◆□▶ ◆□▶ ▲□▶ ▲□▶ □ のQ@

Bak-Sneppen Backwards

Mackenzie Simper University of Utah

Joint with Tom Alberts, Ga Yeong Lee (CalTech)

May 10, 2016 Frontier Probability Days University of Utah

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

The Model

The Bak-Sneppen model is a simplified model of evolution incorporating natural selection and spatial interaction between species, first introduced in 1993.

Let N = number of species. The model is a discrete-time Markov chain on the space $[0, 1]^N$.

(ロ) (同) (三) (三) (三) (○) (○)

The Model

The Bak-Sneppen model is a simplified model of evolution incorporating natural selection and spatial interaction between species, first introduced in 1993.

Let N = number of species. The model is a discrete-time Markov chain on the space $[0, 1]^N$.

1. The species are arranged in a circle and each has a fitness value in [0, 1]. Typically, the initial fitnesses are chosen uniformly at random.

(ロ) (同) (三) (三) (三) (○) (○)

The Model

The Bak-Sneppen model is a simplified model of evolution incorporating natural selection and spatial interaction between species, first introduced in 1993.

Let N = number of species. The model is a discrete-time Markov chain on the space $[0, 1]^N$.

- 1. The species are arranged in a circle and each has a fitness value in [0, 1]. Typically, the initial fitnesses are chosen uniformly at random.
- 2. *Natural Selection* At each time step, the species with minimum fitness is replaced with a new uniform random fitness.

The Model

The Bak-Sneppen model is a simplified model of evolution incorporating natural selection and spatial interaction between species, first introduced in 1993.

Let N = number of species. The model is a discrete-time Markov chain on the space $[0, 1]^N$.

- 1. The species are arranged in a circle and each has a fitness value in [0, 1]. Typically, the initial fitnesses are chosen uniformly at random.
- 2. *Natural Selection* At each time step, the species with minimum fitness is replaced with a new uniform random fitness.
- 3. *Spacial Interaction* The two neighbors of the minimum fitness species are also replaced with new uniform random fitnesses.

▲□▶▲□▶▲□▶▲□▶ □ のQ@

Example – N = 7 Initial Configuration

 $\mathbf{F}_0 = (0.9058, 0.1270, 0.6234, 0.9134, 0.0975, 0.2785, 0.8147)$



Example

$\mathbf{F}_0 = (0.9058, 0.1270, 0.6234, 0.9134, 0.0975, 0.2785, 0.8147)$



Example

$\mathbf{F}_1 = (0.9058, 0.1270, 0.6234, 0.1419, 0.8003, 0.4854, 0.8147)$



Example

 $\mathbf{F}_1 = (0.9058, \frac{0.1270}{0.6234}, 0.1419, 0.8003, 0.4854, 0.8147)$



◆□ ▶ ◆□ ▶ ◆ □ ▶ ◆ □ ▶ ● □ ● ● ● ●

(ロ) (同) (三) (三) (三) (○) (○)

Bacteria Experiments

- Lenski and collaborators cultivated populations of E. coli for several years, 10,000+ generations. The E. coli long-term experimental evolution project site: mxyo.css.msu.edu/ecoli.
- Other researchers were able to show that that Bak-Sneppen model qualitatively reproduces some of the results on the relative fitness of the bacteria.

I. Bose and I. Chaudhuri. Bacterial evolution and Bak-Sneppen model. Internat. J. Modern Phys. C, 12(5):675-685, 2001.

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

Main Goal

We would like to find the *stationary distribution* for the fitnesses of the species, π_N .

◆□▶ ◆□▶ ▲□▶ ▲□▶ □ のQ@

Main Goal

We would like to find the *stationary distribution* for the fitnesses of the species, π_N .

Conjecture: As the number of species approaches infinity, the individual species' fitnesses become independent and uniformly distributed on [f_c , 1], where $f_c \approx 2/3$.

Introduction

Our Work

N = 1000, *T* = 0



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ ─臣 ─のへで

Introduction

Our Work

N = 1000, T = 10,000



◆□▶ ◆□▶ ◆豆▶ ◆豆▶ □豆 の々で

▲□▶ ▲□▶ ▲□▶ ▲□▶ = 三 のへで



The conjecture is still open. Our work has focused on finding the stationary distributions for *small* numbers of species.

A D F A 同 F A E F A E F A Q A

Past Work

The conjecture is still open. Our work has focused on finding the stationary distributions for *small* numbers of species.

- Recently, Eckhard Schlemm (Imperial College London) derived the stationary distribution for 4 species (2012) and differential equations for the stationary distribution for 5 species (2015). Note that 4 species is the smallest non-trivial number of species you could consider.
- He uses the **power method** or **von Mises** iteration.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

The Time-Reversed Process

- **F**₀, **F**₁, **F**₂, ... is the forwards chain. Start with **F**₀ according to the stationary distribution.
- Pick a large integer *T*. Reversed chain: \mathbf{F}_{T-k} , $0 \le k \le T$.
- Why is this useful? The forwards chain and backward chain have the same stationary distribution.

