TÉCNICO LISBOA

Neutral population-genetics evolution for Wright-Fisher and Moran models
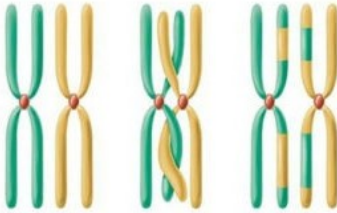
Leonor Botelho de Sousa Barata

December 12, 2019
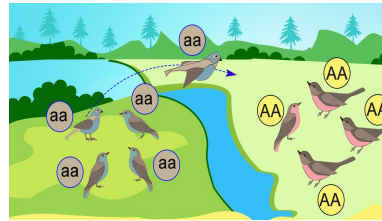
Population genetics: an overview

Population genetics: study genetic differences that influence single or multiple populations.
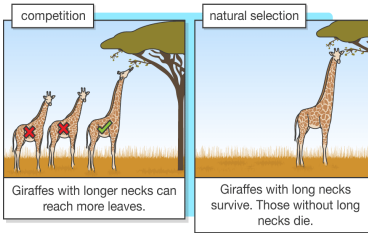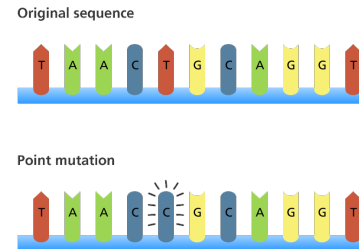
Recombination

Migration

Selection

Mutation

Neutrality: Survival of a given gene is by chance, instead of the *stronger* gene (selected by nature) prevailing in the population.

Wright-Fisher model

Assumptions

- **Haploid** population: Type of individual $=$ Allele type;
- Population size is constant and equal to N;
- Only 2 types of individuals: A and a;
- **Random reproduction**: Each individual of the offspring selects **randomly** the parent from the previous generation and adopts the type of the parent, **independently** of the other individuals.
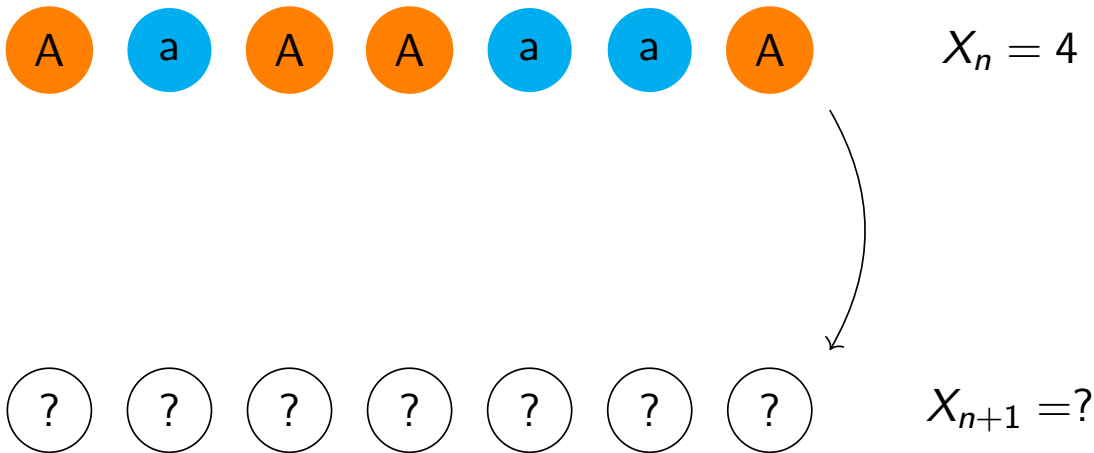
$X_n =$ Number of individuals of type A in the nth generation.
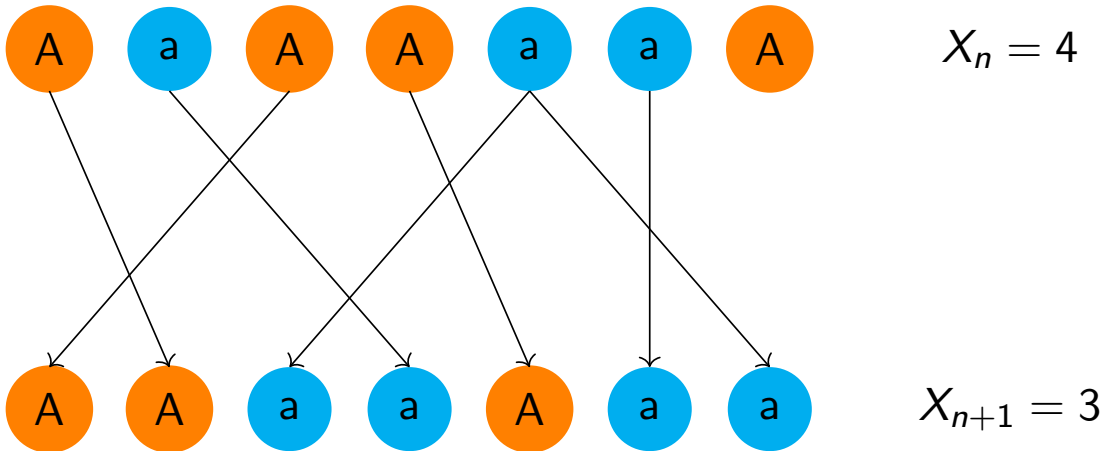
Wright-Fisher model

