Some mathematical models from population genetics

4: Spatial models

Alison Etheridge

University of Oxford

joint work with Jochen Blath (TU Berlin), Mark Meredith (Oxford)
Nick Barton (Edinburgh), Frantz Depaulis (Paris VI)

Lessons learned so far

The key message of our balancing selection example is that if we want to study the genealogy of a sample from a structured population, then fluctuations in background frequencies matter.

A central question then is how should we model fluctuations of spatially distributed populations.

Reminder: Feller's rescaling

Galton Watson process, offspring generating function $\Phi(s)$. Assume $\Phi''(1) < \infty$.

Reminder: Feller's rescaling

Galton Watson process, offspring generating function $\Phi(s)$. Assume $\Phi''(1) < \infty$.

Large population, long timescales, measured in units of size N.

Write Z_n for the population size after n generations.

Reminder: Feller's rescaling

Galton Watson process, offspring generating function $\Phi(s)$. Assume $\Phi''(1) < \infty$.

Large population, long timescales, measured in units of size N.

Write Z_n for the population size after n generations.

If $\Phi'(1) = 1 + \overline{a}$, then $\mathbb{E}[Z_N] = (1+\overline{a})^N Z_0$, so for non-trivial limit assume $\overline{a} = \frac{a}{N}$.

Reminder: Feller's rescaling

Galton Watson process, offspring generating function $\Phi(s)$. Assume $\Phi''(1) < \infty$.

Large population, long timescales, measured in units of size N.

Write Z_n for the population size after n generations.

If $\Phi'(1) = 1 + \overline{a}$, then $\mathbb{E}[Z_N] = (1 + \overline{a})^N Z_0$, so for non-trivial limit assume $\overline{a} = \frac{a}{N}$.

If $\left\{\frac{Z_0}{N}\right\}_{N \geq 1}$ converges, so does $\left\{\frac{Z_{\lfloor Nt \rfloor}}{N}\right\}_{N \geq 1}$.

Limit process: $dX_t = aX_t dt + \sqrt{\gamma X_t} dB_t.$

Spatially distributed populations

Populations dispersed in $\mathbb{R}^d$ or $\mathbb{Z}^d$,

Spatially distributed populations

Populations dispersed in $\mathbb{R}^d$ or $\mathbb{Z}^d$,

Galton-Watson branching process $\rightsquigarrow$ branching Brownian motion/ branching random walk.

Offspring born where parent died.

Spatially distributed populations

Populations dispersed in $\mathbb{R}^d$ or $\mathbb{Z}^d$,

Galton-Watson branching process $\rightsquigarrow$ branching Brownian motion/ branching random walk.

Offspring born where parent died.

Feller rescaling: individual represented by atom of mass $\frac{1}{N}$, time in units of size N.

The limiting processes

The Dawson-Watanabe superprocess:

For positive, twice differentiable test functions ϕ,

$$\langle \phi, X_t \rangle - \langle \phi, X_0 \rangle - \int_0^t \langle D\Delta\phi, X_s \rangle ds - \int_0^t \langle a\phi, X_s \rangle ds$$

is a martingale with quadratic variation $\int_0^t \langle \gamma\phi^2, X_s \rangle ds$.

The limiting processes

The Dawson-Watanabe superprocess:
For positive, twice differentiable test functions ϕ,

$$\langle \phi, X_t \rangle - \langle \phi, X_0 \rangle - \int_0^t \langle D\Delta\phi, X_s \rangle ds - \int_0^t \langle a\phi, X_s \rangle ds$$

is a martingale with quadratic variation $\int_0^t \langle \gamma \phi^2, X_s \rangle ds$.

Super-random walk:

$$dX_i(t) = \sum_j m_{ij} \left(X_j(t) - X_i(t) \right) dt + aX_i(t)dt + \sqrt{\gamma X_i(t)} dW_i(t), \quad i \in \mathbb{Z}^d,$$

where $\{W_i(t), t \geq 0\}_{i \in \mathbb{Z}^d}$ is a collection of independent Brownian motions and $X_i(t)$ is the size of the population in deme i at time t.

Clumping and extinction.

$$\mathbb{E}\left[\langle \phi, X_t \rangle\right] = e^{at} \langle T_t \phi, X_0 \rangle,$$

Take $a = 0$.

$$var\left(\langle \phi, X_t \rangle\right) = \int_0^t \langle \gamma T_{t-s}\left((T_s \phi)^2 \right), X_0 \rangle ds.$$

In one and two dimensions grows without bound.

