The Λ-asymmetric Moran model

Noemi Kurt (Goethe-Universität Frankfurt)

based on joint work with Adrián González Casanova and José Luis Peréz

Bath, June 2023

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Motivation: Leading questions

How to model populations with skewed offspring distribution in the presence of selection?

How can we understand the selective advantage of one subpopulation with respect to another, if they have very different reproductive behaviour?

Motivation: A quick and incomplete look at the literature

- Standard and very classical population genetics models: Kingman-coalescent universality class: Wright-Fisher model, Moran model (including selection)
- Populations with skewed offspring distribution: Genealogy follows multiple merger coalescents ($\Lambda-$coalescents)?
  - [Pitman 99, Sagitov 99] introduced Λ-coalescents
  - For some discussion on the role of Λ-coalescents in modelling skewed offspring distributions see e.g. [Eldon and Wakeley 2007, Birkner and Blath 2009] (and many others)
  - For recent results and state of the art see [Árnason, Koskela, Halldórsdóttir, Eldon 2023]
- Moran Model with Λ-type selection [Etheridge, Griffiths, Taylor 2010]
- Subpopulations with different reproductive mechansism [Gillespie 1973, 1974]

The classical continous time Moran model

- Population of fixed population size N
- Overlapping generations, continuous time
- Each individual independently has an exponential clock, when it rings, the individual reproduces, and the (unique) child replaces a uniformly chosen individual in the population.
- Assuming there are two inheritable types $-, +$ of individuals, we can count the relative frequency of type $-$ at time t. The frequency process $X_t, t \geq 0$ converges (after rescaling by N) to the Wright-Fisher diffusion.
- The geneaolgy of the Moran model is given by the Kingman coalescent.
- Classical Moran model with selection: Individuals of type $-$ reproduce at rate 1, individuals of type $+$ reproduce at rate $1 + s_N, s_N > 0$.
- Genealogy of the Moran model with selection is provided by the ancestral selection graph, [Krone and Neuhauser 1997].

$\Lambda-$asymmetric Moran model

- Fixed population size N, continous time
- Two types $-,+$, where $+$ has a selective advantage over $-$ (to be explained later)
- Two finite measures Λ^-, Λ^+ on $[0,1]$ governing reproduction
- An individual of type $-$ reproduces independently of everybody else at rate $N^{-1}\|\Lambda^-\|$. Upon reproduction, a random number $Y^- \in [0,1]$ distributed according to $\Lambda^-/\|\Lambda^-\|$ determines the number of offspring in the following way: Each of the $N-1$ non-reproducing individuals dies independently with probability Y^- and is replaced by a child of the reproducing individual.
- An individual of type $+$ reproduces in an analogous way, with Λ^+ instead of Λ^-.
- Types are inherited.

Λ−asymmetric Moran model

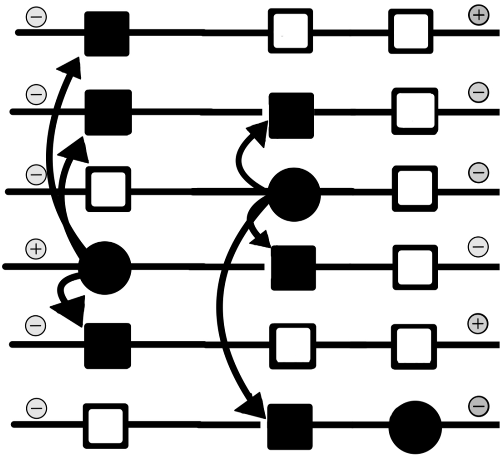


Figure: A realisation of the Λ-asymmetric frequency process. Filled dots represent the reproducing individuals, filled squares the offspring.

Frequency process

The number of offspring at a reproductive event is binomial with parameters Y^- resp. Y^+ and $N-1$.

Denote by X_t^N the relative frequency at time t of individuals of type $-$. This gives a continuous time Markov chain with state space $\{0, 1/N, ..., (N-1)/N, 1\}$ and transitions

$$x \mapsto \begin{cases} x + \frac{k}{N} & \text{at rate } x \int_0^1 \binom{(1-x)N}{k} y^k (1-y)^{(1-x)N-k} \Lambda^-(dy), \\ x - \frac{k}{N} & \text{at rate } (1-x) \int_0^1 \binom{xN}{k} y^k (1-y)^{xN-k} \Lambda^+(dy). \end{cases}$$

Generator:

$$\begin{aligned} \mathcal{B}^N f(x) =& x\|\Lambda^-\|\mathbb{E}\left[f\left(x + \frac{1}{N}\text{Binom}(N(1-x), Y^-)\right) - f(x)\right] \\ &+ (1-x)\|\Lambda^+\|\mathbb{E}\left[f\left(x - \frac{1}{N}\text{Binom}(Nx, Y^+)\right) - f(x)\right]. \end{aligned}$$

Expectation is taken with respect to the random variables Y^- resp. Y^+, distributed according to $\Lambda^-/\|\Lambda^-\|$ resp. $\Lambda^+/\|\Lambda^+\|$.

