Approximate Markov Model for a Finite Population Genetic Algorithms

Adam Prügel-Bennett
ISIS Research Group
Electronics and Computer Science
University of Southampton

Approximate Markov Model for a Finite Population Genetic Algorithms

Work in Progress

Adam Prügel-Bennett
ISIS Research Group
Electronics and Computer Science
University of Southampton

Outline

1. **Introduction**
   - Motivation
   - Cumulant Approach

2. Approximate Markov Model
   - New Approach
   - Results
   - Future Work

Aims

- Modelling GA on a problem with a small number of states
- Want a ‘plug & play’ model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state

Aims

- Modelling GA on a problem with a small number of states
- Want a 'plug & play' model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state

Aims

- Modelling GA on a problem with a small number of states
- Want a 'plug & play' model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state

Aims

- Modelling GA on a problem with a small number of states
- Want a ‘plug & play’ model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state

Aims

- Modelling GA on a problem with a small number of states
- Want a ‘plug & play’ model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state

Aims

- Modelling GA on a problem with a small number of states
- Want a 'plug & play' model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state

Aims

- Modelling GA on a problem with a small number of states
- Want a 'plug & play' model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state

Aims

- Modelling GA on a problem with a small number of states
- Want a 'plug & play' model
- The size of the search space, $\mathcal{S}$, might be $|\mathcal{S}| \approx 100$
- These states might correspond to many configurations
- E.g. **ones-max** where each state corresponds to a set of points in the search space grouped according to their Hamming distance from the all zeros string
- More interesting variants of ones-max, e.g. hurdle problem
- To model non-trivial problems we can amalgamate connected configurations with the same cost into a single state—**coarse grained model of a problem**

Other Heuristics Search

- For some other heuristics search such as descent with a variable mutation rate or simulated annealing we can build an exact Markov Chain model
- This allows us to study the algorithm
- Can compute the expected performance
- Can compute the gradient in the expected performance as a function of the parameters
- Find optimal parameters

Other Heuristics Search

- For some other heuristics search such as descent with a variable mutation rate or simulated annealing we can build an exact Markov Chain model
- This allows us to study the algorithm
- Can compute the expected performance
- Can compute the gradient in the expected performance as a function of the parameters
- Find optimal parameters

Other Heuristics Search

- For some other heuristics search such as descent with a variable mutation rate or simulated annealing we can build an exact Markov Chain model
- This allows us to study the algorithm
- Can compute the expected performance
- Can compute the gradient in the expected performance as a function of the parameters
- Find optimal parameters

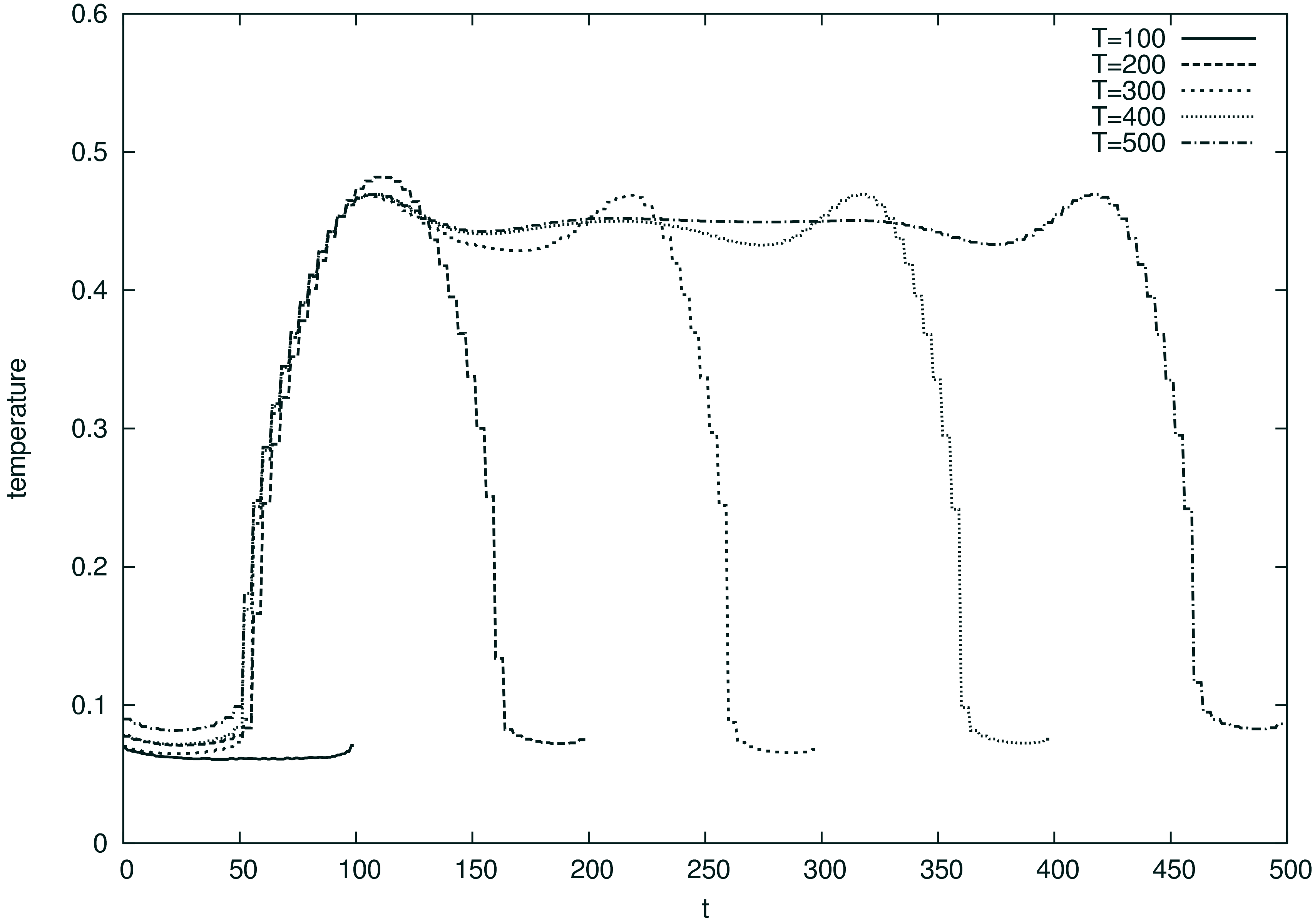
Other Heuristics Search

- For some other heuristics search such as descent with a variable mutation rate or simulated annealing we can build an exact Markov Chain model
- This allows us to study the algorithm
- Can compute the expected performance
- Can compute the gradient in the expected performance as a function of the parameters
- Find optimal parameters

Other Heuristics Search

- For some other heuristics search such as descent with a variable mutation rate or simulated annealing we can build an exact Markov Chain model
- This allows us to study the algorithm
- Can compute the expected performance
- Can compute the gradient in the expected performance as a function of the parameters
- Find optimal parameters

Max-Sat: Annealing Schedules with Best-So-Far



Exact Markov Model for a GA

- Describe the population by a vector

$$\boldsymbol{N} = (N_1, N_2, \cdots N_m) \qquad\qquad m = |\mathcal{S}|$$

  N_i is the number of individuals in state i.

- The probability of a population at time, t, is given by $\mathbb{P}\big(\boldsymbol{N}, t\big)$

- The evolution of $\mathbb{P}\big(\boldsymbol{N}, t\big)$ can be described in terms of a Markov Chain

Exact Markov Model for a GA

- Describe the population by a vector

$$\boldsymbol{N} = (N_1, N_2, \cdots N_m) \qquad\qquad m = |\mathcal{S}|$$

  N_i is the number of individuals in state i.

- The probability of a population at time, t, is given by $\mathbb{P}(\boldsymbol{N}, t)$

- The evolution of $\mathbb{P}(\boldsymbol{N}, t)$ can be described in terms of a Markov Chain

Exact Markov Model for a GA

- Describe the population by a vector

$$\boldsymbol{N} = (N_1, N_2, \cdots N_m) \qquad\qquad m = |\mathcal{S}|$$

  N_i is the number of individuals in state i.

- The probability of a population at time, t, is given by $\mathbb{P}(\boldsymbol{N}, t)$

- The evolution of $\mathbb{P}(\boldsymbol{N}, t)$ can be described in terms of a Markov Chain

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate
- Crossover usually prevents coarse graining problems—care about correlation between members of the population
- There are $\binom{m+p-1}{p-1}$ different possible populations, $\boldsymbol{N}$

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate

- Crossover usually prevents coarse graining problems—care about correlation between members of the population

- There are $\binom{m+p-1}{p-1}$ different possible populations, $\boldsymbol{N}$

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate
- Crossover usually prevents coarse graining problems—care about correlation between members of the population
- There are $\binom{m+p-1}{p-1}$ different possible populations, $\boldsymbol{N}$

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate

- Crossover usually prevents coarse graining problems—care about correlation between members of the population

- There are $\binom{m+p-1}{p-1}$ different possible populations, $\boldsymbol{N}$

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate
- Crossover usually prevents coarse graining problems—care about correlation between members of the population

I will ignore crossover in this talk!

- There are $\binom{m+p-1}{p-1}$ different possible populations, $\boldsymbol{N}$

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate

- Crossover usually prevents coarse graining problems—care about correlation between members of the population

  I will ignore crossover in this talk!

- There are $\binom{m+p-1}{p-1}$ different possible populations, N

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate
- Crossover usually prevents coarse graining problems—care about correlation between members of the population

I will ignore crossover in this talk!

- There are $\binom{m+p-1}{p-1}$ different possible populations, $\boldsymbol{N}$
  - ⋆ For $m = 10$, $p = 10$ the number of possible populations is $92\,378$

Problems with GAs

- Very useful to do this for a GA
  - ⋆ Provide a way to understand the interactions of many parameters
  - ⋆ Find optimal annealing schedule for mutation rate

- Crossover usually prevents coarse graining problems—care about correlation between members of the population

  I will ignore crossover in this talk!

- There are $\binom{m+p-1}{p-1}$ different possible populations, $\boldsymbol{N}$
  - ⋆ For $m = 10$, $p = 10$ the number of possible populations is $92\,378$
  - ⋆ For $m = 100$, $p = 100$ the number of possible populations is 4.5×10^{58}

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population

- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$

- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population
- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$
- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population
- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$
- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population
- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$
- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest
  - ⋆ What population size should I use?

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population
- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$
- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest
  - ⋆ What population size should I use?
  - ⋆ What selection rate should I use?

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population
- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$
- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest
  - ⋆ What population size should I use?
  - ⋆ What selection rate should I use?
  - ⋆ What mutation rate should I use?

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population
- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$
- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest
  - ⋆ What population size should I use?
  - ⋆ What selection rate should I use?
  - ⋆ What mutation rate should I use?
  - ⋆ How long does it take to find the global optima?

Infinite Population Approximation

- One way to avoid the problem caused by the large number of different possible populations is to consider an infinite population
- Because there are no fluctuations in the dynamics of an infinite population it suffices to consider the mean occupancy of each possible state $\mu_i = N_i/p$
- The infinite population approximation is very easy to work with but doesn't answer the questions of practical interest
  - ⋆ What population size should I use?
  - ⋆ What selection rate should I use?
  - ⋆ What mutation rate should I use?
  - ⋆ How long does it take to find the global optima?

  Most practical issues arise because we are using a finite population

Statistical Mechanics Approach

- Approach developed in collaboration with Jonathan Shapiro, Magnus Rattray and Alex Rogers
- For simple problems the evolutionary dynamics can be modelled using statistical properties of distribution of fitness
- Principally these are the cumulants of the distribution of fitness
- Can also include fluctuation to describe an ensemble
- Possible to include more "macroscopic" variables to capture features of more complex problems
- Can also include crossover

Statistical Mechanics Approach

- Approach developed in collaboration with Jonathan Shapiro, Magnus Rattray and Alex Rogers
- For simple problems the evolutionary dynamics can be modelled using statistical properties of distribution of fitness
- Principally these are the cumulants of the distribution of fitness
- Can also include fluctuation to describe an ensemble
- Possible to include more “macroscopic” variables to capture features of more complex problems
- Can also include crossover

Statistical Mechanics Approach

- Approach developed in collaboration with Jonathan Shapiro, Magnus Rattray and Alex Rogers
- For simple problems the evolutionary dynamics can be modelled using statistical properties of distribution of fitness
- Principally these are the cumulants of the distribution of fitness
- Can also include fluctuation to describe an ensemble
- Possible to include more “macroscopic” variables to capture features of more complex problems
- Can also include crossover

Statistical Mechanics Approach

- Approach developed in collaboration with Jonathan Shapiro, Magnus Rattray and Alex Rogers
- For simple problems the evolutionary dynamics can be modelled using statistical properties of distribution of fitness
- Principally these are the cumulants of the distribution of fitness
- Can also include fluctuation to describe an ensemble
- Possible to include more "macroscopic" variables to capture features of more complex problems
- Can also include crossover

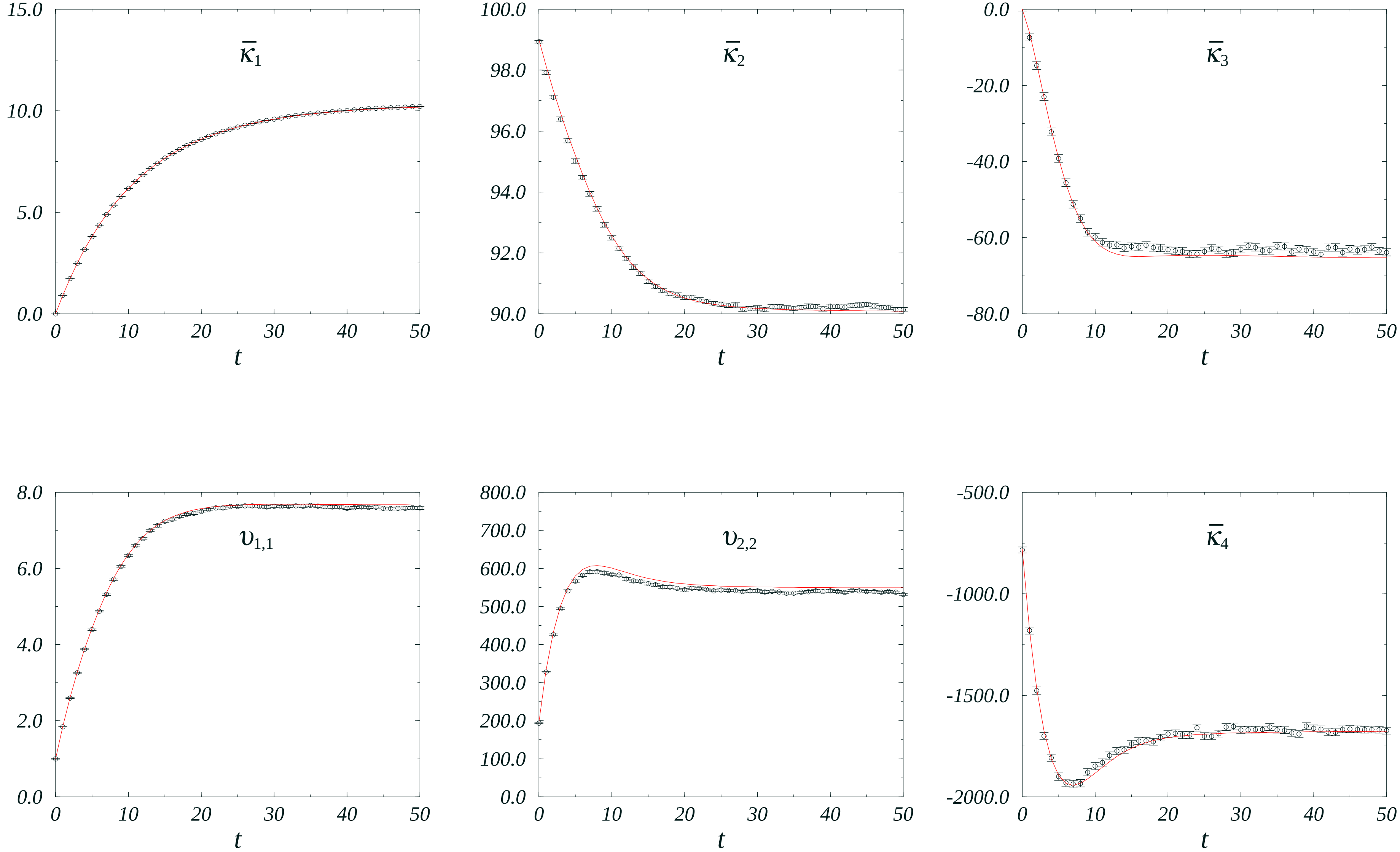
Statistical Mechanics Approach

- Approach developed in collaboration with Jonathan Shapiro, Magnus Rattray and Alex Rogers
- For simple problems the evolutionary dynamics can be modelled using statistical properties of distribution of fitness
- Principally these are the cumulants of the distribution of fitness
- Can also include fluctuation to describe an ensemble
- Possible to include more “macroscopic” variables to capture features of more complex problems
- Can also include crossover

