# The Seed Bank Model

Joint work with J. Blath, N. Kurt and D. Spanò

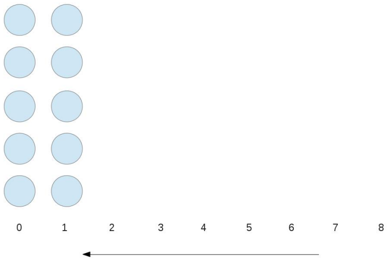
RTG1845
Berlin Mathematical School
Technische Universität Berlin
UNAM

05-04-2013

# Content

1 Introduction
2 On the ancestral process of long-range seed bank models
3 Application to Biology

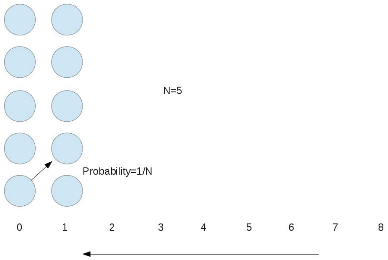
# The Wright-Fisher Model




## Description

- A generation consists of $N$ individuals. Each individual in generation $i$ selects a parent uniformly in the generation $i-1$.

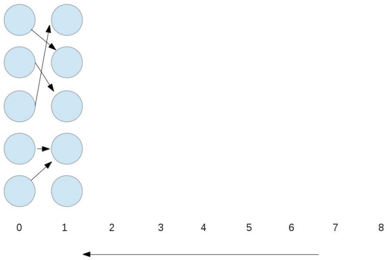
# The Wright-Fisher Model




## Description

- A generation consists of $N$ individuals. Each individual in generation $i$ selects a parent uniformly in the generation $i-1$.

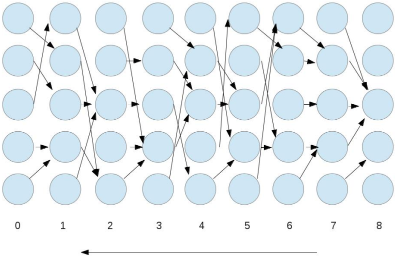
# The Wright-Fisher Model




## Description

- A generation consists of $N$ individuals. Each individual in generation $i$ selects a parent uniformly in the generation $i-1$.

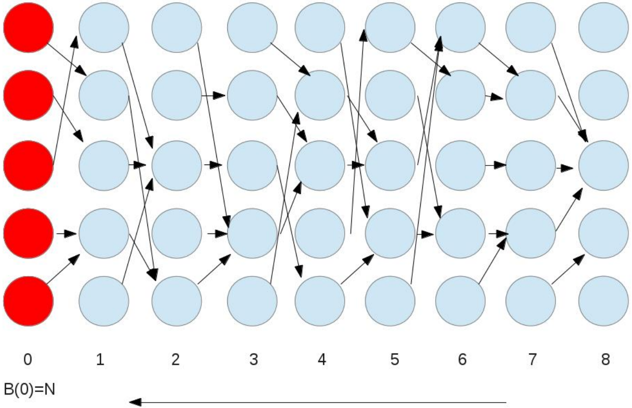
# The Wright-Fisher Model



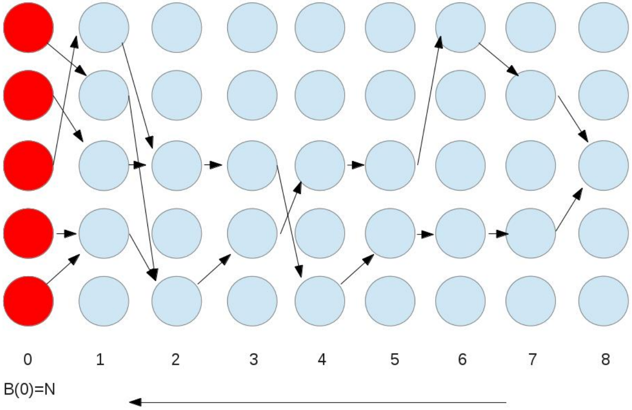

## Description

- A generation consists of $N$ individuals. Each individual in generation $i$ selects a parent uniformly in the generation $i-1$.

