensional space.
- In each generation, each individual produces a Poisson number of offspring (average one).
- Offspring are scattered in a Gaussian distribution around their parent.



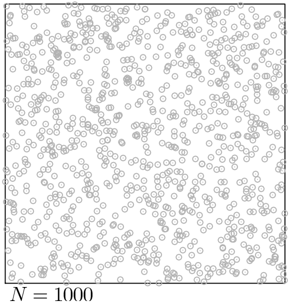


Mitch Gooding
Jerome Kelleher

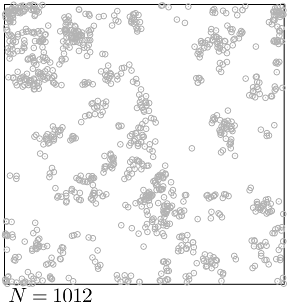
The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher

$t = 0$

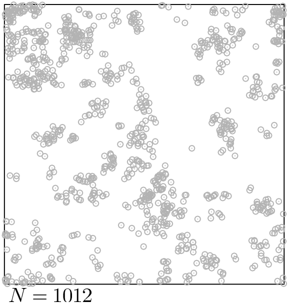


$t = 10$

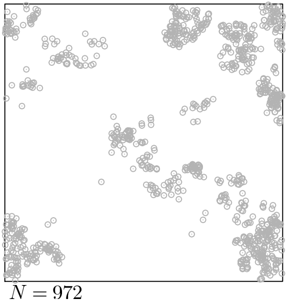


The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher



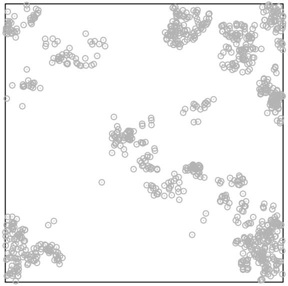
$t = 10$



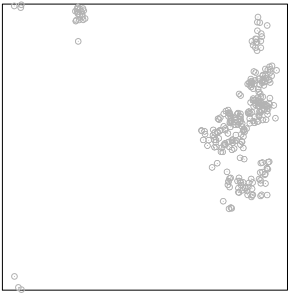
$t = 100$

The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher



$t = 100$ $N = 972$

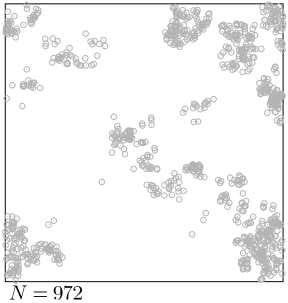


$t = 1000$ $N = 293$

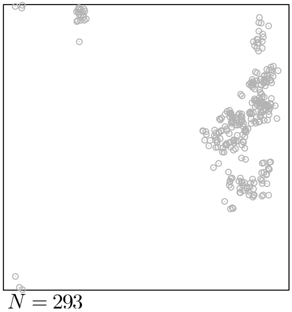
The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher

$t = 100$



$t = 1000$



In $d = 1, 2$ population exhibits clumping/extinction

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1$, 2, independent reproduction $\implies$ clumping;

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1$, 2, independent reproduction $\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1$, 2, independent reproduction $\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

What about modifying the stepping stone model?

$$dp_t(x) = \frac{1}{2}\Delta p_t(x) + \sqrt{\frac{1}{2N_e} p_t(x)(1 - p_t(x))}\, dW(t, x)$$

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1$, 2, independent reproduction $\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

What about modifying the stepping stone model?

$$dp_t(x) = \frac{1}{2}\Delta p_t(x) + \sqrt{\frac{1}{2N_e}p_t(x)(1-p_t(x))}dW(t,x)$$

In 2D the diffusion limit fails over small scales

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1,\ 2$, independent reproduction $\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

What about modifying the stepping stone model?

$$dp_t(x) = \frac{1}{2}\Delta p_t(x) + \sqrt{\frac{1}{2N_e} p_t(x)(1 - p_t(x))}\,dW(t, x)$$

In 2D the diffusion limit fails over small scales ... and so does the obvious backwards model.

Biological problems

Genetic diversity much lower than expected from census numbers

Biological problems

Genetic diversity much lower than expected from census numbers

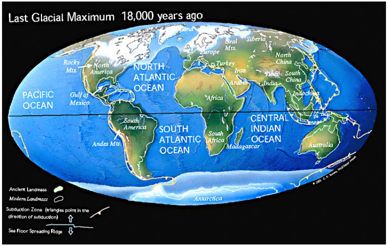
Allele frequencies correlated over long distances

Biological problems

Genetic diversity much lower than expected from census numbers

Allele frequencies correlated over long distances

Demographic history of many species dominated by large scale extinction-recolonisation events

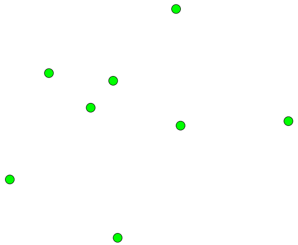


Small neighbourhood size

In a spatial continuum, a single individual can be parent to a significant proportion of the local population.