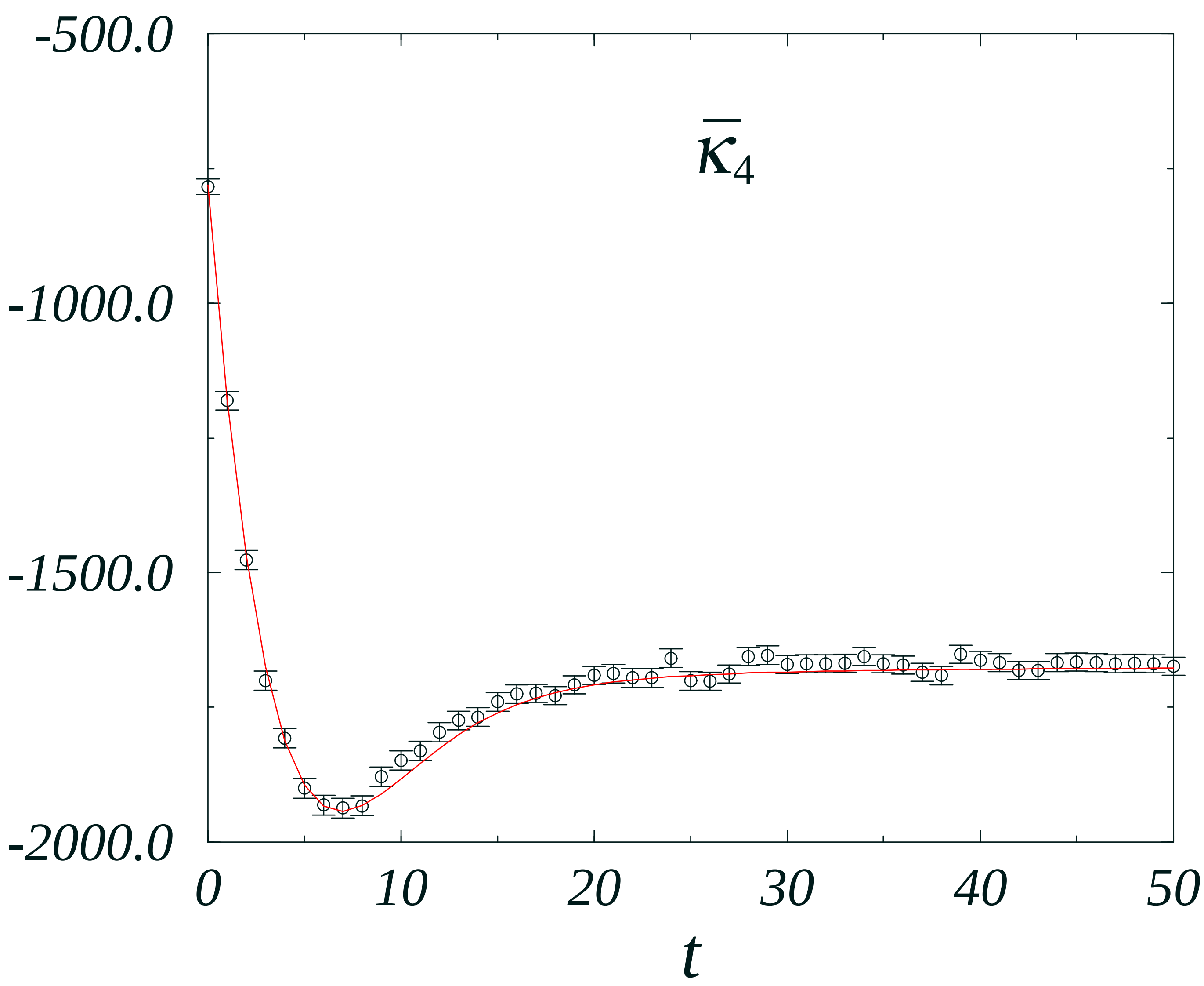
Statistical Mechanics Approach

- Approach developed in collaboration with Jonathan Shapiro, Magnus Rattray and Alex Rogers
- For simple problems the evolutionary dynamics can be modelled using statistical properties of distribution of fitness
- Principally these are the cumulants of the distribution of fitness
- Can also include fluctuation to describe an ensemble
- Possible to include more “macroscopic” variables to capture features of more complex problems
- Can also include crossover

Example Dynamics: Ones-max



Asexual Evolution



Drawback

- Although we can model more complex problems approach is not very general

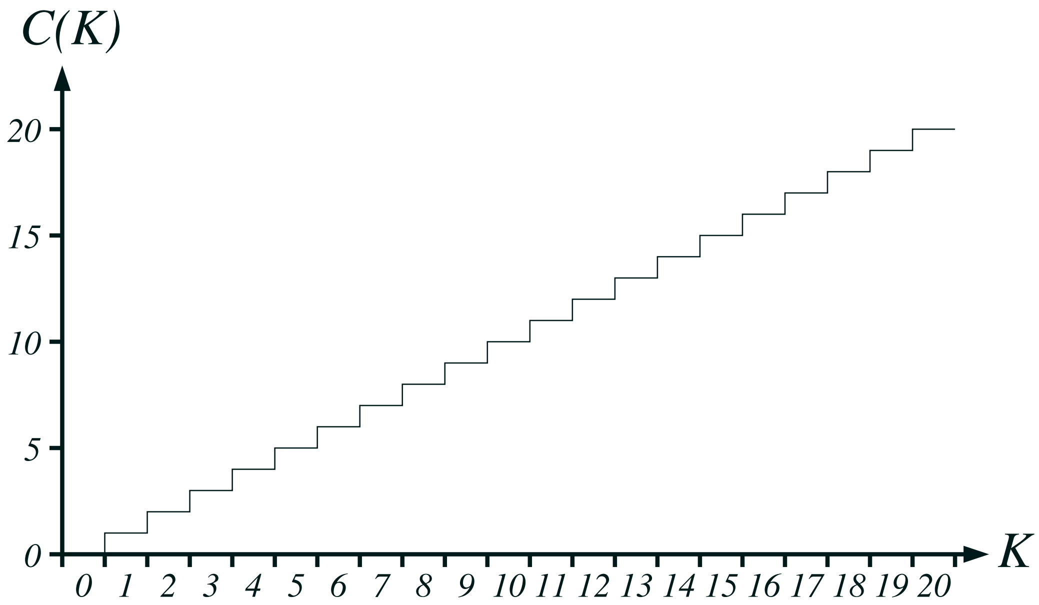
Drawback

- Although we can model more complex problems approach is not very general
- Difficult to use when there is a non-trivial correspondence between fitness and connectivity

Drawback

- Although we can model more complex problems approach is not very general

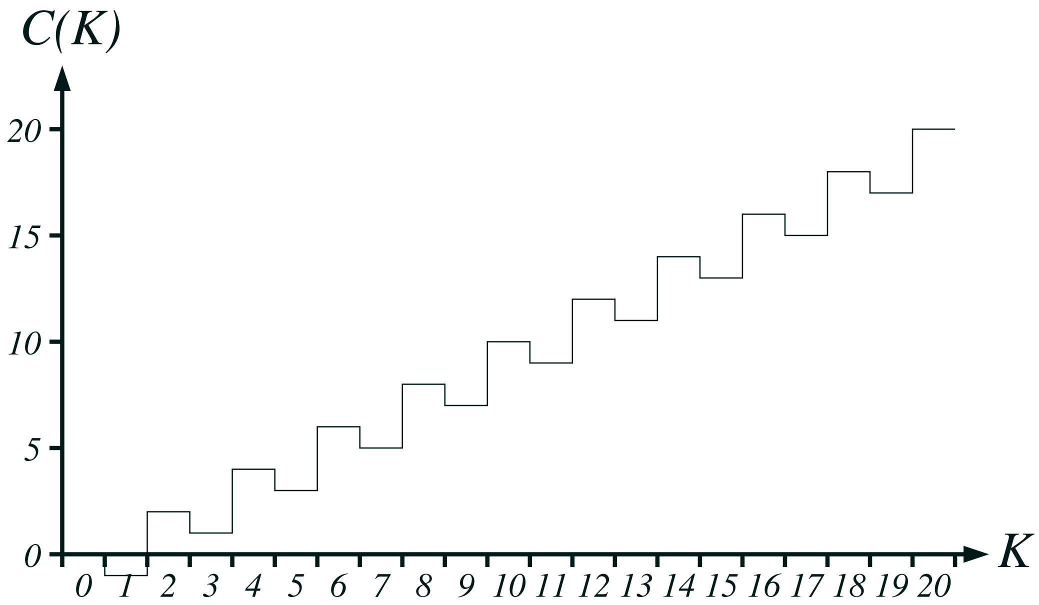
- Difficult to use when there is a non-trivial correspondence between fitness and connectivity



Drawback

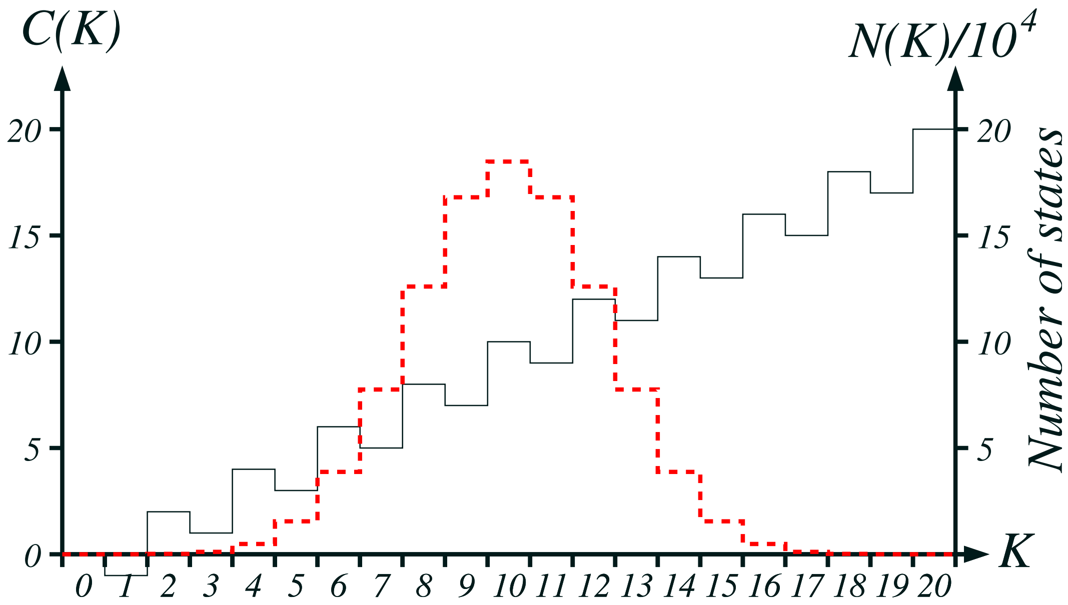
- Although we can model more complex problems approach is not very general

- Difficult to use when there is a non-trivial correspondence between fitness and connectivity



Drawback

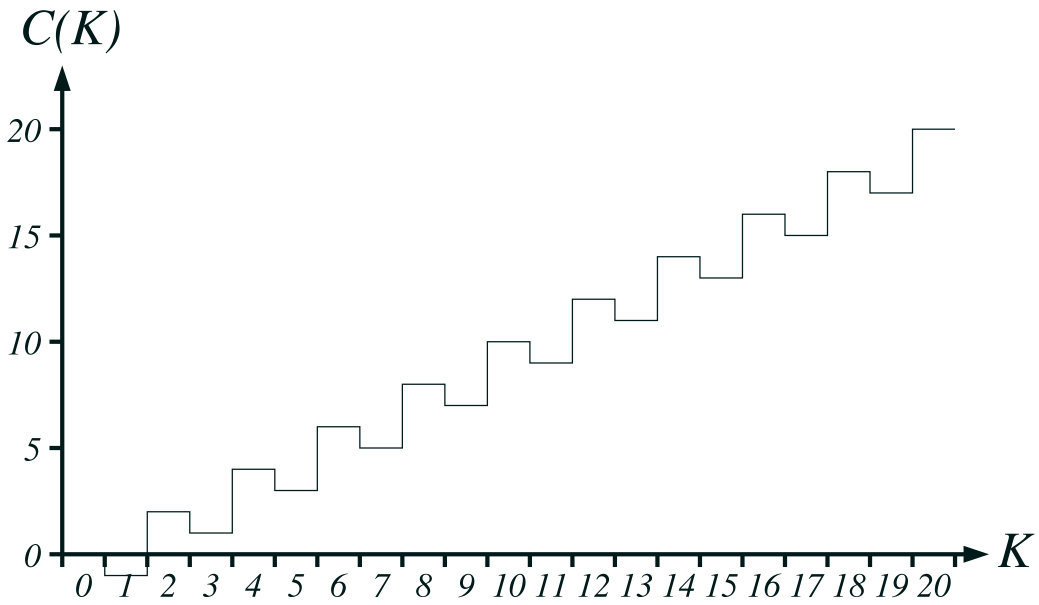
- Although we can model more complex problems approach is not very general
- Difficult to use when there is a non-trivial correspondence between fitness and connectivity



Drawback

- Although we can model more complex problems approach is not very general

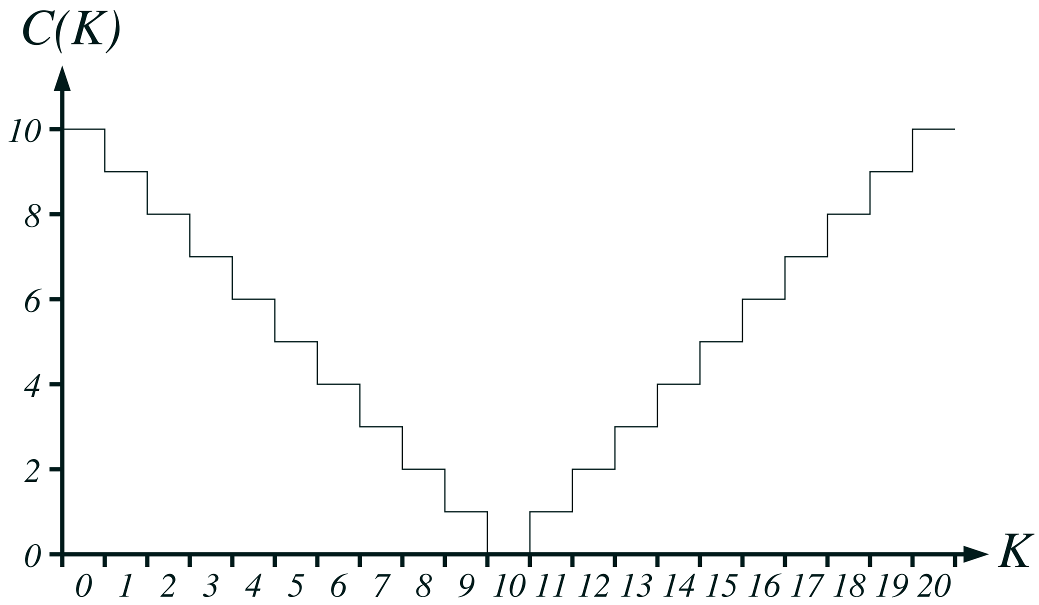
- Difficult to use when there is a non-trivial correspondence between fitness and connectivity



Drawback

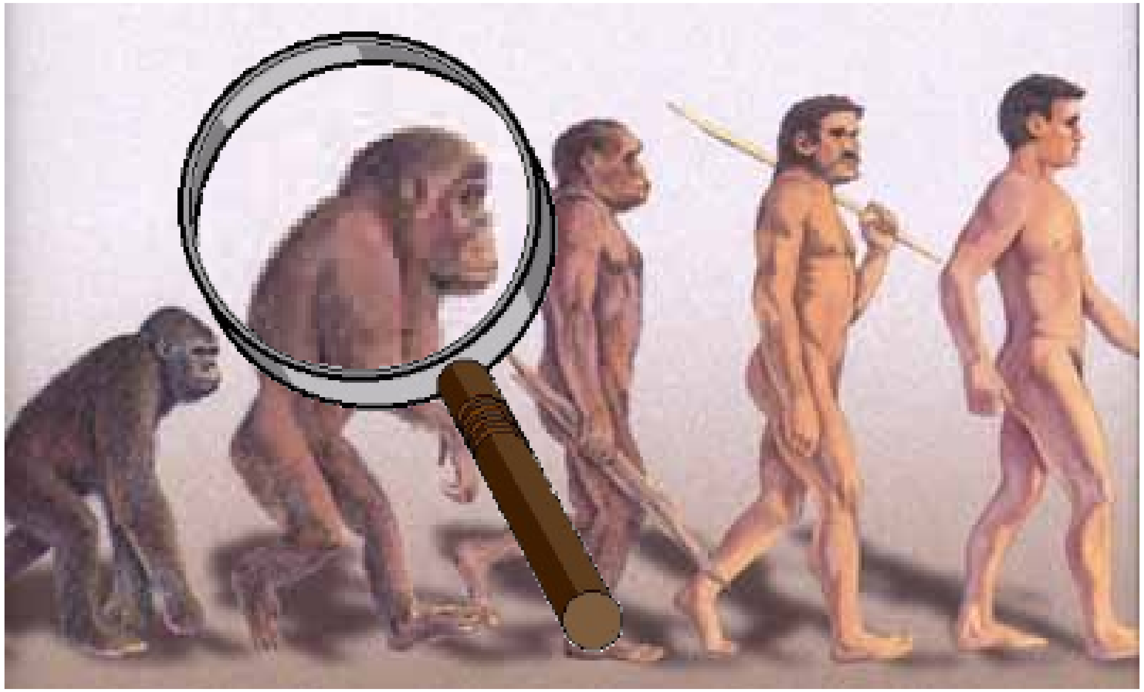
- Although we can model more complex problems approach is not very general

- Difficult to use when there is a non-trivial correspondence between fitness and connectivity



Outline

1. Introduction
   - Motivation
   - Cumulant Approach
2. **Approximate Markov Model**
   - New Approach
   - Results
   - Future Work



Approximate Markov Model

- Instead of describing the full probability distribution $\mathbb{P}(\boldsymbol{N}, t)$ we keep a 'small' set of statistics
  - ⋆ Average occupation $\boldsymbol{\mu} = \langle \boldsymbol{N} \rangle / p$
  - ⋆ Occupation covariance $\mathbf{C} = \left(\langle \boldsymbol{N}\, \boldsymbol{N}^T \rangle - \langle \boldsymbol{N} \rangle \langle \boldsymbol{N}^T \rangle \right) / p^2$
- These are *insufficient statistics* to compute the effect of selection, so we are forced to make approximations
- Can be considered an extension to the infinite population model which considers only $\boldsymbol{\mu} = \langle \boldsymbol{N} \rangle / p$

Approximate Markov Model

- Instead of describing the full probability distribution $\mathbb{P}(\boldsymbol{N}, t)$ we keep a 'small' set of statistics
  - ⋆ Average occupation $\boldsymbol{\mu} = \langle \boldsymbol{N} \rangle / p$
  - ⋆ Occupation covariance $\mathbf{C} = \left(\langle \boldsymbol{N}\, \boldsymbol{N}^T \rangle - \langle \boldsymbol{N} \rangle \langle \boldsymbol{N}^T \rangle \right) / p^2$
- These are *insufficient statistics* to compute the effect of selection, so we are forced to make approximations
- Can be considered an extension to the infinite population model which considers only $\boldsymbol{\mu} = \langle \boldsymbol{N} \rangle / p$

Approximate Markov Model

- Instead of describing the full probability distribution $\mathbb{P}\big(\boldsymbol{N}, t\big)$ we keep a 'small' set of statistics
  - ⋆ Average occupation $\boldsymbol{\mu} = \left\langle \boldsymbol{N} \right\rangle / p$
  - ⋆ Occupation covariance $\mathbf{C} = \left(\left\langle \boldsymbol{N}\, \boldsymbol{N}^T \right\rangle - \left\langle \boldsymbol{N} \right\rangle \left\langle \boldsymbol{N}^T \right\rangle \right) / p^2$
- These are *insufficient statistics* to compute the effect of selection, so we are forced to make approximations
- Can be considered an extension to the infinite population model which considers only $\boldsymbol{\mu} = \left\langle \boldsymbol{N} \right\rangle / p$