$X_n = 4$

Wright-Fisher model

$X_n = 4$

$X_{n+1} = ?$

Wright-Fisher model

$X_n = 4$

$X_{n+1} = 3$

In each generation the whole population is replaced: **non-overlapping generations**.

Wright-Fisher model

- $p_{i,j} := \mathbb{P}(X_{n+1} = j | X_n = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$;

- $(X_{n+1} | X_n = i) \sim Binomial\left(n = N, p = \frac{i}{N}\right)$;

- $\{X_n\}_{n \in \mathbb{N}}$ is a homogeneous DTMC with state space $S = \{0, 1, \ldots, N-1, N\}$;

- 0 and N are absorbing: lost of genetic variability.

Moran model

Assumptions

- **Haploid** population of constant size N;
- Only 2 types of individuals: A and a;
- **Continuous**-time process;
- The reproduction rate of each individual is 1;
- **Random reproduction**: At each transition, one individual is chosen to die and replaced by an existing one.

$$X_t = \text{ Number of individuals of type } A \text{ at time } t.$$

Moran model

$\{X_t\}_{t \geq 0}$ is a homogeneous CTMC with state space $S = \{0, 1, \ldots, N-1, N\}$.

Rates of the process

- $b_i = q_{i,i+1} = i\frac{N-i}{N}$;
- $d_i = q_{i,i-1} = (N-i)\frac{i}{N}$;
- $b_i = d_i$.

0 and N are absorbing: lost of genetic variability.

