

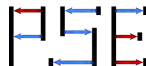
The seed-bank coalescent

Jochen Blath

Joint work (in progress) with

Adrián González Casanova, Noemi Kurt, Maite Wilke Berenguer
(all TU Berlin)

EURANDOM, August 2014



PROBABILISTIC STRUCTURES
IN EVOLUTION

DFG SPP 1590

Seed-banks in population genetics

Many species produce *seeds* or other *dormant forms* (e.g. spores) which introduce strong age-structure / *seed-banks* in population genetic models.

- Seed-banks can act as *buffer* against evolutionary forces such as random genetic drift and selection; ‘bet-hedging’ against environmental changes (see e.g. [ŽIVKOVIĆ & TELLIER 2011] for an overview).
- Their presence typically leads to significantly *increased genetic variability* resp. effective population size.
- Classical mechanisms such as *fixation* and *extinction* of traits become *more complex*: Genetic types can disappear from the active population while returning later due to the germination of seeds or activation of dormant forms.

Seed-banks in population genetics

Many species produce *seeds* or other *dormant forms* (e.g. spores) which introduce strong age-structure / *seed-banks* in population genetic models.

- Seed-banks can act as *buffer* against evolutionary forces such as random genetic drift and selection; ‘bet-hedging’ against environmental changes (see e.g. [ŽIVKOVIĆ & TELLIER 2011] for an overview).
- Their presence typically leads to significantly *increased genetic variability* resp. effective population size.
- Classical mechanisms such as *fixation* and *extinction* of traits become *more complex*: Genetic types can disappear from the active population while returning later due to the germination of seeds or activation of dormant forms.

Seed-banks in population genetics

Many species produce *seeds* or other *dormant forms* (e.g. spores) which introduce strong age-structure / *seed-banks* in population genetic models.

- Seed-banks can act as *buffer* against evolutionary forces such as random genetic drift and selection; ‘bet-hedging’ against environmental changes (see e.g. [ŽIVKOVIĆ & TELLIER 2011] for an overview).
- Their presence typically leads to significantly *increased genetic variability* resp. effective population size.
- Classical mechanisms such as *fixation* and *extinction* of traits become *more complex*: Genetic types can disappear from the active population while returning later due to the germination of seeds or activation of dormant forms.

Seed-banks in population genetics

Many species produce *seeds* or other *dormant forms* (e.g. spores) which introduce strong age-structure / *seed-banks* in population genetic models.

- Seed-banks can act as *buffer* against evolutionary forces such as random genetic drift and selection; ‘bet-hedging’ against environmental changes (see e.g. [ŽIVKOVIĆ & TELLIER 2011] for an overview).
- Their presence typically leads to significantly *increased genetic variability* resp. effective population size.
- Classical mechanisms such as *fixation* and *extinction* of traits become *more complex*: Genetic types can disappear from the active population while returning later due to the germination of seeds or activation of dormant forms.

Seed-banks in bacterial population genetics

Seed-bank effects have also been suggested to play an important role in bacterial evolution ([LENNON & JONES 2011], [GONZÁLEZ ET AL 2014]):

- For example, some bacteria can produce endospores or cysts that remain viable for (in principle arbitrarily) many generations.
- Further, bacteria may also store resp. retrieve genetic material from their 'environment' (e.g. via horizontal gene transfer, say with the help of phages), thus creating a highly *extended gene pool* maintaining enormous genetic variability.

Seed-banks in bacterial population genetics

Seed-bank effects have also been suggested to play an important role in bacterial evolution ([LENNON & JONES 2011], [GONZÁLEZ ET AL 2014]):

- For example, some bacteria can produce endospores or cysts that remain viable for (in principle arbitrarily) many generations.
- Further, bacteria may also store resp. retrieve genetic material from their 'environment' (e.g. via horizontal gene transfer, say with the help of phages), thus creating a highly *extended gene pool* maintaining enormous genetic variability.

Modeling and investigating seed-banks

Despite a series of empirical studies and several theoretical works (including e.g. [KAJ, KRONE & LASCOUX 2001], [TELLIER, LAURENT, LAINER, PAVLIDIS, STEPHAN 2011], [ŽIVKOVIĆ & TELLIER 2011]), the mathematical modeling of seed-banks in population genetics is still incomplete.

Probabilistic models (such as the *Kingman coalescent*) have proved to be efficient and robust tools in quantitative population genetics.

Aim: Include seed-bank effects in classical *Wright-Fisher population models* and classify their scaling limits.

Modeling and investigating seed-banks

Despite a series of empirical studies and several theoretical works (including e.g. [KAJ, KRONE & LASCOUX 2001], [TELLIER, LAURENT, LAINER, PAVLIDIS, STEPHAN 2011], [ŽIVKOVIĆ & TELLIER 2011]), the mathematical modeling of seed-banks in population genetics is still incomplete.

Probabilistic models (such as the *Kingman coalescent*) have proved to be efficient and robust tools in quantitative population genetics.

Aim: Include seed-bank effects in classical *Wright-Fisher population models* and classify their scaling limits.

Known results, I

- [KAJ, KRONE & LASCOUX 2001] include *weak* seed-banks effects in a classical Wright-Fisher model: In a population of size N , each individual independently picks its parent uniformly from a randomly chosen previous generation, say B (with law μ), back in time.
- They show that if μ is supported on $\{1, 2, \dots, m\}$ (independent of N), then the ancestral process converges, after the usual scaling, to a time changed *Kingman coalescent*, with rates multiplied by $\beta^2 := 1/\mathbb{E}[B]^2$.
- An increase of $E[B]$ thus decelerates the coalescent, leading to an increase in the effective population size. However, as observed by [ZT 2011], since the overall coalescent tree structure is retained, this leaves the relative allele frequencies within a sample unchanged. We speak of a *weak seed-bank effect*.
- [ZT 2011] combine the above framework with fluctuating population size.

Known results, I

- [KAJ, KRONE & LASCOUX 2001] include *weak* seed-banks effects in a classical Wright-Fisher model: In a population of size N , each individual independently picks its parent uniformly from a randomly chosen previous generation, say B (with law μ), back in time.
- They show that if μ is supported on $\{1, 2, \dots, m\}$ (independent of N), then the ancestral process converges, after the usual scaling, to a time changed *Kingman coalescent*, with rates multiplied by $\beta^2 := 1/\mathbb{E}[B]^2$.
- An increase of $E[B]$ thus decelerates the coalescent, leading to an increase in the effective population size. However, as observed by [ZT 2011], since the overall coalescent tree structure is retained, this leaves the relative allele frequencies within a sample unchanged. We speak of a *weak seed-bank effect*.
- [ZT 2011] combine the above framework with fluctuating population size.

