A survey of fitness-based models for biological evolution

Iddo Ben-Ari, University of Connecticut

U of Rochester, February 2018

Introduction

Toy models for time evolution of a system consisting of a population of “species”.

Common features

- Population is asymptotically large.
- Fitness-based models:
  - “At birth” each species is assigned a random “fitness” independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model (‘93)
- A model presented by Guiol Machado and Schinazi (‘11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of "species".

Common features

- Population is asymptotically large.
- Fitness-based models:
  - "At birth" each species is assigned a random "fitness" independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model ('93)
- A model presented by Guiol Machado and Schinazi ('11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of "species".

Common features

- Population is asymptotically large.
- Fitness-based models:
  - "At birth" each species is assigned a random "fitness" independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model ('93)
- A model presented by Guiol Machado and Schinazi ('11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of "species".

Common features

- Population is asymptotically large.
- Fitness-based models:
  - "At birth" each species is assigned a random "fitness" independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model ('93)
- A model presented by Guiol Machado and Schinazi ('11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of “species”.

Common features

- Population is asymptotically large.
- Fitness-based models:
  - “At birth” each species is assigned a random “fitness” independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model (‘93)
- A model presented by Guiol Machado and Schinazi (‘11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of “species”.

Common features

- Population is asymptotically large.
- Fitness-based models:
  - “At birth” each species is assigned a random “fitness” independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model (‘93)
- A model presented by Guiol Machado and Schinazi (‘11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of "species".

Common features

- Population is asymptotically large.
- Fitness-based models:
  - "At birth" each species is assigned a random "fitness" independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model ('93)
- A model presented by Guiol Machado and Schinazi ('11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of "species".

Common features

- Population is asymptotically large.
- Fitness-based models:
  - "At birth" each species is assigned a random "fitness" independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model ('93)
- A model presented by Guiol Machado and Schinazi ('11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of "species".

Common features

- Population is asymptotically large.
- Fitness-based models:
  - "At birth" each species is assigned a random "fitness" independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model ('93)
- A model presented by Guiol Machado and Schinazi ('11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of "species".

Common features

- Population is asymptotically large.
- Fitness-based models:
  - "At birth" each species is assigned a random "fitness" independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model ('93)
- A model presented by Guiol Machado and Schinazi ('11)
- Variations of the above.

Introduction

Toy models for time evolution of a system consisting of a population of “species”.

Common features

- Population is asymptotically large.
- Fitness-based models:
  - “At birth” each species is assigned a random “fitness” independent of past.
  - Time evolution eliminates species with lowest fitness from the system.

What is the asymptotic fitness distribution ?

The Models

- Bak-Sneppen model (‘93)
- A model presented by Guiol Machado and Schinazi (‘11)
- Variations of the above.

Bak Sneppen

Bak Sneppen

One of the first models claimed through numerical simulations to exhibit self-organized criticality.

A discrete time ergodic Markov processes with

- N species arranged on the vertices of a cycle (or any finite connected graph).
- Each is a assigned an initial fitness, IID U$[0, 1]$.
- Evolution: at each time, the species with lowest fitness and its neighbors are replaced by new species with IID U$[0, 1]$ fitnesses.

Watch simulation

Simulations suggest

$$\pi_N \underset{N\to\infty}{\to} \text{IID U}[p_c, 1], \text{ where } p_c \sim 2/3,$$

and π_N is the stationary distribution.

Open.

Bak Sneppen

One of the first models claimed through numerical simulations to exhibit self-organized criticality.

A discrete time ergodic Markov processes with

- N species arranged on the vertices of a cycle (or any finite connected graph).
- Each is a assigned an initial fitness, IID $\mathrm{U}[0,1]$.
- Evolution: at each time, the species with lowest fitness and its neighbors are replaced by new species with IID $\mathrm{U}[0,1]$ fitnesses.

Watch simulation

Simulations suggest

$$\pi_N \underset{N\to\infty}{\to} \text{IID } \mathrm{U}[p_c, 1], \text{ where } p_c \sim 2/3,$$

and π_N is the stationary distribution.

Open.

Bak Sneppen

One of the first models claimed through numerical simulations to exhibit self-organized criticality.

A discrete time ergodic Markov processes with

- N species arranged on the vertices of a cycle (or any finite connected graph).
- Each is a assigned an initial fitness, IID U$[0,1]$.
- Evolution: at each time, the species with lowest fitness and its neighbors are replaced by new species with IID U$[0,1]$ fitnesses.

Watch simulation

Simulations suggest

$$\pi_N \underset{N\to\infty}{\to} \text{IID } \mathrm{U}[p_c, 1], \text{ where } p_c \sim 2/3,$$

and π_N is the stationary distribution.

Open.

Bak Sneppen

One of the first models claimed through numerical simulations to exhibit self-organized criticality.

A discrete time ergodic Markov processes with

- N species arranged on the vertices of a cycle (or any finite connected graph).
- Each is a assigned an initial fitness, IID U$[0, 1]$.
- Evolution: at each time, the species with lowest fitness and its neighbors are replaced by new species with IID U$[0, 1]$ fitnesses.

Watch simulation

Simulations suggest

$$\pi_N \underset{N\to\infty}{\longrightarrow} \text{IID U}[p_c, 1], \text{ where } p_c \sim 2/3,$$

and π_N is the stationary distribution.

Open.

Bak Sneppen

One of the first models claimed through numerical simulations to exhibit self-organized criticality.

A discrete time ergodic Markov processes with

- N species arranged on the vertices of a cycle (or any finite connected graph).
- Each is a assigned an initial fitness, IID U$[0, 1]$.
- Evolution: at each time, the species with lowest fitness and its neighbors are replaced by new species with IID U$[0, 1]$ fitnesses.

Watch simulation

Simulations suggest

$$\pi_N \underset{N\to\infty}{\longrightarrow} \text{IID } \mathrm{U}[p_c, 1], \text{ where } p_c \sim 2/3,$$

and π_N is the stationary distribution.

Open.

Bak Sneppen

One of the first models claimed through numerical simulations to exhibit self-organized criticality.

A discrete time ergodic Markov processes with

- N species arranged on the vertices of a cycle (or any finite connected graph).
- Each is a assigned an initial fitness, IID U$[0, 1]$.
- Evolution: at each time, the species with lowest fitness and its neighbors are replaced by new species with IID U$[0, 1]$ fitnesses.

Watch simulation

Simulations suggest

$$\pi_N \underset{N\to\infty}{\to} \text{IID U}[p_c, 1], \text{ where } p_c \sim 2/3,$$

and π_N is the stationary distribution.

Open.

Bak Sneppen

One of the first models claimed through numerical simulations to exhibit self-organized criticality.

A discrete time ergodic Markov processes with

- N species arranged on the vertices of a cycle (or any finite connected graph).
- Each is a assigned an initial fitness, IID U$[0, 1]$.
- Evolution: at each time, the species with lowest fitness and its neighbors are replaced by new species with IID U$[0, 1]$ fitnesses.

Watch simulation

Simulations suggest

$$\pi_N \underset{N\to\infty}{\to} \text{IID U}[p_c, 1], \text{ where } p_c \sim 2/3,$$

and π_N is the stationary distribution.

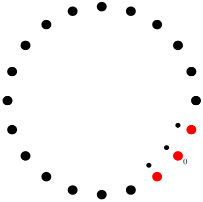
Open.

Bak Sneppen Avalanches

An **avalanche** from threshold p is a part of the path from time all fitnesses are $\geq p$ until next time this happens.

The avalanches provide a natural regenerative structure for the process.

- Evolution of avalanche depends on the past only through the location of site with lowest fitness when started.
- As a result, the sequence of durations of avalanches are IID, and so is the number of vertices affected during each avalanche, AKA the range of the avalanche.

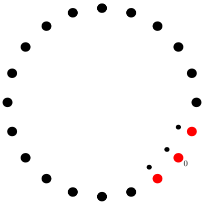


Bak Sneppen Avalanches

An **avalanche** from threshold p is a part of the path from time all fitnesses are $\geq p$ until next time this happens.