Approximate Markov Model

- Instead of describing the full probability distribution $\mathbb{P}(\boldsymbol{N}, t)$ we keep a 'small' set of statistics
  - ⋆ Average occupation $\boldsymbol{\mu} = \langle \boldsymbol{N} \rangle / p$
  - ⋆ Occupation covariance $\mathbf{C} = \left(\langle \boldsymbol{N}\, \boldsymbol{N}^T \rangle - \langle \boldsymbol{N} \rangle \langle \boldsymbol{N}^T \rangle \right) / p^2$
- These are *insufficient statistics* to compute the effect of selection, so we are forced to make approximations
- Can be considered an extension to the infinite population model which considers only $\boldsymbol{\mu} = \langle \boldsymbol{N} \rangle / p$

Modelling a Simple GA

- We consider a population undergoing selection and mutation only
  - ⋆ We select a member of the population with a probability proportional to a selection strength s_i (which is some function of its cost/fitness)
  - ⋆ The probability of selecting from a state i is therefore

$$\frac{s_i N_i}{\sum_j s_j N_j}$$

  - ⋆ The probability of mutating from state j to state i is given by W_{ij}
  - ⋆ Keep the population size, p, fixed

Modelling a Simple GA

- We consider a population undergoing selection and mutation only
  - ⋆ We select a member of the population with a probability proportional to a selection strength s_i (which is some function of its cost/fitness)
  - ⋆ The probability of selecting from a state i is therefore

$$\frac{s_i N_i}{\sum_j s_j N_j}$$

  - ⋆ The probability of mutating from state j to state i is given by W_{ij}
  - ⋆ Keep the population size, p, fixed

Modelling a Simple GA

- We consider a population undergoing selection and mutation only
  - ⋆ We select a member of the population with a probability proportional to a selection strength s_i (which is some function of its cost/fitness)
  - ⋆ The probability of selecting from a state i is therefore

$$\frac{s_i N_i}{\sum_j s_j N_j}$$

  - ⋆ The probability of mutating from state j to state i is given by W_{ij}
  - ⋆ Keep the population size, p, fixed

Modelling a Simple GA

- We consider a population undergoing selection and mutation only
  - ⋆ We select a member of the population with a probability proportional to a selection strength s_i (which is some function of its cost/fitness)
  - ⋆ The probability of selecting from a state i is therefore

$$\frac{s_i N_i}{\sum_j s_j N_j}$$

  - ⋆ The probability of mutating from state j to state i is given by W_{ij}
  - ⋆ Keep the population size, p, fixed

Modelling a Simple GA

- We consider a population undergoing selection and mutation only
  - ⋆ We select a member of the population with a probability proportional to a selection strength s_i (which is some function of its cost/fitness)
  - ⋆ The probability of selecting from a state i is therefore

$$\frac{s_i N_i}{\sum_j s_j N_j}$$

  - ⋆ The probability of mutating from state j to state i is given by W_{ij}
  - ⋆ Keep the population size, p, fixed

Occupancy Distributions

- We can represent the probability of a population as

$$\mathbb{P}(\boldsymbol{N}, t) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t)\, \mathbb{P}(\boldsymbol{N}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

- where $\mathbb{P}(\boldsymbol{N}|\boldsymbol{x})$ is the multinomial

$$\mathbb{P}(\boldsymbol{N} = \boldsymbol{n}|\boldsymbol{x}) = p! \prod_{i \in \mathcal{S}} \frac{(x_i)^{n_i}}{n_i!} \left[\!\left[\sum_{i \in \mathcal{S}} n_i = p \right]\!\right]$$

- With the constraint $\sum_i X_i = 1$

- It slightly easier working with $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ and its mean and covariance rather than $\mathbb{P}(\boldsymbol{N}, t)$

Occupancy Distributions

- We can represent the probability of a population as

$$\mathbb{P}(\boldsymbol{N}, t) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t)\, \mathbb{P}(\boldsymbol{N}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

- where $\mathbb{P}(\boldsymbol{N}|\boldsymbol{x})$ is the multinomial

$$\mathbb{P}(\boldsymbol{N} = \boldsymbol{n}|\boldsymbol{x}) = p! \prod_{i \in \mathcal{S}} \frac{(x_i)^{n_i}}{n_i!} \left[\!\left[\sum_{i \in \mathcal{S}} n_i = p \right]\!\right]$$

- With the constraint $\sum_i X_i = 1$

- It slightly easier working with $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ and its mean and covariance rather than $\mathbb{P}(\boldsymbol{N}, t)$

Occupancy Distributions

- We can represent the probability of a population as

$$\mathbb{P}(\boldsymbol{N}, t) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t)\, \mathbb{P}(\boldsymbol{N}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

- where $\mathbb{P}(\boldsymbol{N}|\boldsymbol{x})$ is the multinomial

$$\mathbb{P}(\boldsymbol{N} = \boldsymbol{n}|\boldsymbol{x}) = p! \prod_{i \in \mathcal{S}} \frac{(x_i)^{n_i}}{n_i!} \left[\!\!\left[\sum_{i \in \mathcal{S}} n_i = p \right]\!\!\right]$$

- With the constraint $\sum_i X_i = 1$
- It slightly easier working with $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ and its mean and covariance rather than $\mathbb{P}(\boldsymbol{N}, t)$

Occupancy Distributions

- We can represent the probability of a population as

$$\mathbb{P}(\boldsymbol{N}, t) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t)\, \mathbb{P}(\boldsymbol{N}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

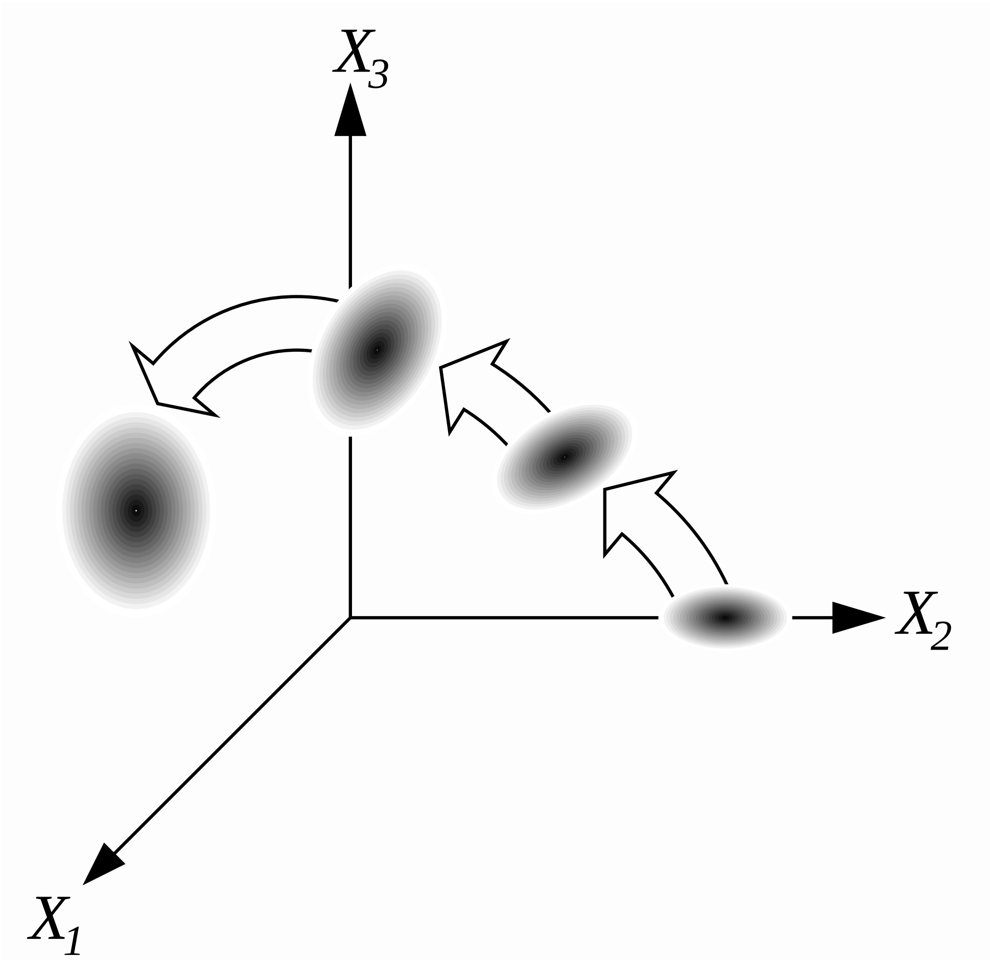
- where $\mathbb{P}(\boldsymbol{N}|\boldsymbol{x})$ is the multinomial

$$\mathbb{P}(\boldsymbol{N} = \boldsymbol{n}|\boldsymbol{x}) = p! \prod_{i \in \mathcal{S}} \frac{(x_i)^{n_i}}{n_i!} \left[\!\!\left[\sum_{i \in \mathcal{S}} n_i = p \right]\!\!\right]$$

- With the constraint $\sum_i X_i = 1$
- It slightly easier working with $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ and its mean and covariance rather than $\mathbb{P}(\boldsymbol{N}, t)$

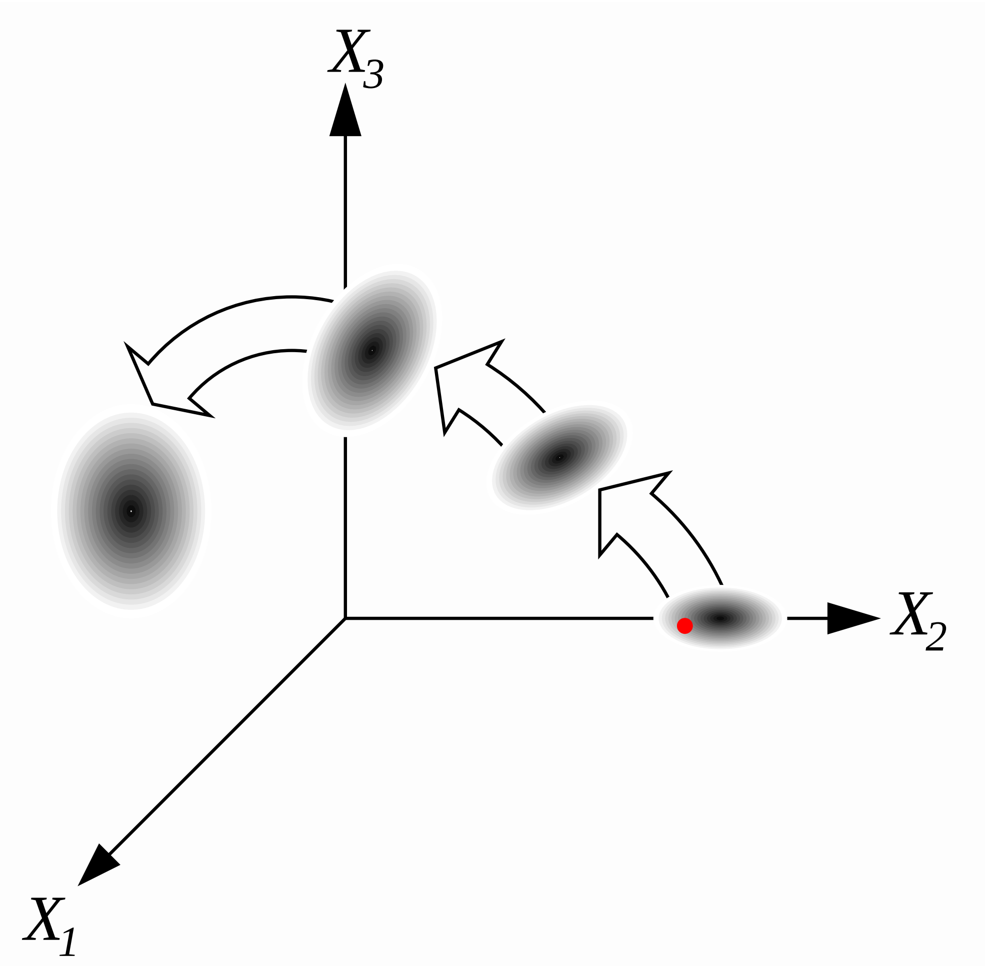
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



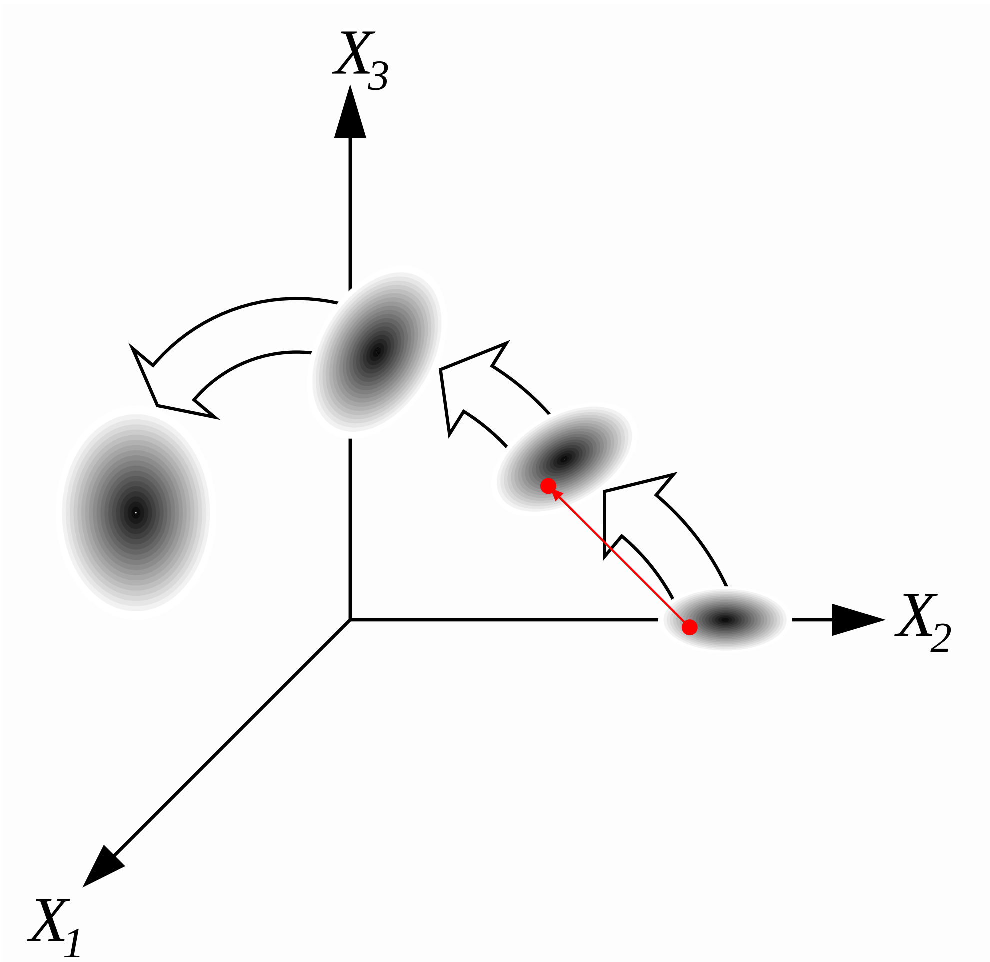
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



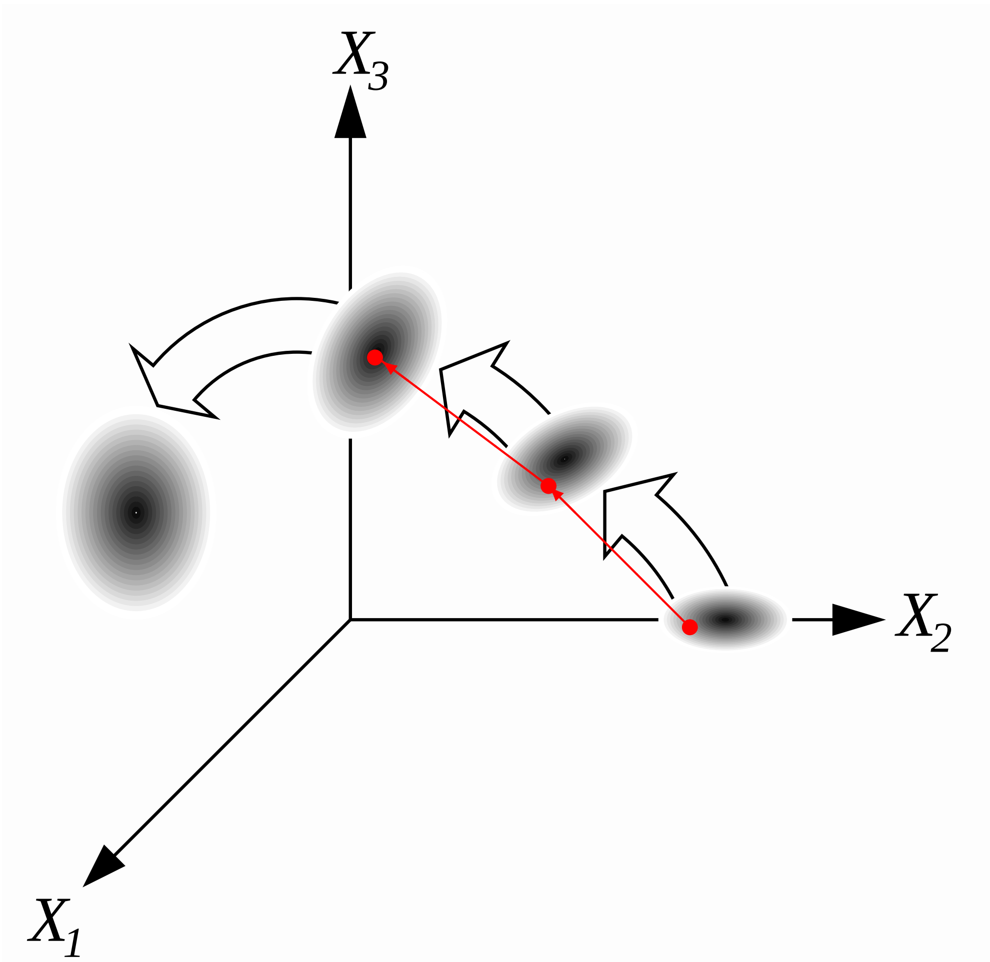
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