Where is selection in this model?

To say that one type has a selective advantage over the other only makes sense if there is some kind of order that allows to compare the measures Λ^- and Λ^+. For example, we would like to treat (at least) the two cases:

1. (Faster reproduction) if $\Lambda^+ = (1+\alpha)\Lambda^-$ for some $\alpha > 0$.
2. (Bigger reproductive events) There exists a function $s : [0,1] \mapsto [0,1]$ such that $s(x) - x \geq 0$ and $\Lambda^-(s(A)) = \Lambda^+(A)$.

In the first case we have in particular $\|\Lambda^+\| = (1+\alpha)\|\Lambda^-\|$. An example of the second case is $\Lambda^- = \delta_a$ and $\Lambda^+ = \delta_b$, with $0 \leq a \leq b \leq 1$.

(Definition) In general we say that $\Lambda^- \leq \Lambda^+$ in the partial order of adaptation if $\Lambda^-[x,1] \leq \Lambda^+[x,1]$ for every $x \in [0,1]$.

Both of the above cases are covered by this.

The magical coupling

Coupling Lemma

Let $\Delta = \{(y,z) \in [0,1]^2 : y+z \in [0,1]\}$ and consider two finite measures Λ^+, Λ^- on $[0,1]$. If $\Lambda^- \leq \Lambda^+$ then there exists a finite measure Λ^1 on Δ and two finite measures $\Lambda^{+,1}$ and $\Lambda^{+,2}$ on $[0,1]$ such that $\Lambda^+ = \Lambda^{+,1} + \Lambda^{+,2}$, and such that the following are satisfied:

- $\Lambda^-(A) = \Lambda^1(\{(y,z) : y \in A\})$ for any $A \in \mathcal{B}([0,1])$.
- $\Lambda^{+,1}(A) = \Lambda^1(\{(y,z) : y+z \in A\})$ for any $A \in \mathcal{B}([0,1])$.
- $\Lambda^+(A) = \Lambda(\{(y,z) : y+z \in A\})$, where the measure Λ on Δ is defined by

$$\Lambda(dy,dz) = \Lambda^1(dy,dz) + \delta_0(dy) \otimes \Lambda^{+,2}(dz).$$

In particular, if $\|\Lambda^-\| = \|\Lambda^+\|$, then we can take $\Lambda^+ = \Lambda^{+,1}$, $\Lambda = \Lambda^1$, and the measure ρ on $[0,1]^2$ defined by

$$\rho(A \times B) = \Lambda(\{(y,z) : y \in A, y+z \in B\}), \qquad A, B \in \mathcal{B}([0,1]),$$

is a coupling of Λ^- and Λ^+ such that $\rho\{(y,z) : y > z\} = 0$.

Remark: There is a nice connection to the theory of optimal transport.

Applying the coupling

For any measurable function $f : [0,1] \mapsto [0,1]$ such that $f(0) = 0$,

$$\int_\Delta f(y)\Lambda(dy,dz) = \int_{[0,1]} f(y)\Lambda^-(dy), \text{ and } \int_\Delta f(y+z)\Lambda(dy,dz) = \int_{[0,1]} f(z)\Lambda^+(dz).$$

Therefore the generator of the frequency process of the $\Lambda-$asymmetric Moran model becomes

$$\mathcal{B}^N f(x) = x \int_\Delta \mathbb{E}\left[f\left(x + \frac{1}{N}\text{Binom}(N(1-x),y)\right) - f(x)\right]\Lambda(dy,dz)$$
$$+ (1-x)\int_\Delta \mathbb{E}\left[f\left(x - \frac{1}{N}\text{Binom}(Nx, y+z)\right) - f(x)\right]\Lambda(dy,dz).$$

We only need one measure Λ now, not two.