The avalanches provide a natural regenerative structure for the process.

- Evolution of avalanche depends on the past only through the location of site with lowest fitness when started.
- As a result, the sequence of durations of avalanches are IID, and so is the number of vertices affected during each avalanche, AKA the range of the avalanche.

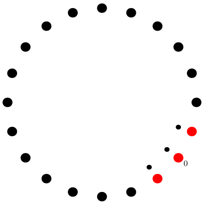


Bak Sneppen Avalanches

An **avalanche** from threshold p is a part of the path from time all fitnesses are $\geq p$ until next time this happens.

The avalanches provide a natural regenerative structure for the process.

- Evolution of avalanche depends on the past only through the location of site with lowest fitness when started.
- As a result, the sequence of durations of avalanches are IID, and so is the number of vertices affected during each avalanche, AKA the range of the avalanche.

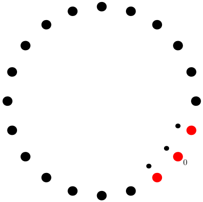


Bak Sneppen Avalanches

An **avalanche** from threshold p is a part of the path from time all fitnesses are $\geq p$ until next time this happens.

The avalanches provide a natural regenerative structure for the process.

- Evolution of avalanche depends on the past only through the location of site with lowest fitness when started.
- As a result, the sequence of durations of avalanches are IID, and so is the number of vertices affected during each avalanche, AKA the range of the avalanche.



Bak-Sneppen Avalanche Statistics

Notation

$$D_N(p) = \text{Duration of avalanche from threshold p}$$
$$R_N(p) = \text{Range of avalance from threshold p}$$
$$P_N(p) = P(R_N(p) = N)$$

Consider an avalanche from threshold p on $\mathbb{Z}$ with initial fitness configuration

$$\ldots, 1, 1, \ldots, \underset{\substack{\uparrow \\ \text{origin}}}{p}, 1, 1, \ldots$$

As before, let

$$D_\infty(p) = \text{Duration of avalanche}$$
$$R_\infty(p) = \text{Range of avalanche}$$
$$P_\infty(p) = P(R_\infty(p) = \infty).$$

Theorem 1 (Meester-Znamenski '04)

$ED_N(p) \to ED_\infty(p),\ ER_N(p) \to ER_\infty(p), P_N(p) \to P_\infty(p).$

- Asymptotic properties can be studied by considering the infinite system.
- Main idea: embedding in and coupling of finite system in infinite system.

Bak-Sneppen Avalanche Statistics

Notation

$$\begin{aligned} D_N(p) &= \text{Duration of avalanche from threshold p} \\ R_N(p) &= \text{Range of avalance from threshold p} \\ P_N(p) &= P(R_N(p) = N) \end{aligned}$$

Consider an avalanche from threshold p on $\mathbb{Z}$ with initial fitness configuration

$$\ldots, 1, 1, \ldots, \underset{\substack{\uparrow \\ \text{origin}}}{p}, 1, 1, \ldots$$

As before, let

$$\begin{aligned} D_\infty(p) &= \text{Duration of avalanche} \\ R_\infty(p) &= \text{Range of avalanche} \\ P_\infty(p) &= P(R_\infty(p) = \infty). \end{aligned}$$

Theorem 1 (Meester-Znamenski '04)

$ED_N(p) \to ED_\infty(p)$, $ER_N(p) \to ER_\infty(p)$, $P_N(p) \to P_\infty(p)$.

- Asymptotic properties can be studied by considering the infinite system.
- Main idea: embedding in and coupling of finite system in infinite system.

Bak-Sneppen Avalanche Statistics

Notation

$$\begin{aligned} D_N(p) &= \text{Duration of avalanche from threshold p} \\ R_N(p) &= \text{Range of avalance from threshold p} \\ P_N(p) &= P(R_N(p) = N) \end{aligned}$$

Consider an avalanche from threshold p on $\mathbb{Z}$ with initial fitness configuration

$$\ldots, 1, 1, \ldots, \underset{\text{origin}}{\underset{\uparrow}{p}}, 1, 1, \ldots$$

As before, let

$$\begin{aligned} D_\infty(p) &= \text{Duration of avalanche} \\ R_\infty(p) &= \text{Range of avalanche} \\ P_\infty(p) &= P(R_\infty(p) = \infty). \end{aligned}$$

Theorem 1 (Meester-Znamenski '04)

$ED_N(p) \to ED_\infty(p)$, $ER_N(p) \to ER_\infty(p)$, $P_N(p) \to P_\infty(p)$.

- Asymptotic properties can be studied by considering the infinite system.
- Main idea: embedding in and coupling of finite system in infinite system.

Bak-Sneppen Avalanche Statistics

Notation

$$\begin{aligned} D_N(p) &= \text{Duration of avalanche from threshold p} \\ R_N(p) &= \text{Range of avalance from threshold p} \\ P_N(p) &= P(R_N(p) = N) \end{aligned}$$

Consider an avalanche from threshold p on $\mathbb{Z}$ with initial fitness configuration

$$\ldots, 1, 1, \ldots, \underset{\substack{\uparrow \\ \text{origin}}}{p}, 1, 1, \ldots$$

As before, let

$$\begin{aligned} D_\infty(p) &= \text{Duration of avalanche} \\ R_\infty(p) &= \text{Range of avalanche} \\ P_\infty(p) &= P(R_\infty(p) = \infty). \end{aligned}$$

Theorem 1 (Meester-Znamenski '04)

$ED_N(p) \to ED_\infty(p),\ ER_N(p) \to ER_\infty(p), P_N(p) \to P_\infty(p).$

- Asymptotic properties can be studied by considering the infinite system.
- Main idea: embedding in and coupling of finite system in infinite system.

Bak-Sneppen Avalanche Statistics

Notation

$$D_N(p) = \text{Duration of avalanche from threshold p}$$
$$R_N(p) = \text{Range of avalance from threshold p}$$
$$P_N(p) = P(R_N(p) = N)$$

Consider an avalanche from threshold p on $\mathbb{Z}$ with initial fitness configuration

$$\ldots, 1, 1, \ldots, \underset{\substack{\uparrow \\ \text{origin}}}{p}, 1, 1, \ldots$$

As before, let

$$D_\infty(p) = \text{Duration of avalanche}$$
$$R_\infty(p) = \text{Range of avalanche}$$
$$P_\infty(p) = P(R_\infty(p) = \infty).$$

Theorem 1 (Meester-Znamenski '04)

$ED_N(p) \to ED_\infty(p),\ ER_N(p) \to ER_\infty(p), P_N(p) \to P_\infty(p).$

- Asymptotic properties can be studied by considering the infinite system.
- Main idea: embedding in and coupling of finite system in infinite system.

Bak-Sneppen Avalanche Statistics

Notation

$$D_N(p) = \text{Duration of avalanche from threshold p}$$
$$R_N(p) = \text{Range of avalance from threshold p}$$
$$P_N(p) = P(R_N(p) = N)$$

Consider an avalanche from threshold p on $\mathbb{Z}$ with initial fitness configuration

$$\ldots, 1, 1, \ldots, \underset{\substack{\uparrow \\ \text{origin}}}{p}, 1, 1, \ldots$$

As before, let

$$D_\infty(p) = \text{Duration of avalanche}$$
$$R_\infty(p) = \text{Range of avalanche}$$
$$P_\infty(p) = P(R_\infty(p) = \infty).$$

Theorem 1 (Meester-Znamenski '04)

$ED_N(p) \to ED_\infty(p),\ ER_N(p) \to ER_\infty(p), P_N(p) \to P_\infty(p).$

- Asymptotic properties can be studied by considering the infinite system.
- Main idea: embedding in and coupling of finite system in infinite system.