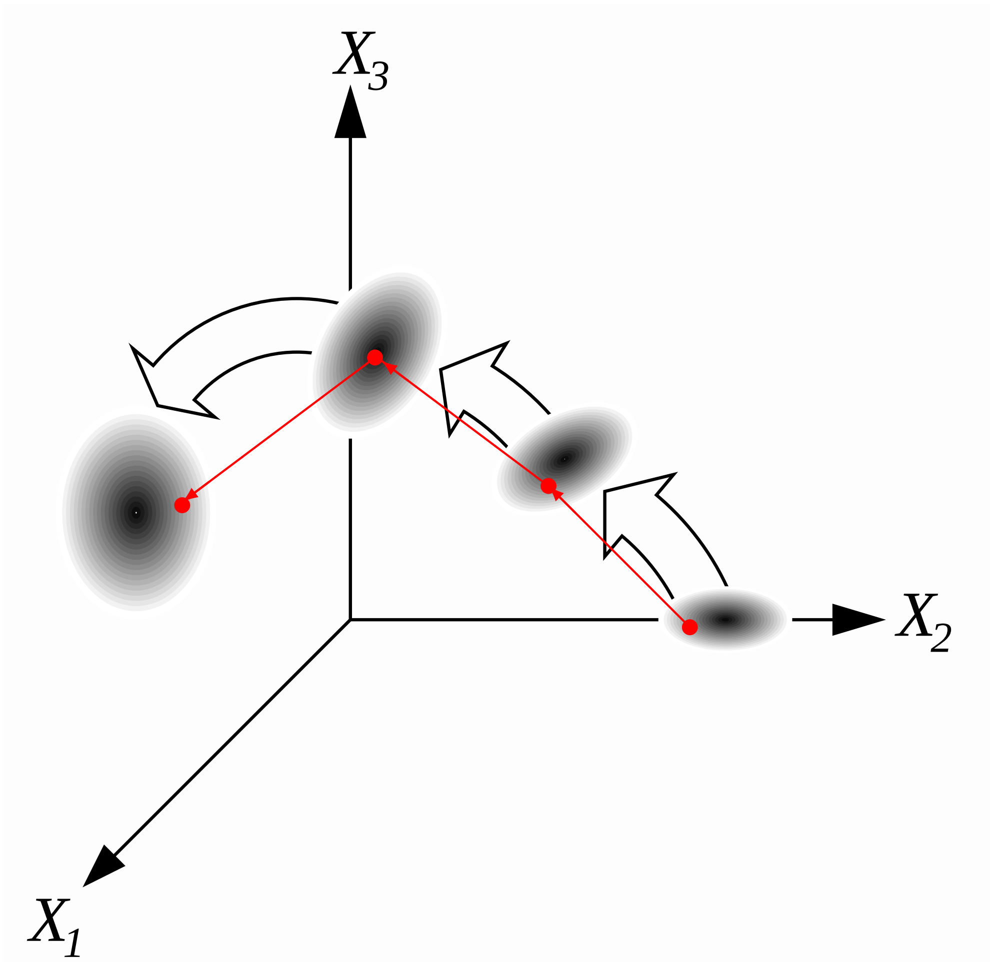
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



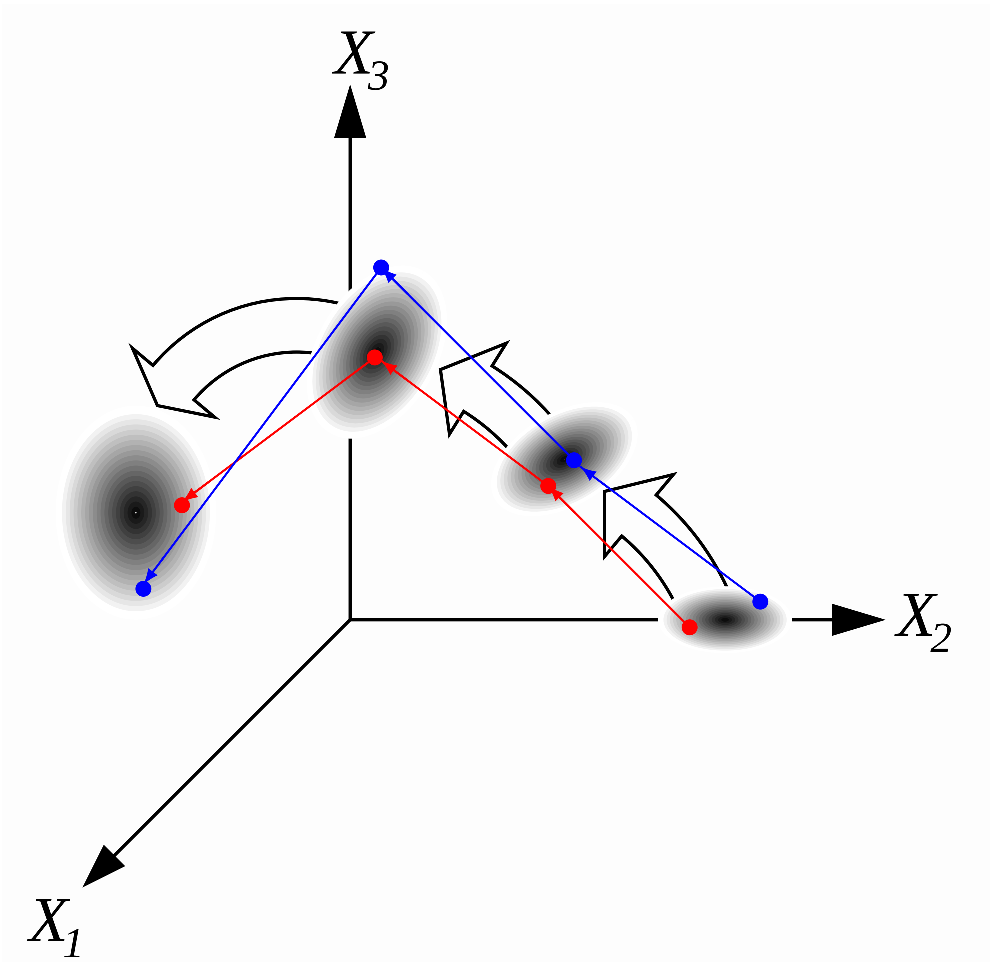
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



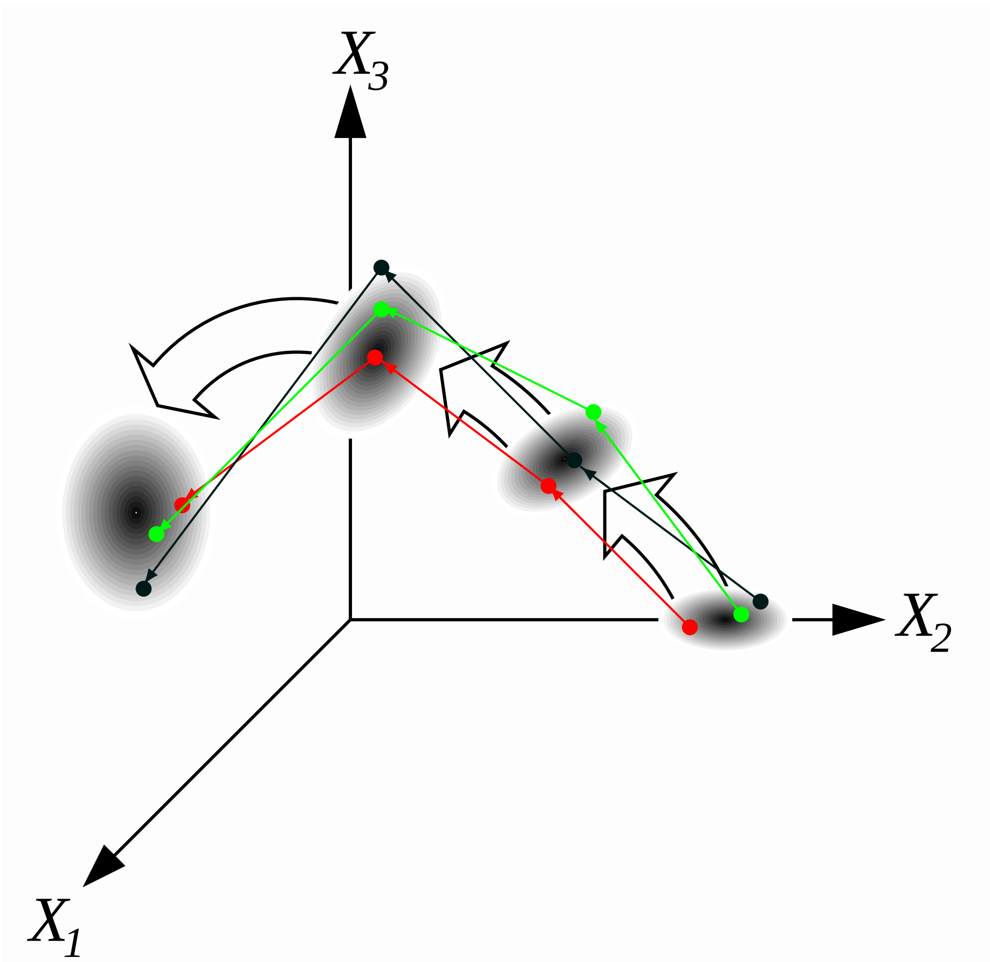
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



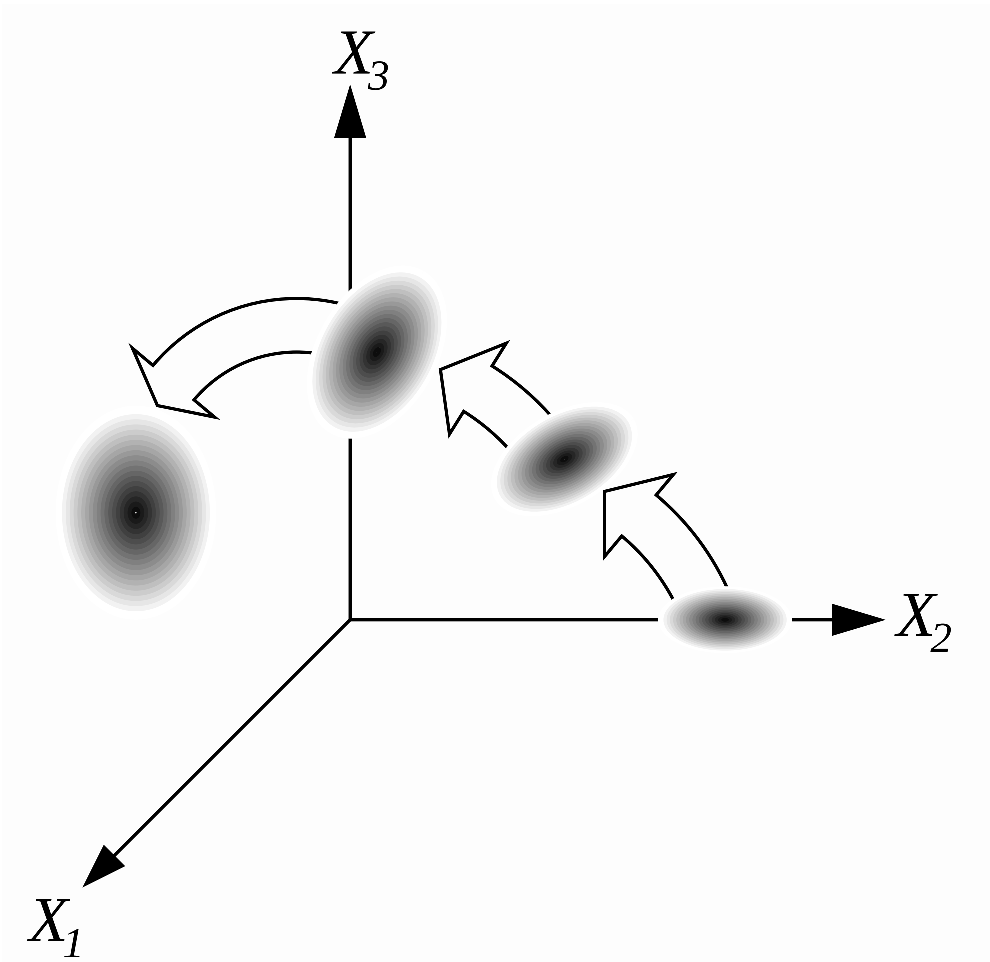
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



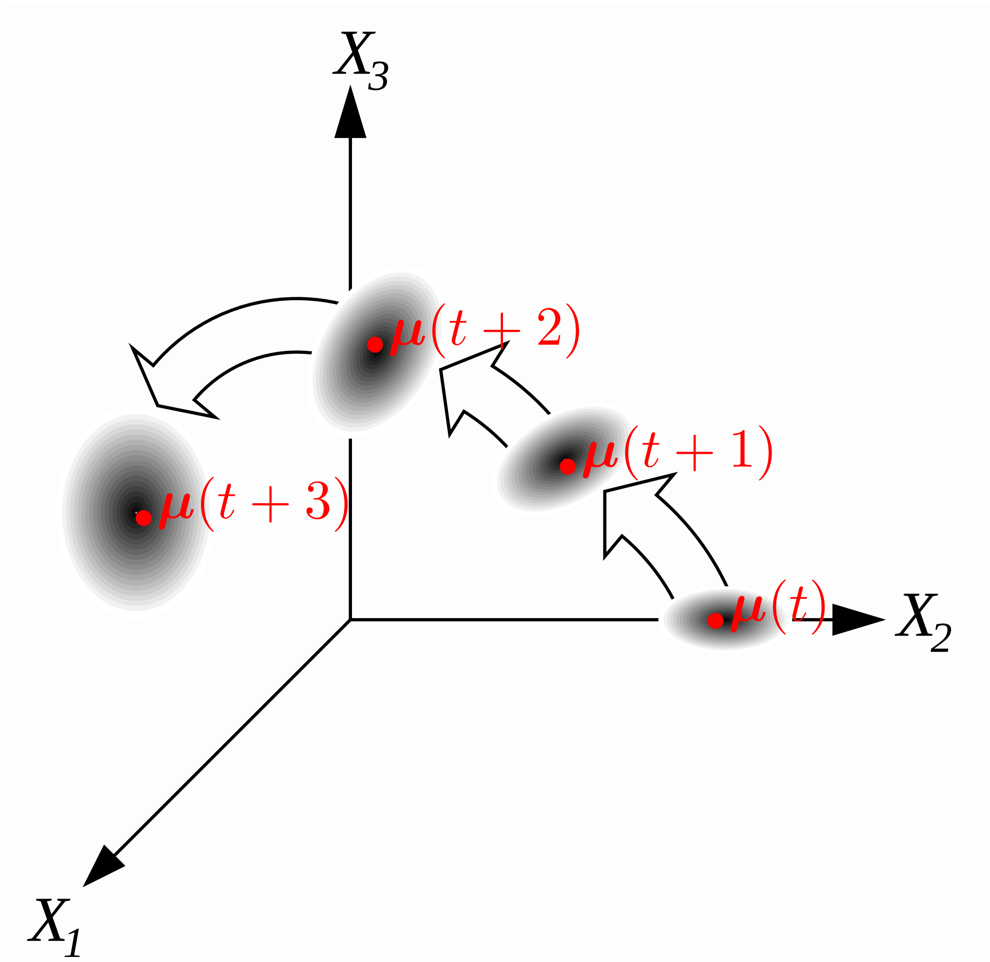
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



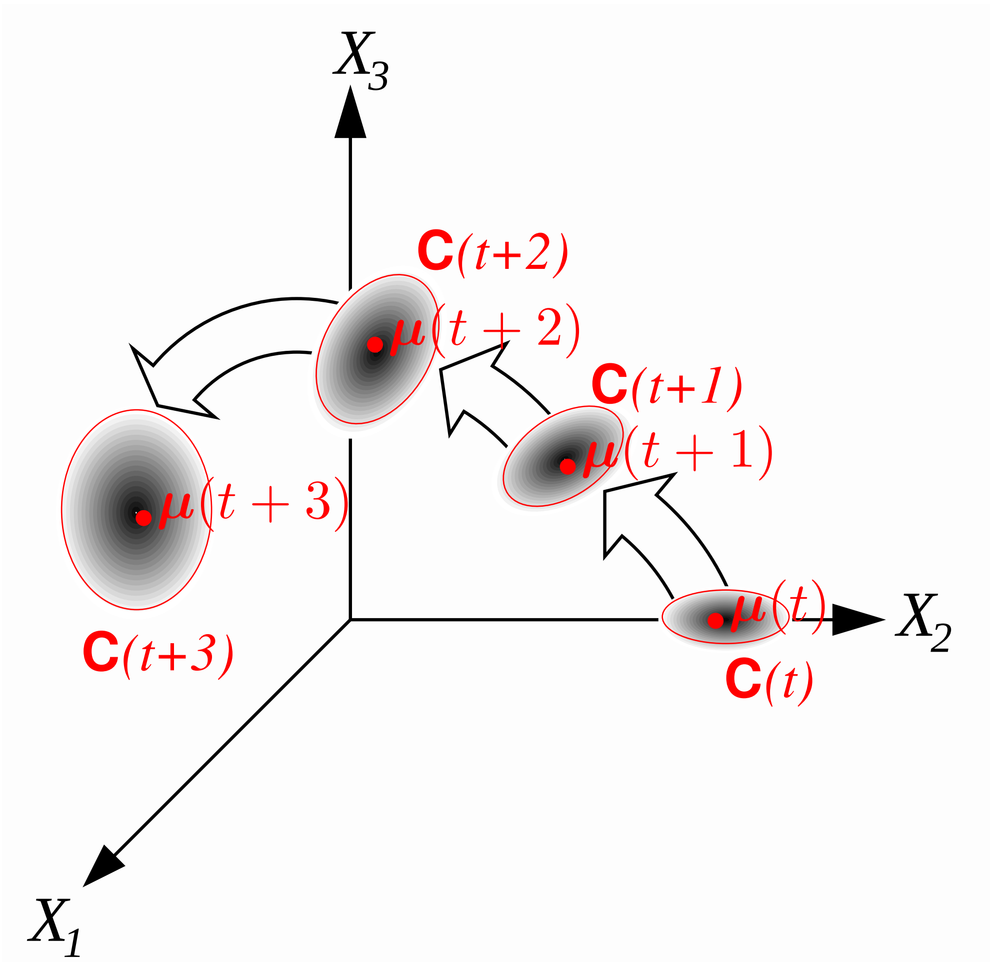
Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



Distribution Ensemble

- $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$ describes an ensemble of distributions $\boldsymbol{X}$



Four Phase Dynamics

- An ensemble of GAs can be modelled as follows

  1. Draw a distribution, $\boldsymbol{X}$, from $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$
  2. Draw a sample, $\boldsymbol{N}$, of size p (the population size) from $\boldsymbol{X}$
  3. The occupation probability after selection is given by

  $$X_i^s = \frac{N_i s_i}{\sum_j N_j s_j}$$

  4. The occupation probability after mutation is given by

  $$\boldsymbol{X}^{sm} = \boldsymbol{W}\boldsymbol{X}^s$$

  where $\boldsymbol{W}$ is the mutation matrix with elements $W_{i,j}$

- To compute $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$ we have to average over
  - ⋆ all possible ways of drawing a finite population $\boldsymbol{N}$
  - ⋆ over the distribution $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