Controlling the population

Exogenously specify total population size $\rightsquigarrow$ **Fleming-Viot superprocess**.

Controlling the population

Exogenously specify total population size $\rightsquigarrow$ **Fleming-Viot superprocess**.

As we saw in the first lecture, in $\mathbb{Z}^d$ can specify the population size *locally* $\rightsquigarrow$ **The Classical Stepping Stone Model**.

Controlling the population

Exogenously specify total population size $\rightsquigarrow$ **Fleming-Viot superprocess**.

As we saw in the first lecture, in $\mathbb{Z}^d$ can specify the population size *locally* $\rightsquigarrow$ **The Classical Stepping Stone Model**.

Populations should be regulated by local rules.

Controlling the population

Exogenously specify total population size ⇝ **Fleming-Viot superprocess**.

As we saw in the first lecture, in $\mathbb{Z}^d$ can specify the population size *locally* ⇝ **The Classical Stepping Stone Model**.

Populations should be regulated by local rules.

Individuals living in locally crowded regions will have a lower reproductive success than those living in sparsely populated regions.

Locally regulated populations

$$a(s,x) = \alpha\big(M - \langle h(x,y), X_s(dy)\rangle\big).$$

For simplicity $h(x,y) = h(\|x-y\|)$.

Locally regulated populations

$$a(s,x) = \alpha\big(M - \langle h(x,y), X_s(dy)\rangle\big).$$

For simplicity $h(x,y) = h(\|x-y\|)$. For infinite initial measures, to prevent immediate catastrophe, $\int_0^\infty h(r) r^{d-1} dr < \infty$.

Locally regulated populations

$$a(s,x) = \alpha\big(M - \langle h(x,y), X_s(dy)\rangle\big).$$

For simplicity $h(x,y) = h(\|x-y\|)$. For infinite initial measures, to prevent immediate catastrophe, $\int_0^\infty h(r) r^{d-1} dr < \infty$.

The stepping-stone version of the Bolker-Pacala model: In the super-random walk setting the corresponding model is

$$dX_t(i) = \sum_j m_{ij}\left(X_t(j) - X_t(i)\right)dt + \alpha\left(M - \sum_j \lambda_{ij} X_t(j)\right) X_t(i)dt$$
$$+ \sqrt{\gamma X_t(i)}\, dB_t^{(i)}.$$

Locally regulated populations

$$a(s,x) = \alpha\big(M - \langle h(x,y), X_s(dy)\rangle\big).$$

For simplicity $h(x,y) = h(\|x-y\|)$. For infinite initial measures, to prevent immediate catastrophe, $\int_0^\infty h(r) r^{d-1} dr < \infty$.

The stepping-stone version of the Bolker-Pacala model: In the super-random walk setting the corresponding model is

$$dX_t(i) = \sum_j m_{ij}\left(X_t(j) - X_t(i)\right) dt + \alpha\left(M - \sum_j \lambda_{ij} X_t(j)\right) X_t(i) dt + \sqrt{\gamma X_t(i)}\, dB_t^{(i)}.$$

Note that moment equations not closed.

Survival and Extinction

Theorem

For each fixed interaction kernel h and $\gamma, K > 0$ there exists $\alpha_0 = \alpha_0(K, \gamma, h)$ such that for $\alpha > \alpha_0$, the superprocess version of the Bolker-Pacala model with parameters $(h, K/\alpha, \alpha, \gamma)$ started from any finite initial measure dies out in finite time. If h also satisfies $\int h(r)r^{d-1}dr < \infty$, then when started from any tempered initial measure (with $p > d$) the process with these parameters suffers local extinction.

Let $\alpha > 0$ be fixed.

- If $r^{2-\delta}h(r)$ is unbounded for some $\delta > 0$, then for each fixed $\gamma > 0$, there is an $M_0 > 0$ such that for $M < M_0$ the superprocess version of the Bolker-Pacala model with parameters (h, M, α, γ) started from any finite initial measure dies out in finite time. If also $\int h(r)r^{d-1}dr < \infty$, so that in particular $d = 1$, then when started from any tempered initial measure (with $p > 1$) the process with these parameters suffers local extinction.