Critical Thresholds

Define

$$\begin{aligned} p_D &= \inf\{p : ED_\infty(p) = \infty\} \\ p_R &= \inf\{p : ER_\infty(p) = \infty\} \\ p_P &= \inf\{p : P_\infty(p) > 0\} \end{aligned}$$

Theorem 2 (Meester-Znamenski '03,Meester-Znamenski '04)

1. $0 < p_D = p_R \le p_P < 1 - e^{-68}$.
2. *If* $p_R = p_P$, *then* $\pi_N \underset{N\to\infty}{\to}$ *IID* $U[p_P, 1]$.

Letting F be the fitness at some distinguished site 0, then

Proposition 1

1. $\pi_N(F \le p_D) \to 0$.
2. $\pi_N(F \in \cdot | F > p_P) \to U[p_P, 1]$.

This was not stated in the paper, but follows from the proofs.

Critical Thresholds

Define

$$p_D = \inf\{p : ED_\infty(p) = \infty\}$$
$$p_R = \inf\{p : ER_\infty(p) = \infty\}$$
$$p_P = \inf\{p : P_\infty(p) > 0\}$$

Theorem 2 (Meester-Znamenski '03,Meester-Znamenski '04)

1. $0 < p_D = p_R \leq p_P < 1 - e^{-68}$.
2. *If* $p_R = p_P$, *then* $\pi_N \underset{N\to\infty}{\to}$ *IID* $U[p_P, 1]$.

Letting F be the fitness at some distinguished site 0, then

Proposition 1

1. $\pi_N(F \leq p_D) \to 0$.
2. $\pi_N(F \in \cdot | F > p_P) \to U[p_P, 1]$.

This was not stated in the paper, but follows from the proofs.

Critical Thresholds

Define

$$\begin{aligned} p_D &= \inf\{p : ED_\infty(p) = \infty\} \\ p_R &= \inf\{p : ER_\infty(p) = \infty\} \\ p_P &= \inf\{p : P_\infty(p) > 0\} \end{aligned}$$

Theorem 2 (Meester-Znamenski '03,Meester-Znamenski '04)

1. $0 < p_D = p_R \le p_P < 1 - e^{-68}$.
2. *If* $p_R = p_P$*, then* $\pi_N \underset{N\to\infty}{\to}$ *IID* $U[p_P, 1]$.

Letting F be the fitness at some distinguished site 0, then

Proposition 1

1. $\pi_N(F \le p_D) \to 0$.
2. $\pi_N(F \in \cdot | F > p_P) \to U[p_P, 1]$.

This was not stated in the paper, but follows from the proofs.

Critical Thresholds

Define

$$\begin{aligned} p_D &= \inf\{p : ED_\infty(p) = \infty\} \\ p_R &= \inf\{p : ER_\infty(p) = \infty\} \\ p_P &= \inf\{p : P_\infty(p) > 0\} \end{aligned}$$

Theorem 2 (Meester-Znamenski '03, Meester-Znamenski '04)

1. $0 < p_D = p_R \leq p_P < 1 - e^{-68}$.
2. *If* $p_R = p_P$*, then* $\pi_N \underset{N\to\infty}{\to}$ *IID* $U[p_P, 1]$.

Letting F be the fitness at some distinguished site 0, then

Proposition 1

1. $\pi_N(F \leq p_D) \to 0$.
2. $\pi_N(F \in \cdot | F > p_P) \to U[p_P, 1]$.

This was not stated in the paper, but follows from the proofs.

Bak Sneppen, a little more

Proposition 2 (B. WIP)

Let

$$\rho = \inf_{p,d} \sum_{k=1}^{\infty} \frac{1}{d_k p_k},$$

where the infimum is taken over all probability distributions $(p_1, p_2, \ldots)$ *on* $\mathbb{N}$ *and all-integer nondecreasing valued sequences* $(d_k)_{k\in\mathbb{N}}$ *with the growth constraint* $d_1 = 1, d_{k+1} < 2d_k$ *for* $k > 1$*. Then*

$$p_P \leq 1 - e^{-\rho},$$

- Simulations give $\rho < 11.3$.
- We need to get to $-\ln \frac{1}{3} = 1.09861228867$.

Theorem 3 (B. WIP)

If $P(R_\infty(p) > r) \geq cr^{-\alpha}$ *for some* $\alpha < 1$*, then* $p_P \leq p$*.*

- Roughly speaking, if $R_\infty(p)$ "little" short of being integrable, then p is already above p_P.

Bak Sneppen, a little more

Proposition 2 (B. WIP)

Let

$$\rho = \inf_{p,d} \sum_{k=1}^{\infty} \frac{1}{d_k p_k},$$

where the infimum is taken over all probability distributions $(p_1, p_2, \dots)$ *on* $\mathbb{N}$ *and all-integer nondecreasing valued sequences* $(d_k)_{k\in\mathbb{N}}$ *with the growth constraint* $d_1 = 1, d_{k+1} < 2d_k$ *for* $k > 1$*. Then*

$$p_P \leq 1 - e^{-\rho},$$

- Simulations give $\rho < 11.3$.
- We need to get to $-\ln \frac{1}{3} = 1.09861228867$.

Theorem 3 (B. WIP)

If $P(R_\infty(p) > r) \geq cr^{-\alpha}$ *for some* $\alpha < 1$*, then* $p_P \leq p$*.*

- Roughly speaking, if $R_\infty(p)$ "little" short of being integrable, then p is already above p_P.

Bak Sneppen, a little more

Proposition 2 (B. WIP)

Let

$$\rho = \inf_{p,d} \sum_{k=1}^{\infty} \frac{1}{d_k p_k},$$

where the infimum is taken over all probability distributions $(p_1, p_2, \dots)$ *on* $\mathbb{N}$ *and all-integer nondecreasing valued sequences* $(d_k)_{k\in\mathbb{N}}$ *with the growth constraint* $d_1 = 1, d_{k+1} < 2d_k$ *for* $k > 1$*. Then*

$$p_P \leq 1 - e^{-\rho},$$

- Simulations give $\rho < 11.3$.
- We need to get to $-\ln\frac{1}{3} = 1.09861228867$.

Theorem 3 (B. WIP)

If $P(R_\infty(p) > r) \geq cr^{-\alpha}$ *for some* $\alpha < 1$*, then* $p_P \leq p$*.*

- Roughly speaking, if $R_\infty(p)$ "little" short of being integrable, then p is already above p_P.

Bak Sneppen, a little more

Proposition 2 (B. WIP)

Let

$$\rho = \inf_{p,d} \sum_{k=1}^{\infty} \frac{1}{d_k p_k},$$

where the infimum is taken over all probability distributions $(p_1, p_2, \dots)$ *on* $\mathbb{N}$ *and all-integer nondecreasing valued sequences* $(d_k)_{k\in\mathbb{N}}$ *with the growth constraint* $d_1 = 1, d_{k+1} < 2d_k$ *for* $k > 1$*. Then*

$$p_P \leq 1 - e^{-\rho},$$

- Simulations give $\rho < 11.3$.
- We need to get to $-\ln\frac{1}{3} = 1.09861228867$.

Theorem 3 (B. WIP)

If $P(R_\infty(p) > r) \geq cr^{-\alpha}$ *for some* $\alpha < 1$*, then* $p_P \leq p$*.*

- Roughly speaking, if $R_\infty(p)$ "little" short of being integrable, then p is already above p_P.

Bak Sneppen, a little more

Proposition 2 (B. WIP)

Let

$$\rho = \inf_{p,d} \sum_{k=1}^{\infty} \frac{1}{d_k p_k},$$

where the infimum is taken over all probability distributions $(p_1, p_2, \dots)$ *on* $\mathbb{N}$ *and all-integer nondecreasing valued sequences* $(d_k)_{k\in\mathbb{N}}$ *with the growth constraint* $d_1 = 1, d_{k+1} < 2d_k$ *for* $k > 1$*. Then*

$$p_P \leq 1 - e^{-\rho},$$

- Simulations give $\rho < 11.3$.
- We need to get to $-\ln\frac{1}{3} = 1.09861228867$.