The Time-Reversed Process

The update rule for the time-reversed process must:

- Choose a species at random
- Replace the fitnesses of the chosen species and its two neighbors, subject to the constraint that the chosen species is given the minimum fitness among all others.

Both the probabilities for choosing each species and the new fitness values depend on the:

- current configuration of fitnesses
- the unknown stationary distribution, π_N

We can use the latter property to determine properties of π_N .

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

Set-up

• Markov operator *P* on functions $f : [0, 1]^N \to \mathbb{R}$:

$$Pf(x) = \mathbb{E}[f(\mathbf{F}_{k+1})|\mathbf{F}_k = x]$$

• Inner product:

$$\langle f, g \rangle_{\pi_N} = \mathbb{E}_{\pi_N}[f(X)g(X)]$$

The Markov operator for time-reversed process, is the adjoint Q:

$$\langle f, Pg \rangle_{\pi_N} = \langle Qf, g \rangle_{\pi_N}$$

A D F A 同 F A E F A E F A Q A

For Bak-Sneppen it's easy to write down P

- If *i* is the position of the minimum fitness, define *E_i* as the set of positions of species who are replaced.
- Thus $E_i = \{i 1, i, i + 1\}.$
- Define $\mathbf{L}_i \mathbf{y} + \mathbf{L}_i^c \mathbf{x} = (x_1, ..., x_{i-2}, y_{i-1}, y_i, y_{i+1}, x_{i+2}, ..., x_N).$

$$Pf(\mathbf{x}) = \sum_{i \in \mathbb{Z}_N} \mathbf{1} \{ \operatorname{argmin}(\mathbf{x}) = i \} \int_{[0,1]^N} f(\mathbf{x} + L_i(\mathbf{u} - \mathbf{x})) \, d\mathbf{u}.$$

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

Result

Forwards operator:

$$Pf(\mathbf{x}) = \sum_{i \in \mathbb{Z}_N} \mathbf{1} \{ \operatorname{argmin}(\mathbf{x}) = i \} \int_{[0,1]^N} f(\mathbf{x} + L_i(\mathbf{u} - \mathbf{x})) \, d\mathbf{u}.$$

Result

Forwards operator:

$$Pf(\mathbf{x}) = \sum_{i \in \mathbb{Z}_N} \mathbf{1} \{ \operatorname{argmin}(\mathbf{x}) = i \} \int_{[0,1]^N} f(\mathbf{x} + L_i(\mathbf{u} - \mathbf{x})) \, d\mathbf{u}.$$

Adjoint operator:

$$Qf(\mathbf{x}) = \frac{1}{\pi_{N}(\mathbf{x})} \sum_{i \in \mathbb{Z}_{N}} \int_{[0,1]^{N}} f(\mathbf{L}_{i}\mathbf{y} + \mathbf{L}_{i}^{c}\mathbf{x}) \mathbf{1} \{ \operatorname{argmin}(\mathbf{L}_{i}\mathbf{y} + \mathbf{L}_{i}^{c}\mathbf{x}) = i \}$$
$$\pi_{N}(\mathbf{L}_{i}\mathbf{y} + \mathbf{L}_{i}^{c}\mathbf{x}) d\mathbf{y}$$

Introduction

Our Work

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ



▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

4 Species

Functional equation for N = 4 species:

$$\pi_4(\mathbf{x}) = \sum_{i=1}^4 \int_{[0,1]^3} [\mathbf{1}\{\operatorname{argmin}(y_{i-1}, y_i, y_{i+1}, x_{i+2}) = i\} \\ \pi_4(y_{i-1}, y_i, y_{i+1}, x_{i+2})] \, dy_{i-1} \, dy_i \, dy_{i+1}$$

Keep in mind all index operations are mod 4.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

General Models

- We can look at more general models in which more than 3 species are replaced.
- If *i* is the position of the minimum fitness, define *E_i* as the set of positions of species who are replaced.
- Traditional model: $E_i = \{i 1, i, i + 1\}.$

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● のへぐ

Functional equation for π_N

Reminder: $\mathbf{L}_i \mathbf{y} + \mathbf{L}_i^c \mathbf{x} = (x_1, ..., x_{i-2}, y_{i-1}, y_i, y_{i+1}, x_{i+2}, ..., x_N).$

$$\pi_{\mathsf{N}}(\mathbf{x}) = \sum_{i \in \mathbb{Z}/N} \int_{[0,1]^N} \mathbf{1}\{\operatorname{argmin}(\mathbf{L}_i \mathbf{y} + \mathbf{L}_i^c \mathbf{x}) = i\} \pi_{\mathsf{N}}(\mathbf{L}_i \mathbf{y} + \mathbf{L}_i^c \mathbf{x}) d\mathbf{y}.$$

▲□▶ ▲□▶ ▲□▶ ▲□▶ = 三 のへで

Key Point

The stationary distribution is the sum of functions of the number of non-replaced species.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Key Point

The stationary distribution is the sum of functions of the number of non-replaced species.