Known results, I

- [KAJ, KRONE & LASCOUX 2001] include *weak* seed-banks effects in a classical Wright-Fisher model: In a population of size N , each individual independently picks its parent uniformly from a randomly chosen previous generation, say B (with law μ), back in time.
- They show that if μ is supported on $\{1, 2, \dots, m\}$ (independent of N), then the ancestral process converges, after the usual scaling, to a time changed *Kingman coalescent*, with rates multiplied by $\beta^2 := 1/\mathbb{E}[B]^2$.
- An increase of $E[B]$ thus decelerates the coalescent, leading to an increase in the effective population size. However, as observed by [ZT 2011], since the overall coalescent tree structure is retained, this leaves the relative allele frequencies within a sample unchanged. We speak of a *weak seed-bank effect*.
- [ZT 2011] combine the above framework with fluctuating population size.

Known results, I

- [KAJ, KRONE & LASCOUX 2001] include *weak* seed-banks effects in a classical Wright-Fisher model: In a population of size N , each individual independently picks its parent uniformly from a randomly chosen previous generation, say B (with law μ), back in time.
- They show that if μ is supported on $\{1, 2, \dots, m\}$ (independent of N), then the ancestral process converges, after the usual scaling, to a time changed *Kingman coalescent*, with rates multiplied by $\beta^2 := 1/\mathbb{E}[B]^2$.
- An increase of $E[B]$ thus decelerates the coalescent, leading to an increase in the effective population size. However, as observed by [ZT 2011], since the overall coalescent tree structure is retained, this leaves the relative allele frequencies within a sample unchanged. We speak of a *weak seed-bank effect*.
- [ZT 2011] combine the above framework with fluctuating population size.

Known results, II

- More generally, [BGCKS 2013] show that a sufficient condition for convergence to the Kingman coalescent (with similar scaling and rates) is that $E[B] < \infty$ (again, with B resp. μ independent of N).
- Further, they show that a long-range seed-bank effect can completely alter genealogical behaviour: E.g., if the seed-bank age distribution μ is heavy-tailed, say,

$$\mu(k) = L(k)k^{-\alpha},$$

for $k \in \mathbb{N}$, where L is a slowly varying function, then if $\alpha < 1$ the expected time to the most recent common ancestor is infinite, and if $\alpha < 1/2$ two randomly sampled individuals do not have a common ancestor at all w.p.p.

Known results, II

- More generally, [BGCKS 2013] show that a sufficient condition for convergence to the Kingman coalescent (with similar scaling and rates) is that $E[B] < \infty$ (again, with B resp. μ independent of N).
- Further, they show that a long-range seed-bank effect can completely alter genealogical behaviour: E.g., if the seed-bank age distribution μ is heavy-tailed, say,

$$\mu(k) = L(k)k^{-\alpha},$$

for $k \in \mathbb{N}$, where L is a slowly varying function, then if $\alpha < 1$ the expected time to the most recent common ancestor is infinite, and if $\alpha < 1/2$ two randomly sampled individuals do not have a common ancestor at all w.p.p.

How to model strong seed-bank effects

It can be argued that such extreme behaviour seems artificial (although bacterial endospores may stay viable for essentially unlimited numbers of generations).

Instead, as often in mathematical population genetics, one can turn to the case $\mu = \mu(N)$ and scale the seed-bank effect with population size.

In [GC ET AL 2014] and [BEGCK 2014], a seed-bank model with

$$\mu = \mu(N) = (1 - \varepsilon)\delta_1 + \varepsilon\delta_{N^\beta}, \quad \beta > 0, \varepsilon \in (0, 1).$$

is studied, and it is shown that for $\beta < 1/4$ the ancestral process converges, after non-classical time-scaling by $N^{1+2\beta}$, to the Kingman coalescent.

Problem: Such models are highly non-Markovian. Problems with explicit forward dynamics / limits.

How to model strong seed-bank effects

It can be argued that such extreme behaviour seems artificial (although bacterial endospores may stay viable for essentially unlimited numbers of generations).

Instead, as often in mathematical population genetics, one can turn to the case $\mu = \mu(N)$ and scale the seed-bank effect with population size.

In [GC ET AL 2014] and [BEGCK 2014], a seed-bank model with

$$\mu = \mu(N) = (1 - \varepsilon)\delta_1 + \varepsilon\delta_{N^\beta}, \quad \beta > 0, \varepsilon \in (0, 1).$$

is studied, and it is shown that for $\beta < 1/4$ the ancestral process converges, after non-classical time-scaling by $N^{1+2\beta}$, to the Kingman coalescent.

Problem: Such models are highly non-Markovian. Problems with explicit forward dynamics / limits.

How to model strong seed-bank effects

It can be argued that such extreme behaviour seems artificial (although bacterial endospores may stay viable for essentially unlimited numbers of generations).

Instead, as often in mathematical population genetics, one can turn to the case $\mu = \mu(N)$ and scale the seed-bank effect with population size.

In [GC ET AL 2014] and [BEGCK 2014], a seed-bank model with

$$\mu = \mu(N) = (1 - \varepsilon)\delta_1 + \varepsilon\delta_{N^\beta}, \quad \beta > 0, \varepsilon \in (0, 1).$$

is studied, and it is shown that for $\beta < 1/4$ the ancestral process converges, after non-classical time-scaling by $N^{1+2\beta}$, to the Kingman coalescent.

Problem: Such models are highly non-Markovian. Problems with explicit forward dynamics / limits.

A new natural scaling limit?

To sum up, while there are some results for the *weak* seed-bank case, it seems that the right scaling set-up in ‘stronger’ regimes is still missing.

In this talk (work in progress) we investigate a Markovian seed-bank model forward in time and show that it has an interesting new ancestral scaling limit, the *seed-bank coalescent*.

The Wright-Fisher model with geometric seed-bank component - notation

Consider a haploid population of fixed size N reproducing in fixed discrete generations $k = 0, 1, \dots$. Assume that each individual carries a genetic type from some type-space E , say $E = \{a, A\}$.

Further, assume that the population also sustains a *seed-bank* of constant size $M = M(N)$ in each generation, which consists of the dormant individuals. For simplicity, we refer to the N 'active' individuals as *plants* and to the M dormant individuals as *seeds*.

Given $N, M \in \mathbb{N}$, let $\varepsilon \in [0, 1]$ such that $\varepsilon N \leq M$ and set $\delta := \varepsilon N / M$ (i.e., $\delta M = \varepsilon N$), and assume for convenience that all involved products and fractions are integers.

Let $[N] := \{1, \dots, N\}$ and $[N]_0 := [N] \cup \{0\}$.

The Wright-Fisher model with geometric seed-bank component - notation

Consider a haploid population of fixed size N reproducing in fixed discrete generations $k = 0, 1, \dots$. Assume that each individual carries a genetic type from some type-space E , say $E = \{a, A\}$.

Further, assume that the population also sustains a *seed-bank* of constant size $M = M(N)$ in each generation, which consists of the dormant individuals. For simplicity, we refer to the N 'active' individuals as *plants* and to the M dormant individuals as *seeds*.