Four Phase Dynamics

- An ensemble of GAs can be modelled as follows

  1. Draw a distribution, $\boldsymbol{X}$, from $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$
  2. Draw a sample, $\boldsymbol{N}$, of size p (the population size) from $\boldsymbol{X}$
  3. The occupation probability after selection is given by

$$X_i^s = \frac{N_i s_i}{\sum_j N_j s_j}$$

  4. The occupation probability after mutation is given by

$$\boldsymbol{X}^{sm} = \boldsymbol{W}\boldsymbol{X}^s$$

  where $\boldsymbol{W}$ is the mutation matrix with elements $W_{i,j}$

- To compute $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$ we have to average over
  - ⋆ all possible ways of drawing a finite population $\boldsymbol{N}$
  - ⋆ over the distribution $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

Four Phase Dynamics

- An ensemble of GAs can be modelled as follows

  1. Draw a distribution, $\boldsymbol{X}$, from $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$
  2. Draw a sample, $\boldsymbol{N}$, of size p (the population size) from $\boldsymbol{X}$
  3. The occupation probability after selection is given by

  $$X_i^s = \frac{N_i s_i}{\sum_j N_j s_j}$$

  4. The occupation probability after mutation is given by

  $$\boldsymbol{X}^{sm} = \boldsymbol{W}\boldsymbol{X}^s$$

  where $\boldsymbol{W}$ is the mutation matrix with elements $W_{i,j}$

- To compute $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$ we have to average over

  ⋆ all possible ways of drawing a finite population $\boldsymbol{N}$
  ⋆ over the distribution $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

Four Phase Dynamics

- An ensemble of GAs can be modelled as follows

1. Draw a distribution, $\boldsymbol{X}$, from $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$
2. Draw a sample, $\boldsymbol{N}$, of size p (the population size) from $\boldsymbol{X}$
3. The occupation probability after selection is given by

$$X_i^s = \frac{N_i s_i}{\sum_j N_j s_j}$$

4. The occupation probability after mutation is given by

$$\boldsymbol{X}^{sm} = \boldsymbol{W}\boldsymbol{X}^s$$

   where $\boldsymbol{W}$ is the mutation matrix with elements $W_{i,j}$

- To compute $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$ we have to average over
  - ⋆ all possible ways of drawing a finite population $\boldsymbol{N}$
  - ⋆ over the distribution $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

Four Phase Dynamics

- An ensemble of GAs can be modelled as follows

1. Draw a distribution, $\boldsymbol{X}$, from $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$
2. Draw a sample, $\boldsymbol{N}$, of size p (the population size) from $\boldsymbol{X}$
3. The occupation probability after selection is given by

$$X_i^s = \frac{N_i s_i}{\sum_j N_j s_j}$$

4. The occupation probability after mutation is given by

$$\boldsymbol{X}^{sm} = \boldsymbol{W}\boldsymbol{X}^s$$

where $\boldsymbol{W}$ is the mutation matrix with elements $W_{i,j}$

- To compute $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$ we have to average over
  - ⋆ all possible ways of drawing a finite population $\boldsymbol{N}$
  - ⋆ over the distribution $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

Four Phase Dynamics

- An ensemble of GAs can be modelled as follows

  1. Draw a distribution, $\boldsymbol{X}$, from $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$
  2. Draw a sample, $\boldsymbol{N}$, of size p (the population size) from $\boldsymbol{X}$
  3. The occupation probability after selection is given by

$$X_i^s = \frac{N_i s_i}{\sum_j N_j s_j}$$

  4. The occupation probability after mutation is given by

$$\boldsymbol{X}^{sm} = \boldsymbol{W}\boldsymbol{X}^s$$

  where $\boldsymbol{W}$ is the mutation matrix with elements $W_{i,j}$

- To compute $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$ we have to average over
  - ⋆ all possible ways of drawing a finite population $\boldsymbol{N}$
  - ⋆ over the distribution $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

Four Phase Dynamics

- An ensemble of GAs can be modelled as follows

  1. Draw a distribution, $\boldsymbol{X}$, from $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$
  2. Draw a sample, $\boldsymbol{N}$, of size p (the population size) from $\boldsymbol{X}$
  3. The occupation probability after selection is given by

$$X_i^s = \frac{N_i s_i}{\sum_j N_j s_j}$$

  4. The occupation probability after mutation is given by

$$\boldsymbol{X}^{sm} = \boldsymbol{W}\boldsymbol{X}^s$$

  where $\boldsymbol{W}$ is the mutation matrix with elements $W_{i,j}$

- To compute $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$ we have to average over
  - ⋆ all possible ways of drawing a finite population $\boldsymbol{N}$
  - ⋆ over the distribution $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

The Hard Part

- The only hard part is computing the distribution after selection
- We will only attempt to compute the mean and covariance of $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$
- In particular we want to compute

$$\mu_i^s = \langle X_i^s \rangle = \left\langle \frac{N_i s_i}{\sum_j N_j s_j} \right\rangle$$

$$C_{ij}^s = \langle X_i^s X_j^s \rangle - \langle X_i^s \rangle \langle X_j^s \rangle = \left\langle \frac{N_i N_j s_i s_j}{\left(\sum_k N_k s_k\right)^2} \right\rangle - \mu_i^s \mu_j^s$$

where $\langle \cdots \rangle$ denotes average over sampling and $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

The Hard Part

- The only hard part is computing the distribution after selection
- We will only attempt to compute the mean and covariance of $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$
- In particular we want to compute

$$\mu_i^s = \left\langle X_i^s \right\rangle = \left\langle \frac{N_i s_i}{\sum_j N_j s_j} \right\rangle$$

$$C_{ij}^s = \left\langle X_i^s X_j^s \right\rangle - \left\langle X_i^s \right\rangle \left\langle X_j^s \right\rangle = \left\langle \frac{N_i N_j s_i s_j}{\left(\sum_k N_k s_k\right)^2} \right\rangle - \mu_i^s \mu_j^s$$

where $\langle \cdots \rangle$ denotes average over sampling and $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

The Hard Part

- The only hard part is computing the distribution after selection
- We will only attempt to compute the mean and covariance of $f_{\boldsymbol{X}}(\boldsymbol{x}, t+1)$
- In particular we want to compute

$$\mu_i^s = \left\langle X_i^s \right\rangle = \left\langle \frac{N_i s_i}{\sum_j N_j s_j} \right\rangle$$

$$C_{ij}^s = \left\langle X_i^s X_j^s \right\rangle - \left\langle X_i^s \right\rangle \left\langle X_j^s \right\rangle = \left\langle \frac{N_i N_j s_i s_j}{\left(\sum_k N_k s_k\right)^2} \right\rangle - \mu_i^s \mu_j^s$$

where $\langle \cdots \rangle$ denotes average over sampling and $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

Stochastic Average

- A simple approach is to use the *stochastic expansion*

$$\left\langle f(\boldsymbol{N})\right\rangle = \left\langle f(\left\langle \boldsymbol{N}\right\rangle + \delta\boldsymbol{N})\right\rangle \qquad \delta\boldsymbol{N} = \boldsymbol{N} - \left\langle \boldsymbol{N}\right\rangle$$

$$\approx f(\left\langle \boldsymbol{N}\right\rangle) + \sum_i \left\langle \delta N_i\right\rangle \frac{\partial f(\left\langle \boldsymbol{N}\right\rangle)}{\partial \mu_i} + \frac{1}{2}\sum_{i,j}\left\langle \delta N_i \delta N_j\right\rangle \frac{\partial^2 f(\left\langle \boldsymbol{N}\right\rangle)}{\partial \mu_i \partial \mu_j}$$

$$= f(p\boldsymbol{\mu}) + \frac{1}{2}\sum_{i,j} p^2 C_{ij} \frac{\partial^2 f(p\boldsymbol{\mu})}{\partial \mu_i \partial \mu_j}$$

- ⋆ $\boldsymbol{\mu} = \left\langle \boldsymbol{N}\right\rangle / p$
- ⋆ $C_{ij} = \left\langle \delta N_i \delta N_j\right\rangle / p^2$

Stochastic Average

- A simple approach is to use the *stochastic expansion*

$$\left\langle f(\boldsymbol{N})\right\rangle = \left\langle f(\left\langle \boldsymbol{N}\right\rangle + \delta\boldsymbol{N})\right\rangle \qquad \delta\boldsymbol{N} = \boldsymbol{N} - \left\langle \boldsymbol{N}\right\rangle$$

$$\approx f(\left\langle \boldsymbol{N}\right\rangle) + \sum_i \left\langle \delta N_i\right\rangle \frac{\partial f(\left\langle \boldsymbol{N}\right\rangle)}{\partial \mu_i} + \frac{1}{2}\sum_{i,j} \left\langle \delta N_i \delta N_j\right\rangle \frac{\partial^2 f(\left\langle \boldsymbol{N}\right\rangle)}{\partial \mu_i \partial \mu_j}$$

$$= f(p\boldsymbol{\mu}) + \frac{1}{2}\sum_{i,j} p^2 C_{ij} \frac{\partial^2 f(p\boldsymbol{\mu})}{\partial \mu_i \partial \mu_j}$$

- ⋆ $\boldsymbol{\mu} = \left\langle \boldsymbol{N}\right\rangle / p$
- ⋆ $C_{ij} = \left\langle \delta N_i \delta N_j\right\rangle / p^2$

Stochastic Average

- A simple approach is to use the *stochastic expansion*

$$\left\langle f(\boldsymbol{N})\right\rangle = \left\langle f(\left\langle \boldsymbol{N}\right\rangle + \delta\boldsymbol{N})\right\rangle \qquad \delta\boldsymbol{N} = \boldsymbol{N} - \left\langle \boldsymbol{N}\right\rangle$$

$$\approx f(\left\langle \boldsymbol{N}\right\rangle) + \sum_i \left\langle \delta N_i\right\rangle \frac{\partial f(\left\langle \boldsymbol{N}\right\rangle)}{\partial \mu_i} + \frac{1}{2}\sum_{i,j}\left\langle \delta N_i \delta N_j\right\rangle \frac{\partial^2 f(\left\langle \boldsymbol{N}\right\rangle)}{\partial \mu_i \partial \mu_j}$$

$$= f(p\boldsymbol{\mu}) + \frac{1}{2}\sum_{i,j} p^2 C_{ij} \frac{\partial^2 f(p\boldsymbol{\mu})}{\partial \mu_i \partial \mu_j}$$

- ⋆ $\boldsymbol{\mu} = \left\langle \boldsymbol{N}\right\rangle / p$
- ⋆ $C_{ij} = \left\langle \delta N_i \delta N_j \right\rangle / p^2$

Stochastic Expansion Approximation

- This lead to simple update equation for $\boldsymbol{\mu}$ and $\mathbf{C}$

$$\mu_i^s(t) \approx \frac{s_i}{m}\mu_i\left(1 + \sum_{k,l} s_k\, s_l\, c_{kl}(t)\right) - \frac{s_i}{m^2}\sum_k s_k\, c_{ik}(t)$$

$$C_{ij}^s(t) \approx \frac{s_i\, s_j}{m^2}\left(C_{ij}(t) + \frac{\mu_i\mu_j}{m^2}\sum_{k,l} s_k\, s_l\, C_{kl}(t) - \frac{\mu_i}{m}\sum_k s_k\, C_{jk}(t) - \frac{\mu_j}{m}\sum_{k,l} s_k\, C_{ik}(t)\right)$$

- $m = \sum_i s_i \mu_i$
- Simple update accurate for short times

Stochastic Expansion Approximation

- This lead to simple update equation for $\boldsymbol{\mu}$ and $\mathbf{C}$

$$\mu_i^s(t) \approx \frac{s_i}{m}\mu_i \left(1 + \sum_{k,l} s_k\, s_l\, c_{kl}(t) \right) - \frac{s_i}{m^2} \sum_k s_k\, c_{ik}(t)$$

$$C_{ij}^s(t) \approx \frac{s_i\, s_j}{m^2} \left(C_{ij}(t) + \frac{\mu_i \mu_j}{m^2} \sum_{k,l} s_k\, s_l\, C_{kl}(t) - \frac{\mu_i}{m} \sum_k s_k\, C_{jk}(t) - \frac{\mu_j}{m} \sum_{k,l} s_k\, C_{ik}(t) \right)$$

- $m = \sum_i s_i \mu_i$

- Simple update accurate for short times

Can We Do Better?

- The quantities we are interested in can be obtained from

$$G(\boldsymbol{S}) = \left\langle \log\left(\sum_i s_i N_i\right)\right\rangle$$

$$\mu_i^s = \left\langle X_i^s \right\rangle = \left\langle \frac{N_i s_i}{\sum_j N_j s_j} \right\rangle = s_i \frac{\partial G(\boldsymbol{S})}{\partial s_i}$$

$$C_{ij}^s = \left\langle X_i^s X_j^s \right\rangle - \left\langle X_i^s \right\rangle \left\langle X_j^s \right\rangle = \left\langle \frac{N_i N_j s_i s_j}{\left(\sum_k N_k s_k\right)^2} \right\rangle - \mu_i^s \mu_j^s$$

$$= -s_i s_j \frac{\partial^2 G(\boldsymbol{S})}{\partial s_i \partial s_j} - \mu_i^s \mu_j^s$$

Can We Do Better?

- The quantities we are interested in can be obtained from

$$G(\boldsymbol{S}) = \left\langle \log\left(\sum_i s_i N_i\right)\right\rangle$$

$$\mu_i^s = \langle X_i^s \rangle = \left\langle \frac{N_i s_i}{\sum_j N_j s_j} \right\rangle = s_i \frac{\partial G(\boldsymbol{S})}{\partial s_i}$$

$$C_{ij}^s = \langle X_i^s X_j^s \rangle - \langle X_i^s \rangle \langle X_j^s \rangle = \left\langle \frac{N_i N_j s_i s_j}{\left(\sum_k N_k s_k\right)^2} \right\rangle - \mu_i^s \mu_j^s$$

$$= -s_i s_j \frac{\partial^2 G(\boldsymbol{S})}{\partial s_i \partial s_j} - \mu_i^s \mu_j^s$$

Averaging Logarithm

- Using Euler's representation of the logarithm

$$\log(A) = \int_0^\infty \frac{\mathrm{e}^{-z} - \mathrm{e}^{-zA}}{z} \, \mathrm{d}z$$

- we get

$$G(\boldsymbol{S}) = \left\langle \log\left(\sum_i s_i N_i \right) \right\rangle = \int_0^\infty \frac{\mathrm{e}^{-z} - g(\boldsymbol{S}, z)}{z} \, \mathrm{d}z$$

- where

$$g(\boldsymbol{S}, z) = \left\langle \mathrm{e}^{-z \sum_j s_j N_j} \right\rangle$$

Averaging Logarithm

- Using Euler's representation of the logarithm

$$\log(A) = \int_0^\infty \frac{\mathrm{e}^{-z} - \mathrm{e}^{-zA}}{z} \mathrm{d}z$$

- we get

$$G(\boldsymbol{S}) = \left\langle \log\left(\sum_i s_i N_i\right)\right\rangle = \int_0^\infty \frac{\mathrm{e}^{-z} - g(\boldsymbol{S}, z)}{z} \mathrm{d}z$$

- where

$$g(\boldsymbol{S}, z) = \left\langle \mathrm{e}^{-z\sum_j s_j N_j}\right\rangle$$

Averaging Logarithm

- Using Euler's representation of the logarithm

$$\log(A) = \int_0^\infty \frac{\mathrm{e}^{-z} - \mathrm{e}^{-zA}}{z} \, \mathrm{d}z$$

- we get

$$G(\boldsymbol{S}) = \left\langle \log\left(\sum_i s_i N_i \right) \right\rangle = \int_0^\infty \frac{\mathrm{e}^{-z} - g(\boldsymbol{S}, z)}{z} \, \mathrm{d}z$$

- where

$$g(\boldsymbol{S}, z) = \left\langle \mathrm{e}^{-z \sum_j s_j N_j} \right\rangle$$

Sampling

- In sampling we draw a population, $\boldsymbol{N}$, from a multinomial distribution

$$\mathbb{P}(\boldsymbol{N}=\boldsymbol{n}|\boldsymbol{x}) = \prod_{i=1}^{m} \frac{x_i^{n_i}}{n_i!} \left[\!\!\left[\sum_{i=1}^{m} n_i = p \right]\!\!\right]$$

- Now

$$\sum_{\boldsymbol{n}} \mathbb{P}(\boldsymbol{N}=\boldsymbol{n}|\boldsymbol{x}) \mathrm{e}^{-z\sum_j s_j n_j} = \left(\sum_i x_i \mathrm{e}^{-z\sum_j s_j} \right)^p$$

- so

$$g(\boldsymbol{S}, z) = \left\langle \mathrm{e}^{-z\sum_j s_j N_j} \right\rangle = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

Sampling

- In sampling we draw a population, $\boldsymbol{N}$, from a multinomial distribution

$$\mathbb{P}(\boldsymbol{N} = \boldsymbol{n}|\boldsymbol{x}) = \prod_{i=1}^{m} \frac{x_i^{n_i}}{n_i!} \left[\!\left[\sum_{i=1}^{m} n_i = p \right]\!\right]$$

- Now

$$\sum_{\boldsymbol{n}} \mathbb{P}(\boldsymbol{N} = \boldsymbol{n}|\boldsymbol{x}) \mathrm{e}^{-z\sum_j s_j n_j} = \left(\sum_i x_i \mathrm{e}^{-z\sum_j s_j} \right)^p$$

- so

$$g(\boldsymbol{S}, z) = \left\langle \mathrm{e}^{-z\sum_j s_j N_j} \right\rangle = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

Sampling

- In sampling we draw a population, $\boldsymbol{N}$, from a multinomial distribution

$$\mathbb{P}(\boldsymbol{N}=\boldsymbol{n}|\boldsymbol{x}) = \prod_{i=1}^{m} \frac{x_i^{n_i}}{n_i!} \left[\!\!\left[\sum_{i=1}^{m} n_i = p \right]\!\!\right]$$

- Now

$$\sum_{\boldsymbol{n}} \mathbb{P}(\boldsymbol{N}=\boldsymbol{n}|\boldsymbol{x}) \mathrm{e}^{-z\sum_j s_j n_j} = \left(\sum_i x_i \mathrm{e}^{-z\sum_j s_j} \right)^p$$

- so

$$g(\boldsymbol{S}, z) = \left\langle \mathrm{e}^{-z\sum_j s_j N_j} \right\rangle = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

Averaging over $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

- We need to compute

$$g(\boldsymbol{S}, z) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

- So far we have made no approximations!

- For $p = 2$ we can compute this exactly

$$g(\boldsymbol{S}, z) = \left\langle \left(\sum_i X_i \mathrm{e}^{-z s_i} \right)^2 \right\rangle = \sum_{i,j} \left(\mu_i \, \mu_j + C_{ij} \right) \mathrm{e}^{-z(s_i + s_j)}$$

- We can also do the integral over z analytically

Averaging over $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

- We need to compute

$$g(\boldsymbol{S}, z) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

- So far we have made no approximations!