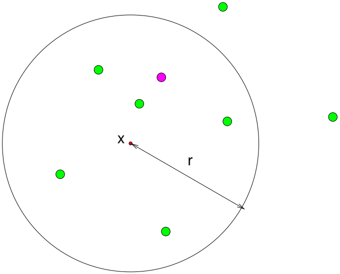
An individual based model

- Start with Poisson intensity λdx.
  Events rate $dt \otimes dx \otimes \xi(dr, du)$.
  Throw down ball $B(x, r)$.



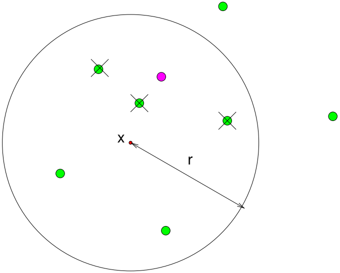
An individual based model

- Start with Poisson intensity λdx.
  Events rate $dt \otimes dx \otimes \xi(dr, du)$.
  Throw down ball $B(x, r)$.
- If region empty, do nothing,
  otherwise:
- Choose parent from $B(x, r)$,



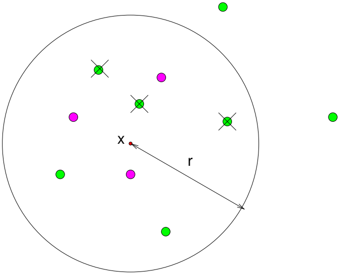
An individual based model

- Start with Poisson intensity λdx.
  Events rate $dt \otimes dx \otimes \xi(dr, du)$.
  Throw down ball $B(x, r)$.
- If region empty, do nothing,
  otherwise:
- Choose parent from $B(x, r)$,
- Each individual in region dies with probability u,



An individual based model

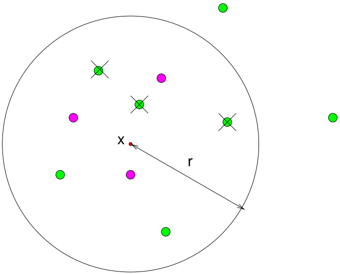
- Start with Poisson intensity λdx. Events rate $dt \otimes dx \otimes \xi(dr, du)$. Throw down ball $B(x, r)$.
- If region empty, do nothing, otherwise:
- Choose parent from $B(x, r)$,
- Each individual in region dies with probability u,
- New individuals born according to Poisson intensity $\lambda u \mathbf{1}_{B_r(x)}$.



An individual based model

- Start with Poisson intensity λdx.
  Events rate $dt \otimes dx \otimes \xi(dr, du)$.
  Throw down ball $B(x, r)$.
- If region empty, do nothing,
  otherwise:
- Choose parent from $B(x, r)$,
- Each individual in region dies with probability u,
- New individuals born according to Poisson intensity $\lambda u \mathbf{1}_{B_r(x)}$.

Offspring inherit type of parent



$\lambda \to \infty$ limit (no space)

Start from $\texttt{Poiss}(\lambda)$

If first reproduction event has 'impact' u

- $\texttt{Poiss}((1-u)\lambda)$ 'survivors';
- $\texttt{Poiss}(u\lambda)$ offspring.

As $\lambda \to \infty$ proportion u of individuals die and are replaced by offspring of the type of the parent.

The Λ-Fleming-Viot process

State $\{\rho(t, \cdot) \in \mathcal{M}_1(K), t \geq 0\}$. K space of genetic types.

- Poisson Point Process Π intensity $dt \otimes F(du)$
- if $(t, u) \in \Pi$, individual sampled at random from population at time $t-$ (i.e. choose $k \sim \rho(t-)$)
- proportion u of population replaced by offspring of chosen individual

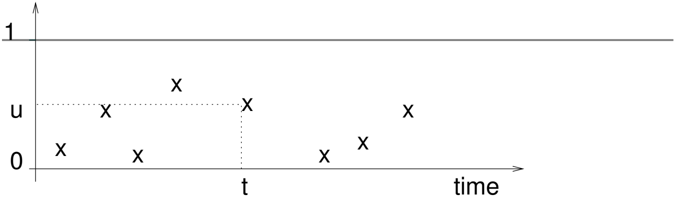
$$\rho(t, \cdot) = (1-u)\rho(t-, \cdot) + u\delta_k.$$

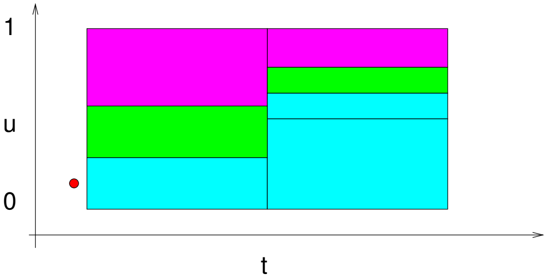
$F(du) = \frac{\Lambda(du)}{u^2}$, Λ finite measure on $[0, 1]$.