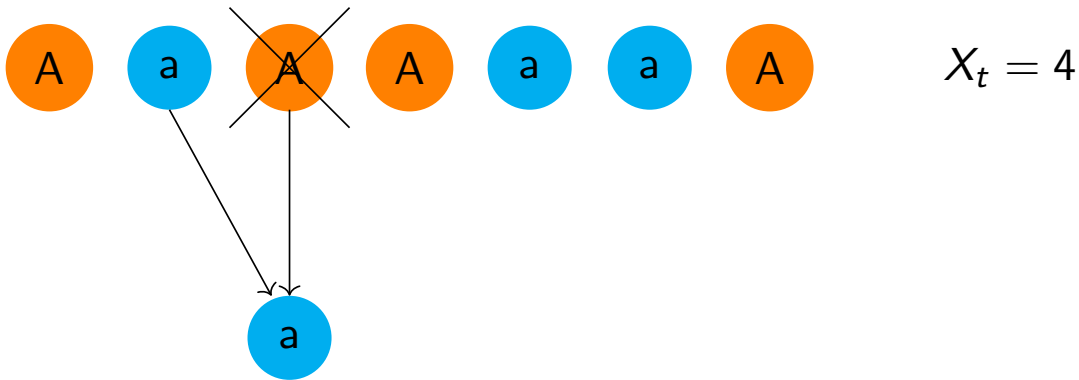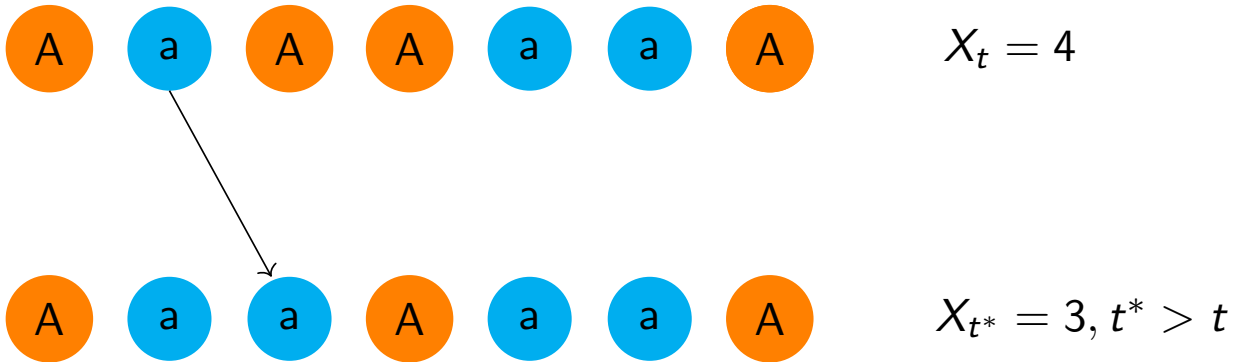# Moran model

$X_t = 4$

Moran model

$X_t = 4$

Exponential clock for a transition to occur

Moran model

$X_t = 4$

Moran model

$X_t = 4$

$X_{t^*} = 3, t^* > t$

The population is not entirely replaced: **overlapping generations**.

Parallel: Wright-Fisher vs Moran model

	Wright-Fisher	Moran
Time	**Discrete**-time	**Continuous**-time
Generations	**Non**-overlapping	Overlapping
Martingale	Yes	Yes
Fixation time τ finite a.s.	$\mathbb{P}(\tau < \infty) = 1$	$\mathbb{P}(\tau < \infty) = 1$
Fixation probability of type A	$\mathbb{P}_i(X_\tau = N) = \frac{i}{N}$	$\mathbb{P}_i(X_\tau = N) = \frac{i}{N}$
Expected fixation time of either type A or type a, $p = \frac{i}{N}$	$\mathbb{E}_i(\tau) = -2N\Big(p\log(p) + (1-p)\log(1-p)\Big)$	$\mathbb{E}_i(\tau) = -N\Big(p\log(p) + (1-p)\log(1-p)\Big)$
Conditional expected time to fixate type A, $p = \frac{i}{N}$	$\mathbb{E}_i(\tau \mid T_N < T_0) = -2N\frac{1-p}{p}\log(1-p)$	$\mathbb{E}_i(\tau \mid T_N < T_0) = -N\frac{1-p}{p}\log(1-p)$
Conditional expected time to fixate type a, $p = \frac{i}{N}$	$\mathbb{E}_i(\tau \mid T_0 < T_N) = -2N\frac{p}{1-p}\log(p)$	$\mathbb{E}_i(\tau \mid T_0 < T_N) = -N\frac{p}{1-p}\log(p)$
Expected heterozygosity, $H_n = \frac{X_n}{N}\frac{N-X_n}{N-1}$	$\mathbb{E}(H_n) = (1 - \frac{1}{N})^n \mathbb{E}(H_0)$ $\approx e^{-\frac{n}{N}}\mathbb{E}(H_0)$	$\mathbb{E}(H_t) = e^{-\frac{2t}{N}}\mathbb{E}(H_0)$

Infinite population limit

- Expected time to reach fixation given initial state $X_0 = i$,

$$\mathbb{E}_i(\tau)$$

  is of **order of the population size N** in both processes.