- For $p = 2$ we can compute this exactly

$$g(\boldsymbol{S}, z) = \left\langle \left(\sum_i X_i \mathrm{e}^{-z s_i} \right)^2 \right\rangle = \sum_{i,j} \left(\mu_i \, \mu_j + C_{ij} \right) \mathrm{e}^{-z(s_i + s_j)}$$

- We can also do the integral over z analytically

Averaging over $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

- We need to compute

$$g(\boldsymbol{S}, z) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

- So far we have made no approximations!

- For $p = 2$ we can compute this exactly

$$g(\boldsymbol{S}, z) = \left\langle \left(\sum_i X_i \mathrm{e}^{-z s_i} \right)^2 \right\rangle = \sum_{i,j} \left(\mu_i \, \mu_j + C_{ij} \right) \mathrm{e}^{-z(s_i + s_j)}$$

- We can also do the integral over z analytically

Averaging over $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

- We need to compute

$$g(\boldsymbol{S}, z) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

- So far we have made no approximations!
- For $p = 2$ we can compute this exactly

$$g(\boldsymbol{S}, z) = \left\langle \left(\sum_i X_i \mathrm{e}^{-z s_i} \right)^2 \right\rangle = \sum_{i,j} \left(\mu_i \, \mu_j + C_{ij} \right) \mathrm{e}^{-z(s_i + s_j)}$$

- We can also do the integral over z analytically

Averaging over $f_{\boldsymbol{X}}(\boldsymbol{x}, t)$

- We need to compute

$$g(\boldsymbol{S}, z) = \int f_{\boldsymbol{X}}(\boldsymbol{x}, t) \left(\sum_i x_i \mathrm{e}^{-z s_i} \right)^p \mathrm{d}\boldsymbol{x}$$

- So far we have made no approximations!

- For $p = 2$ we can compute this exactly

$$g(\boldsymbol{S}, z) = \left\langle \left(\sum_i X_i \mathrm{e}^{-z s_i} \right)^2 \right\rangle = \sum_{i,j} \left(\mu_i \, \mu_j + C_{ij} \right) \mathrm{e}^{-z(s_i + s_j)}$$

- We can also do the integral over z analytically

Exact Update Equations for $p = 2$

- The update equations for $p = 2$ are

$$\mu_i^s = 2 \sum_j \frac{\mu_i \mu_j + C_{ij} s_i}{s_i + s_j}$$

$$C_{ij}^s = \frac{2(\mu_i \mu_j + C_{ij}) s_i s_j}{(s_i + s_j)^2} + \sum_k \frac{2(\mu_i \mu_k + C_{ik}) s_i^2}{(s_i + s_k)^2} [\![i = j]\!] - \mu_i \mu_j$$

$$\boldsymbol{\mu}' = \mathbf{W} \boldsymbol{\mu}^s$$

$$\mathbf{C}' = \mathbf{W} \mathbf{C}^s \mathbf{W}^T$$

- This is a very elegant way of computing the dynamics for a GA with population size 2

Exact Update Equations for $p = 2$

- The update equations for $p = 2$ are

$$\mu_i^s = 2 \sum_j \frac{\mu_i \mu_j + C_{ij} s_i}{s_i + s_j}$$

$$C_{ij}^s = \frac{2(\mu_i \mu_j + C_{ij}) s_i s_j}{(s_i + s_j)^2} + \sum_k \frac{2(\mu_i \mu_k + C_{ik}) s_i^2}{(s_i + s_k)^2} [\![i = j]\!] - \mu_i \mu_j$$

$$\boldsymbol{\mu}' = \mathbf{W} \boldsymbol{\mu}^s$$

$$\mathbf{C}' = \mathbf{W} \mathbf{C}^s \mathbf{W}^T$$

- This is a very elegant way of computing the dynamics for a GA with population size 2

Exact Update Equations for $p = 2$

- The update equations for $p = 2$ are

$$\mu_i^s = 2\sum_j \frac{\mu_i\mu_j + C_{ij}s_i}{s_i + s_j}$$

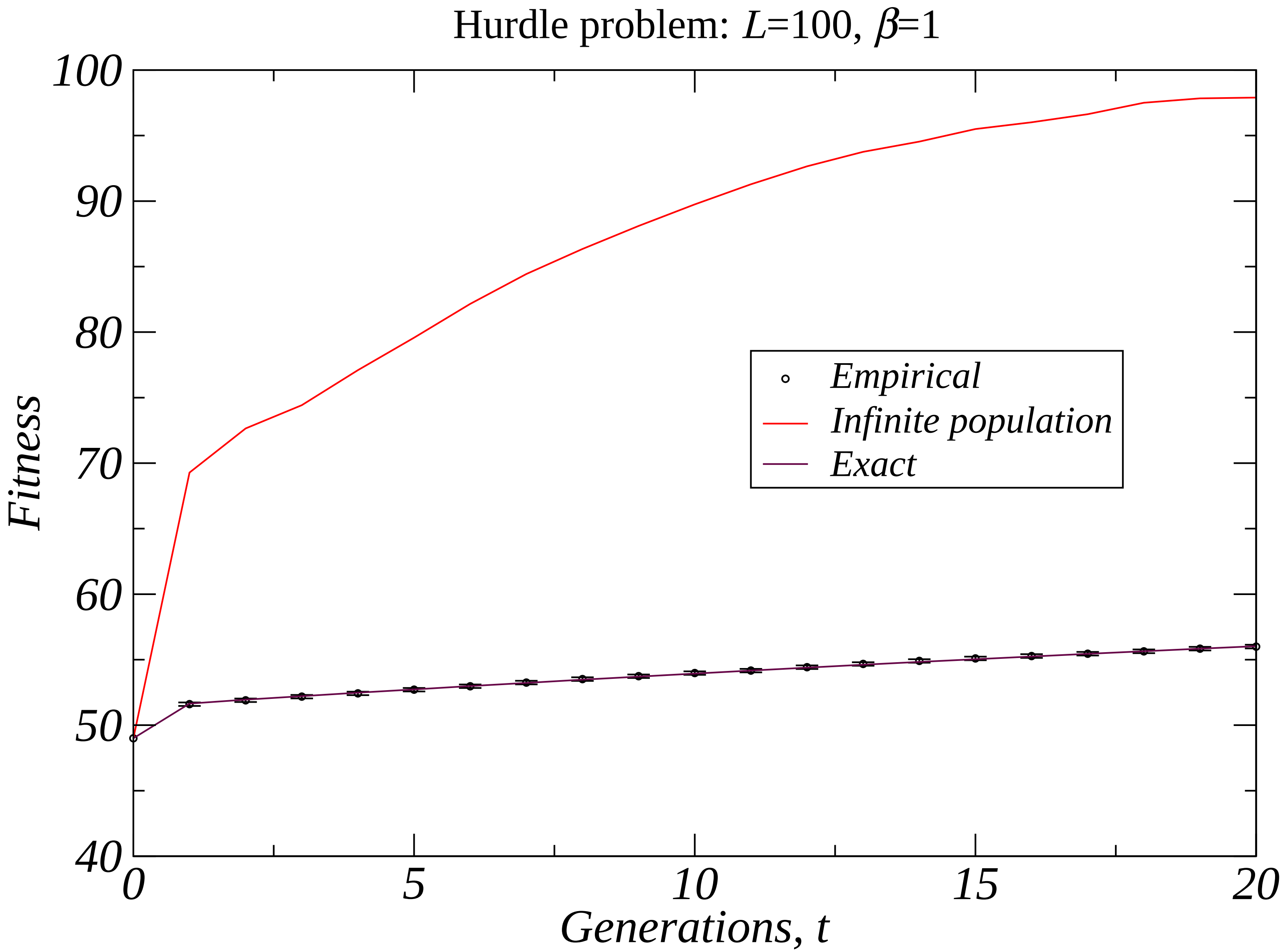
$$C_{ij}^s = \frac{2(\mu_i\mu_j + C_{ij})s_i s_j}{(s_i + s_j)^2} + \sum_k \frac{2(\mu_i\mu_k + C_{ik})s_i^2}{(s_i + s_k)^2} [\![i = j]\!] - \mu_i\mu_j$$

$$\boldsymbol{\mu}' = \mathbf{W}\boldsymbol{\mu}^s$$

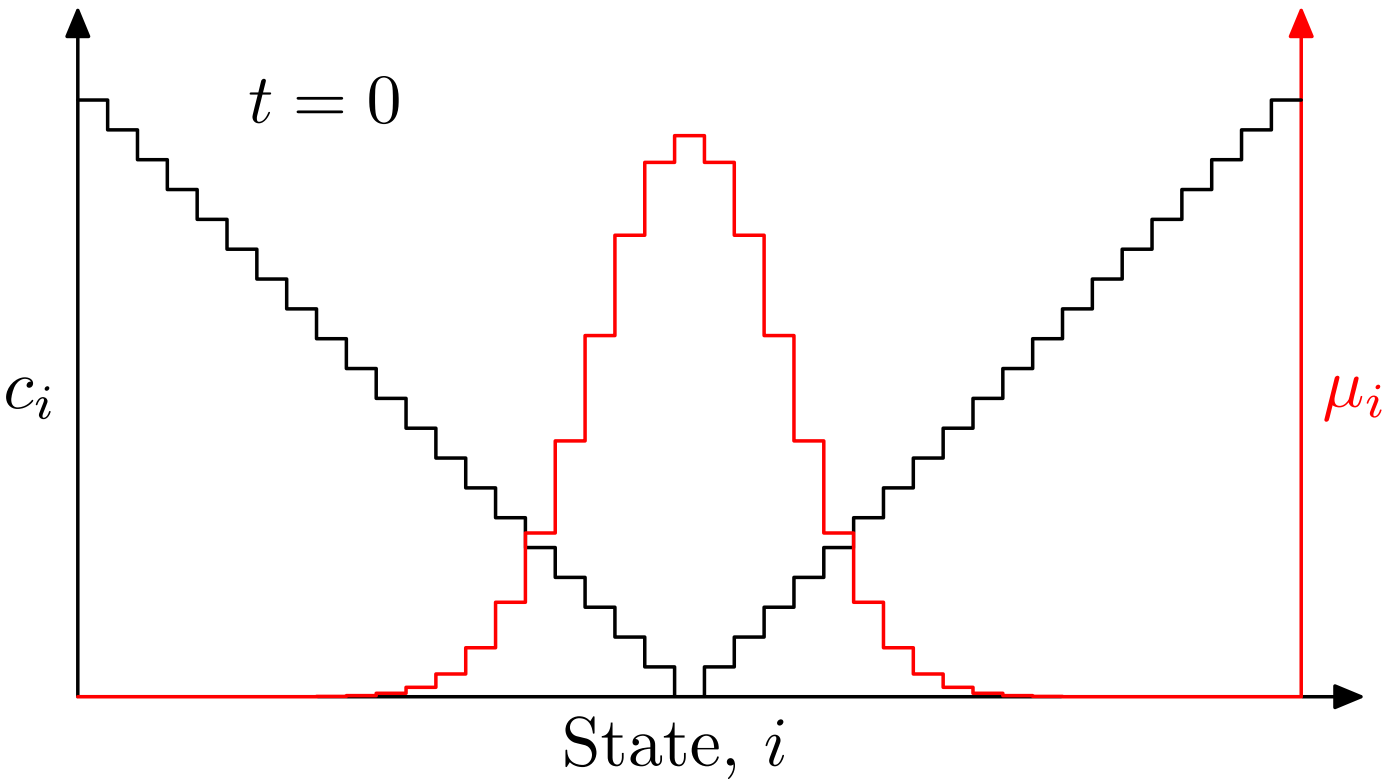
$$\mathbf{C}' = \mathbf{W}\mathbf{C}^s\mathbf{W}^T$$

- This is a very elegant way of computing the dynamics for a GA with population size 2