Donnelly & Kurtz (1999)

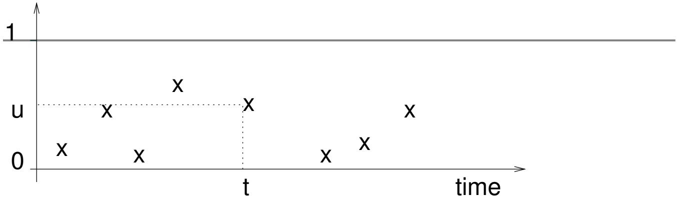
('Generalised Fleming-Viot process', Bertoin & Le Gall 2003)

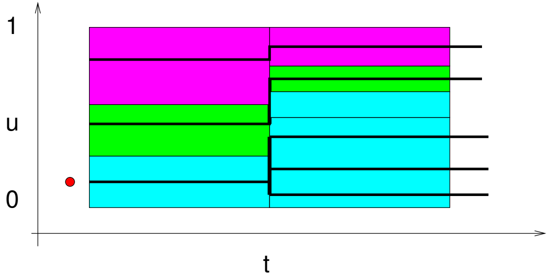
The Λ-Fleming-Viot process





The Λ-Fleming-Viot process





Λ-coalescents

Donnelly & Kurtz (1999), Pitman (1999), Sagitov (1999)

If there are currently n ancestral lineages, each transition involving j of them merging happens at rate

$$\beta_{n,j} = \int_0^1 u^j (1-u)^{n-j} \, \frac{\Lambda(du)}{u^2}$$

- Λ a finite measure on $[0,1]$
- Kingman's coalescent, $\Lambda = \delta_0$

The spatial Λ-Fleming-Viot process

State $\{\rho(t, x, \cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$.

The spatial Λ-Fleming-Viot process

Barton - E - Véber and friends

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

The spatial Λ-Fleming-Viot process

Barton - E - Véber and friends

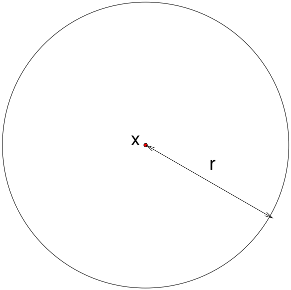
State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

Dynamics: for each $(t,x,r,u) \in \Pi$,

The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

Dynamics: for each $(t,x,r,u) \in \Pi$,

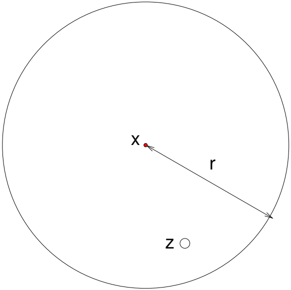


The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

Dynamics: for each $(t,x,r,u) \in \Pi$,

- $z \sim U(B_r(x))$



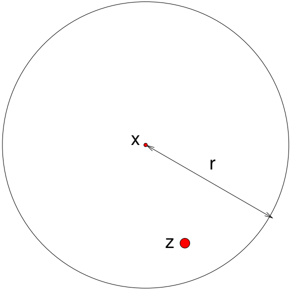
The spatial Λ-Fleming-Viot process

Barton - E - Véber and friends

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

Dynamics: for each $(t,x,r,u) \in \Pi$,

- $z \sim U(B_r(x))$
- $k \sim \rho(t-,z,\cdot)$.



The spatial Λ-Fleming-Viot process

Barton - E - Véber and friends

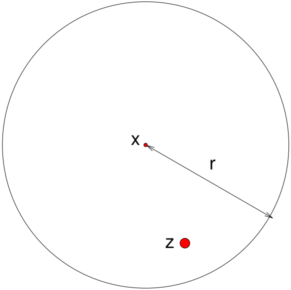
State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

Dynamics: for each $(t,x,r,u) \in \Pi$,

- $z \sim U(B_r(x))$
- $k \sim \rho(t-,z,\cdot)$.

For all $y \in B_r(x)$,

$$\rho(t,y,\cdot) = (1-u)\rho(t-,y,\cdot) + u\delta_k.$$

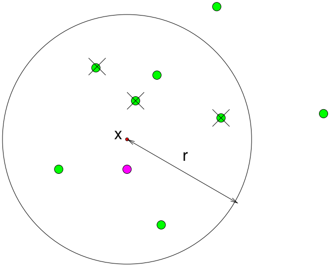


Backwards in time

- A *single* ancestral lineage evolves in series of jumps with intensity

$$dt \otimes \int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2}\, u\, \xi(dr, du) dx$$

on $\mathbb{R}_+ \times \mathbb{R}^2$ where $L_r(x) = |B_r(0) \cap B_r(x)|$.