Given $N, M \in \mathbb{N}$, let $\varepsilon \in [0, 1]$ such that $\varepsilon N \leq M$ and set $\delta := \varepsilon N / M$ (i.e., $\delta M = \varepsilon N$), and assume for convenience that all involved products and fractions are integers.

Let $[N] := \{1, \dots, N\}$ and $[N]_0 := [N] \cup \{0\}$.

The Wright-Fisher model with geometric seed-bank component - notation

Consider a haploid population of fixed size N reproducing in fixed discrete generations $k = 0, 1, \dots$. Assume that each individual carries a genetic type from some type-space E , say $E = \{a, A\}$.

Further, assume that the population also sustains a *seed-bank* of constant size $M = M(N)$ in each generation, which consists of the dormant individuals. For simplicity, we refer to the N 'active' individuals as *plants* and to the M dormant individuals as *seeds*.

Given $N, M \in \mathbb{N}$, let $\varepsilon \in [0, 1]$ such that $\varepsilon N \leq M$ and set $\delta := \varepsilon N / M$ (i.e., $\delta M = \varepsilon N$), and assume for convenience that all involved products and fractions are integers.

Let $[N] := \{1, \dots, N\}$ and $[N]_0 := [N] \cup \{0\}$.

The Wright-Fisher model with geometric seed-bank component - dynamics

- The N *plants* from generation 0 produce $(1 - \varepsilon)N$ *plants* in generation 1 by multinomial sampling with equal weights (ordinary WF dynamics).
- Additionally, $\delta M = \varepsilon N$ uniformly sampled *seeds* from the seed-bank of size M in generation 0 'germinate', that is, they turn into exactly one *plant* in generation 1 each, and thus leave the seed-bank.
- The *plants* from generation 0 are thus replaced by these $(1 - \varepsilon)N + \delta M = N$ new active individuals, forming the *plants* in generation 1.
- For the seed-bank, the N *plants* from generation 0 produce $\delta M = \varepsilon N$ *seeds* by multinomial sampling, replacing those *seeds* that germinated.
- The remaining $(1 - \delta)M$ *seeds* from generation 0 remain inactive and stay in the seed-bank.
- Throughout reproduction, offspring and seeds copy/maintain the genetic type of the parent.

Thus, in generation 1, we have again N *plants* and M *seeds*. This probabilistic mechanism is then to be repeated independently in generations $k = 2, 3, \dots$

The Wright-Fisher model with geometric seed-bank component - dynamics

- The N *plants* from generation 0 produce $(1 - \varepsilon)N$ *plants* in generation 1 by multinomial sampling with equal weights (ordinary WF dynamics).
- Additionally, $\delta M = \varepsilon N$ uniformly sampled *seeds* from the seed-bank of size M in generation 0 'germinate', that is, they turn into exactly one *plant* in generation 1 each, and thus leave the seed-bank.
- The *plants* from generation 0 are thus replaced by these $(1 - \varepsilon)N + \delta M = N$ new active individuals, forming the *plants* in generation 1.
- For the seed-bank, the N *plants* from generation 0 produce $\delta M = \varepsilon N$ *seeds* by multinomial sampling, replacing those *seeds* that germinated.
- The remaining $(1 - \delta)M$ *seeds* from generation 0 remain inactive and stay in the seed-bank.
- Throughout reproduction, offspring and seeds copy/maintain the genetic type of the parent.

Thus, in generation 1, we have again N *plants* and M *seeds*. This probabilistic mechanism is then to be repeated independently in generations $k = 2, 3, \dots$

The Wright-Fisher model with geometric seed-bank component - dynamics

- The N *plants* from generation 0 produce $(1 - \varepsilon)N$ *plants* in generation 1 by multinomial sampling with equal weights (ordinary WF dynamics).
- Additionally, $\delta M = \varepsilon N$ uniformly sampled *seeds* from the seed-bank of size M in generation 0 'germinate', that is, they turn into exactly one *plant* in generation 1 each, and thus leave the seed-bank.
- The *plants* from generation 0 are thus replaced by these $(1 - \varepsilon)N + \delta M = N$ new active individuals, forming the *plants* in generation 1.
- For the seed-bank, the N *plants* from generation 0 produce $\delta M = \varepsilon N$ *seeds* by multinomial sampling, replacing those *seeds* that germinated.
- The remaining $(1 - \delta)M$ *seeds* from generation 0 remain inactive and stay in the seed-bank.
- Throughout reproduction, offspring and seeds copy/maintain the genetic type of the parent.

Thus, in generation 1, we have again N *plants* and M *seeds*. This probabilistic mechanism is then to be repeated independently in generations $k = 2, 3, \dots$

The Wright-Fisher model with geometric seed-bank component - dynamics

- The N *plants* from generation 0 produce $(1 - \varepsilon)N$ *plants* in generation 1 by multinomial sampling with equal weights (ordinary WF dynamics).
- Additionally, $\delta M = \varepsilon N$ uniformly sampled *seeds* from the seed-bank of size M in generation 0 'germinate', that is, they turn into exactly one *plant* in generation 1 each, and thus leave the seed-bank.
- The *plants* from generation 0 are thus replaced by these $(1 - \varepsilon)N + \delta M = N$ new active individuals, forming the *plants* in generation 1.
- For the seed-bank, the N *plants* from generation 0 produce $\delta M = \varepsilon N$ *seeds* by multinomial sampling, replacing those *seeds* that germinated.
- The remaining $(1 - \delta)M$ *seeds* from generation 0 remain inactive and stay in the seed-bank.
- Throughout reproduction, offspring and seeds copy/maintain the genetic type of the parent.

Thus, in generation 1, we have again N *plants* and M *seeds*. This probabilistic mechanism is then to be repeated independently in generations $k = 2, 3, \dots$

The Wright-Fisher model with geometric seed-bank component - dynamics

- The N *plants* from generation 0 produce $(1 - \varepsilon)N$ *plants* in generation 1 by multinomial sampling with equal weights (ordinary WF dynamics).
- Additionally, $\delta M = \varepsilon N$ uniformly sampled *seeds* from the seed-bank of size M in generation 0 'germinate', that is, they turn into exactly one *plant* in generation 1 each, and thus leave the seed-bank.
- The *plants* from generation 0 are thus replaced by these $(1 - \varepsilon)N + \delta M = N$ new active individuals, forming the *plants* in generation 1.
- For the seed-bank, the N *plants* from generation 0 produce $\delta M = \varepsilon N$ *seeds* by multinomial sampling, replacing those *seeds* that germinated.
- The remaining $(1 - \delta)M$ *seeds* from generation 0 remain inactive and stay in the seed-bank.
- Throughout reproduction, offspring and seeds copy/maintain the genetic type of the parent.