- Motivates the transformation of the Wright-Fisher and Moran processes

$$\tilde{X}_t^N = \frac{X_{\lceil Nt \rceil}}{N}.$$

Infinite population limit

- The Wright-Fisher diffusion, $\tilde{X}_t$, is an Itô diffusion process satisfying strongly the SDE:

$$d\tilde{X}_t = \sqrt{\tilde{X}_t(1 - \tilde{X}_t)}dW_t, t \geq 0$$

  and $\left(\mathcal{L}_{\tilde{X}_t}f\right)(x) = \frac{1}{2}x(1-x)f''(x)$.

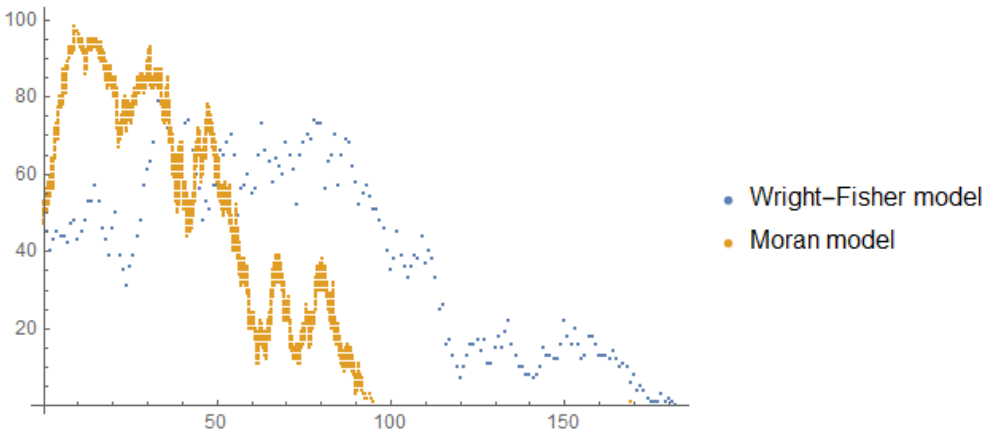 Proving **weak convergence** of the generators:

- For the Wright-Fisher transformation,
  $\left(\mathcal{L}_{\widetilde{WF}}f\right)(x) \rightarrow \frac{1}{2}x(1-x)f''(x) = \left(\mathcal{L}_{\tilde{X}_t}f\right)(x)$.

- For the Moran transformation,
  $\left(\mathcal{L}_{\widetilde{M}}f\right)(x) \rightarrow x(1-x)f''(x) = 2\left(\mathcal{L}_{\tilde{X}_t}f\right)(x)$.

- Moran model evolves **twice** as fast comparing to the Wright-Fisher model.
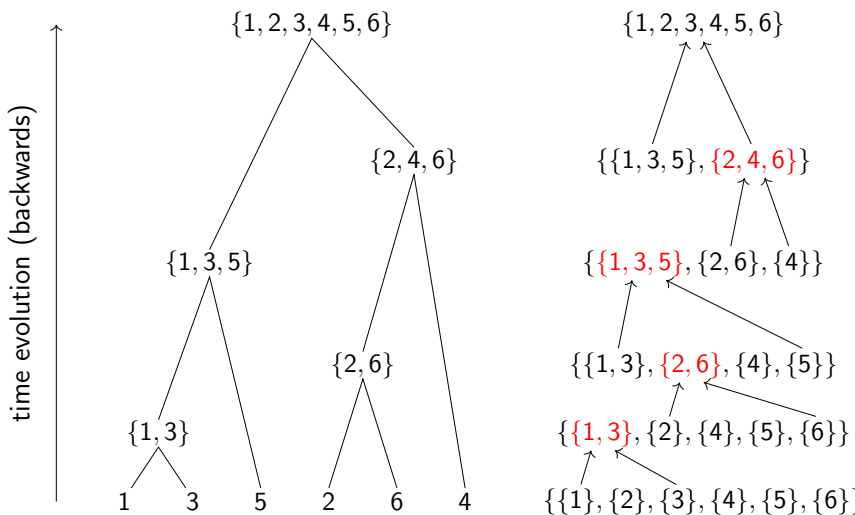
Wright-Fisher vs Moran model

Wright-Fisher vs Moran simulation with $N = 100, X_0 = 50$.