Theorem 3 (B. WIP)

If $P(R_\infty(p) > r) \geq cr^{-\alpha}$ *for some* $\alpha < 1$*, then* $p_P \leq p$*.*

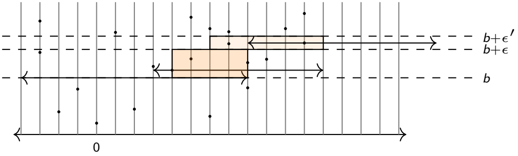
- Roughly speaking, if $R_\infty(p)$ "little" short of being integrable, then p is already above p_P.

Bak Sneppen – on proofs

- Tool: Graphical representation of avalanches on $\mathbb{Z}$, due to Meester and his coauthors.
- Switch from uniform fitnesses to Exp(1). This allows for Poisson process techniques.

At end of avalanche from threshold b, fitness of sites in its range IID b+Exp(1).

- To each site attach a rate-1 Poisson process, processes are independent.
- Suppose the avalanche from threshold b starting from the origin has the range given by the arrow.
- Fitness distribution of sites in range coincides with the first arrivals of the Poisson processes above b.
- The range of avalanche from threshold $b + \epsilon$ will be at least $\frac{3}{2} \times R_b$, if at least one of the avalanches in the orange region extends to the right at least as R_b did.
- Allows to approach through thinning of a Poisson Point Process.
- For large enough b, one can show that exists an infinite cascade of such avalanches below fitness $b + \epsilon$.



Local Bak-Sneppen

Joint with R.C. Silva

Two Geometries

What would be a "proper" tractable analog for Bak-Sneppen ?

The difficulty in the Bak-Sneppen model stems from the following:

- Use **complete graph** geometry to locate the global minimum.
- Use **"nearest neighbor"** geometry to determine at what vertices species will be replaced.

A first attempt at this question would be

- Use one geometry.

The complete graph geometry is trivial so we're left with the latter.

Two Geometries

What would be a “proper” tractable analog for Bak-Sneppen ?

The difficulty in the Bak-Sneppen model stems from the following:

- Use **complete graph** geometry to locate the global minimum.
- Use **“nearest neighbor”** geometry to determine at what vertices species will be replaced.

A first attempt at this question would be

- Use one geometry.

The complete graph geometry is trivial so we’re left with the latter.

Two Geometries

What would be a "proper" tractable analog for Bak-Sneppen ?

The difficulty in the Bak-Sneppen model stems from the following:

- Use **complete graph** geometry to locate the global minimum.
- Use **"nearest neighbor"** geometry to determine at what vertices species will be replaced.

A first attempt at this question would be

- Use one geometry.

The complete graph geometry is trivial so we're left with the latter.

Two Geometries

What would be a "proper" tractable analog for Bak-Sneppen ?

The difficulty in the Bak-Sneppen model stems from the following:

- Use **complete graph** geometry to locate the global minimum.
- Use **"nearest neighbor"** geometry to determine at what vertices species will be replaced.

A first attempt at this question would be

- Use one geometry.

The complete graph geometry is trivial so we're left with the latter.

Two Geometries

What would be a “proper” tractable analog for Bak-Sneppen ?

The difficulty in the Bak-Sneppen model stems from the following:

- Use **complete graph** geometry to locate the global minimum.
- Use **“nearest neighbor”** geometry to determine at what vertices species will be replaced.

A first attempt at this question would be

- Use one geometry.

The complete graph geometry is trivial so we’re left with the latter.

Two Geometries

What would be a "proper" tractable analog for Bak-Sneppen ?

The difficulty in the Bak-Sneppen model stems from the following:

- Use **complete graph** geometry to locate the global minimum.
- Use **"nearest neighbor"** geometry to determine at what vertices species will be replaced.

A first attempt at this question would be

- Use one geometry.

The complete graph geometry is trivial so we're left with the latter.

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Local Bak-Sneppen

Consider a finite, connected (undirected) graph $G = (V, E)$.

Initially

- Assign IID U$[0, 1]$ fitnesses to each $v \in V$.
- Set X_0 as the vertex with lowest fitness.

Time evolution

- Given X_n, set X_{n+1} to be the vertex with minimal fitness among $u \sim X_n$ and X_n itself.
- Set fitness of all elements in neighborhood of X_{n+1} as IID U$[0, 1]$, independent of past.

Observe

- Markov chain on state space = product of V and $[0, 1]$-valued functions on V.
- Chain is ergodic.

What can we say about this new process ?

Stationary Distribution for Local Bak-Sneppen

Notation

- For $v \in V$, $A_v = \{u \in V : \{u, v\} \in E \text{ or } u = v\}$.
- **Random walk on** G: from $v \in V$ move to uniformly sampled $u \in A_v$.
- Stationary distribution: $\mu(v) = \frac{|A_v|}{\sum_{u \in V} |A_u|}$.
- If $U_1, \ldots, U_n$ are IID U$[0,1]$, then set U$(n, [0,1])$ as the distribution

$$P(U_1 \in \cdot | U_1 > \min\{U_2, \ldots, U_n\}).$$

Theorem 4 (Silva-B.)

- *Let* $(Z_n^u : n \in \mathbb{Z}_+)$ *be independent random walks on* G *with* $Z_0^u = u$.
- *Sample* X *independently according to* μ.
- *Set*

$$\tau_v = \inf\{n \in \mathbb{Z}_+ : Z_n^X \in A_v\}, \text{ and } V_i = \{v \in V : \tau_v = i\}.$$

Given $V_0, V_1, \ldots,$ *assign fitnesses at each* $v \in V$*, which are independent and are*

1. $U[0,1]$ *for* $v \in A_X = V_0$.
2. $U(|A_{Z_i^X}|, [0,1])$ *if* $v \in V_i,\ i > 0$.

Then the joint distribution of X *and the fitnesses is stationary for the local Bak-Sneppen.*

Stationary Distribution for Local Bak-Sneppen

Notation

- For $v \in V$, $A_v = \{u \in V : \{u, v\} \in E \text{ or } u = v\}$.
- **Random walk on** G: from $v \in V$ move to uniformly sampled $u \in A_v$.
- Stationary distribution: $\mu(v) = \frac{|A_v|}{\sum_{u \in V} |A_u|}$.
- If $U_1, \ldots, U_n$ are IID U[0, 1], then set U(n, [0, 1]) as the distribution

$$P(U_1 \in \cdot | U_1 > \min\{U_2, \ldots, U_n\}).$$

Theorem 4 (Silva-B.)

- *Let* $(Z_n^u : n \in \mathbb{Z}_+)$ *be independent random walks on* G *with* $Z_0^u = u$.
- *Sample* X *independently according to* μ.
- *Set*

$$\tau_v = \inf\{n \in \mathbb{Z}_+ : Z_n^X \in A_v\}, \text{ and } V_i = \{v \in V : \tau_v = i\}.$$

Given $V_0, V_1, \ldots,$ *assign fitnesses at each* $v \in V$*, which are independent and are*

1. $U[0, 1]$ *for* $v \in A_X = V_0$.
2. $U(|A_{Z_i^X}|, [0, 1])$ *if* $v \in V_i, \; i > 0$.

Then the joint distribution of X *and the fitnesses is stationary for the local Bak-Sneppen.*

Stationary Distribution for Local Bak-Sneppen

Notation

- For $v \in V$, $A_v = \{u \in V : \{u, v\} \in E \text{ or } u = v\}$.
- **Random walk on** G: from $v \in V$ move to uniformly sampled $u \in A_v$.
- Stationary distribution: $\mu(v) = \frac{|A_v|}{\sum_{u \in V} |A_u|}$.
- If $U_1, \dots, U_n$ are IID U$[0,1]$, then set U$(n, [0,1])$ as the distribution

$$P(U_1 \in \cdot | U_1 > \min\{U_2, \dots, U_n\}).$$

Theorem 4 (Silva-B.)

- *Let* $(Z_n^u : n \in \mathbb{Z}_+)$ *be independent random walks on* G *with* $Z_0^u = u$.
- *Sample* X *independently according to* μ.
- *Set*

$$\tau_v = \inf\{n \in \mathbb{Z}_+ : Z_n^X \in A_v\}, \text{ and } V_i = \{v \in V : \tau_v = i\}.$$

Given $V_0, V_1, \dots,$ *assign fitnesses at each* $v \in V$, *which are independent and are*

1. $U[0,1]$ *for* $v \in A_X = V_0$.
2. $U(|A_{Z_i^X}|, [0,1])$ *if* $v \in V_i, \ i > 0$.