Thus, in generation 1, we have again N *plants* and M *seeds*. This probabilistic mechanism is then to be repeated independently in generations $k = 2, 3, \dots$

The Wright-Fisher model with geometric seed-bank component - dynamics

- The N *plants* from generation 0 produce $(1 - \varepsilon)N$ *plants* in generation 1 by multinomial sampling with equal weights (ordinary WF dynamics).
- Additionally, $\delta M = \varepsilon N$ uniformly sampled *seeds* from the seed-bank of size M in generation 0 'germinate', that is, they turn into exactly one *plant* in generation 1 each, and thus leave the seed-bank.
- The *plants* from generation 0 are thus replaced by these $(1 - \varepsilon)N + \delta M = N$ new active individuals, forming the *plants* in generation 1.
- For the seed-bank, the N *plants* from generation 0 produce $\delta M = \varepsilon N$ *seeds* by multinomial sampling, replacing those *seeds* that germinated.
- The remaining $(1 - \delta)M$ *seeds* from generation 0 remain inactive and stay in the seed-bank.
- Throughout reproduction, offspring and seeds copy/maintain the genetic type of the parent.

Thus, in generation 1, we have again N *plants* and M *seeds*. This probabilistic mechanism is then to be repeated independently in generations $k = 2, 3, \dots$

Picture: WF with seed-bank dynamics

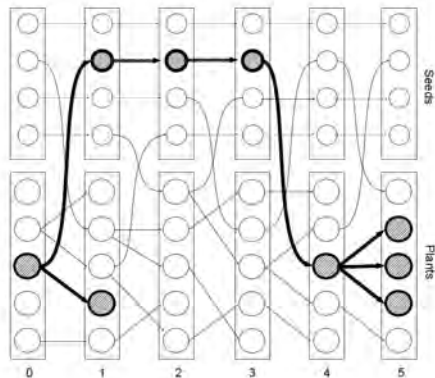


Figure: Here, the genetic type of the third plant in generation 0 (highlighted in grey) is lost after one generation, but returns in generation four via the seed-bank, which acts as a buffer against genetic drift and maintains genetic variability.

Formal definition, notation

Definition 1.1 (Wright-Fisher model with geometric seed-bank component)

Fix pop.-size $N \in \mathbb{N}$, seed-bank size $M = M(N)$, genetic type space E and parameters δ, ε as before. Given initial type configurations $\xi_0 \in E^N$ and $\eta_0 \in E^M$, let

$$\xi_k := (\xi_k(i))_{i \in [N]}, \quad k \in \mathbb{N},$$

be the random genetic type configuration in E^N of the *plants* in generation k (obtained from the above mechanism), and

$$\eta_k := (\eta_k(j))_{j \in [M]}, \quad k \in \mathbb{N},$$

be the genetic type configuration of the *seeds* in E^M . We call the discrete-time Markov chain $(\xi_k, \eta_k)_{k \in \mathbb{N}_0}$ with values in $E^N \times E^M$ the *type configuration process* of the *Wright-Fisher model with geometric seed-bank component*.

Remarks, scaling regime

Note that the offspring distribution of active individuals (both for the number of plants and for the number of seeds) is exchangeable within their respective sub-population.

Further, one immediately sees that the time that a seed stays in the seed-bank is iid geometric with success parameter δ .

We will later let ε, δ (and M) scale with N , and in particular assume that $\varepsilon = \varepsilon(N) = c/N$ and $N = K \cdot M(N)$ for some $c, K \in (0, \infty)$. In this case, the seed-bank age distribution is geometric with parameter cK/N .

Remarks, scaling regime

Note that the offspring distribution of active individuals (both for the number of plants and for the number of seeds) is exchangeable within their respective sub-population.

Further, one immediately sees that the time that a seed stays in the seed-bank is iid geometric with success parameter δ .

We will later let ε, δ (and M) scale with N , and in particular assume that $\varepsilon = \varepsilon(N) = c/N$ and $N = K \cdot M(N)$ for some $c, K \in (0, \infty)$. In this case, the seed-bank age distribution is geometric with parameter cK/N .

Frequency chains of the geometric seed-bank model

We now specialise to the bi-allelic case $E = \{a, A\}$.

Definition 1.2

Frequency chains Let

$$X_k^N := \frac{1}{N} \sum_{i \in [N]} \mathbf{1}_{\{\xi_k(i)=a\}} \quad \text{and} \quad Y_k^M := \frac{1}{M} \sum_{j \in [M]} \mathbf{1}_{\{\eta_k(j)=a\}}, \quad k \in \mathbb{N}_0. \quad (1)$$

Both are discrete-time Markov chains taking values in

$$I^N = \left\{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\right\} \quad \text{resp.} \quad I^M = \left\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\right\}.$$

Let $\mathbb{P}_{x,y}$ be the law of (X^N, Y^M) with initial frequencies (x, y) :

$$\mathbb{P}_{x,y}(\cdot) := \mathbb{P}(\cdot \mid X_0^N = x, Y_0^M = y) \quad \text{for} \quad (x, y) \in I^N \times I^M$$

(with analogous notation for the expectation, variance etc).

Transition probabilities of the geometric seed-bank model

The corresponding time-homogeneous transition probabilities can now be characterized as follows:

Proposition 1.3

Let $c = \varepsilon N = \delta M$. For (x, y) resp. $(\bar{x}, \bar{y}) \in I^N \times I^M$, we have

$$\begin{aligned} p_{x,y} &:= \mathbb{P}_{x,y}(X_1^N = \bar{x}, Y_1^M = \bar{y}) \\ &= \sum_{i=0}^c \mathbb{P}_{x,y}(Z = i) \mathbb{P}_{x,y}(U = \bar{x}N - i) \mathbb{P}_{x,y}(V = (\bar{y} - y)M + i) \end{aligned}$$

where Z, U, V are r.v. independent under $\mathbb{P}_{x,y}$, with distributions

$$\mathcal{L}_{x,y}(Z) = \text{Hyp}_{M,c,yM}, \quad \mathcal{L}_{x,y}(U) = \text{Bin}_{N-c,x}, \quad \mathcal{L}_{x,y}(V) = \text{Bin}_{c,x}.$$

Here, $\text{Hyp}_{M,c,yM}$ denotes the hypergeometric distribution with parameters M, c, yM and $\text{Bin}_{c,x}$ is the binomial distribution with parameters c and x .

Illustration

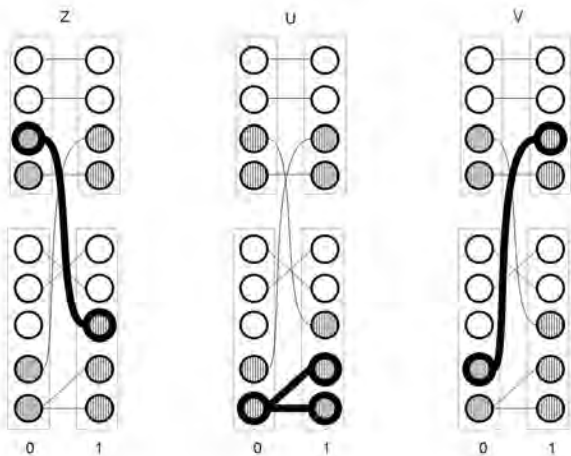


Figure: Here, $Z = 1$, $U = 2$ and $V = 1$

Interpretation of transition probabilities

Remark 1.4

The random variables introduced in Proposition 1.3 have a simple interpretation:

- Z is the number of plants in generation 1 that are offspring of a seed of type a in generation 0. This corresponds to the number of seeds of type a that germinate (noting that, in contrast to plants, the ‘offspring’ of a germinating seed is always precisely one plant and the seed vanishes).
- U is the number of plants in generation 1 that are offspring of plants of type a in generation 0.
- V is the number of seeds in generation 1 that are **produced by** plants of type a in generation 0.