- Suppose that the population $\{X_t(i)\}_{i\in\mathbb{Z}^d, t\geq 0}$ evolves according to the stepping stone version of the Bolker-Pacala model, then if $m_{ij} > c\lambda_{ij}$, for some $c > 0$, then there exists $M_1 > 0$ such that for $M > M_1$ the process survives for all time with (strictly) positive probability (started from any non-trivial initial condition).

- Suppose that the population $\{X_t(i)\}_{i\in\mathbb{Z}^d, t\geq 0}$ evolves according to the stepping stone version of the Bolker-Pacala model, then if $m_{ij} > c\lambda_{ij}$, for some $c > 0$, then there exists $M_1 > 0$ such that for $M > M_1$ the process survives for all time with (strictly) positive probability (started from any non-trivial initial condition).

Hutzenthaler & Wakolbinger prove an ergodic theorem and also show that if M is too small, the process dies out.

Rescaling:

Take $d = 2$. Define X^θ by

$$\langle \phi, X_t^\theta \rangle = \left\langle \frac{1}{\theta^2} \phi\left(\frac{x}{\theta}\right), X_{\theta^2 t}(dx) \right\rangle .$$

Notation $h^\theta(r) = \theta^2 h(\theta r)$.

Rescaling:

Take $d = 2$. Define X^θ by

$$\langle \phi, X_t^\theta \rangle = \left\langle \frac{1}{\theta^2} \phi \left(\frac{x}{\theta} \right), X_{\theta^2 t}(dx) \right\rangle .$$

Notation $h^\theta(r) = \theta^2 h(\theta r)$.

$$\langle \phi, X_t^\theta \rangle - \langle \phi, X_0^\theta \rangle - \int_0^t \langle D\Delta\phi, X_s^\theta \rangle ds$$
$$- \int_0^t \left\langle \theta^2 \alpha \left(M - \langle h^\theta(\|x - y\|), X_s^\theta(dy) \rangle \right) \phi(x), X_s^\theta(dx) \right\rangle ds$$

a martingale with quadratic variation

$$\int_0^t \langle \gamma \phi^2, X_s^\theta \rangle ds.$$

Notes:

If $r^2h(r) \to \infty$ as $r \to \infty$, then h^θ grows without bound as $\theta \to \infty$, suggesting extinction.

Notes:

If $r^2 h(r) \to \infty$ as $r \to \infty$, then h^θ grows without bound as $\theta \to \infty$, suggesting extinction.

In $d = 2$, for classical Dawson-Watanabe superprocess, if x is typical point in support of X_t, then

$$\lim_{r \downarrow 0} \mathbb{E}^{(x)} \left[\frac{\langle \chi_{B(x,r)}, X_t \rangle}{r^2 \log(1/r)} \right] = k$$

for a constant k (independent of x and t).

Notes:

If $r^2 h(r) \to \infty$ as $r \to \infty$, then h^θ grows without bound as $\theta \to \infty$, suggesting extinction.

In $d = 2$, for classical Dawson-Watanabe superprocess, if x is typical point in support of X_t, then

$$\lim_{r \downarrow 0} \mathbb{E}^{(x)} \left[\frac{\langle \chi_{B(x,r)}, X_t \rangle}{r^2 \log(1/r)} \right] = k$$

for a constant k (independent of x and t).

$$\langle h^\theta(\|x - y\|), X_s(dy) \rangle \sim \log \theta.$$

Survival in two dimensions reflects successful eradication of clumping.

Competing species

Strategies for survival:

Competing species

Strategies for survival:

- colonise relatively unpopulated areas quickly,

Competing species

Strategies for survival:

- colonise relatively unpopulated areas quickly,
- quickly exploit resources in those areas,

Competing species

Strategies for survival:

- colonise relatively unpopulated areas quickly,
- quickly exploit resources in those areas,
- tolerate local competition.

Model I

$$
\begin{aligned}
dX_i(t) &= \sum_{j\in\mathbb{Z}^d} m_{ij}\left(X_j(t)-X_i(t)\right)dt \\
&+\alpha\left(M-\sum_{j\in\mathbb{Z}^d}\lambda_{ij}X_j(t)-\sum_{j\in\mathbb{Z}^d}\gamma_{ij}Y_j(t)\right)X_i(t)dt+\sqrt{X_i(t)}dB_i(t),
\end{aligned}
$$