Then the joint distribution of X *and the fitnesses is stationary for the local Bak-Sneppen.*

Stationary Distribution for Local Bak-Sneppen

Notation

- For $v \in V$, $A_v = \{u \in V : \{u, v\} \in E \text{ or } u = v\}$.
- **Random walk on** G: from $v \in V$ move to uniformly sampled $u \in A_v$.
- Stationary distribution: $\mu(v) = \frac{|A_v|}{\sum_{u \in V} |A_u|}$.
- If $U_1, \ldots, U_n$ are IID U[0, 1], then set U$(n, [0, 1])$ as the distribution

$$P(U_1 \in \cdot | U_1 > \min\{U_2, \ldots, U_n\}).$$

Theorem 4 (Silva-B.)

- *Let $(Z_n^u : n \in \mathbb{Z}_+)$ be independent random walks on G with $Z_0^u = u$.*
- *Sample X independently according to μ.*
- *Set*

$$\tau_v = \inf\{n \in \mathbb{Z}_+ : Z_n^X \in A_v\}, \text{ and } V_i = \{v \in V : \tau_v = i\}.$$

Given $V_0, V_1, \ldots,$ assign fitnesses at each $v \in V$, which are independent and are

1. *$U[0, 1]$ for $v \in A_X = V_0$.*
2. *$U(|A_{Z_i^X}|, [0, 1])$ if $v \in V_i$, $i > 0$.*

Then the joint distribution of X and the fitnesses is stationary for the local Bak-Sneppen.

Stationary Distribution for Local Bak-Sneppen

Notation

- For $v \in V$, $A_v = \{u \in V : \{u, v\} \in E \text{ or } u = v\}$.
- **Random walk on** G: from $v \in V$ move to uniformly sampled $u \in A_v$.
- Stationary distribution: $\mu(v) = \frac{|A_v|}{\sum_{u \in V} |A_u|}$.
- If $U_1, \ldots, U_n$ are IID U$[0,1]$, then set U$(n, [0,1])$ as the distribution

$$P(U_1 \in \cdot | U_1 > \min\{U_2, \ldots, U_n\}).$$

Theorem 4 (Silva-B.)

- *Let* $(Z_n^u : n \in \mathbb{Z}_+)$ *be independent random walks on* G *with* $Z_0^u = u$.
- *Sample* X *independently according to* μ.
- *Set*

$$\tau_v = \inf\{n \in \mathbb{Z}_+ : Z_n^X \in A_v\}, \text{ and } V_i = \{v \in V : \tau_v = i\}.$$

Given $V_0, V_1, \ldots,$ *assign fitnesses at each* $v \in V$, *which are independent and are*

1. $U[0,1]$ *for* $v \in A_X = V_0$.
2. $U(|A_{Z_i^X}|, [0,1])$ *if* $v \in V_i$, $i > 0$.

Then the joint distribution of X *and the fitnesses is stationary for the local Bak-Sneppen.*

Stationary Distribution for Local Bak-Sneppen

Notation

- For $v \in V$, $A_v = \{u \in V : \{u, v\} \in E \text{ or } u = v\}$.
- **Random walk on** G: from $v \in V$ move to uniformly sampled $u \in A_v$.
- Stationary distribution: $\mu(v) = \frac{|A_v|}{\sum_{u \in V} |A_u|}$.
- If $U_1, \ldots, U_n$ are IID U$[0,1]$, then set U$(n, [0,1])$ as the distribution

$$P(U_1 \in \cdot | U_1 > \min\{U_2, \ldots, U_n\}).$$

Theorem 4 (Silva-B.)

- *Let* $(Z_n^u : n \in \mathbb{Z}_+)$ *be independent random walks on* G *with* $Z_0^u = u$*.*
- *Sample* X *independently according to* μ*.*
- *Set*

$$\tau_v = \inf\{n \in \mathbb{Z}_+ : Z_n^X \in A_v\}, \text{ and } V_i = \{v \in V : \tau_v = i\}.$$

Given $V_0, V_1, \ldots,$ *assign fitnesses at each* $v \in V$*, which are independent and are*

1. $U[0,1]$ *for* $v \in A_X = V_0$*.*
2. $U(|A_{Z_i^X}|, [0,1])$ *if* $v \in V_i, \ i > 0$*.*

Then the joint distribution of X *and the fitnesses is stationary for the local Bak-Sneppen.*

Local Bak-Sneppen for Regular Graphs

Corollary 1

If G is d-regular, then the stationary distribution for the local Bak-Sneppen satisfies:

- *X is uniform.*
- *Given X, the fitnesses are independent and*
    1. *$U[0,1]$ for vertices in A_X.*
    2. *$U(d+1,[0,1])$ for all other vertices.*

Now send size to infinity

Corollary 2

Suppose that $(V_n : n \in \mathbb{N})$ is an increasing sequence of finite sets. For each n, let $G_n = (V_n, E_n)$ be a d-regular connected graph.
Then

- *The fitnesses under the stationary distribution for the local Bak-Sneppen on G_n converge weakly to an IID measure with marginal $U(d+1,[0,1])$.*

- IID structure, as expected for Bak-Sneppen.
- Unlike Bak-Sneppen: no threshold.

Local Bak-Sneppen for Regular Graphs

Corollary 1

If G is d-regular, then the stationary distribution for the local Bak-Sneppen satisfies:

- *X is uniform.*
- *Given X, the fitnesses are independent and*
  1. *$U[0,1]$ for vertices in A_X.*
  2. *$U(d+1,[0,1])$ for all other vertices.*

Now send size to infinity

Corollary 2

Suppose that $(V_n : n \in \mathbb{N})$ is an increasing sequence of finite sets. For each n, let $G_n = (V_n, E_n)$ be a d-regular connected graph.
Then

- *The fitnesses under the stationary distribution for the local Bak-Sneppen on G_n converge weakly to an IID measure with marginal $U(d+1,[0,1])$.*

- IID structure, as expected for Bak-Sneppen.
- Unlike Bak-Sneppen: no threshold.

Local Bak-Sneppen for Regular Graphs

Corollary 1

If G is d-regular, then the stationary distribution for the local Bak-Sneppen satisfies:

- *X is uniform.*
- *Given X, the fitnesses are independent and*
  1. *$U[0,1]$ for vertices in A_X.*
  2. *$U(d+1,[0,1])$ for all other vertices.*

Now send size to infinity

Corollary 2

Suppose that $(V_n : n \in \mathbb{N})$ is an increasing sequence of finite sets. For each n, let $G_n = (V_n, E_n)$ be a d-regular connected graph.
Then

- *The fitnesses under the stationary distribution for the local Bak-Sneppen on G_n converge weakly to an IID measure with marginal $U(d+1,[0,1])$.*

- IID structure, as expected for Bak-Sneppen.
- Unlike Bak-Sneppen: no threshold.

Local Bak-Sneppen for Regular Graphs

Corollary 1

If G is d-regular, then the stationary distribution for the local Bak-Sneppen satisfies:

- *X is uniform.*
- *Given X, the fitnesses are independent and*
  1. *$U[0,1]$ for vertices in A_X.*
  2. *$U(d+1,[0,1])$ for all other vertices.*

Now send size to infinity

Corollary 2

Suppose that $(V_n : n \in \mathbb{N})$ is an increasing sequence of finite sets. For each n, let $G_n = (V_n, E_n)$ be a d-regular connected graph.
Then

- *The fitnesses under the stationary distribution for the local Bak-Sneppen on G_n converge weakly to an IID measure with marginal $U(d+1,[0,1])$.*

- IID structure, as expected for Bak-Sneppen.
- Unlike Bak-Sneppen: no threshold.