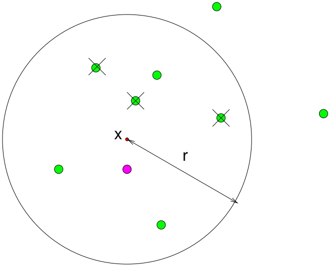
Backwards in time

- A *single* ancestral lineage evolves in series of jumps with intensity

$$dt \otimes \int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2}\, u\, \xi(dr, du)dx$$

  on $\mathbb{R}_+ \times \mathbb{R}^2$ where $L_r(x) = |B_r(0) \cap B_r(x)|$.

- Lineages can coalesce when hit by same 'event'.



Backwards in time

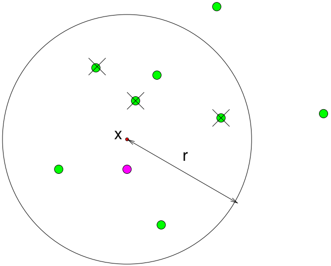
- A *single* ancestral lineage evolves in series of jumps with intensity

$$dt \otimes \int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2}\, u\, \xi(dr, du)dx$$

on $\mathbb{R}_+ \times \mathbb{R}^2$ where $L_r(x) = |B_r(0) \cap B_r(x)|$.

- Lineages can coalesce when hit by same 'event'.

Note: If $\xi(dr, du) = \mu(dr) \otimes \delta_u$, rate of jumps $\propto u$.



Duality

Suppose $K = \{a, A\}$. Define $w(t, x) = \rho(t, x, \{a\})$ to be the proportion of the population at site x at time t that are of type a.

Duality

Suppose $K = \{a, A\}$. Define $w(t, x) = \rho(t, x, \{a\})$ to be the proportion of the population at site x at time t that are of type a.

(Only really defined up to a set of Lebesgue measure zero)

Duality

Suppose $K = \{a, A\}$. Define $w(t, x) = \rho(t, x, \{a\})$ to be the proportion of the population at site x at time t that are of type a.

(Only really defined up to a set of Lebesgue measure zero)

- Sample N_0 individuals from locations $\{X_i(0)\}_{i=1}^{N_0}$ from the present day population;
- Let $\{X_i(t)\}_{i=1}^{N_t}$ denote the positions of the random number of individuals ancestral to the sample at time t before the present

Duality

Suppose $K = \{a, A\}$. Define $w(t, x) = \rho(t, x, \{a\})$ to be the proportion of the population at site x at time t that are of type a.

(Only really defined up to a set of Lebesgue measure zero)

- Sample N_0 individuals from locations $\{X_i(0)\}_{i=1}^{N_0}$ from the present day population;
- Let $\{X_i(t)\}_{i=1}^{N_t}$ denote the positions of the random number of individuals ancestral to the sample at time t before the present

$$\mathbb{E}\left[\prod_{i=1}^{N_0} w\big(t, X_i(0)\big)\right] = \mathbb{E}\left[\prod_{i=1}^{N_t} w\big(0, X_i(t)\big)\right].$$

Direct analogue of our duality in the stepping stone model

Duality

Suppose $K = \{a, A\}$. Define $w(t,x) = \rho(t,x,\{a\})$ to be the proportion of the population at site x at time t that are of type a.

(Only really defined up to a set of Lebesgue measure zero)

- Sample N_0 individuals from locations $\{X_i(0)\}_{i=1}^{N_0}$ from the present day population;
- Let $\{X_i(t)\}_{i=1}^{N_t}$ denote the positions of the random number of individuals ancestral to the sample at time t before the present

$$\mathbb{E}\left[\prod_{i=1}^{N_0} w\big(t, X_i(0)\big)\right] = \mathbb{E}\left[\prod_{i=1}^{N_t} w\big(0, X_i(t)\big)\right].$$

(actually have to sample from random positions and integrate to circumvent issues with sets of Lebesgue measure zero)

Direct analogue of our duality in the stepping stone model

A framework for modelling

- Different spaces,
- Different shapes of event,
- Non-uniform replacement,
- Non-constant density,
- Multiple parents,

A framework for modelling

- Different spaces,
- Different shapes of event,
- Non-uniform replacement,
- Non-constant density,
- Multiple parents,
- Selection,
- Recombination,

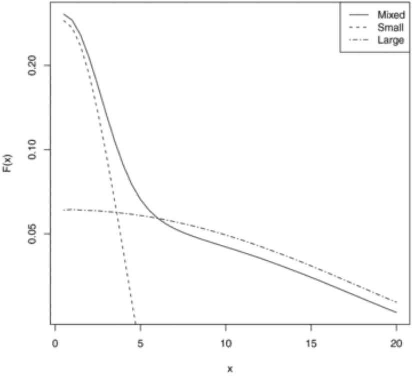
A framework for modelling

- Different spaces,
- Different shapes of event,
- Non-uniform replacement,
- Non-constant density,
- Multiple parents,
- Selection,
- Recombination,

Robust results? $\rightsquigarrow$ Scaling limits.