Model I

$$dX_i(t) = \sum_{j\in\mathbb{Z}^d} m_{ij}\left(X_j(t) - X_i(t)\right) dt$$

$$+\,\alpha\left(M - \sum_{j\in\mathbb{Z}^d} \lambda_{ij} X_j(t) - \sum_{j\in\mathbb{Z}^d} \gamma_{ij} Y_j(t)\right) X_i(t)dt + \sqrt{X_i(t)}dB_i(t),$$

$$dY_i(t) = \sum_{j\in\mathbb{Z}^d} m'_{ij}\left(Y_j(t) - Y_i(t)\right) dt$$

$$+\,\alpha'\left(M' - \sum_{j\in\mathbb{Z}^d} \lambda'_{ij} Y_j(t) - \sum_{j\in\mathbb{Z}^d} \gamma'_{ij} X_j(t)\right) Y_i(t)dt + \sqrt{Y_i(t)}d\tilde{B}_i(t).$$

Model II

Simplify our previous model:

Model II

Simplify our previous model:

- competition is only within-site

Model II

Simplify our previous model:

- competition is only within-site
- migration mechanism is the same for both populations

Model II

Simplify our previous model:

- competition is only within-site
- migration mechanism is the same for both populations
- total population size in each site is fixed

Model II

Simplify our previous model:

- competition is only within-site
- migration mechanism is the same for both populations
- total population size in each site is fixed

Write

$$p_i(t) = \frac{X_i(t)}{(X_i(t) + Y_i(t))} = \frac{X_i(t)}{N}.$$

$$dp_i(t) = \sum_{j \in \mathbb{Z}^d} m_{ij}\left(p_j(t) - p_i(t)\right) dt$$

$$+ s p_i(t)\left(1 - p_i(t)\right)\left(1 - \mu p_i(t)\right) dt + \sqrt{\frac{1}{N} p_i(t)\left(1 - p_i(t)\right)}\, dW_i(t),$$

where

$$s = \alpha M - \alpha' M' + \left(\alpha' \lambda'_{ii} - \alpha \gamma_{ii}\right) N,$$

$$\mu = \frac{\left(\alpha' \lambda'_{ii} - \alpha \gamma_{ii}\right) N + \left(\alpha \lambda_{ii} - \alpha' \gamma'_{ii}\right) N}{\alpha M - \alpha' M' + \left(\alpha' \lambda'_{ii} - \alpha \gamma_{ii}\right) N}$$

$$dp_i(t) = \sum_{j\in\mathbb{Z}^d} m_{ij}\,(p_j(t) - p_i(t))\,dt$$

$$+\, sp_i(t)\,(1 - p_i(t))\,(1 - \mu p_i(t))\,dt + \sqrt{\frac{1}{N} p_i(t)\,(1 - p_i(t))}\,dW_i(t),$$

where

$$s = \alpha M - \alpha' M' + (\alpha' \lambda'_{ii} - \alpha\gamma_{ii})\,N,$$

$$\mu = \frac{(\alpha'\lambda'_{ii} - \alpha\gamma_{ii})\,N + (\alpha\lambda_{ii} - \alpha'\gamma'_{ii})\,N}{\alpha M - \alpha' M' + (\alpha'\lambda'_{ii} - \alpha\gamma_{ii})\,N}$$

‘Selection in favour of heterozygosity’ when $\mu > 1$, $s > 0$,

$(\alpha\lambda_{ii} - \alpha'\gamma'_{ii})\,N > \alpha M - \alpha' M'$, and $(\alpha'\lambda'_{ii} - \alpha\gamma_{ii})\,N > \alpha' M' - \alpha M$.

The symmetric case

In the case when the two populations evolve symmetrically, Model II reduces to

$$dp_i(t) = \sum_j m_{ij} \left(p_j(t) - p_i(t)\right) dt$$

$$+ s p_i(t) \left(1 - p_i(t)\right) \left(1 - 2p_i(t)\right) dt + \sqrt{\frac{1}{N} p_i(t) \left(1 - p_i(t)\right)}\, dW_i(t),$$

The symmetric case

In the case when the two populations evolve symmetrically, Model II reduces to

$$dp_i(t) = \sum_j m_{ij}\left(p_j(t) - p_i(t)\right)dt$$

$$+\, sp_i(t)\left(1 - p_i(t)\right)\left(1 - 2p_i(t)\right)dt + \sqrt{\frac{1}{N}p_i(t)\left(1 - p_i(t)\right)}\,dW_i(t),$$

For general s there is no convenient moment dual, but we find an alternative duality with a system of *branching annihilating random walks*.