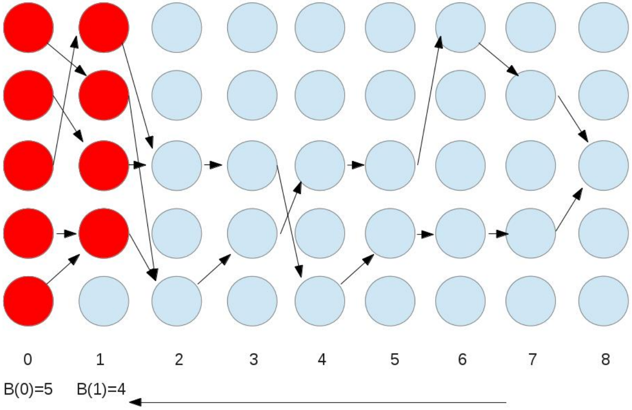
# The Wright-Fisher Model



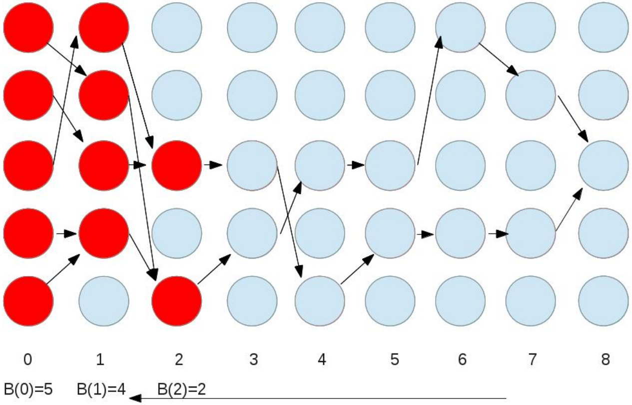
# The Wright-Fisher Model



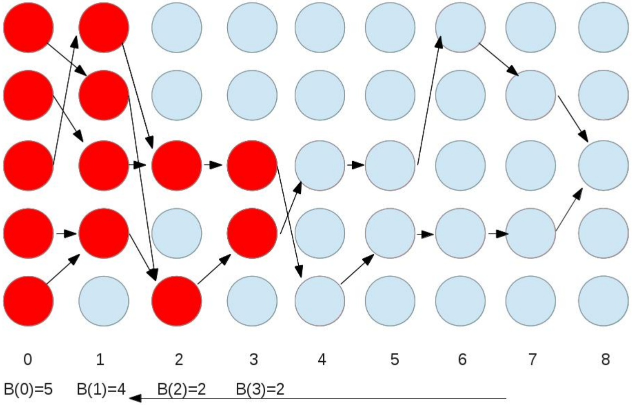
# The Wright-Fisher Model



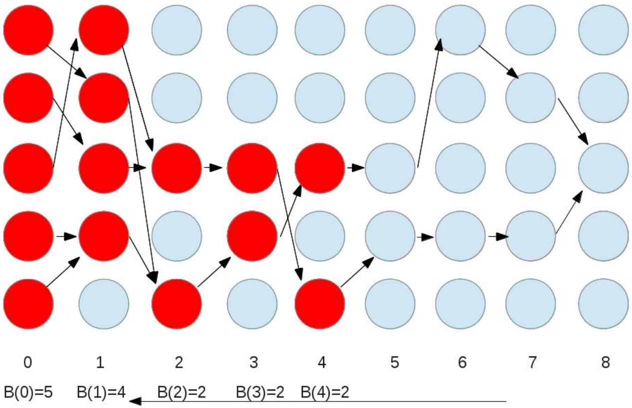
# The Wright-Fisher Model



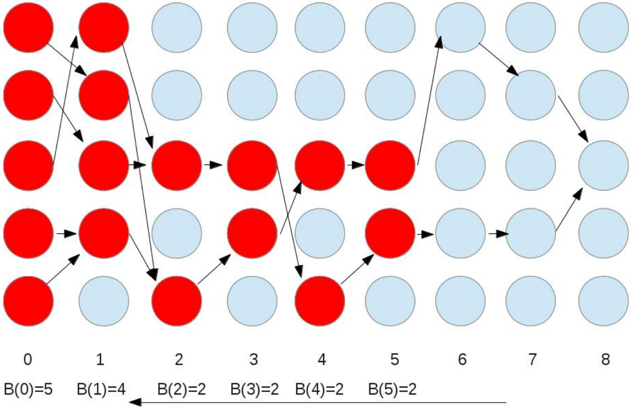
# The Wright-Fisher Model



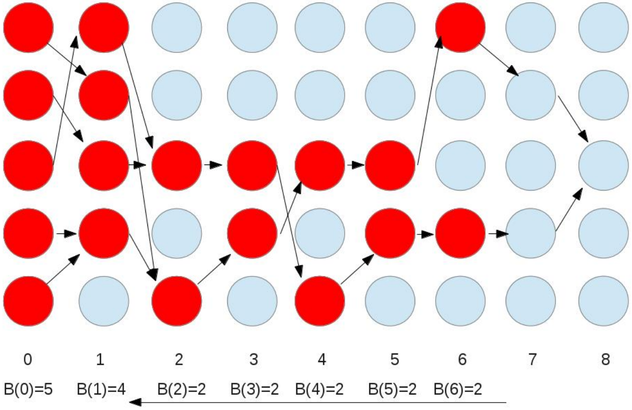
# The Wright-Fisher Model



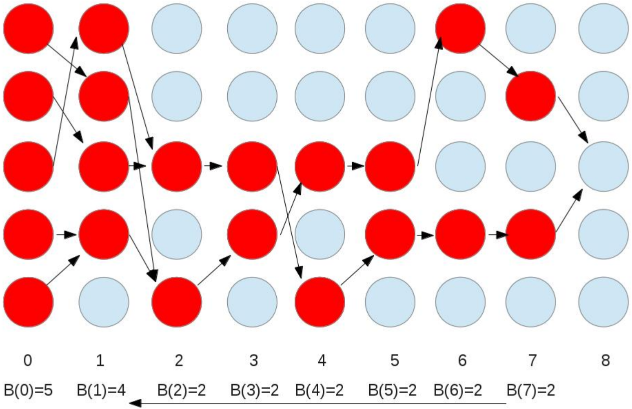
# The Wright-Fisher Model



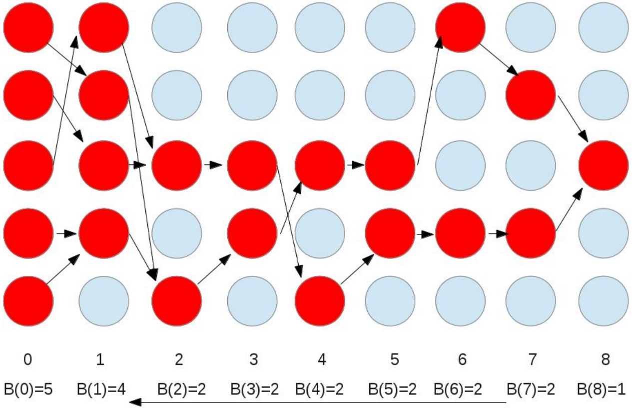