The n-coalescent

n-coalescent: CTMC which allows as a genealogical description from a **backward-time perspective** of a population with finite size equal to n. In particular, it allows us to identify the MRCA.

Figure 1: Example of a realization of a n-coalescent, with $n = 6$.

- Each step: an element of the collection of all partitions of $\{1, \ldots, n\}$;

- 1 coalescent event per transition;

- Rate of transition from i to $i - 1$: $\binom{i}{2}$.

The n-coalescent and the Kingman coalescent

- Kingman (1982)

  n-coalescent $\begin{cases} \text{jump chain: which are the lineages} \\ \text{pure-death process: number of lineages} \end{cases}$

- **Kingman coalescent**: limit $n \to \infty$ of the n-coalescent, allows the genealogical description of a infinite size population.

- *Coming down from infinity*: the Kingman coalescent will always reach a finite number of lineages.

# Kingman coalescent and its suitability for Wright-Fisher and Moran models
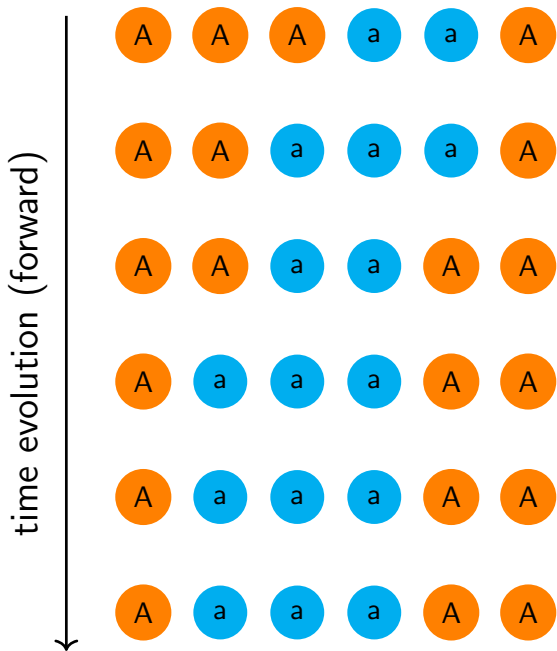
Kingman coalescent for the Wright-Fisher model

When the time is accelerated by a factor of N, the Kingman coalescent describes the genealogy of a sub-population of fixed size n under the Wright-Fisher model, where the total population size is N and $n << N$.
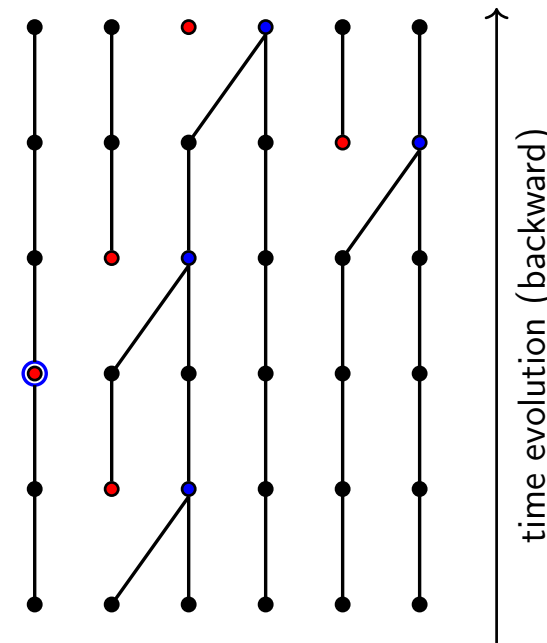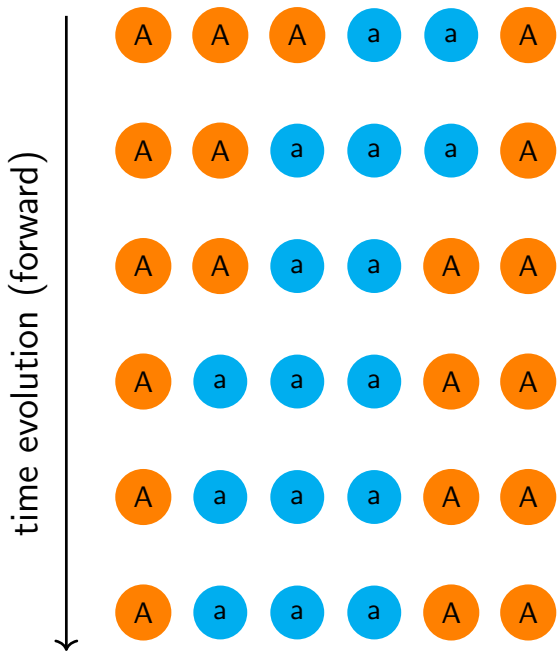
Kingman coalescent for the Moran model