Branching annihilating random walk

The Markov process $\{n_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$, in which $n_i(t) \in \mathbb{Z}_+$, with dynamics

$$\begin{array}{ll} \left\{\begin{array}{l} n_i \mapsto n_i - 1, \\ n_j \mapsto n_j + 1 \end{array}\right. & \text{at rate } n_i m_{ij} \\ n_i \mapsto n_i + m & \text{at rate } s n_i \\ n_i \mapsto n_i - 2 & \text{at rate } \frac{1}{2} n_i (n_i - 1) \end{array}$$

is called a *branching annihilating random walk with offspring number* m *and branching rate* s.

Branching annihilating random walk

The Markov process $\{n_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$, in which $n_i(t) \in \mathbb{Z}_+$, with dynamics

$$\begin{cases} n_i \mapsto n_i - 1, \\ n_j \mapsto n_j + 1 \end{cases} \quad \text{at rate } n_i m_{ij}$$

$$n_i \mapsto n_i + m \qquad \text{at rate } s n_i$$

$$n_i \mapsto n_i - 2 \qquad \text{at rate } \tfrac{1}{2} n_i (n_i - 1)$$

is called a *branching annihilating random walk with offspring number* m *and branching rate* s.

Duality: Set $w_i = 1 - 2p_i$ and let $\underline{n}_t$ be branching annihilating random walk with offspring number two, then for $s > 0$

$$\mathbb{E}\left[\underline{w}(t)^{\underline{n}(0)}\right] = \mathbb{E}\left[\underline{w}(0)^{\underline{n}(t)}\right].$$

Conjectures for Model II

Based on results of Cardy and Täuber, we conjecture:

For Model II with $\mu = 2$

- In $d = 1$, there is a critical value $s_0 > 0$ such that the populations will both persist for all time with positive probability if and only if $s > s_0$,
- In $d = 2$, there is positive probability that both populations will persist for all time if and only if $s > 0$,
- In $d \geq 3$, this probability is positive if and only if $s \geq 0$.

It would be odd if the case $\mu = 2$ were pathological.

Conjectures for Model I

Let $m_{ij} = m'_{ij}$, $\alpha = \alpha'$, $M = M'$ fixed, and $\lambda_{ij} = \lambda'_{ij}, \gamma_{ij} = \gamma'_{ij}$. Parameters such that each population can survive in absence of the other.

1. If $\lambda_{ij} < \gamma_{ij}$ for all j, then eventually only one population will be present.
2. If $\lambda_{ij} > \gamma_{ij}$ for all j, then if $d \geq 2$, with positive probability both populations will exihibit longterm coexistence.
   In one dimension the same result will hold true provided that $\lambda_{ij} - \gamma_{ij}$ is sufficiently large.
3. If $\lambda_{ij} = \gamma_{ij}$ and $d \geq 3$ positive probability coexistence.
   If $d \leq 2$ then with probability one, one population will die out.

Heteromyopia

Does space promote coexistence?

Heteromyopia

Does space promote coexistence? Murrell & Law 2003: asymmetry in interaction.

- Overall strength of interspecific and intraspecific competition is the same ($\sum_j \lambda_{ij} = \sum_j \gamma_{ij}$) but distance over which sense heterospecific neighbours (competitors) is shorter than that over which sense conspecific neighbours.

Heteromyopia

Does space promote coexistence? Murrell & Law 2003: asymmetry in interaction.

- Overall strength of interspecific and intraspecific competition is the same ($\sum_j \lambda_{ij} = \sum_j \gamma_{ij}$) but distance over which sense heterospecific neighbours (competitors) is shorter than that over which sense conspecific neighbours.

- Analogue in our setting: symmetric version of Model I with $\lambda_{ij} = \lambda\left(\|i-j\|\right)$, $\gamma_{ij} = \gamma\left(\|i-j\|\right)$, where the functions λ and γ are monotone decreasing and $\sum_j \lambda_{ij} = \sum_j \gamma_{ij}$, but the range of λ_{ij} is greater than that of γ_{ij}.

Heteromyopia

Does space promote coexistence? Murrell & Law 2003: asymmetry in interaction.

- Overall strength of interspecific and intraspecific competition is the same ($\sum_j \lambda_{ij} = \sum_j \gamma_{ij}$) but distance over which sense heterospecific neighbours (competitors) is shorter than that over which sense conspecific neighbours.