The result now follows by construction since (in law),

$$X_1^N = \frac{U + Z}{N}, \quad \text{and} \quad Y_1^M = y + \frac{V - Z}{M}.$$

The limiting generator of the allele frequency processes

For $c, K \in (0, \infty)$ assume

$$\varepsilon = \varepsilon(N) = \frac{c}{N}, \quad M = M(N) = \frac{N}{K}, \quad \text{and} \quad \delta = \delta(N) = \frac{c}{M(N)} = \frac{cK}{N}.$$

Proposition 1.5

For the above parameter choices, and test functions $f \in C^{(3)}([0, 1]^2)$, and for $(x, y) \in I^N \times I^M$ consider the **discrete generator** A^N of $(X_k^N, Y_k^M)_{k \in \mathbb{N}}$:

$$A^N f(x, y) := N \mathbb{E}_{x, y} \left[f(X_1^N, Y_1^M) - f(x, y) \right].$$

Then,

$$\begin{aligned} & \lim_{N \rightarrow \infty} A^N f(x, y) \\ &= c(y-x) \frac{\partial f}{\partial x}(x, y) + cK(x-y) \frac{\partial f}{\partial y}(x, y) + \frac{1}{2}x(1-x) \frac{\partial^2 f}{\partial x^2}(x, y). \end{aligned}$$

The scaling-limit of the allele frequency process

The result follows from the usual Taylor approximation for $A^N f$ (lengthy details omitted). We arrive at the following

Corollary 1.6 (Wright-Fisher diffusion with seed-bank component)

Under the conditions of Proposition 1.5,

$$(X_{\lfloor Nt \rfloor}^N, Y_{\lfloor Nt \rfloor}^N)_{t \geq 0} \Rightarrow (X_t, Y_t)_{t \geq 0}$$

on $D_{[0, \infty)}([0, 1]^2)$, where $(X_t, Y_t)_{t \geq 0}$ is a 2-dimensional diffusion solving

$$\begin{aligned} dX_t &= c(Y_t - X_t)dt + \sqrt{X_t(1 - X_t)}dB_t, \\ dY_t &= cK(X_t - Y_t)dt, \end{aligned} \tag{2}$$

with $X_0 = x, Y_0 = y$.

Seed-bank size

Note that if the seed-bank is large (K close to 0), then Y changes only slowly, in line with intuition.

If $N/M(N) \rightarrow 0$, this would lead to a constant contribution from an 'infinite' seed-bank, whereas in the regime $M(N)/N \rightarrow 0$, the influence of the seed-bank should vanish asymptotically.

Before we can further investigate the system (2), we derive an important tool.

Seed-bank size

Note that if the seed-bank is large (K close to 0), then Y changes only slowly, in line with intuition.

If $N/M(N) \rightarrow 0$, this would lead to a constant contribution from an 'infinite' seed-bank, whereas in the regime $M(N)/N \rightarrow 0$, the influence of the seed-bank should vanish asymptotically.

Before we can further investigate the system (2), we derive an important tool.

The dual of the seed-bank frequency process

The classical Wright-Fisher diffusion is known to be *dual* to the *block counting process* of the Kingman-coalescent.

Such dual processes are often extremely useful in the analysis of the underlying system, and it is easy to see that our Wright-Fisher diffusion with geometric seed-bank component also has a nice dual.

The dual of the seed-bank frequency process

Definition 1.7

We define the *block counting process of the seed-bank coalescent* $(N_t, M_t)_{t \geq 0}$ to be the continuous time Markov chain started in $(N_0, M_0) \in \mathbb{N}^2$ with transitions

$$\begin{aligned}(n, m) &\mapsto (n - 1, m + 1) && \text{at rate} && cn \\(n, m) &\mapsto (n + 1, m - 1) && \text{at rate} && cKm \\(n, m) &\mapsto (n - 1, m) && \text{at rate} && \binom{n}{2}\end{aligned}$$

Denote by $\mathbb{P}^{n,m}$ the distribution of $(N_t, M_t)_{t \geq 0}$ if started in $(N_0, M_0) = (n, m)$, and denote the corresponding expected value by $\mathbb{E}^{n,m}$.

Moment duality

It is easy to see that *eventually*, $N_t + M_t = 1$ (as $t \rightarrow \infty$), since the sum $M + N$ is a pure death process.

We now show that $(N_t, M_t)_{t \geq 0}$ is the *moment dual* of $(X_t, Y_t)_{t \geq 0}$.

Theorem 1.8

For every $(x, y) \in [0, 1]^2$ and every $n, m \in \mathbb{N}$,

$$\mathbb{E}_{x,y} [X_t^n Y_t^m] = \mathbb{E}^{n,m} [x^{N_t} y^{M_t}].$$

Proof of Theorem 1.8

Let $f(x, y) := x^n y^m =: \bar{f}(n, m)$. Apply the generator A of (X_t, Y_t) to f :

$$Af(x, y)$$

$$\begin{aligned} &= c(y-x) \frac{df}{dx} f(x, y) + \frac{1}{2} x(1-x) \frac{d^2 f}{dx^2} f(x, y) + cK(x-y) \frac{df}{dy} f(x, y) \\ &= c(y-x) n x^{n-1} y^m + \frac{1}{2} x(1-x) n(n-1) x^{n-2} y^m + cK(x-y) x^n m y^{m-1} \\ &= cn(x^{n-1} y^{m+1} - x^n y^m) + \binom{n}{2} (x^{n-1} y^m - x^n y^m) \\ &\quad + cKm(x^{n+1} y^{m-1} - x^n y^m). \end{aligned}$$

Note that the *rhs* is the precisely the generator of $(N_t, M_t)_{t \geq 0}$, denoted by \bar{A} , applied to \bar{f} . So we conclude that

$$Af(x, y) = \bar{A}\bar{f}(n, m)$$

and the duality follows from standard arguments.

The long-time behavior (in law)

The long-term behaviour of our system (2) is not obvious. While a classical Wright Fisher diffusion $\{Z_t\}$, given by

$$dZ_t = \sqrt{Z_t(1 - Z_t)}dB_t, \quad Z_0 = z \in [0, 1],$$

will get absorbed at the boundaries a.s. after finite time (in fact with finite expectation), hitting 1 with probability z , this is more involved for our frequency process in the presence of a strong seed-bank.

Obviously, $(0, 0)$ and $(1, 1)$ are absorbing states for (2).