# The Wright-Fisher Model



# The Wright-Fisher Model



# The Wright-Fisher Model



# Scaling

Fix a sample of size n. Let the number of individuals go to infinity.
Measure the time in terms of the number of individuals per generation.

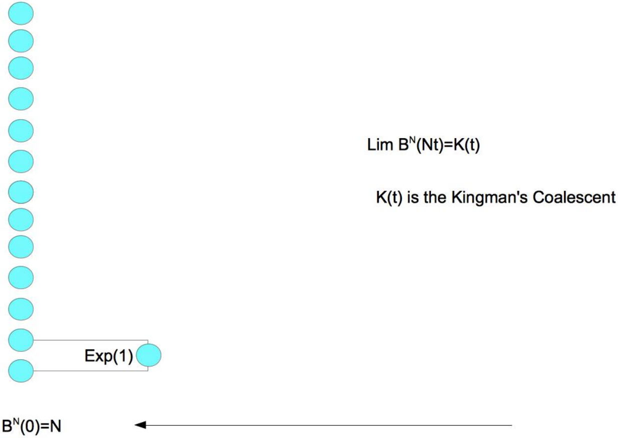
$$B^N([Nt]) \Rightarrow K(t)$$

# (The block counting process of) The Kingman coalescent

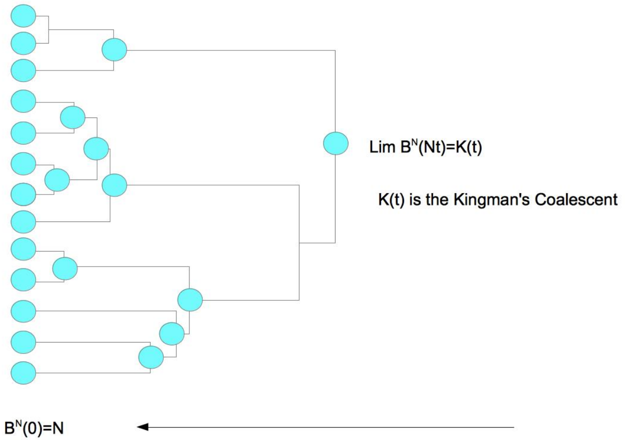
**Description**

Each pair of blocks coalesce at rate 1, independently of the others.