- Analogue in our setting: symmetric version of Model I with $\lambda_{ij} = \lambda\left(\|i-j\|\right)$, $\gamma_{ij} = \gamma\left(\|i-j\|\right)$, where the functions λ and γ are monotone decreasing and $\sum_j \lambda_{ij} = \sum_j \gamma_{ij}$, but the range of λ_{ij} is greater than that of γ_{ij}.

Small scales $\rightsquigarrow$ homozygous advantage.

Larger scales $\rightsquigarrow$ heterozygous advantage.

What about genetics?

Recall from 1st lecture that for a neutral subdivided population with allelic types a, A, the proportion of type a alleles is

$$dp_i(t) = \sum_j \frac{N_j}{N_i} m_{ij} \left(p_j(t) - p_i(t)\right) dt + \sqrt{\frac{\gamma}{N_i} p_i(t) \left(1 - p_i(t)\right)} dW_i(t).$$

What about genetics?

Recall from 1st lecture that for a neutral subdivided population with allelic types a, A, the proportion of type a alleles is

$$dp_i(t) = \sum_j \frac{N_j}{N_i} m_{ij} \left(p_j(t) - p_i(t) \right) dt + \sqrt{\frac{\gamma}{N_i} p_i(t) \left(1 - p_i(t) \right)} dW_i(t).$$

The genealogy of a sample from the population is given by a system of coalescing random walks in a *random environment*.

What about genetics?

Recall from 1st lecture that for a neutral subdivided population with allelic types a, A, the proportion of type a alleles is

$$dp_i(t) = \sum_j \frac{N_j}{N_i} m_{ij} \left(p_j(t) - p_i(t)\right) dt + \sqrt{\frac{\gamma}{N_i} p_i(t) \left(1 - p_i(t)\right)} dW_i(t).$$

The genealogy of a sample from the population is given by a system of coalescing random walks in a *random environment*. What do we need to know about $\{N_i\}$ (or its continuous counterpart) to make good approximations to the genealogy?

What about genetics?

Recall from 1st lecture that for a neutral subdivided population with allelic types a, A, the proportion of type a alleles is

$$dp_i(t) = \sum_j \frac{N_j}{N_i} m_{ij} \left(p_j(t) - p_i(t)\right) dt + \sqrt{\frac{\gamma}{N_i} p_i(t) \left(1 - p_i(t)\right)} dW_i(t).$$

The genealogy of a sample from the population is given by a system of coalescing random walks in a *random environment*. What do we need to know about $\{N_i\}$ (or its continuous counterpart) to make good approximations to the genealogy?

Hidden assumption: the population size in each deme is large.

An approach of Malécot

Discrete time:
Infinite alleles model, write $F(y)$ for the probability of identity in state of two genes separated by y.

- Ancestral lineages follow independent Brownian motions,
- *local population density*, δ, a constant,
- probability two lineages currently at separation y (a vector in $\mathbb{R}^2$ in the most interesting setting) have a common ancestor in the previous generation is $\frac{1}{\delta}\int g_1(y-z)g_1(z)dz$, where g_1 is a Gaussian density.

A recursion for identity

Writing k for the mutation probability

$$F(y) = (1-k)^2 \Big(\frac{1-F(0)}{\delta} \int g_1(y-z) g_1(z) dz \\ + \int g_1(x) g_1(x') F(y+x'-x) dx dx' \Big).$$

A recursion for identity

Writing k for the mutation probability

$$F(y) = (1-k)^2 \Big(\frac{1-F(0)}{\delta} \int g_1(y-z) g_1(z) dz \\ + \int g_1(x) g_1(x') F(y+x'-x) dx dx' \Big).$$

Continuous time

Many authors: lineages currently at separation y coalesce at instantaneous rate $\gamma g_\alpha(y)$.

A recursion for identity

Writing k for the mutation probability

$$F(y) = (1-k)^2\Big(\frac{1-F(0)}{\delta}\int g_1(y-z)g_1(z)dz \\ + \int g_1(x)g_1(x')F(y+x'-x)dxdx'\Big).$$

Continuous time

Many authors: lineages currently at separation y coalesce at instantaneous rate $\gamma g_\alpha(y)$.

Problem: There is no consistent forwards in time population model.