The long-time behavior (in law)

The long-term behaviour of our system (2) is not obvious. While a classical Wright Fisher diffusion $\{Z_t\}$, given by

$$dZ_t = \sqrt{Z_t(1 - Z_t)}dB_t, \quad Z_0 = z \in [0, 1],$$

will get absorbed at the boundaries a.s. after finite time (in fact with finite expectation), hitting 1 with probability z , this is more involved for our frequency process in the presence of a strong seed-bank.

Obviously, $(0, 0)$ and $(1, 1)$ are absorbing states for (2).

The long-time behavior (in law)

Nevertheless, one can still compute its 'fixation probability' as $t \rightarrow \infty$, in a suitable sense (in law). We prepare this with a moment computation.

Proposition 1.9

All mixed moments of $(X_t, Y_t)_{t \geq 0}$ solving (2) converge to the **same** finite limit depending only on x, y, K . More precisely, for each fixed $n, m \in \mathbb{N}$, we have

$$\lim_{t \rightarrow \infty} \mathbb{E}_{x,y}[X_t^n Y_t^m] = \frac{y + xK}{1 + K}. \quad (3)$$

Proof of Proposition 1.9

Let $(N_t, M_t)_{t \geq 0}$ as in Definition 1.7, started in (n, m) . Let T be the first time at which there is only one particle left in the system $(N_t, M_t)_{t \geq 0}$, that is,

$$T := \inf \{t > 0 : N_t + M_t = 1\}.$$

Note that for any finite initial configuration (n, m) , the positive stopping time T is almost surely finite and has finite expectation.

Now, by our Duality-Theorem 1.8,

$$\begin{aligned} \mathbb{E}_{x,y} [X_t^n Y_t^m] &= \mathbb{E}^{n,m} [x^{N_t} y^{M_t}] \\ &= \mathbb{E}^{n,m} [x^{N_t} y^{M_t} \mid T \leq t] \mathbb{P}^{n,m} \{T \leq t\} \\ &\quad + \mathbb{E}^{n,m} [x^{N_t} y^{M_t} \mid T > t] \mathbb{P}^{n,m} \{T > t\} \\ &= \left(x \mathbb{P}^{n,m} \{N_t = 1\} + y \mathbb{P}^{n,m} \{M_t = 1\} \right) \mathbb{P}^{n,m} \{T \leq t\} \\ &\quad + \mathbb{E}^{n,m} [x^{N_t} y^{M_t} \mid T > t] \mathbb{P}^{n,m} \{T > t\}. \end{aligned}$$

Proof of Proposition 1.9

Let $(N_t, M_t)_{t \geq 0}$ as in Definition 1.7, started in (n, m) . Let T be the first time at which there is only one particle left in the system $(N_t, M_t)_{t \geq 0}$, that is,

$$T := \inf \{t > 0 : N_t + M_t = 1\}.$$

Note that for any finite initial configuration (n, m) , the positive stopping time T is almost surely finite and has finite expectation.

Now, by our Duality-Theorem 1.8,

$$\begin{aligned} \mathbb{E}_{x,y} [X_t^n Y_t^m] &= \mathbb{E}^{n,m} [x^{N_t} y^{M_t}] \\ &= \mathbb{E}^{n,m} [x^{N_t} y^{M_t} \mid T \leq t] \mathbb{P}^{n,m} \{T \leq t\} \\ &\quad + \mathbb{E}^{n,m} [x^{N_t} y^{M_t} \mid T > t] \mathbb{P}^{n,m} \{T > t\} \\ &= \left(x \mathbb{P}^{n,m} \{N_t = 1\} + y \mathbb{P}^{n,m} \{M_t = 1\} \right) \mathbb{P}^{n,m} \{T \leq t\} \\ &\quad + \mathbb{E}^{n,m} [x^{N_t} y^{M_t} \mid T > t] \mathbb{P}^{n,m} \{T > t\}. \end{aligned}$$

Proof of Proposition 1.9, continued

Note that the last term on the *rhs* vanishes as $t \rightarrow \infty$, since the expectation is bounded by 1.

Further, the stationary distribution of a single particle, jumping between the two states 'plant' and 'seed' at rate c resp. cK , is given by $(K/(1+K), 1/(1+K))$, hence we arrive at

$$\lim_{t \rightarrow \infty} \mathbb{E}_{x,y} [X_t^n Y_t^m] = \frac{xK}{1+K} + \frac{y}{1+K},$$

independent of the initial choice of n, m , as required.

Proof of Proposition 1.9, continued

Note that the last term on the *rhs* vanishes as $t \rightarrow \infty$, since the expectation is bounded by 1.

Further, the stationary distribution of a single particle, jumping between the two states 'plant' and 'seed' at rate c resp. cK , is given by $(K/(1+K), 1/(1+K))$, hence we arrive at

$$\lim_{t \rightarrow \infty} \mathbb{E}_{x,y} [X_t^n Y_t^m] = \frac{xK}{1+K} + \frac{y}{1+K},$$

independent of the initial choice of n, m , as required.

The long-time behavior (in law), II

Corollary 1.10 (Fixation in law)

Given c, K and $(X_0, Y_0) = (x, y) \in [0, 1]^2$, we have that

$$\lim_{t \rightarrow \infty} \mathcal{L}(X_t, Y_t) = \frac{y + xK}{1 + K} \delta_{(1,1)} + \frac{1 + (1 - x)K - y}{1 + K} \delta_{(0,0)}.$$

Note that this is in line with the classical results for the Wright-Fisher diffusion:

As $K \rightarrow \infty$ (that is, the seed-bank becomes small compared to the plant population), the fixation probability of a alleles approaches x .

Further, if K becomes small (so that the seed-bank population dominates the plant population), the fixation probability is governed by the initial fraction y of a -alleles in the seed-bank.

The long-time behavior (in law), II

Corollary 1.10 (Fixation in law)

Given c, K and $(X_0, Y_0) = (x, y) \in [0, 1]^2$, we have that

$$\lim_{t \rightarrow \infty} \mathcal{L}(X_t, Y_t) = \frac{y + xK}{1 + K} \delta_{(1,1)} + \frac{1 + (1 - x)K - y}{1 + K} \delta_{(0,0)}.$$

Note that this is in line with the classical results for the Wright-Fisher diffusion:

As $K \rightarrow \infty$ (that is, the seed-bank becomes small compared to the plant population), the fixation probability of a alleles approaches x .

Further, if K becomes small (so that the seed-bank population dominates the plant population), the fixation probability is governed by the initial fraction y of a -alleles in the seed-bank.

The long-time behavior (in law), II

Corollary 1.10 (Fixation in law)

Given c, K and $(X_0, Y_0) = (x, y) \in [0, 1]^2$, we have that

$$\lim_{t \rightarrow \infty} \mathcal{L}(X_t, Y_t) = \frac{y + xK}{1 + K} \delta_{(1,1)} + \frac{1 + (1 - x)K - y}{1 + K} \delta_{(0,0)}.$$

Note that this is in line with the classical results for the Wright-Fisher diffusion:

As $K \rightarrow \infty$ (that is, the seed-bank becomes small compared to the plant population), the fixation probability of a alleles approaches x .