Local Bak-Sneppen for Regular Graphs

Corollary 1

If G is d-regular, then the stationary distribution for the local Bak-Sneppen satisfies:

- *X is uniform.*
- *Given X, the fitnesses are independent and*
  1. *$U[0,1]$ for vertices in A_X.*
  2. *$U(d+1,[0,1])$ for all other vertices.*

Now send size to infinity

Corollary 2

Suppose that $(V_n : n \in \mathbb{N})$ is an increasing sequence of finite sets. For each n, let $G_n = (V_n, E_n)$ be a d-regular connected graph.
Then

- *The fitnesses under the stationary distribution for the local Bak-Sneppen on G_n converge weakly to an IID measure with marginal $U(d+1,[0,1])$.*

- IID structure, as expected for Bak-Sneppen.
- Unlike Bak-Sneppen: no threshold.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID $\mathrm{U}[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID U$[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID $U[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID $U[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID $U[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID $U[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \le f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID U$[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID U$[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID U$[0, 1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID $U[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

Guiol-Machado-Schinazi

Why ?

- Have population growth be part of model, not external parameter.
- Tractability.

Construction

- The population size is a reflected random walk on $\mathbb{Z}_+$ (that is random walk minus its running minimum).
- When population increases, AKA birth (possibly multiple), new individuals are assigned IID U$[0,1]$ fitnesses.
- When population decreases, the individual with lowest fitness is eliminated.

What is the asymptotic fitness distribution ?

More precisely, letting $\hat{F}_n(f)$ denote the empirical fitness distribution

$$\hat{F}_n(f) = \begin{cases} \text{prop. with fitness } \leq f & \text{if pop. size is } > 0 \\ \text{CDF of } \delta_0 & \text{otherwise.} \end{cases}$$

Understand limit (LLN) and fluctuations (CLT) of $\hat{F}_n$ as $n \to \infty$.

GMS Law of Large Numbers

Notation.

- $I \overset{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0, 1)$. Then

$$\hat{F}_n \to F_\infty := CDF\ of\ U[f_c, 1],\ uniformly,\ a.s.$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS Law of Large Numbers

Notation.

- $I \overset{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0,1)$. Then

$$\hat{F}_n \to F_\infty := \text{CDF of } U[f_c, 1], \text{ uniformly, a.s.}$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS Law of Large Numbers

Notation.

- $I \stackrel{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0, 1)$. Then

$$\hat{F}_n \to F_\infty := \text{CDF of } U[f_c, 1], \text{ uniformly, a.s.}$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS Law of Large Numbers

Notation.

- $I \stackrel{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0, 1)$. Then

$$\hat{F}_n \to F_\infty := \text{CDF of } U[f_c, 1], \text{ uniformly, a.s.}$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS Law of Large Numbers

Notation.

- $I \overset{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0, 1)$. Then

$$\hat{F}_n \to F_\infty := \textit{CDF of } U[f_c, 1], \textit{ uniformly, a.s.}$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS Law of Large Numbers

Notation.

- $I \overset{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0, 1)$. Then

$$\hat{F}_n \to F_\infty := CDF\ of\ U[f_c, 1],\ \ uniformly,\ a.s.$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS Law of Large Numbers

Notation.

- $I \stackrel{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0, 1)$. Then

$$\hat{F}_n \to F_\infty := CDF\ of\ U[f_c, 1],\ uniformly,\ a.s.$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS Law of Large Numbers

Notation.

- $I \stackrel{\text{dist}}{=}$ Incement of random walk.
- $I = I_+ - I_-$ where,
  $I_+ = \max(I, 0)$ is the positive increment; and
  $I_- = max(-I, 0)$ the nagative increment.

Assumptions.

- $E|I| < \infty$.
- Transience: $EI_+ > EI_-$.

Theorem 5 (GMS, Volkov-Skevi, B.)

Let $f_c = EI_-/EI_+ \in [0, 1)$. Then

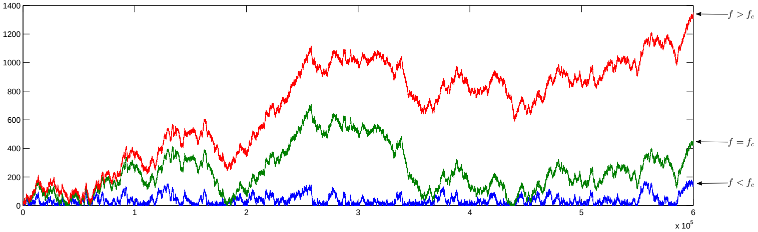
$$\hat{F}_n \to F_\infty := \text{CDF of } U[f_c, 1], \text{ uniformly, a.s.}$$

If I is deterministic (that is population grows deterministically), this is Glivenko-Cantelli.

GMS LLN

Idea of Proof

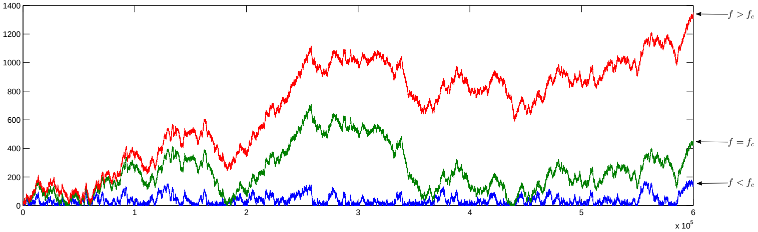
- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_+ - EI_-$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS LLN

Idea of Proof

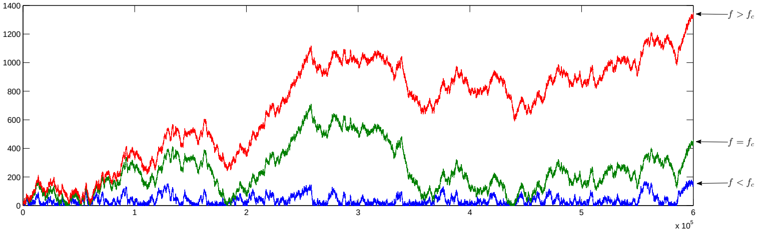
- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_{+} - EI_{-}$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS LLN

Idea of Proof

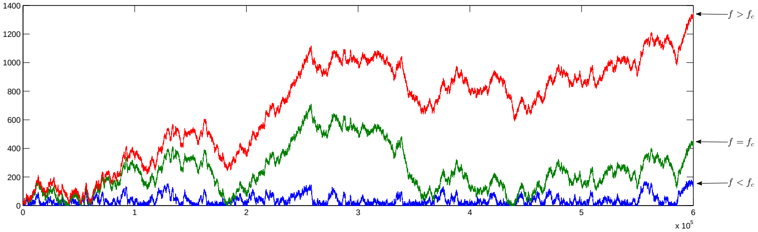
- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_+ - EI_-$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS LLN

Idea of Proof

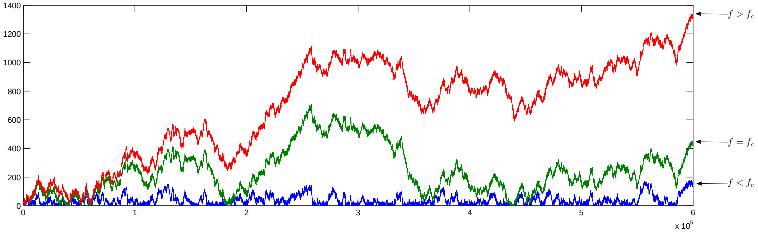
- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_{+} - EI_{-}$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS LLN

Idea of Proof

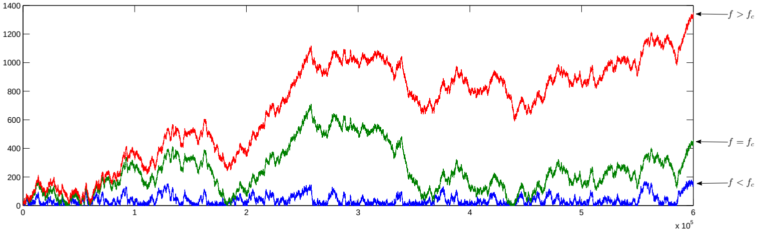
- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_{+} - EI_{-}$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS LLN