Example: N = 6, $E_i = \{i - 1, i, i + 1, i + 2\}$ (replace 4 species): If

$$\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6),$$

then

$$\pi_6(\mathbf{x}) = g(x_1, x_2) + g(x_2, x_3) + g(x_3, x_4) + g(x_4, x_5) + g(x_5, x_6) + g(x_6, x_1)$$

for some function $g: [0, 1]^2 \to \mathbb{R}$.

(ロ) (同) (三) (三) (三) (○) (○)

Replace N - 1 species

If we replace N - 1 species, the stationary distribution is the sum of functions of one variable, which results in an easy differential equation to solve:

$$\pi_N(\mathbf{x}) = \frac{(N-1)(N-2)}{N} \sum_{i \in \mathbb{Z}/N} \frac{1 - (1-x_i)^{N-1}}{((1-x_i)^{N-1} + N - 2)^2}$$

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

Replace N - 1 species

If we replace N - 1 species, the stationary distribution is the sum of functions of one variable, which results in an easy differential equation to solve:

$$\pi_{N}(\mathbf{x}) = \frac{(N-1)(N-2)}{N} \sum_{i \in \mathbb{Z}/N} \frac{1 - (1 - x_{i})^{N-1}}{((1 - x_{i})^{N-1} + N - 2)^{2}}$$

$$N = 4:$$

$$\pi_{4}(\mathbf{x}) = \frac{3}{2} \sum_{i \in \mathbb{Z}/4} \frac{x_{i}(3 - x_{i}(3 - x_{i}))}{(3 - x_{i}(3 - x_{i}(3 - x_{i}))^{2}}$$

This matches Schlemm's result.

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

Replace N - 2 species

If we replace N - 2 species, the stationary distribution is the sum of functions of two variables:

$$\pi_N(\mathbf{x}) = \sum_{i \in \mathbb{Z}/N} q(x_i, x_{i+1}).$$

(ロ) (同) (三) (三) (三) (○) (○)

Replace N - 2 species

If we replace N - 2 species, the stationary distribution is the sum of functions of two variables:

$$\pi_N(\mathbf{x}) = \sum_{i \in \mathbb{Z}/N} q(x_i, x_{i+1}).$$

We guessed the ansatz:

$$q_N(u,v) = B_N(u)G_N(v) + A_N(v),$$

for v < u. This leads to a coupled system of three second-order ODEs that A_n, B_N, G_N must satisfying. These involve hypergeometric functions of the second kind.

▲□▶▲□▶▲□▶▲□▶ □ のQ@

Larger numbers of non-replaced species

- For larger numbers, it becomes very challenging.
- We don't expect that our method can be used to tackle the big conjecture.

Introduction

Our Work

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Reduced Markov Chain

The non-replaced species form a Markov chain by themselves, i.e. in the traditional model where 3 species are replaced, you only need to keep track of N - 3 values to keep track of the dynamics of all *N* species.



◆□▶ ◆□▶ ▲□▶ ▲□▶ □ のQ@

Future Questions

1. We've seen that you can reduce the full chain to only keep track of the non-replaced species. Are there any other reductions?



(ロ) (同) (三) (三) (三) (三) (○) (○)

Future Questions

- 1. We've seen that you can reduce the full chain to only keep track of the non-replaced species. Are there any other reductions?
- 2. When you replace N 1 species, as $N \to \infty$, the fitnesses remain independent and uniformly distributed on [0, 1]. How many species must be replaced to see something other than the uniform distribution on [0, 1]?

Introduction

Future Questions

- 1. We've seen that you can reduce the full chain to only keep track of the non-replaced species. Are there any other reductions?
- 2. When you replace N 1 species, as $N \to \infty$, the fitnesses remain independent and uniformly distributed on [0, 1]. How many species must be replaced to see something other than the uniform distribution on [0, 1]?
- 3. For small numbers of replaced species, it appears from simulations that as $N \to \infty$, the fitnesses become independent and uniformly distributed on $[f_c, 1]$, where f_c decreases as the number of replaced species increases.

Replace 1



◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ○ □ ○ のへぐ

Our Work

Replace 2



◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ○ □ ○ ○ ○ ○

Our Work

Replace 3



Our Work

Replace 4



Introduction

Our Work

Graph of *f*_c



・ロト・(四ト・(川下・(日下))

Introduction

Our Work

Thank you

Questions?