Example: Wright and Malécot again

The effect of mixed events on F(x, μ). A mixture of rare large events and frequent small events



Evol, Volume 64, Issue 9, 1 September 2010, Pages 2701–2715, https://doi.org/10.1111/j.1558-5646.2010.01019.x

Recap: The Wright-Fisher model with selection

Two types a, A, relative fitnesses $\begin{array}{c|c} a & A \\ \hline 1-s & 1 \end{array}$

During reproduction, each individual produces large number of juveniles from which next generation sampled.

If proportion A-alleles in parental population is q, proportion in pool of juveniles is

$$q^* = \frac{q}{1 - s(1-q)} \approx q + sq(1-q).$$

Population size N (fixed).

Establishment of a favourable allele (Fisher 1930)

While rare, No. offspring of a favoured individual $\sim \texttt{Binom}(N, (1+s)/N) \approx \texttt{Poiss}(1+s)$.

Branching process approximation: probability extinction satisfies

$$x = \exp(-(1+s)(1-x)),$$

Survival probability, $y = 1 - x$,

$$y = 1 - \exp(-(1+s)y) = (1+s)y - \frac{1}{2}(1+s)^2 y^2 + \mathcal{O}(y^3).$$

Rearranging:

$$\frac{1}{2}(1+s)^2 y^2 = sy \quad \Longrightarrow \quad y \approx 2s.$$

Does space matter?

Maryuama (1970),

- subdivided population, demes (large) constant size;
- selection acts independently in each deme;
- contribution of each deme to next generation proportional to size.

$\rightsquigarrow$ Fixation probability independent of population subdivision.

Does space matter?

Maryuama (1970),

- subdivided population, demes (large) constant size;
- selection acts independently in each deme;
- contribution of each deme to next generation proportional to size.

$\rightsquigarrow$ Fixation probability independent of population subdivision.

Barton (1993), this is no longer true if one adds extinction-recolonisation events to colonies.

Does space matter?

Maryuama (1970),

- subdivided population, demes (large) constant size;
- selection acts independently in each deme;
- contribution of each deme to next generation proportional to size.

$\rightsquigarrow$ Fixation probability independent of population subdivision.

Barton (1993), this is no longer true if one adds extinction-recolonisation events to colonies.

Aim to investigate the interaction of natural selection and spatial structure in the framework of the spatial Lambda-Fleming-Viot process.

Introducing selection to the SLFV

WARNING: There are lots of ways to do this.

Here we mimic what we did for the Wright-Fisher/Moran models in the first lecture.

Introducing selection to the SLFV

$K = \{a, A\}$, $w(t,x) = \rho(t,x,a)$ proportion of type a

- (i) Two types, a, A. Weight type a by $(1-s)$. If a reproduction event affects a region $B(x,r)$ in which current proportion of a-alleles is $\overline{w}$, then probability offspring are type a is

$$\frac{(1-s)\overline{w}}{1-s\overline{w}}$$

Introducing selection to the SLFV

$K = \{a, A\}$, $w(t,x) = \rho(t,x,a)$ proportion of type a

- (i) Two types, a, A. Weight type a by $(1-s)$. If a reproduction event affects a region $B(x,r)$ in which current proportion of a-alleles is $\overline{w}$, then probability offspring are type a is
$$\frac{(1-s)\overline{w}}{1-s\overline{w}} = \overline{w}(1-s) + s\overline{w}^2 + \mathcal{O}(s^2).$$

Introducing selection to the SLFV

$K = \{a, A\}$, $w(t,x) = \rho(t,x,a)$ proportion of type a

- (i) Two types, a, A. Weight type a by $(1-s)$. If a reproduction event affects a region $B(x,r)$ in which current proportion of a-alleles is $\overline{w}$, then probability offspring are type a is
$$\frac{(1-s)\overline{w}}{1-s\overline{w}} = \overline{w}(1-s) + s\overline{w}^2 + \mathcal{O}(s^2).$$
- (ii) Neutral events rate $\propto (1-s)$, selective events rate $\propto s$. At selective reproduction events, sample two potential parents. If types aa, then an a reproduces, otherwise an A does.