Idea of Proof

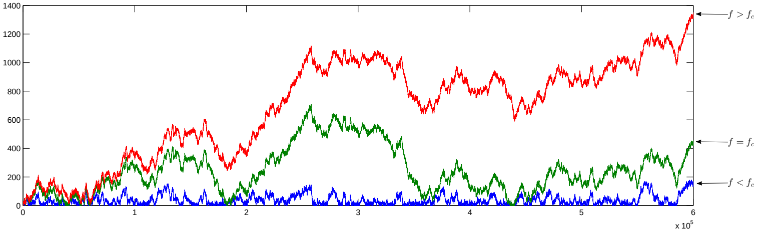
- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_+ - EI_-$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS LLN

Idea of Proof

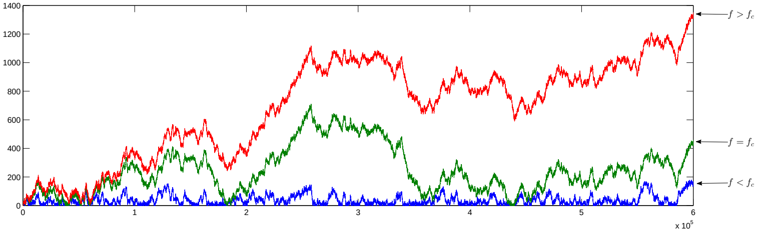
- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_+ - EI_-$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS LLN

Idea of Proof

- Size of population with fitness $\leq f$ is reflected random walk with drift $fEI_{+} - EI_{-}$.
  - Transient if $f > f_c$
  - Null recurrent if $f = f_c$
  - Positive recurrent if $f < f_c$
- If $f > f_c$, there exists finite time after which there will always be a species with lower fitness.
- Therefore, the proportion of species with fitness $> f_c$ which will be eliminated tends to 0.



GMS Central Limit Theorem

Let

$$\hat{\Delta}_n = \hat{F}_n - F_\infty .$$

Then we know that $\hat{\Delta}_n$ converges to 0, uniformly, a.s.

Next, we look at fluctuations.

Assumption

$$E(I^2) < \infty.$$

Processes appearing in limit

- W_1 standard BM, and the corresponding bridge Br_1:

$$\mathrm{Br}_1(f) := W_1(f) - fW_1(1).$$

- If $f_c = 0$, choose $\widetilde{W_1} \equiv 0$.
- If $f_c > 0$: $\widetilde{W_1}$ standard BM derived from W_1 as follows :
  - $U \sim U[f_c, 1]$, independent of W_1.
  - An "interval" $\widetilde{A}_t$ of length $(1 - f_c)t$, shifted by U.
  - $\widetilde{W_1}(t) := \dfrac{1}{\sqrt{f_c(1 - f_c)}} \left((1 - f_c)W_1(f_c t) + f_c \int \mathbf{1}_{\widetilde{A}_t}(s) dW_1(s) \right).$

- W_2, standard BM, independent of W_1, U.

GMS Central Limit Theorem

Let

$$\hat{\Delta}_n = \hat{F}_n - F_\infty .$$

Then we know that $\hat{\Delta}_n$ converges to 0, uniformly, a.s.

Next, we look at fluctuations.

Assumption

$$E(I^2) < \infty.$$

Processes appearing in limit

- W_1 standard BM, and the corresponding bridge Br_1:

$$\mathrm{Br}_1(f) := W_1(f) - fW_1(1).$$

- If $f_c = 0$, choose $\widetilde{W_1} \equiv 0$.
- If $f_c > 0$: $\widetilde{W_1}$ standard BM derived from W_1 as follows :
  - $U \sim U[f_c, 1]$, independent of W_1.
  - An "interval" $\widetilde{A}_t$ of length $(1 - f_c)t$, shifted by U.
  - $\widetilde{W_1}(t) := \dfrac{1}{\sqrt{f_c(1 - f_c)}} \left((1 - f_c)W_1(f_c t) + f_c \int \mathbf{1}_{\widetilde{A}_t}(s) dW_1(s) \right).$

- W_2, standard BM, independent of W_1, U.

GMS Central Limit Theorem

Let

$$\hat{\Delta}_n = \hat{F}_n - F_\infty .$$

Then we know that $\hat{\Delta}_n$ converges to 0, uniformly, a.s.

Next, we look at fluctuations.

Assumption

$$E(I^2) < \infty.$$

Processes appearing in limit

- W_1 standard BM, and the corresponding bridge Br_1:

$$\mathrm{Br}_1(f) := W_1(f) - fW_1(1).$$

- If $f_c = 0$, choose $\widetilde{W_1} \equiv 0$.
- If $f_c > 0$: $\widetilde{W_1}$ standard BM derived from W_1 as follows :
  - $U \sim U[f_c, 1]$, independent of W_1.
  - An "interval" $\widetilde{A}_t$ of length $(1 - f_c)t$, shifted by U.
  - $\widetilde{W_1}(t) := \frac{1}{\sqrt{f_c(1 - f_c)}} \left((1 - f_c)W_1(f_c t) + f_c \int \mathbf{1}_{\widetilde{A}_t}(s) dW_1(s) \right) .$

- W_2, standard BM, independent of W_1, U.

GMS Central Limit Theorem

Let

$$\hat{\Delta}_n = \hat{F}_n - F_\infty .$$

Then we know that $\hat{\Delta}_n$ converges to 0, uniformly, a.s.

Next, we look at fluctuations.

Assumption

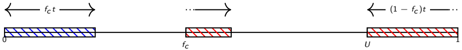
$$E(I^2) < \infty.$$

Processes appearing in limit

- W_1 standard BM, and the corresponding bridge Br_1:

$$\mathrm{Br}_1(f) := W_1(f) - fW_1(1).$$

- If $f_c = 0$, choose $\widetilde{W}_1 \equiv 0$.
- If $f_c > 0$: $\widetilde{W}_1$ standard BM derived from W_1 as follows :
  - $U \sim U[f_c, 1]$, independent of W_1.
  - An "interval" $\widetilde{A}_t$ of length $(1 - f_c)t$, shifted by U.
  - $\widetilde{W}_1(t) := \frac{1}{\sqrt{f_c(1 - f_c)}} \left((1 - f_c)W_1(f_c t) + f_c \int \mathbf{1}_{\widetilde{A}_t}(s) dW_1(s) \right).$



- W_2, standard BM, independent of W_1, U.

GMS CLT

For a path $\omega \in D[0,1]$, let $\Psi(\omega) := \omega(1) - \inf_{0 \le t \le 1} \omega(t)$

Theorem 6 (B.)

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \overbrace{\sigma_1 Br_1 + \sigma_2 W_2(1)(1-F_\infty)}^{\textit{Gaussian process}} \\ \underbrace{\Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2)}_{\textit{Positive RV}} \end{pmatrix},$$

with $\sigma_1 = \sqrt{EI_+}$, $\widetilde{\sigma}_1 = \sqrt{f_c(1-f_c)EI_+}$, $\sigma_2 = \sqrt{f_c^2 E(I_+^2) + E(I_-^2)}$, *and the convergence is* $D(f_c,1] \times \mathbb{R}$.

Marginals

$$\sqrt{n}\widehat{\Delta}_n(f) \Rightarrow \begin{cases} \sigma(f \wedge 1)N(0,1) & f > f_c; \\ \sigma(f_c)|N(0,1)| & f = f_c; \\ 0 & f < f_c, \end{cases}$$

where $\sigma(f) := \frac{1}{E(I_+)}\sqrt{f(1-f)E(I_+) + \left(\frac{1-f}{1-f_c}\right)^2 \left(f_c^2 E(I_+^2) + E(I_-^2)\right)}$

GMS CLT

For a path $\omega \in D[0,1]$, let $\Psi(\omega) := \omega(1) - \inf_{0 \le t \le 1} \omega(t)$

Theorem 6 (B.)