The Kingman coalescent can be used to describe the genealogy of a sub-population of fixed size n in a population of size N described by the Moran model, when time is accelerated by a factor $\frac{N}{2}$.

Genealogy of Wright-Fisher
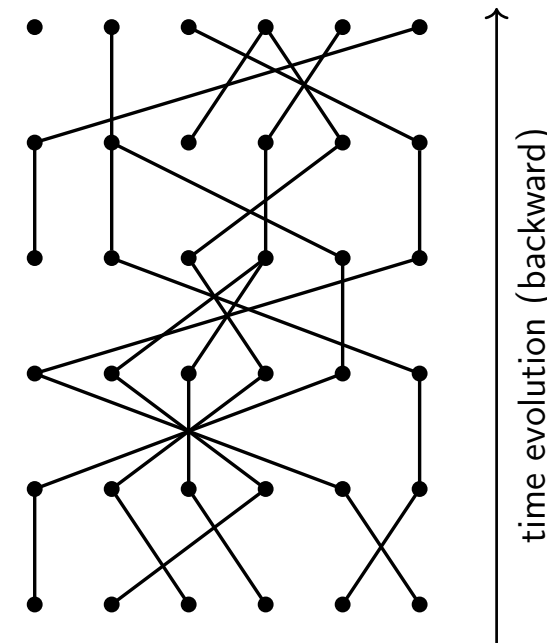
Genealogy of Moran model

Duality between the pure-death process of Kingman coalescent and Wright-Fisher diffusion

Duality allows us to compute genetic measures on the Wright-Fisher diffusion from the pure-death process of Kingman coalescent.