Further, if K becomes small (so that the seed-bank population dominates the plant population), the fixation probability is governed by the initial fraction y of a -alleles in the seed-bank.

Proof of the long-time result

Sketch: It is easy to see that the only two-dimensional random variable, for which all moments are constant equal to $\frac{xK+y}{1+K}$, has law

$$\frac{y + xK}{1 + K} \delta_{(1,1)} + \frac{1 + (1 - x)K - y}{1 + K} \delta_{(0,0)}.$$

Indeed, uniqueness follows from the moment problem, which is uniquely solvable on $[0, 1]^2$. Convergence in law follows from convergence of all moments.

Almost sure behaviour?

Note that the above result does *not* fully explain the pathwise/almost-sure picture.

Indeed, (vaguely speaking) we expect that absorption will not happen in finite time, since the dual *block counting process of the seed-bank coalescent*, started from an infinite initial state, *does not come down from infinity*, which means that the total (infinite) population does not have a most-recent common ancestor (in progress).

Thus, initial genetic variability in an infinite population should never be completely lost.

Almost sure behaviour?

Note that the above result does *not* fully explain the pathwise/almost-sure picture.

Indeed, (vaguely speaking) we expect that absorption will not happen in finite time, since the dual *block counting process of the seed-bank coalescent*, started from an infinite initial state, *does not come down from infinity*, which means that the total (infinite) population does not have a most-recent common ancestor (in progress).

Thus, initial genetic variability in an infinite population should never be completely lost.

The genealogy of a sample

In view of the form of the block counting process, it is now easy to guess the stochastic process describing the limiting gene genealogy of a sample taken from the Wright-Fisher model with geometric seed-bank component.

The genealogy of a sample and partitions with flags

For $k \geq 1$, let \mathcal{P}_k be the set of partitions of $[k]$. For $\pi \in \mathcal{P}_k$ let $|\pi|$ be the number of blocks of the partition π . We define the space of *marked partitions* to be

$$\mathcal{P}_k^{\{p,s\}} = \left\{ \pi = (\pi_1^{a_1}, \dots, \pi_{|\pi|}^{a_{|\pi|}}) : (\pi_i) \in \mathcal{P}_k, (a_i) \in \{s, p\}^{|\pi|} \right\}.$$

This enables us to attach to each partition block a *flag* which can be either 'plant' or 'seed' (p or s), so that we can trace whether an ancestral line is currently in the active or dormant part of the population.

For example, for $k = 5$, an element π of $\mathcal{P}_k^{\{p,s\}}$ is the marked partition

$$\pi = \{ \{1, 3\}^p \{2\}^s \{4, 5\}^p \}.$$

The genealogy of a sample and partitions with flags, II

For two marked partitions $\pi, \pi' \in \mathcal{P}_k^{\{p,s\}}$ we write $\pi \succ \pi'$ if π' can be obtained by merging exactly 2 blocks of π carrying the p -flag, and the resulting block in π' again carries a p -flag. For example

$$\{\{1, 3\}^p \{2\}^s \{4, 5\}^p\} \succ \{\{1, 3, 4, 5\}^p \{2\}^s\}.$$

We use the notation $\pi \bowtie \pi'$ if π' can be obtained by changing the flag of precisely one block of π , for example

$$\{\{1, 3\}^p \{2\}^s \{4, 5\}^p\} \bowtie \{\{1, 3\}^s \{2\}^s \{4, 5\}^p\}.$$

The seed-bank coalescent

Definition 1.11

For $k \geq 1$ and $c, K \in (0, \infty)$ we define the k -*seed-bank coalescent* $(\Pi_t^{(k)})_{t \geq 0}$ with seed-bank intensity c and seed-bank size $1/K$ to be the continuous time pure-jump Markov process with values in $\mathcal{P}_k^{\{p,s\}}$, with transitions:

$\pi \rightarrow \pi'$ at rate 1 if $\pi \succ \pi'$,

$\pi \rightarrow \pi'$ at rate c if $\pi \boxtimes \pi'$ and one p is replaced by one s ,

$\pi \rightarrow \pi'$ at rate cK if $\pi \boxtimes \pi'$ and one s is replaced by one p .

If $c = K = 1$, we speak of the *standard k -seed-bank coalescent*.

The seed-bank coalescent, II

Note that the transitions rates of the seed-bank coalescent are *consistent* when restricting from $k + 1$ to k . Hence, we may extend it to infinite marked partitions.

Definition 1.12

We may define the standard seed-bank coalescent, $(\Pi_t)_{t \geq 0} = (\Pi_t^{(\infty)})_{t \geq 0}$, with values in $\mathcal{P}_\infty^{\{p,s\}}$, as the unique Markov process distributed obtained as projective Kolmogoroff limit as $k \rightarrow \infty$ of the laws of the standard k -seed-bank coalescent (and similarly for the projective limit of the seed-bank coalescent with seed-bank intensity c and seed-bank size $1/K$).

The seed-bank coalescent - Illustration

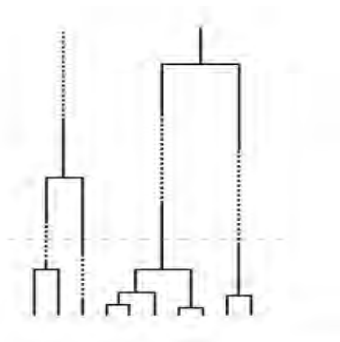


Figure: A possible realisation of the standard 10-seed-bank coalescent. Dotted lines indicate ‘inactive lineages’ (carrying an s -flag, which are prohibited from merging). At the time marked with the dashed horizontal line the process is in state $\{\{1, 2\}^s \{3\}^p \{4, 5, 6, 7, 8\}^p \{9, 10\}^s\}$.

The seed-bank coalescent as scaling limit

The seed-bank coalescent appears as the limiting genealogy of a sample taken from the Wright-Fisher model with geometric seed-bank component in the same way as the Kingman coalescent in the classical Wright-Fisher model.

Indeed, consider the genealogy of a sample of n ($\ll N$) individuals, sampled from present generation 0. Denote by $\Pi_i^{(N,n)} \in \mathcal{P}_n^{\{p,s\}}$ the configuration of the genealogy at generation $-i$, where two individuals belong to the same block of the partition $\Pi_i^{(N,n)}$ if and only if their ancestral lines have met before generation $-i$, and the flag s or p indicates whether the ancestor of generation $-i$ is a plant or a seed.

The seed-bank coalescent as scaling limit

The seed-bank coalescent appears as the limiting genealogy of a sample taken from the Wright-Fisher model with geometric seed-bank component in the same way as the Kingman coalescent in the classical Wright-Fisher model.