Introducing selection to the SLFV

$K = \{a, A\}$, $w(t,x) = \rho(t,x,a)$ proportion of type a

- (i) Two types, a, A. Weight type a by $(1-s)$. If a reproduction event affects a region $B(x,r)$ in which current proportion of a-alleles is $\overline{w}$, then probability offspring are type a is
$$\frac{(1-s)\overline{w}}{1-s\overline{w}} = \overline{w}(1-s) + s\overline{w}^2 + \mathcal{O}(s^2).$$

- (ii) Neutral events rate $\propto (1-s)$, selective events rate $\propto s$. At selective reproduction events, sample two potential parents. If types aa, then an a reproduces, otherwise an A does.

c.f. what we did for Moran model

(Spatial) Ancestral selection graph

Evolution of ancestry due to neutral events as before:

- lineages evolve in a series of jumps;
- they can coalesce when covered by same event.

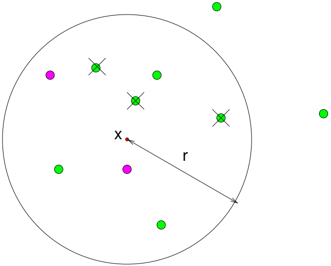
(Spatial) Ancestral selection graph

Evolution of ancestry due to neutral events as before:

- lineages evolve in a series of jumps;
- they can coalesce when covered by same event.

At *selective* events

- *Two 'potential' parents must be traced;*
- Lineages can coalesce when hit by same 'event'.



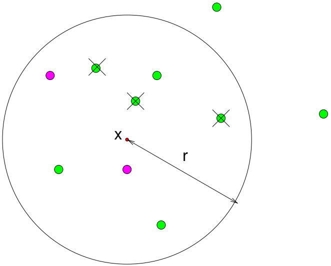
(Spatial) Ancestral selection graph

Evolution of ancestry due to neutral events as before:

- lineages evolve in a series of jumps;
- they can coalesce when covered by same event.

At *selective* events

- *Two 'potential' parents must be traced;*
- Lineages can coalesce when hit by same 'event'.



A sampled individual is type a *iff* all lineages in the corresponding ASG are type a at any previous time.

Zooming out: recovering classical models

The spread of a *favoured* allele is classically modelled through the (stochastic) Fisher-KPP equation:

$$du = \Big(\frac{1}{2}\Delta u + su(1-u)\Big)dt + \mathbf{1}_{d=1}\epsilon\sqrt{u(1-u)}W(dt,dx).$$

Zooming out: recovering classical models

The spread of a *favoured* allele is classically modelled through the (stochastic) Fisher-KPP equation:

$$du = \Big(\frac{1}{2}\Delta u + su(1-u)\Big)dt + \mathbf{1}_{d=1}\epsilon\sqrt{u(1-u)}W(dt,dx).$$

Over sufficiently large spatial and temporal scales, does the proportion of favoured alleles in the SLFV with selection look like a solution to the (stochastic) Fisher-KPP equation?

Zooming out: recovering classical models

The spread of a *favoured* allele is classically modelled through the (stochastic) Fisher-KPP equation:

$$du = \big(\frac{1}{2}\Delta u + su(1-u)\big)dt + \mathbf{1}_{d=1}\epsilon\sqrt{u(1-u)}W(dt, dx).$$

Over sufficiently large spatial and temporal scales, does the proportion of favoured alleles in the SLFV with selection look like a solution to the (stochastic) Fisher-KPP equation?

Key tool: ancestral selection graph.

Zooming out: recovering classical models

The spread of a *favoured* allele is classically modelled through the (stochastic) Fisher-KPP equation:

$$du = \big(\frac{1}{2}\Delta u + su(1-u)\big)dt + \mathbf{1}_{d=1}\epsilon\sqrt{u(1-u)}W(dt, dx).$$

Over sufficiently large spatial and temporal scales, does the proportion of favoured alleles in the SLFV with selection look like a solution to the (stochastic) Fisher-KPP equation?

Key tool: ancestral selection graph.

Stochastic Fisher-KPP is dual to branching and coalescing Brownian motion

Branching Brownian motion and the Fisher-KPP equation

Binary branching BM $\mathbf{X}_t = \{X_t^1, \dots, X_t^{N_t}\}$

- Individuals follow independent Brownian motions
- Individual lifetime $\texttt{Exp}(s)$
- Replaced (at location where die) by two offspring

Branching Brownian motion and the Fisher-KPP equation

Binary branching BM $\mathbf{X}_t = \{X_t^1, \dots, X_t^{N_t}\}$

- Individuals follow independent Brownian motions
- Individual lifetime $\mathtt{Exp}(s)$
- Replaced (at location where die) by two offspring

$$w(t,x) = \mathbb{E}_x\Big[\prod_{i=1}^{N_t} w(0, X_t^i)\Big], \quad \frac{\partial w}{\partial t} = \frac{1}{2}\Delta w + s(w^2 - w)$$