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \overbrace{\sigma_1 Br_1 + \sigma_2 W_2(1)(1-F_\infty)}^{\textit{Gaussian process}} \\ \underbrace{\Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2)}_{\textit{Positive RV}} \end{pmatrix},$$

with $\sigma_1 = \sqrt{EI_+}$, $\widetilde{\sigma}_1 = \sqrt{f_c(1-f_c)EI_+}$, $\sigma_2 = \sqrt{f_c^2 E(I_+^2) + E(I_-^2)}$, *and the convergence is* $D(f_c, 1] \times \mathbb{R}$.

Marginals

$$\sqrt{n}\widehat{\Delta}_n(f) \Rightarrow \begin{cases} \sigma(f \wedge 1)N(0,1) & f > f_c; \\ \sigma(f_c)|N(0,1)| & f = f_c; \\ 0 & f < f_c, \end{cases}$$

where $\sigma(f) := \dfrac{1}{E(I_+)}\sqrt{f(1-f)E(I_+) + \left(\dfrac{1-f}{1-f_c}\right)^2 \left(f_c^2 E(I_+^2) + E(I_-^2)\right)}$

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma_1 W_1 + \sigma_2 W_2}) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma_1} \widetilde{W_1} + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{EI_+}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + \sigma_2 W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + \sigma_2 W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while its limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

GMS with selection

Assume

$$P(I = 1) = p = 1 - P(I = -1).$$

New feature

- At birth the individual obtains
  - w/prob r new U[0, 1] fitness.
  - w/prob $1 - r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1 - p$.

Conclusion

1. Number of sites coincides with GMS with $P(I = 1) = pr, \; P(I = -1) = 1 - p$ and $P(I = 0) = 1 - pr - (1 - p)$.
2. The system is transient if and only if $pr > (1 - p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is $U[f_c, 1]$.

GMS with selection

Assume

$$P(I = 1) = p = 1 - P(I = -1).$$

New feature

- At birth the individual obtains
  - w/prob r new U[0, 1] fitness.
  - w/prob $1 - r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1 - p$.

Conclusion

1. Number of sites coincides with GMS with $P(I = 1) = pr,\ P(I = -1) = 1 - p$ and $P(I = 0) = 1 - pr - (1 - p)$.
2. The system is transient if and only if $pr > (1 - p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is $U[f_c, 1]$.

GMS with selection

Assume
$$P(I = 1) = p = 1 - P(I = -1).$$

New feature

- At birth the individual obtains
  - w/prob r new U$[0, 1]$ fitness.
  - w/prob $1 - r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1 - p$.

Conclusion

1. Number of sites coincides with GMS with $P(I = 1) = pr,\ P(I = -1) = 1 - p$ and $P(I = 0) = 1 - pr - (1 - p)$.
2. The system is transient if and only if $pr > (1 - p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is U$[f_c, 1]$.

GMS with selection

Assume

$$P(I=1) = p = 1 - P(I=-1).$$

New feature

- At birth the individual obtains
  - w/prob r new U$[0,1]$ fitness.
  - w/prob $1-r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1-p$.

Conclusion

1. Number of sites coincides with GMS with $P(I=1) = pr, \; P(I=-1) = 1-p$ and $P(I=0) = 1 - pr - (1-p)$.
2. The system is transient if and only if $pr > (1-p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is U$[f_c, 1]$.

GMS with selection

Assume

$$P(I = 1) = p = 1 - P(I = -1).$$

New feature

- At birth the individual obtains
  - w/prob r new U$[0, 1]$ fitness.
  - w/prob $1 - r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1 - p$.

Conclusion

1. Number of sites coincides with GMS with $P(I = 1) = pr, \; P(I = -1) = 1 - p$ and $P(I = 0) = 1 - pr - (1 - p)$.
2. The system is transient if and only if $pr > (1 - p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is U$[f_c, 1]$.

GMS with selection

Assume

$$P(I = 1) = p = 1 - P(I = -1).$$

New feature

- At birth the individual obtains
  - w/prob r new U$[0, 1]$ fitness.
  - w/prob $1 - r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1 - p$.

Conclusion

1. Number of sites coincides with GMS with $P(I = 1) = pr,\ P(I = -1) = 1 - p$ and $P(I = 0) = 1 - pr - (1 - p)$.
2. The system is transient if and only if $pr > (1 - p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is U$[f_c, 1]$.

GMS with selection

Assume

$$P(I = 1) = p = 1 - P(I = -1).$$

New feature

- At birth the individual obtains
  - w/prob r new U$[0, 1]$ fitness.
  - w/prob $1 - r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1 - p$.

Conclusion

1. Number of sites coincides with GMS with $P(I = 1) = pr, \; P(I = -1) = 1 - p$ and $P(I = 0) = 1 - pr - (1 - p)$.
2. The system is transient if and only if $pr > (1 - p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is U$[f_c, 1]$.

GMS with selection

Assume

$$P(I = 1) = p = 1 - P(I = -1).$$

New feature

- At birth the individual obtains
  - w/prob r new U$[0, 1]$ fitness.
  - w/prob $1 - r$, an existing fitness, uniformly among existing fitnesses, or new one if population is zero.
- At death, eliminate all species with lowest fitness.

We refer to the population with fixed fitness as a **site**.

Observation

- Probability of new site is pr.
- Probability of eliminating a site is $1 - p$.

Conclusion

1. Number of sites coincides with GMS with $P(I = 1) = pr,\; P(I = -1) = 1 - p$ and $P(I = 0) = 1 - pr - (1 - p)$.
2. The system is transient if and only if $pr > (1 - p)$.
3. In this case $f_c = \frac{1-p}{pr}$, and the asymptotic site fitness distribution is U$[f_c, 1]$.

GMS w/Selection

What is site size distribution ?

Let $\hat{H}_n$ denote the empirical distribution of sites and their respective fitness:

$$\hat{H}_n(A \times B) = \frac{\#\text{ sites whose size is in } A \text{ and whose fitness is in } B}{\#\text{ sites}}.$$

Theorem 7 (Schinazi-B. '15)

$$\hat{H}_n \to Geom\left(\frac{pr - (1-p)}{p - (1-p)}\right) \otimes U[f_c, 1],\ \ a.s.$$

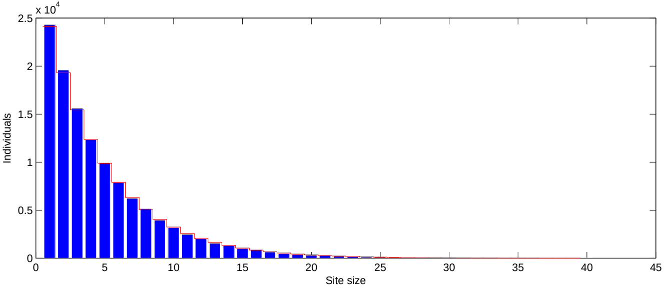


Figure: Empirical dist of site sizes ($p = 0.8$, $r = 0.4$, $n = 10^6$) and corresponding Geom.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)\big(H_\infty(k-1) - H_\infty(k)\big) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

  Then change in the number of sites of size k occurs only at
- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem

Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem

Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem

Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

  Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

GMS w/selection

Why Geometric ?

Fix site size $k > 1$. Consider number of sites of size k with fitness $> f_c$.

Assume the proportion of such sites converges to $H_\infty(k)$.

- Number of such sites grows at speed

$$p(1-r)(H_\infty(k-1) - H_\infty(k)) + o(1) = H_\infty(k) * (pr - (1-p))$$

Then change in the number of sites of size k occurs only at

- At birth
  - Increases by 1 when new individual selects a site of size $k-1$.
  - Decreases by 1 when new individual selects a site of size k.
- At death, but occurs only finitely often.
- Equality because # sites grows at speed $pr - (1-p)$.

This equation guarantees geometric decay.

The problem
Proving that the assumption actually holds.

- Easy calculus exercise if $pr > \frac{1}{2}$.
- Otherwise: use "mean reversion" away from linear curve.

Thank you.