- No sampling consistency

Fourier transform

$$F(y) = (1-k)^2\Big(\frac{1-F(0)}{\delta}\int g_1(y-z)g_1(z)dz$$
$$+\int g_1(x)g_1(x')F(y+x'-x)dxdx'.$$

Writing $f(y)=\frac{1-F(0)}{\delta}\int g_1(y-z)g_1(z)dz$,

$$\tilde{F}(\tilde{y}) = \frac{1}{2\pi}\int e^{iy.\tilde{y}}F(y)dy.$$

$$\tilde{F}(\tilde{y}) = (1-k)^2\Big(\tilde{f}(\tilde{y}) + (2\pi)^2\tilde{g}_1(-\tilde{y})\tilde{g}_1(\tilde{y})\tilde{F}(\tilde{y})\Big)$$

Rearranging,

$$\tilde{F}(\tilde{y}) = \frac{(1-k)^2\tilde{f}(\tilde{y})}{1-(1-k)^2(2\pi)^2\tilde{g}_1(\tilde{y})\tilde{g}_1(-\tilde{y})}.$$

For $y \neq 0$

$$F(y) = \frac{(1-F(0))}{\delta}\int_0^\infty \frac{1}{4\pi\sigma^2 t}(1-k)^{2t}e^{-\|y\|^2/(4\sigma^2 t)}dt,$$

where σ^2 is the variance of the dispersal distribution g_1.

Write $(1-k) = e^{-\mu}$. Since

$$\int_0^\infty e^{-pt} t^{\nu-1} e^{-\frac{\alpha}{4t}} dt \;=\; 2\left(\frac{1}{4}\frac{\alpha}{p}\right)^{\nu/2} K_\nu(\sqrt{\alpha p}), Re\alpha \;>\; 0, Rep \;>\; 0,$$

$$F(y) = \frac{(1-F(0))}{\delta 2\pi\sigma^2} K_0\left(\frac{\|y\|}{\sigma}\sqrt{2\mu}\right).$$

Now assume a local scale κ over which $F(\kappa) \approx F(0)$. Using $K_0(z) \sim \log(1/z)$ as $z \to 0$

$$F(y) \approx \frac{1}{\mathcal{N} + \log(\sigma/\kappa\sqrt{2\mu})} K_0(\sqrt{2\mu}\|y\|/\sigma), \quad \|y\| > \kappa.$$

$\mathcal{N} = 2\pi\delta\sigma^2$ is Wright’s *neighbourhood size*.

Extending Malécot’s formula

Over sufficiently large scales populations may look approximately homogeneous.

Extending Malécot’s formula

Over sufficiently large scales populations may look approximately homogeneous. Assumptions:

- For large timesteps, temporal correlations are negligible.
- For well separated lineages:
  - the chances of coancestry are negligible,
  - movements of lineages uncorrelated,

Extending Malécot’s formula

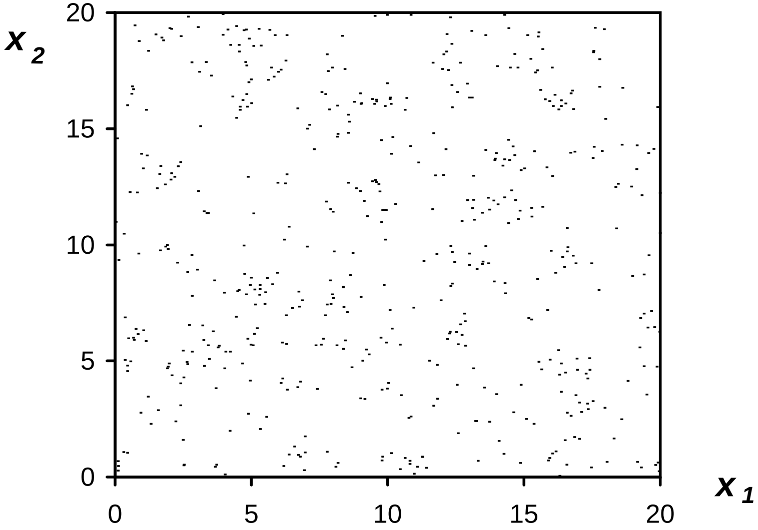
Over sufficiently large scales populations may look approximately homogeneous. Assumptions:

- For large timesteps, temporal correlations are negligible.
- For well separated lineages:
  - the chances of coancestry are negligible,
  - movements of lineages uncorrelated,

Then over all but small scales, Malécot’s formula remains valid if parameters replaced by *effective* parameters (dispersal rate, neighbourhood size and local scale).