Branching Brownian motion and the Fisher-KPP equation

Binary branching BM $\mathbf{X}_t = \{X_t^1, \dots, X_t^{N_t}\}$

- Individuals follow independent Brownian motions
- Individual lifetime $\mathtt{Exp}(s)$
- Replaced (at location where die) by two offspring

$$w(t,x) = \mathbb{E}_x\Big[\prod_{i=1}^{N_t} w(0, X_t^i)\Big], \quad \frac{\partial w}{\partial t} = \frac{1}{2}\Delta w + s(w^2 - w)$$

$S \sim \mathtt{Exp}(s)$= lifetime ancestor started at x, $h \ll 1$,

$$\begin{aligned}\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X_{t+h}^i)\Big] &= \mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X_{t+h}^i)\Big| S < h\Big]\mathbb{P}[S < h] \\ &\quad + \mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X_{t+h}^i)\Big| S > h\Big]\mathbb{P}[S > h]\end{aligned}$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0,X^i_{t+h})\Big| S<h\Big]\mathbb{P}[S<h] = sh\left(\mathbb{E}_x\big[\prod_{i=1}^{N_t} w(0,X^i_t)\big]\right)^2 + O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X^i_{t+h})\Big| S < h\Big]\mathbb{P}[S < h] \;\; = shw(t,x)^2 + O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X^i_{t+h})\Big| S < h\Big]\mathbb{P}[S < h] \; = shw(t,x)^2 + O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X^i_{t+h})\Big| S > h\Big]\mathbb{P}[S > h]$$

$$= (1 - sh)E_x\Big[\mathbb{E}_{B_h}\big[\prod_{i=1}^{N_t} w(0, X^i_t)\big]\Big] + O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0,X^i_{t+h})\Big|S<h\Big]\mathbb{P}[S<h] \ \ = shw(t,x)^2+O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0,X^i_{t+h})\Big|S>h\Big]\mathbb{P}[S>h] = (1-sh)E_x\big[w(B_h,t)\big]+O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X^i_{t+h})\Big| S < h\Big]\mathbb{P}[S < h] \;= shw(t,x)^2 + O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0, X^i_{t+h})\Big| S > h\Big]\mathbb{P}[S > h] = (1-sh)E_x\big[w(B_h, t)\big]+O(h^2)$$

$$\begin{aligned}\lim_{h\to 0}\frac{w(t+h,x)-w(t,x)}{h} &= \lim_{h\to 0}\frac{1}{h}E_x\big[w(B_h,t)-w(x,t)\big] \\ &\quad +s\big(w(t,x)^2 - w(t,x)\big) \\ &= \frac{1}{2}\Delta w(t,x) + sw(t,x)\big(w(t,x)-1\big)\end{aligned}$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0,X^i_{t+h})\Big|S<h\Big]\mathbb{P}[S<h] \;\; = shw(t,x)^2+O(h^2)$$

$$\mathbb{E}_x\Big[\prod_{i=1}^{N_{t+h}} w(0,X^i_{t+h})\Big|S>h\Big]\mathbb{P}[S>h] = (1-sh)E_x\big[w(B_h,t)\big]+O(h^2)$$

$$\begin{aligned}\lim_{h\to 0}\frac{w(t+h,x)-w(t,x)}{h} &= \lim_{h\to 0}\frac{1}{h}E_x\big[w(B_h,t)-w(x,t)\big]\\ &\quad +s\big(w(t,x)^2-w(t,x)\big)\\ &= \frac{1}{2}\Delta w(t,x)+sw(t,x)\big(w(t,x)-1\big)\end{aligned}$$

$u(t,x)=1-w(t,x)$ solves $\frac{\partial u}{\partial t}=\frac{1}{2}\Delta u+su(1-u)$

Convergence of SLFV with selection

- SLFV dual to system of branching and coalescing random walks
- Fisher-KPP equation dual to binary branching Brownian motion

Convergence of SLFV with selection

- SLFV dual to system of branching and coalescing random walks
- Fisher-KPP equation dual to binary branching Brownian motion

Recall Wright-Fisher diffusion (genetic drift) dual to Kingman coalescent
In spatial setting, Wright-Fisher noise reflected in coalescence in dual - lineages coalesce at rate determined by *local time* they spend together, but only makes sense in $d = 1$

Convergence of SLFV with selection

- SLFV dual to system of branching and coalescing random walks
- Fisher-KPP equation dual to binary branching Brownian motion

Recall Wright-Fisher diffusion (genetic drift) dual to Kingman coalescent
In spatial setting, Wright-Fisher noise reflected in coalescence in dual - lineages coalesce at rate determined by *local time* they spend together, but only makes sense in $d = 1$

To identify convergence to (stochastic) Fisher-KPP, show convergence of the dual processes

Brief aside on random walk

Suppose X is simple random walk on $\mathbb{Z}$.

Write τ for hitting time of $\{0, N\}$

Take $X_0 = 1$. Doob's Optional Stopping Theorem says $\mathbb{E}[X_\tau] = X_0$.