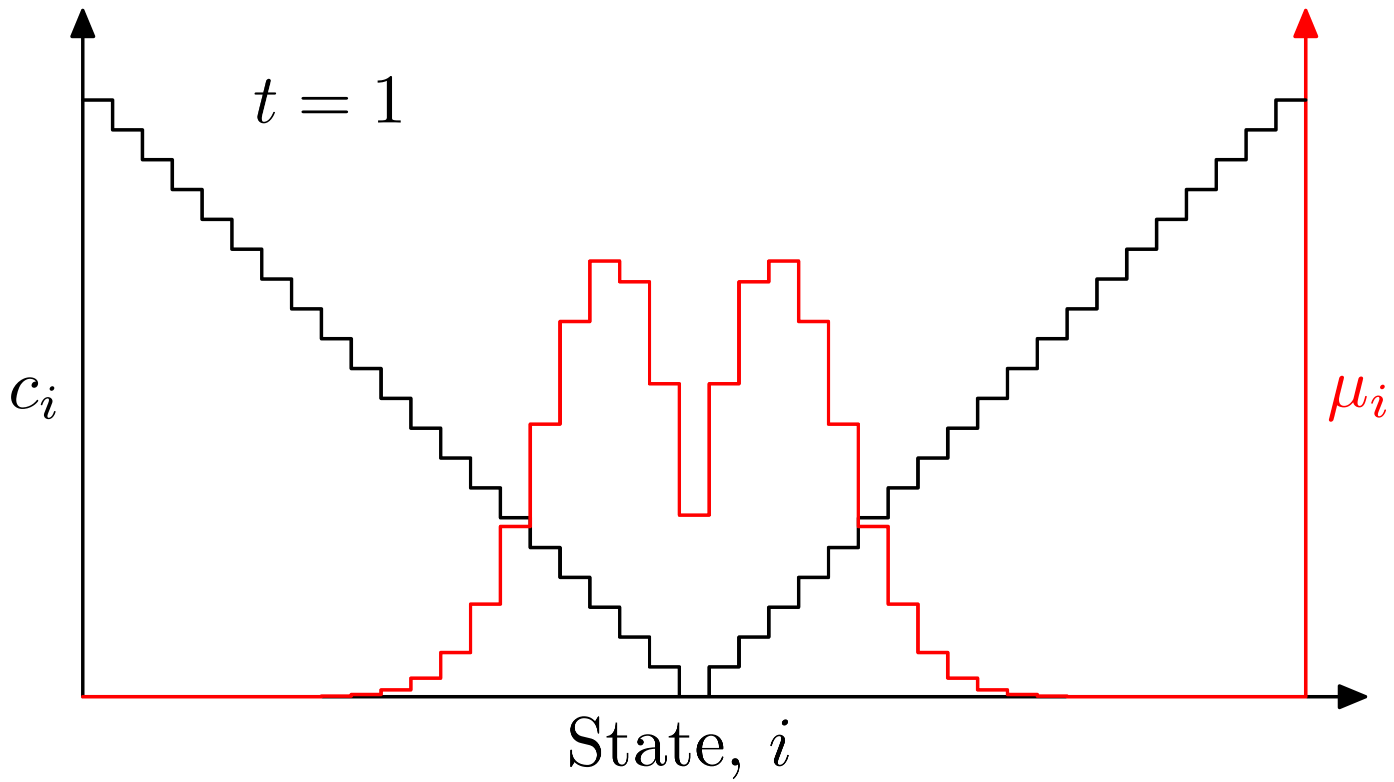
Hurdle Problem $P = 2$



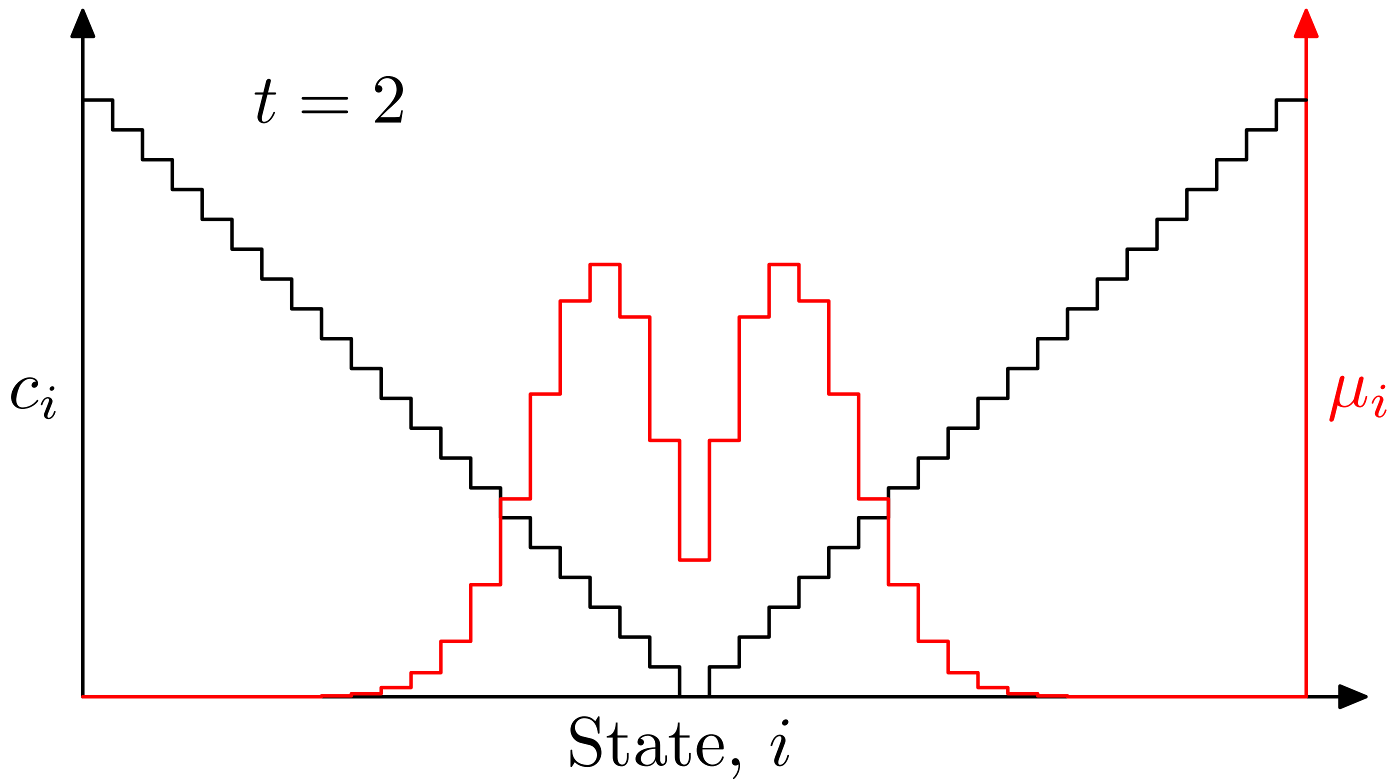
Symmetric Onemax $P = 2$



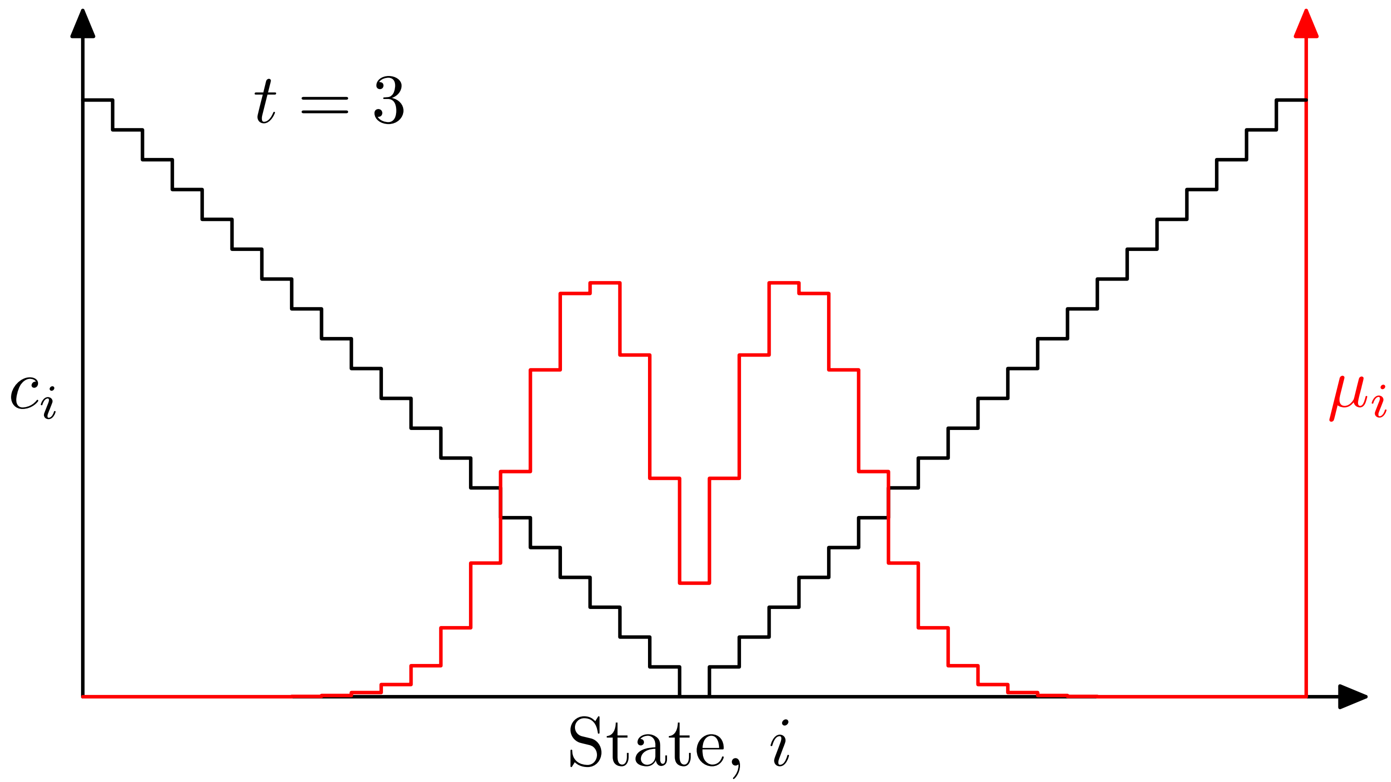
Symmetric Onemax $P = 2$



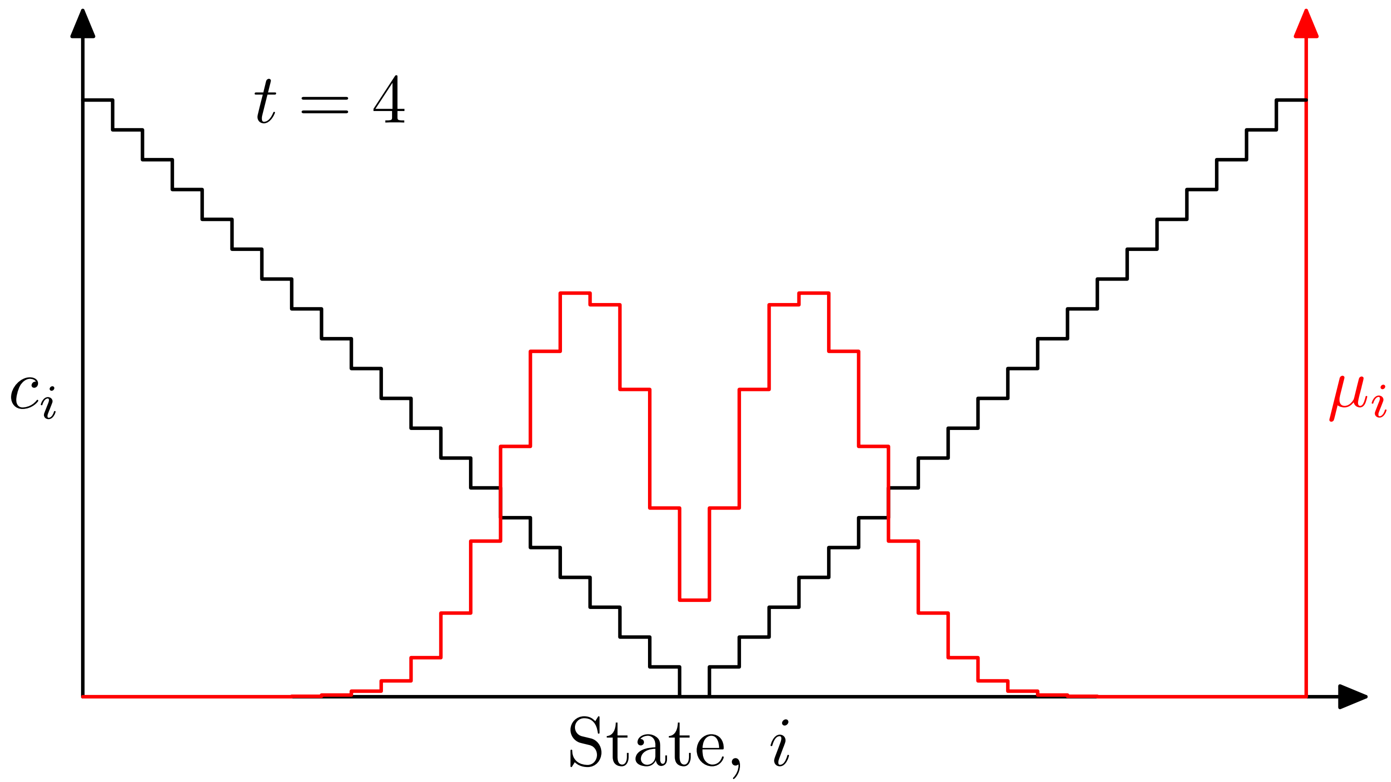
Symmetric Onemax $P = 2$



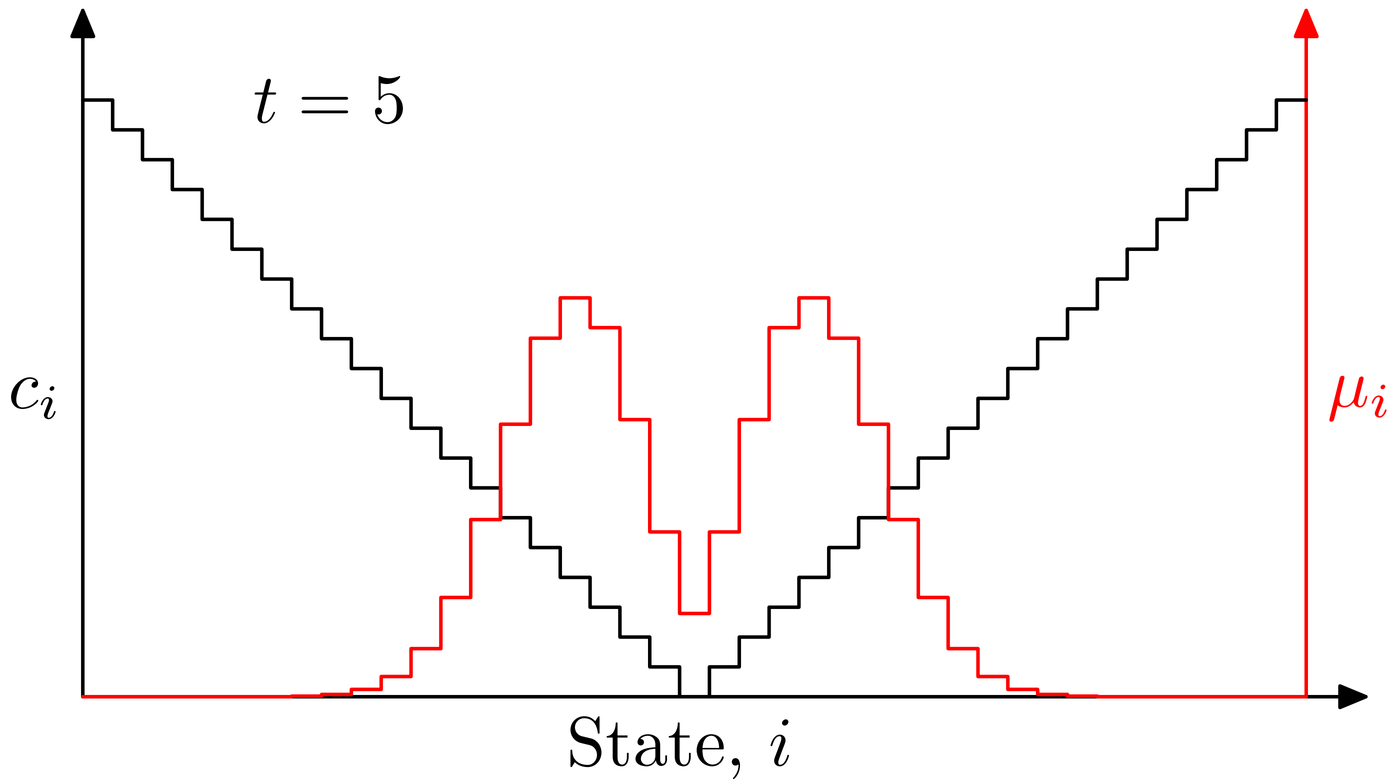
Symmetric Onemax $P = 2$



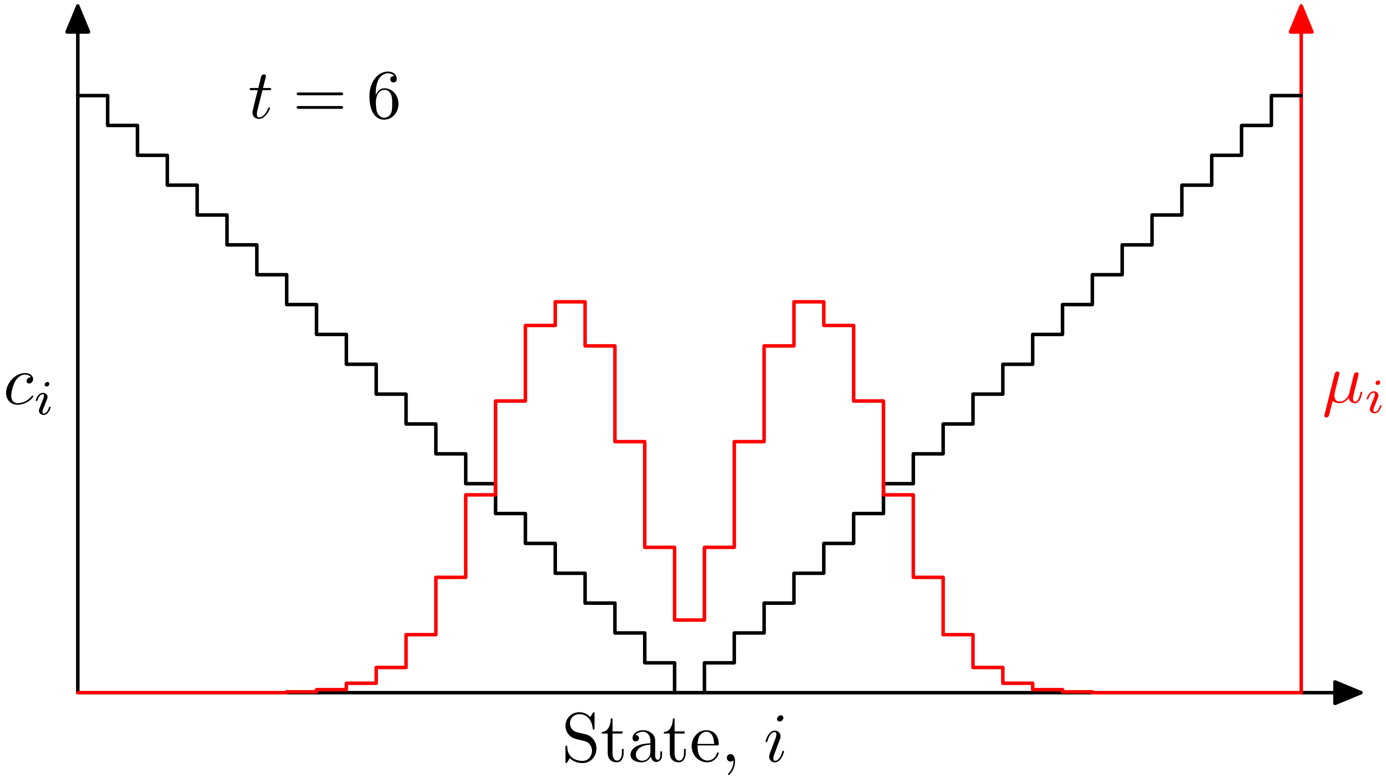
Symmetric Onemax $P = 2$



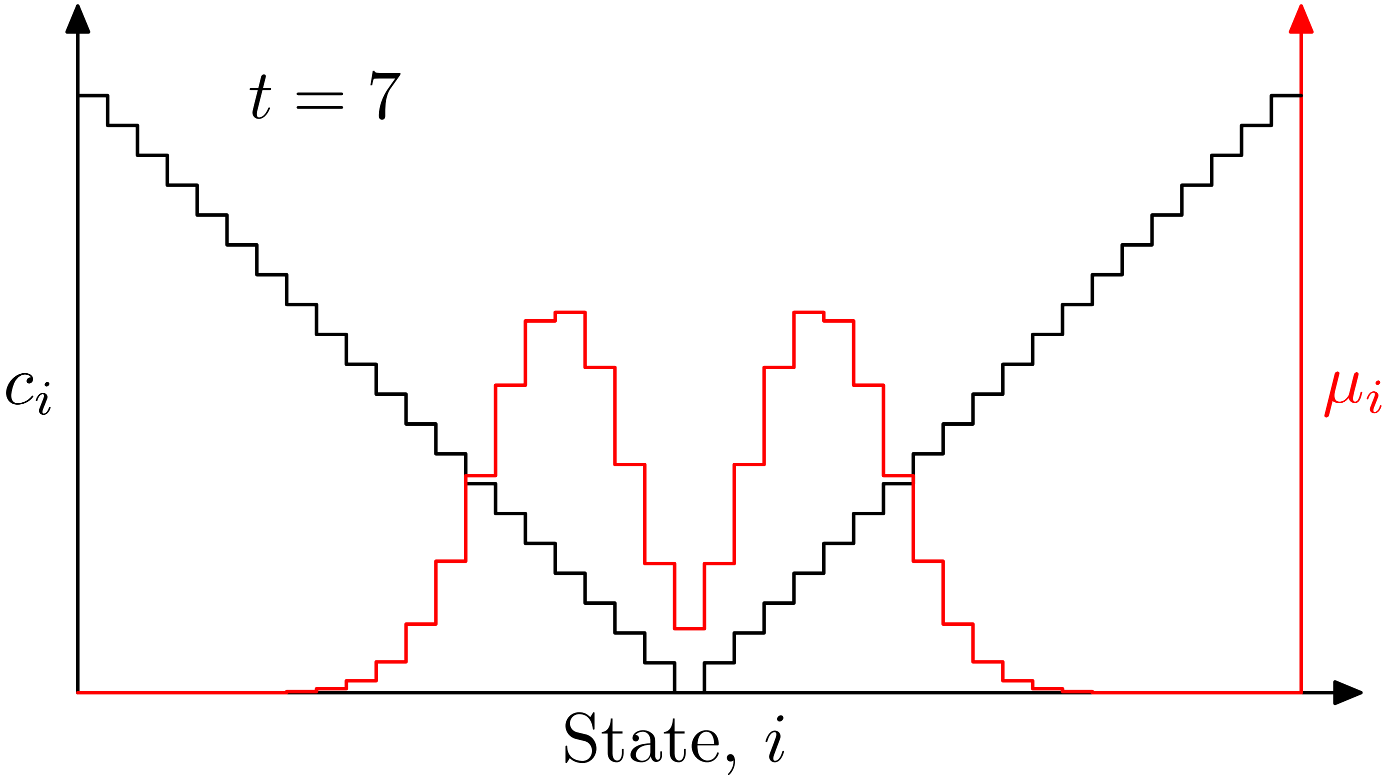