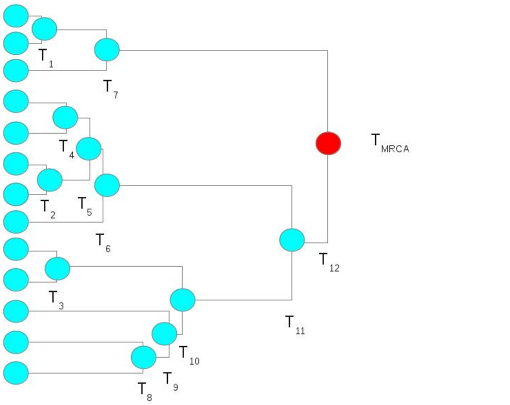
# (The block counting process of) The Kingman coalescent



# (The block counting process of) The Kingman coalescent



# Time to the most recent common ancestor



# Time to the most recent common ancestor

$$E[T_{MRCA}] = \sum_{i=1}^{n-1} E[T_i] = \sum_{i=1}^{n-1} \frac{1}{\binom{n+1-i}{2}} = \sum_{u=2}^{n} \frac{2}{u(u-1)} = 2\big(1 - \frac{1}{n}\big)$$

# The work of Kaj, Krone and Lascoux (2001)

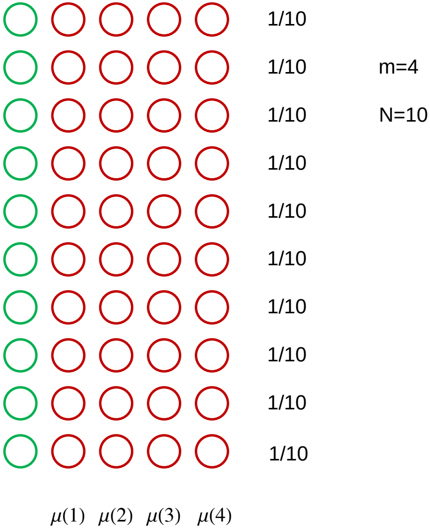
- The seed bank model introduced by Kaj, Krone and Lascoux is a generalization of the Wright-Fisher model. Its biological motivation are species that reproduce using seeds. (Like a cactus.)

# The work of Kaj, Krone and Lascoux (2001)

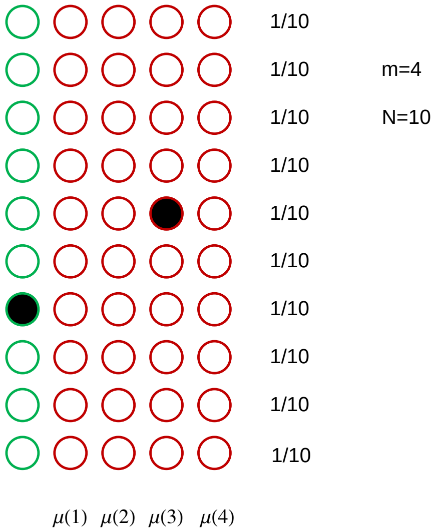
- The seed bank model introduced by Kaj, Krone and Lascoux is a generalization of the Wright-Fisher model. Its biological motivation are species that reproduce using seeds. (Like a cactus.)
- Dynamics: Let $\mu$ be a bounded measure on $\mathbb{N}$. Each individual selects its parent independently by the following 2 steps:
  1. Select the generation of the parent by performing a $\mu$ distributed jump.
  2. Select a parent uniformly among the members of the selected generation.
- Problem: We lose the Markov property.

m=4

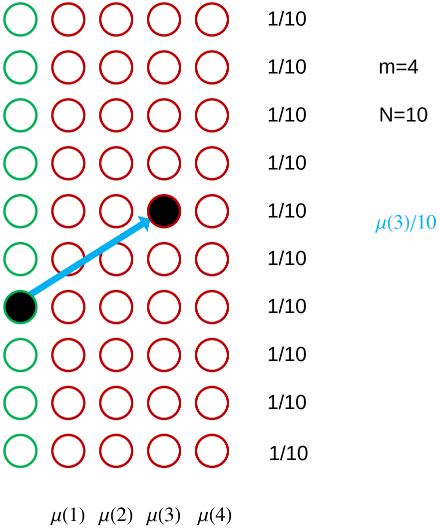
N=10


1/10
1/10
1/10
1/10
1/10
1/10
1/10
1/10
1/10
1/10
m=4
N=10
$\mu(1)$
$\mu(2)$
$\mu(3)$
$\mu(4)$