Indeed, consider the genealogy of a sample of n ($\ll N$) individuals, sampled from present generation 0. Denote by $\Pi_i^{(N,n)} \in \mathcal{P}_n^{\{p,s\}}$ the configuration of the genealogy at generation $-i$, where two individuals belong to the same block of the partition $\Pi_i^{(N,n)}$ if and only if their ancestral lines have met before generation $-i$, and the flag s or p indicates whether the ancestor of generation $-i$ is a plant or a seed.

The seed-bank coalescent as scaling limit

Possible events:

- One (or several) plant individuals become seeds in the previous generation
- One (or several) seeds become plants in the previous generations
- Two (or more) individuals have the same ancestor in the previous generation (which by construction is necessarily a plant), meaning that their ancestral lines merge
- Any possible combination of these three events.

The seed-bank coalescent as scaling limit

It turns out that only three possible transitions play a role in the limit as $N \rightarrow \infty$, and the others have a probability that is of smaller order.

Proposition 1.13

Under the assumption of Proposition 1.5, assume that $n \ll N$ and

$\Pi_0^{(N,n)} = \{\{1\}^p, \dots, \{n\}^p\}$. Then for $\pi, \pi' \in \mathcal{P}^{(p,s)}$,

$$\mathbb{P}(\Pi_{i+1}^{(N,n)} = \pi' \mid \Pi_i^{(N,n)} = \pi) = \begin{cases} \frac{1}{N} + O(N^{-2}) & \text{if } \pi \succ \pi', \\ \frac{c}{N} + O(N^{-2}) & \text{if } \pi \boxtimes \pi' \text{ and a } p \text{ is repld by an } s, \\ \frac{cK}{N} + O(N^{-2}) & \text{if } \pi \boxtimes \pi' \text{ and an } s \text{ is repld by a } p, \\ O(N^{-2}) & \text{otherwise.} \end{cases}$$

The seed-bank coalescent as scaling limit

Standard arguments now give the following:

Corollary 1.14

For any $n \ll N$, under the assumptions of Proposition 1.5, for $(\Pi_{\lfloor Nt \rfloor}^{(N,n)})$ converges weakly as $N \rightarrow \infty$ to the seed-bank coalescent $(\Pi_t^{(n)})$ started with n plants.

Properties of the seed-bank coalescent (in progress)

It is not surprising that the seed-bank coalescent behaves very different from a classical Kingman coalescent.

We illustrate this with two examples.

Coming down from infinity

The notion of *coming down from infinity* was introduced by [PITMAN 1999] and [SCHWEINSBERG 2000]. They say that an exchangeable coalescent process *comes down from infinity* if the corresponding block counting process (of an infinite sample) has finitely many blocks immediately after time 0 (i.e. for all $t > 0$ a.s.).

Further, the coalescent is said to *stay infinite* if the number of blocks is infinite for all $t \geq 0$ a.s.

[SCHWEINSBERG] also gives a necessary and sufficient criterion for so-called *Lambda-coalescents* to come down from infinity. In particular, the Kingman coalescent does come down from infinity.

However, note that the seed-bank coalescent does not belong to the class of Lambda-coalescents, so that Schweinsberg's result does not immediately apply.

Coming down from infinity

The notion of *coming down from infinity* was introduced by [PITMAN 1999] and [SCHWEINSBERG 2000]. They say that an exchangeable coalescent process *comes down from infinity* if the corresponding block counting process (of an infinite sample) has finitely many blocks immediately after time 0 (i.e. for all $t > 0$ a.s.).

Further, the coalescent is said to *stay infinite* if the number of blocks is infinite for all $t \geq 0$ a.s.

[SCHWEINSBERG] also gives a necessary and sufficient criterion for so-called *Lambda-coalescents* to come down from infinity. In particular, the Kingman coalescent does come down from infinity.

However, note that the seed-bank coalescent does not belong to the class of Lambda-coalescents, so that Schweinsberg's result does not immediately apply.

Coming down from infinity

The notion of *coming down from infinity* was introduced by [PITMAN 1999] and [SCHWEINSBERG 2000]. They say that an exchangeable coalescent process *comes down from infinity* if the corresponding block counting process (of an infinite sample) has finitely many blocks immediately after time 0 (i.e. for all $t > 0$ a.s.).

Further, the coalescent is said to *stay infinite* if the number of blocks is infinite for all $t \geq 0$ a.s.

[SCHWEINSBERG] also gives a necessary and sufficient criterion for so-called *Lambda-coalescents* to come down from infinity. In particular, the Kingman coalescent does come down from infinity.

However, note that the seed-bank coalescent does not belong to the class of Lambda-coalescents, so that Schweinsberg's result does not immediately apply.

Coming down from infinity

The notion of *coming down from infinity* was introduced by [PITMAN 1999] and [SCHWEINSBERG 2000]. They say that an exchangeable coalescent process *comes down from infinity* if the corresponding block counting process (of an infinite sample) has finitely many blocks immediately after time 0 (i.e. for all $t > 0$ a.s.).

Further, the coalescent is said to *stay infinite* if the number of blocks is infinite for all $t \geq 0$ a.s.

[SCHWEINSBERG] also gives a necessary and sufficient criterion for so-called *Lambda-coalescents* to come down from infinity. In particular, the Kingman coalescent does come down from infinity.

However, note that the seed-bank coalescent does not belong to the class of Lambda-coalescents, so that Schweinsberg's result does not immediately apply.

Coming down from infinity

Theorem 1.15 (in progress)

*The seed bank coalescent does **not** come down from infinity. In fact, its block-counting process $(N_t, M_t)_{t \geq 0}$ stays infinite, that is, for each infinite starting configuration (N_0, M_0) with $N_0 = \infty$,*

$$\mathbb{P}\{N_t + M_t = \infty, \text{ for all } t > 0\} = 1.$$

Of course, this has to do with lineages immediately jumping into the seed-bank. Proof omitted.

Time to the most recent common ancestor

In view of the previous subsection it is now quite obvious that the seed bank causes a relevant delay in the time to the most recent common ancestor.

Definition 1.16

Let $k \in \mathbb{N} \cup \{\infty\}$. We define the *time to the most recent common ancestor* of a sample of n plants to be

$$T_{MRCA}[n] = \inf\{t > 0 : |\Pi_t^{(k)}| = 1 \text{ given that } \Pi_0 = \{\{1\}^p \dots \{n\}^p\}\},$$

or equivalently

$$T_{MRCA}[n] = \inf\{t > 0 : (N_t, M_t) = (1, 0) \text{ given that } (N_0, M_0) = (n, 0)\}$$

Time to the most recent common ancestor

Theorem 1.17 (in progress)

For all $c, K \in (0, \infty)$, the seed bank coalescent satisfies

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E}[T_{MRCA}[n]]}{\log \log n} > 0. \quad (4)$$

This should be compared with a similar result for the *Bolthausen-Sznitman coalescent* in [GOLDSCHMIDT & MARTIN 2005]

Finally...

... thank you for your attention!