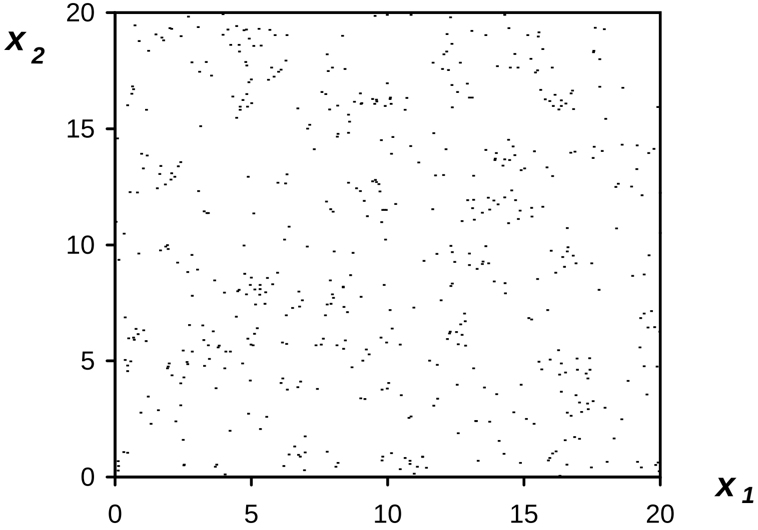
Some comments

Effective parameters may be hard to find.



Some comments

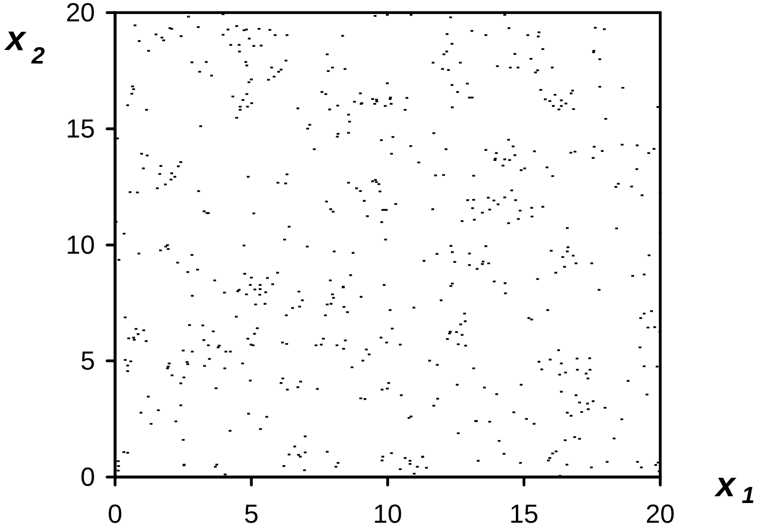
Effective parameters may be hard to find.



No explicit models for which we can calculate the parameters.

Some comments

Effective parameters may be hard to find.

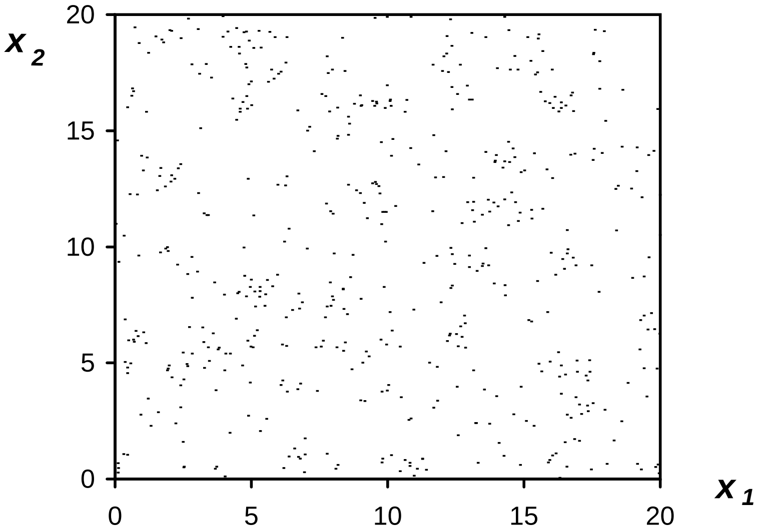


No explicit models for which we can calculate the parameters.

No extension to a sample of size $n > 2$.

Some comments

Effective parameters may be hard to find.



No explicit models for which we can calculate the parameters.
No extension to a sample of size $n > 2$.

... but anyway in a spatial continuum, neighbourhood size could be small and then pairwise coalescences may not dominate.