1/10
1/10
1/10
1/10
1/10
1/10
1/10
1/10
1/10
1/10
m=4
N=10
$\mu(1)$ $\mu(2)$ $\mu(3)$ $\mu(4)$


1/10
1/10
m=4
1/10
N=10
1/10
1/10
1/10
1/10
1/10
1/10
1/10
$\mu(1)$ $\mu(2)$ $\mu(3)$ $\mu(4)$

## The Seed Bank Model

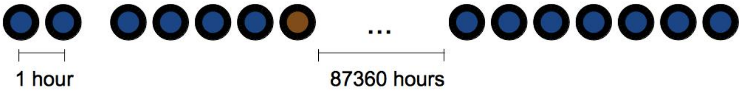


# The work of Kaj, Krone and Lascoux (2001)

- The ancestral process can be described in terms of a finite state Markov Chain.
- Main result: The scaling limit is the Kingman coalescent, under a constant time change.
- Limitation: $\mu$ must be bounded.

# The seed bank model with long-range dependence

- What happens if we remove the boundedness condition of the jump measure $\mu$ in the seed bank model?
- Motivation.



# The seed bank model with long-range dependence

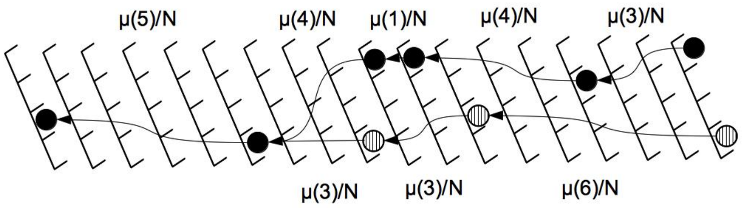
- Answer: It depends on $\mu$.
- We say that $\mu \in \Gamma_\alpha$, if

$$\mu(\{n, \ldots\}) = n^{-\alpha} L(n), \quad \alpha > 0,$$

  where $L(n)$ is a slowly varying function.
- The qualitative behaviour of the model changes drastically depending on $\alpha$.

# Construction of the Renewal Process

The ancestral line $A(v)$ of an individual $v$ is given by a renewal process with interarrival law $\mu$ and an additional uniform choice of an individual. The renewal times correspond to the generation of an ancestor. Ancestral lines of a sample of individuals are coupled renewal processes.



### J. Blath, AGC, N. Kurt, D. Spanò (2011)

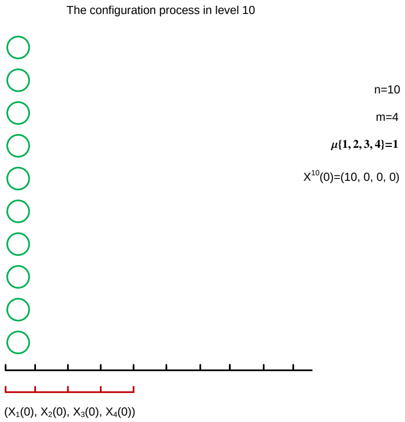
Let $\mu, N$ be fixed and let $v, w$ denote two individuals living at time 0.

(a) If $\alpha > 1$, then $\mathbb{E}[T_{MRCA}] < \infty$
(b) If $\alpha \in (1/2, 1)$, then $\mathbb{P}(A(v) \cap A(w) \neq \emptyset) = 1$ and $\mathbb{E}[T] = \infty$
(c) If $\alpha \in (0, 1/2)$, then $\mathbb{P}(A(v) \cap A(w) \neq \emptyset) < 1$

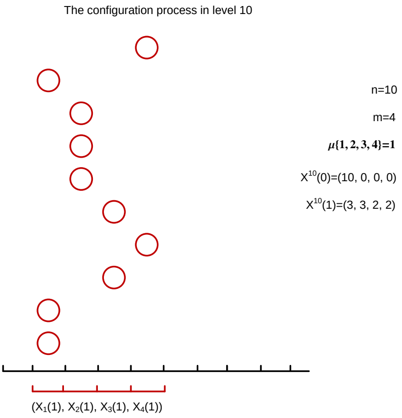
**J. Blath, AGC, N. Kurt, D. Spanò (2011)**