Fixation probability

The fixation probability of type A in the Wright-Fisher diffusion, knowing that the initial fraction of the same type is $x \in [0, 1]$, is given by:
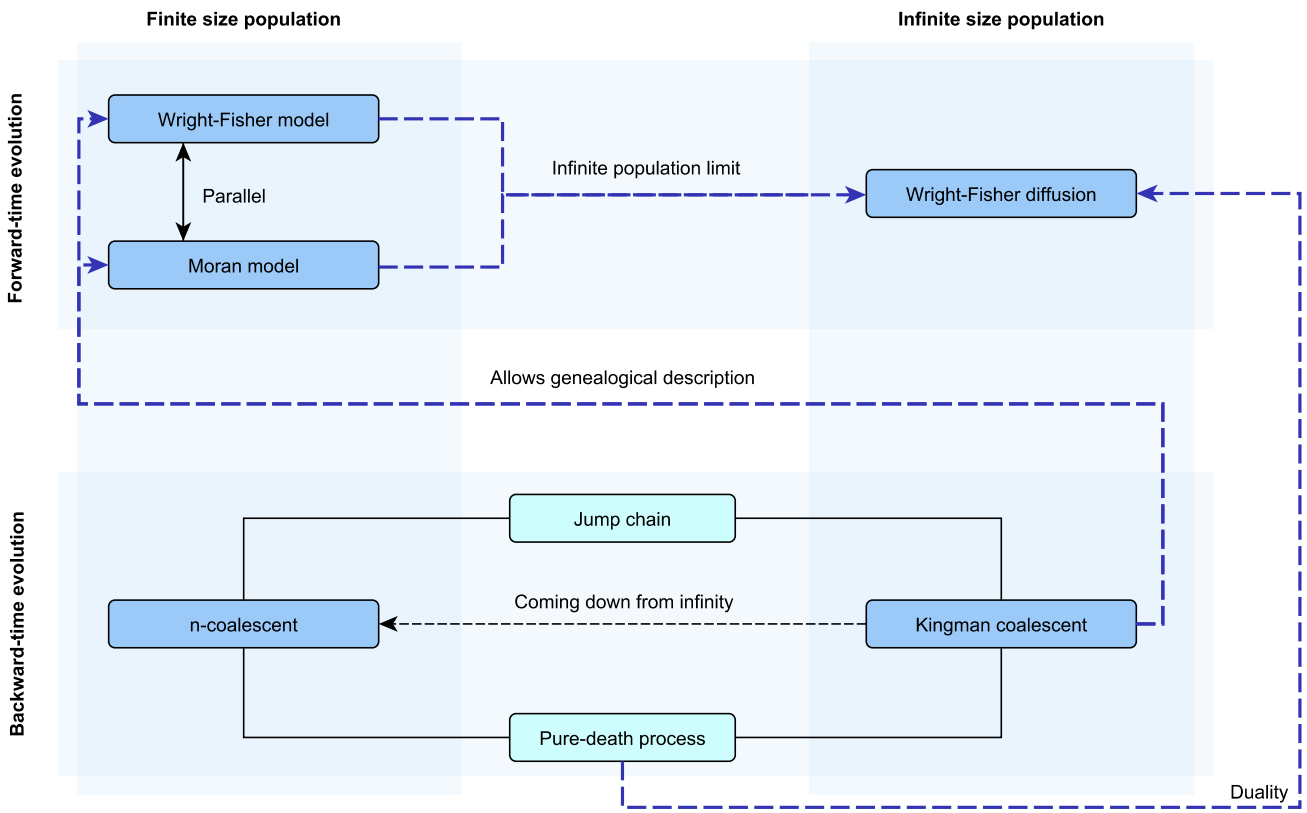
$$\mathbb{P}_x(X_{\tau^*} = 1) = x.$$

Expected heterozygosity

For the Wright-Fisher diffusion, we have the following relation between the expected value of the heterozygosity as a function of its initial value:

$$\mathbb{E}(H_t) = e^{-t}\mathbb{E}(H_0).$$

Conclusions of the work in a diagram

Future work

- Consider general k allele types, instead of just 2 types;
- Consider both mutation and selection parameters simultaneously;
- Adapt resampling for polyploid individuals;
- Consider a more realistic variable population size $N(t)$ that evolves in time, instead of a fixed one.

References

[1] L. Avena, C. da Costa, and F. den Hollander. Stochastic models for genetic evolution, 2019.

[2] C. Darwin. *On the origin of species*. John Murray, 1859.

[3] R. A. Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, 1930.

[4] M. Kimura. *The neutral theory of molecular evolution*. Cambridge University Press, 1983.

[5] M. Kimura et al. Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626, 1968.

[6] J. F. Kingman. On the genealogy of large populations. *Journal of applied probability*, 19:27–43, 1982.

[7] J. F. C. Kingman. The coalescent. *Stochastic processes and their applications*, 13(3):235–248, 1982.

[8] G. Mendel. *Experiments in plant hybridisation*. Harvard University Press, 1965.

[9] P. A. P. Moran. Random processes in genetics. In *Mathematical proceedings of the Cambridge philosophical society*, volume 54, pages 60–71. Cambridge University Press, 1958.

[10] T. G. K. Stewart N. Ethier. *Markov Processes: Characterization and Convergence.* John Wiley & Sons, 1986.

[11] S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97, 1931.