Thus $\mathbb{P}[X_\tau = N] = 1/N$.

Brief aside on random walk

Suppose X is simple random walk on $\mathbb{Z}$.

Write τ for hitting time of $\{0, N\}$

Take $X_0 = 1$. Doob's Optional Stopping Theorem says $\mathbb{E}[X_\tau] = X_0$.

Thus $\mathbb{P}[X_\tau = N] = 1/N$.

So number of excursions away from zero before one reaches N is Geometric with mean N.

Brief aside on random walk

Suppose X is simple random walk on $\mathbb{Z}$.

Write τ for hitting time of $\{0, N\}$

Take $X_0 = 1$. Doob's Optional Stopping Theorem says $\mathbb{E}[X_\tau] = X_0$.

Thus $\mathbb{P}[X_\tau = N] = 1/N$.

So number of excursions away from zero before one reaches N is Geometric with mean N.

In $d = 2$, corresponding quantity has mean $\propto \log N$.

Reminder: parameters in SLFV with selection

- Events driven by Poisson Point Process Π that specifies
  - centre and *radius* event
  - *impact* event
- *selection coefficient* determines proportion of selective events

Reminder: parameters in SLFV with selection

- Events driven by Poisson Point Process Π that specifies
  - centre and *radius* event
  - *impact* event
- *selection coefficient* determines proportion of selective events

For simplicity suppose r, u, s fixed

Reminder: parameters in SLFV with selection

- Events driven by Poisson Point Process Π that specifies
    - centre and *radius* event
    - *impact* event
- *selection coefficient* determines proportion of selective events

For simplicity suppose r, u, s fixed

Dual lineages make jumps of length $O(r)$ at rate proportional to ur^d, and branch at rate proportional to sur^d

lineages can only coalesce when at separation less than $4r$

Scaling limits I: High neighbourhood size

Set $u_n = u/n^\gamma$, $s_n = s/n^\delta$, $w^{(n)}(t,x) = w(nt, n^\beta x)$,
Jump rate nu_n, jump size $1/n^\beta$. Diffusive scaling: $2\beta = 1 - \gamma$

- At 'branching' event, two lineages at separation $\mathcal{O}(1/n^\beta)$.
- Probability separate to $\mathcal{O}(1)$ before come back together is $\mathcal{O}(1/n^\beta)$, $(d = 1)$; $\mathcal{O}(1/\log n)$, $(d = 2)$; $\mathcal{O}(1)$, $(d \geq 3)$.
- If two lineages hit by same event, given one jumps, they coalesce with probability $\mathcal{O}(1/n^\gamma)$.

$d \geq 2$: Probability 'long' excursion before coalesce $\mathcal{O}(1)$;

$d = 1$: Number attempts to reach separation $\mathcal{O}(1)$
$\sim$ number of attempts to coalesce: $\beta = \gamma$;

Selection events rate nu_ns_n $\mathcal{O}(1)$: $1 - \gamma - \delta = 0$.

$$\rightsquigarrow \beta = \gamma = 1/3, \quad \delta = 2/3.$$

Scaling limits I: High neighbourhood size

Fixed impact u and event radius r, selection coefficient s

- Set $u_n = u/n^{1/3}$, $s_n = s/n^{2/3}$, $w^{(n)}(t,x) = w(nt, n^{1/3}x)$,

$$dw = \frac{1}{2}\Delta w dt + sw(1-w)dt + \mathbf{1}_{d=1}\epsilon\sqrt{w(1-w)}W(dt,dx)$$

E. Véber, Yu.

Scaling limits I: High neighbourhood size

Fixed impact u and event radius r, selection coefficient s

- Set $u_n = u/n^{1/3}$, $s_n = s/n^{2/3}$, $w^{(n)}(t,x) = w(nt, n^{1/3}x)$,

$$dw = \frac{1}{2}\Delta w dt + sw(1-w)dt + \mathbf{1}_{d=1}\epsilon\sqrt{w(1-w)}W(dt,dx)$$

E. Véber, Yu.

Here, $u_n \to 0$, corresponding to high *neighbourhood size*.

Scaling limits I: High neighbourhood size

Fixed impact u and event radius r, selection coefficient s

- Set $u_n = u/n^{1/3}$, $s_n = s/n^{2/3}$, $w^{(n)}(t,x) = w(nt, n^{1/3}x)$,

$$dw = \frac{1}{2}\Delta w dt + sw(1-w)dt + \mathbf{1}_{d=1}\epsilon\sqrt{w(1-w)}W(dt,dx)$$

E. Véber, Yu.

Here, $u_n \to 0$, corresponding to high *neighbourhood size*.

$\gamma > \beta$ (even bigger neighbourhood size) $\rightsquigarrow$ deterministic equation in all dimensions