Applying the coupling

- Construction of the Λ-asymmetric ancestral selection graph
- Scaling limits of the forward and backward processes
- Griffiths' representation for the fixation probabilities

Reconsidering the frequency process

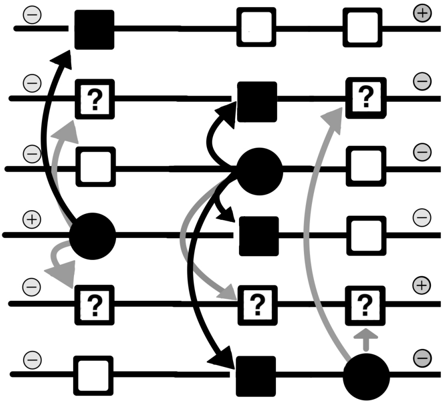


Figure: The realisation of the Λ-asymmetric frequency process from before, now in terms of the coupling construction. Black arrows occur with probability y, grey ones with probability z sampled according to Λ. Individuals of type $+$ may reproduce through any arrow, individuals of type $-$ only through black arrows.

The $\Lambda-$asymmetric ancestral selection graph

$\Lambda-$asymmetric ancestral selection graph, ASG

Consider a Poisson processes M^N with values in $\mathbb{R}_+ \times [0,1] \times [N] \times \Delta \times [0,1]^N$ and intensity measure $dt \times dm \times \Lambda(dy, dz) \times du_1 \times du_2 ... \times du_N$, where dm denotes the uniform measure on $[N]$. Each point $(t, i) \in \mathbb{R} \times [N]$, represents the i-th individual alive at time t.

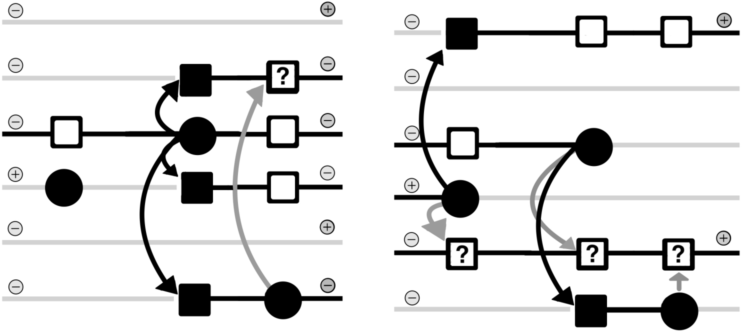
- We say that at time t there is a *neutral arrow* between i and j if there is a point $(t, i, y, z, u_1, u_2, ..., u_N) \in M^N$ such that $u_j \in [0, y]$.
- We say that at time t there is a *selective arrow* between i and j if there is a point $(t, i, y, z, u_1, u_2, ..., u_N) \in M^N$ such that $u_j \in [0, y + z]$.

The ancestral selection graph is then given by $(\mathbb{R}_+ \times [N], M^N)$.

This is the formal construction behind the previous figure.

Ancestral process

From the ancestral selection graph resp. its graphical representation we can consider the ancestry of a sample (going backward in time).



We denote by A_t^N the number of potential ancestors of a sample of n individuals, at backward time t.

Ancestral process

$(A^N_t)_{t\geq 0}$ is a continuous-time Markov chain with values in $[N]$ starting at $A^{N,T}_0 = n$ and transition rates

$$n \mapsto \begin{cases} n-k & \text{at rate } \frac{n}{N}\int_\Delta \binom{n-1}{k} y^k(1-y)^{n-1-k}\Lambda(dy,dz), \quad k=1,...,n-1 \\ n-k+1 & \text{at rate } (1-\frac{n}{N})\int_\Delta \binom{n}{k} y^k(1-y)^{n-k}\Lambda(dy,dz), \quad k=2,...,n \\ n+1 & \text{at rate } (1-\frac{n}{N})\int_\Delta [(1-y)^n-(1-y-z)^n]\Lambda(dy,dz). \end{cases}$$

Duality

The processes $(X^N_t)_{t\geq 0}$ and $(A^N_t)_{t\geq 0}$ are dual with respect to the sampling function $S_0(x,n) = \prod_{i=1}^n \frac{Nx+1-i}{N+1-i}$, that is,

$$\mathbb{E}_x[S_0(X^N_t,n)] = \mathbb{E}_n[S_0(x,A^N_t)] \quad \forall t\geq 0, x\in[N_0]/N, n\in\mathbb{N}.$$

Open Problems

1. Relatively straightforward: Include mutation:
   - Which types of mutation, how to include in the $\Lambda-$asymmetric Moran model
   - Equilibrium frequencies in the model with mutation
   - Stationary distribution, characterisation
2. A more substantial project: Statistical aspects, test for selection etc.
3. Can you detect Λ-selection
4. What else could be considered: Special cases, limits as rates tend to 0 or ∞, connections to other models...

Thank you - and let's get started!