If $E_\mu[\nu] < \infty$ the block counting process induced by our model converges weakly to the Kingman coalescent (constantly time changed),
i.e. $B_N(Nt) \Rightarrow K(\gamma(1)^2 t)$

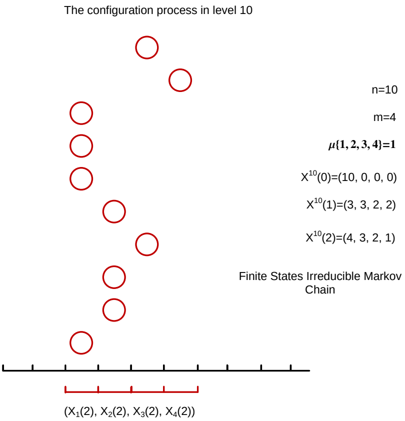
# Ingredient 1: a Markov process



# Ingredient 1: a Markov process



# Ingredient 1: a Markov process



# Ingredient 2: stationary distribution

**Proof**

Let $X^n$ be the configuration process in level n.

- There exists a stationary distribution for $X^n$ if and only if $E_\mu[X] < \infty$.

The stationary distribution is

$$\gamma = \text{mult}\Big(\frac{1}{E_\mu[X]}, \frac{\mu(i>1)}{E_\mu[X]}, \frac{\mu(i>2)}{E_\mu[X]}, \ldots\Big)$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, ...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,\ldots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, \dots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, ...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, \ldots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, ...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, \ldots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,\ldots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, \ldots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,\ldots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1,0,0,\ldots)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 2: stationary distribution

$$\{\text{ball 1 visits urn } k\} = \{X_k^1 = (1, 0, 0, ...)\}$$

$$P_\gamma(\{\text{ball 1 visits urn } k\}) = \frac{1}{E_\mu[X]}$$

# Ingredient 3: coupling argument

## Idea

If particles are always in the stationary distribution

$$P(\text{coalesce in generation } k) = \frac{1}{N(E_\mu[X])^2} = \frac{1}{N}\gamma(1)^2$$

Then consider an artificial system where particles are always in the stationary distribution and couple it with the ancestral process of the seed bank model.

# Ingredient 3: Coupling argument

## The coupling is fast

Let $\tau$ be the first time particles labeled 1 in each system are in the same generation.

$$E[\tau] < \infty$$

## J. Blath, AGC, N. Kurt, D. Spanò (2011)

Let $\mu, N$ be fixed and let $v, w$ denote two individuals living at time 0.

(a) If $\alpha > 1$, then $\mathbb{E}[T_{MRCA}] < \infty$

(b) If $\alpha \in (1/2, 1)$, then $\mathbb{P}(A(v) \cap A(w) \neq \emptyset) = 1$ and $\mathbb{E}[T] = \infty$

(c) If $\alpha \in (0, 1/2)$, then $\mathbb{P}(A(v) \cap A(w) \neq \emptyset) < 1$

Common ancestor will be close to the seed-mutation.
Genetic drift will not cause fixation nor extinction.

# Azotobacter vinelandii

- The seed effect is very important in evolution of bacteria.

# Azotobacter vinelandii

- The seed effect is very important in evolution of bacteria.
- *Azotobacter vinelandii* makes very big jumps. (endospores)

# Azotobacter vinelandii

- The seed effect is very important in evolution of bacteria.
- *Azotobacter vinelandii* makes very big jumps. (endospores)
- *Azotobacter vinelandii* has too many genes of *Pseudomonas* not to be a *Pseudomonas* and too less to be one.

# Azotobacter vinelandii

- The seed effect is very important in evolution of bacteria.
- *Azotobacter vinelandii* makes very big jumps. (endospores)
- *Azotobacter vinelandii* has too many genes of *Pseudomonas* not to be a *Pseudomonas* and too less to be one.
- Our model partially explains this phenomena.

# Azotobacter vinelandii

- The seed effect is very important in evolution of bacteria.
- *Azotobacter vinelandii* makes very big jumps. (endospores)
- *Azotobacter vinelandii* has too many genes of *Pseudomonas* not to be a *Pseudomonas* and too less to be one.
- Our model partially explains this phenomena.
- Challenges.

## References


- Kaj, I., Krone, S., Lascoux, M. 2001. Coalescent theory for seed bank models. J. Appl. Prob. 38:285-300
- J. Blath, A. González Casanova, N. Kurt and D. Spanó. On the ancestral process of long-range seed bank models. To appear in J. of Appl. Prob.


Gracias
Merci!
Danke!