Symmetric Onemax $P = 2$



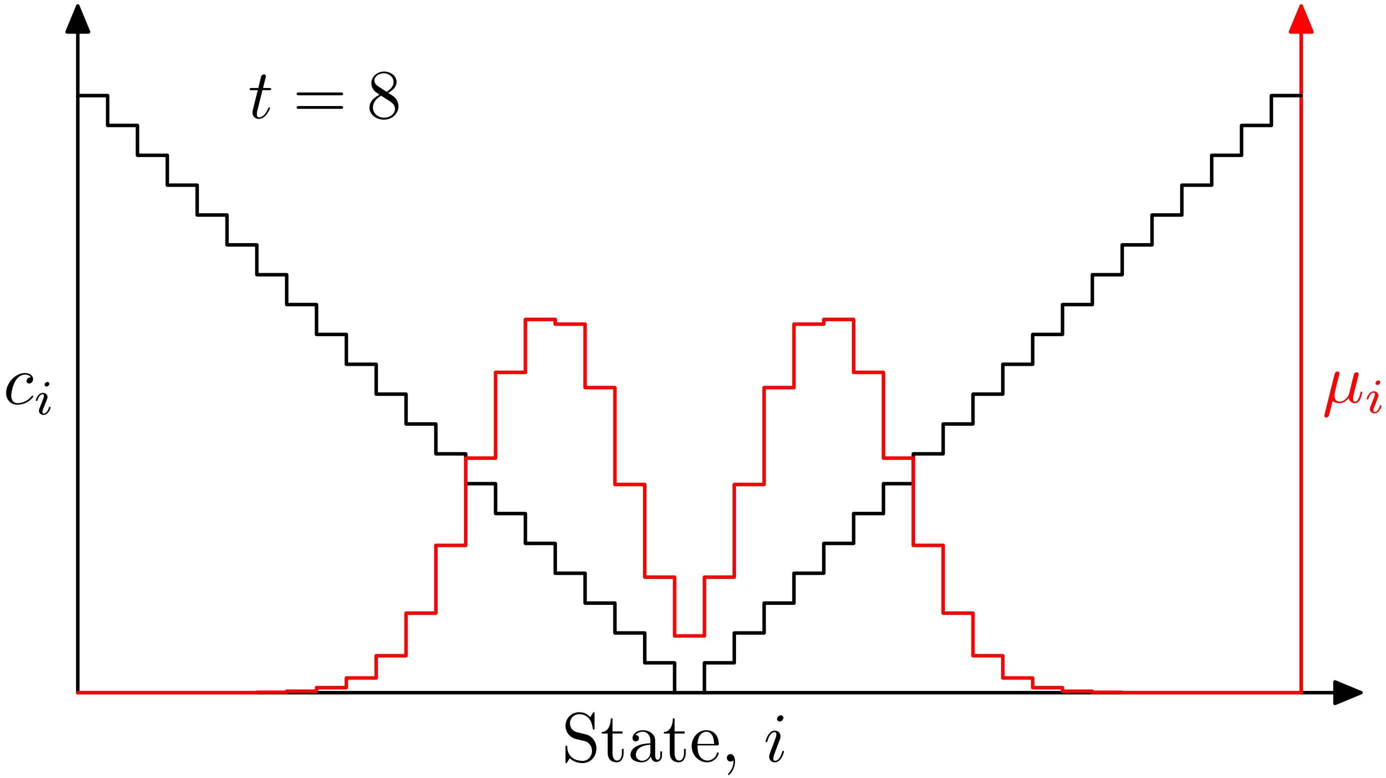
Symmetric Onemax $P = 2$



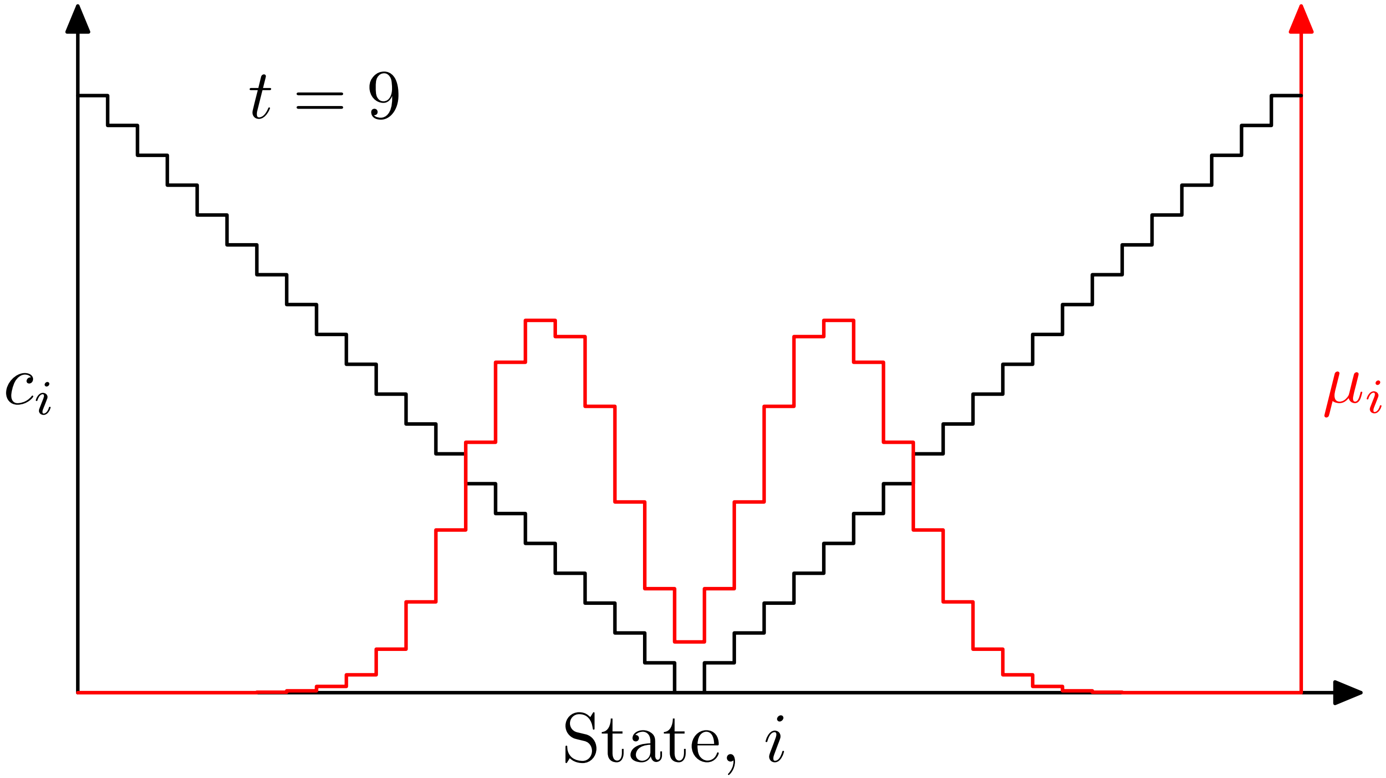
Symmetric Onemax $P = 2$



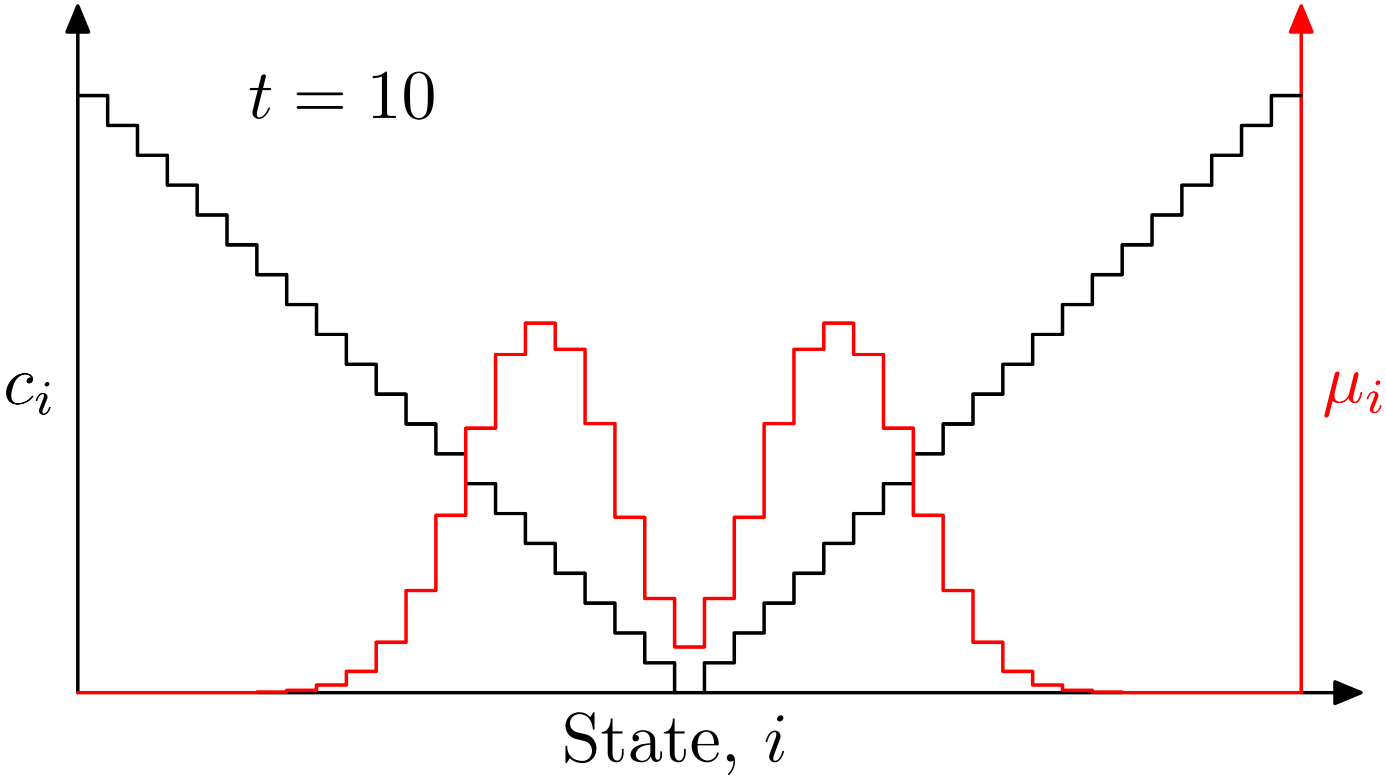
Symmetric Onemax $P = 2$



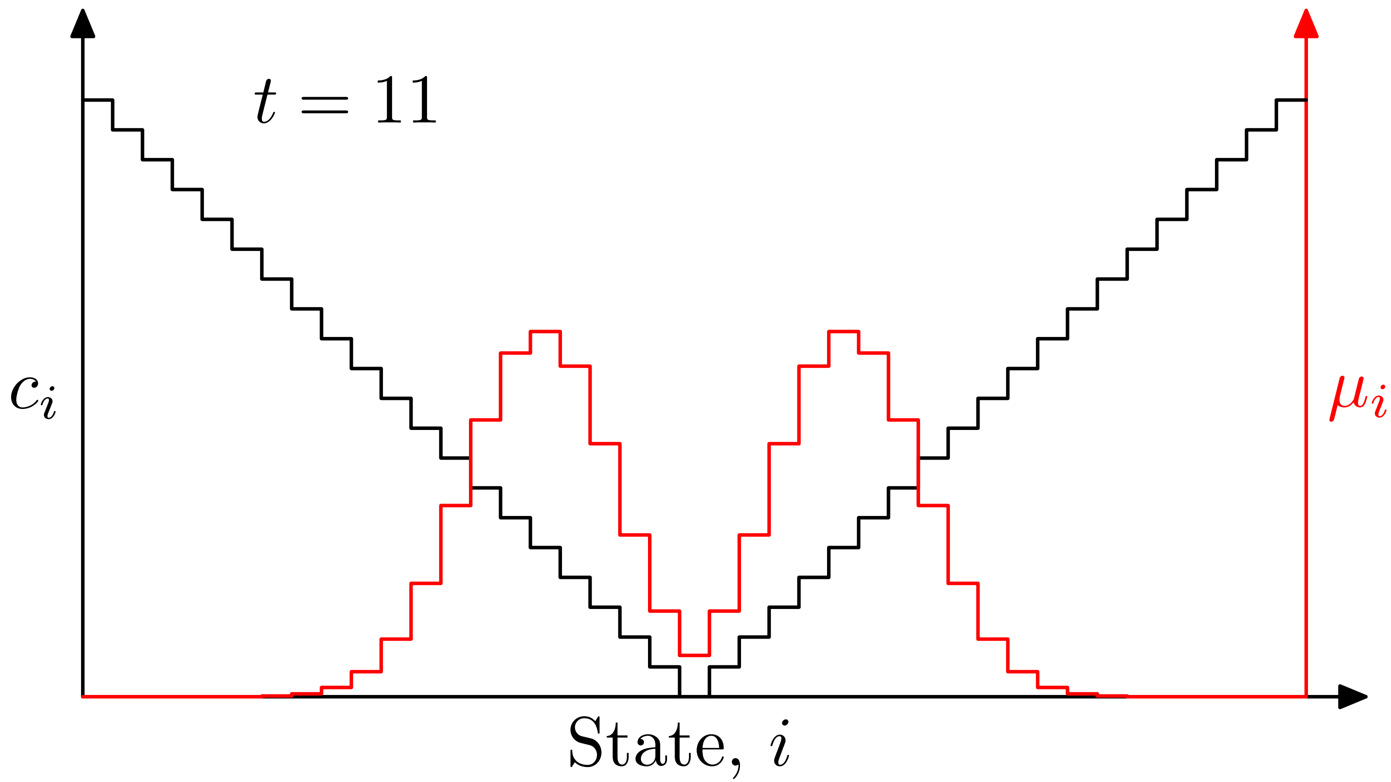
Symmetric Onemax $P = 2$



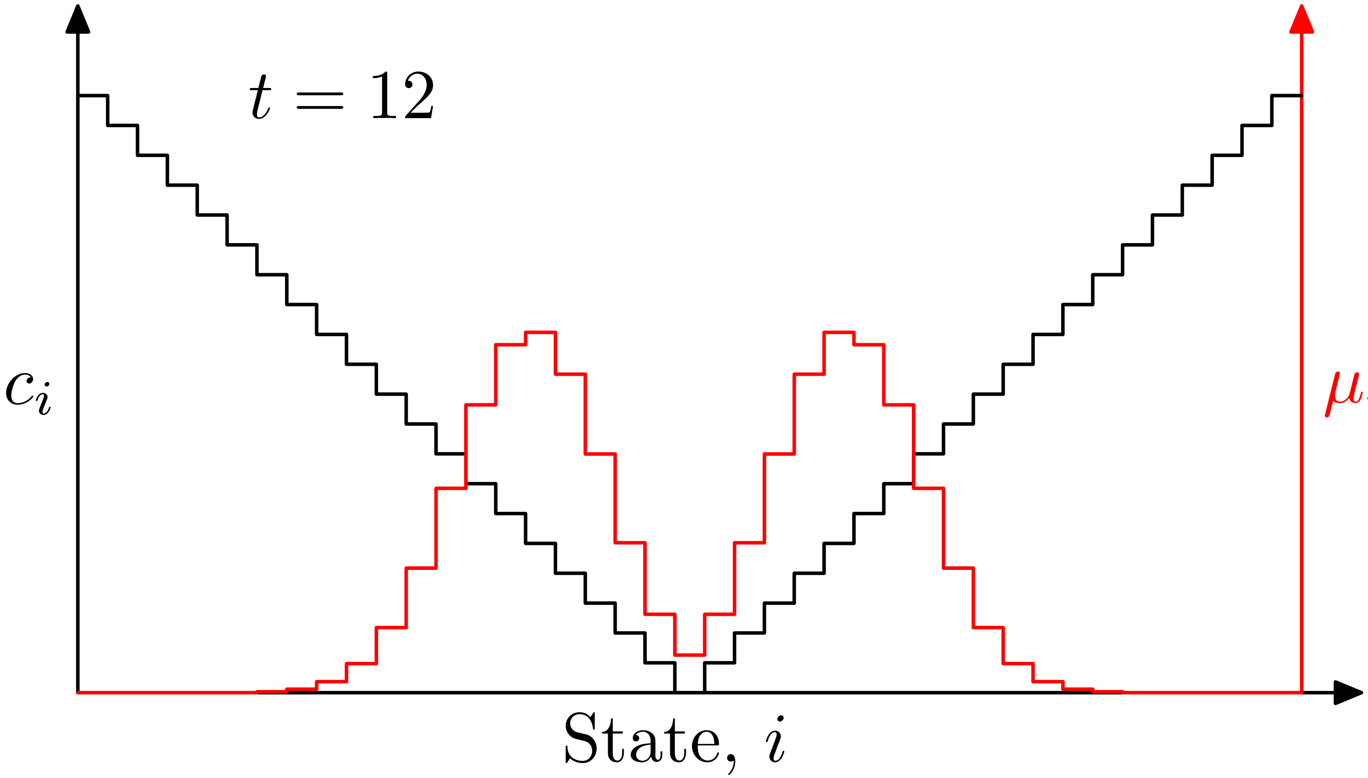
Symmetric Onemax $P = 2$



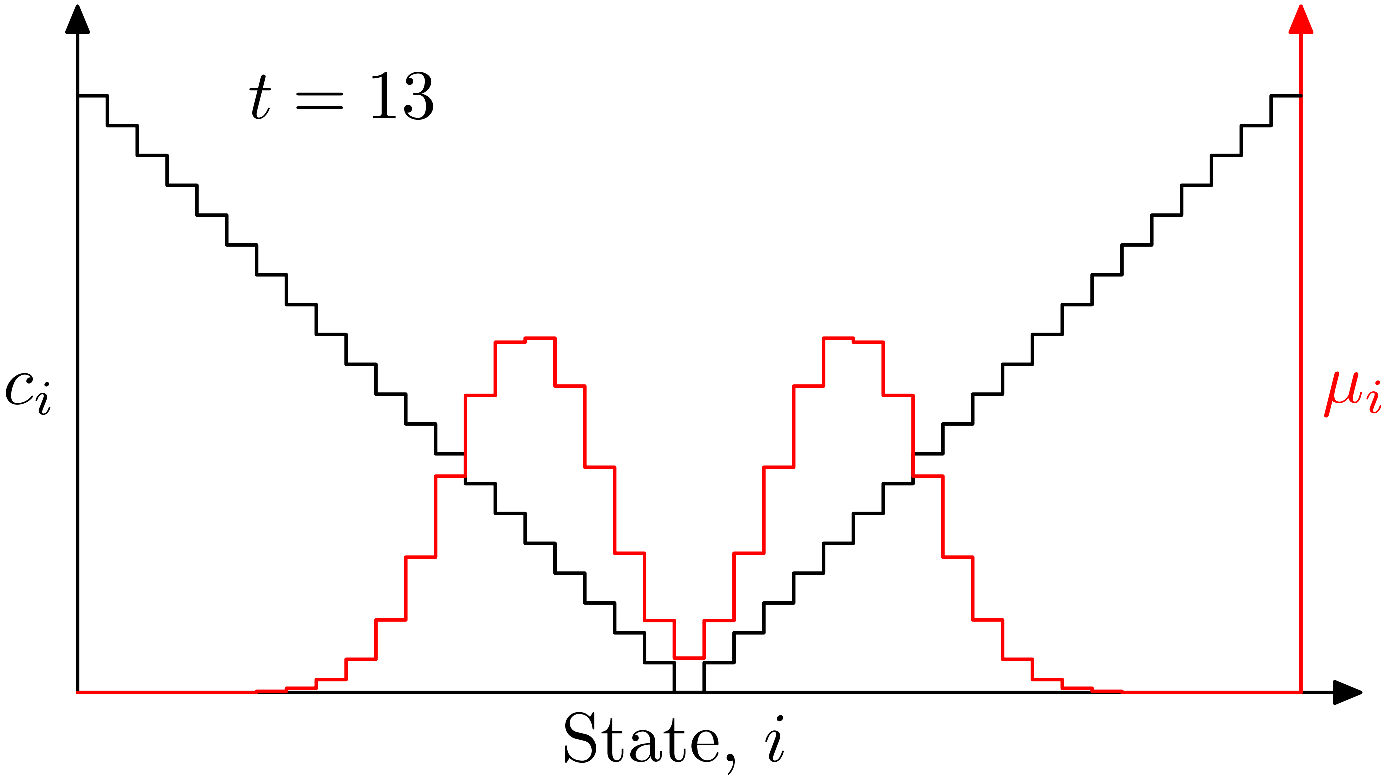
Symmetric Onemax $P = 2$



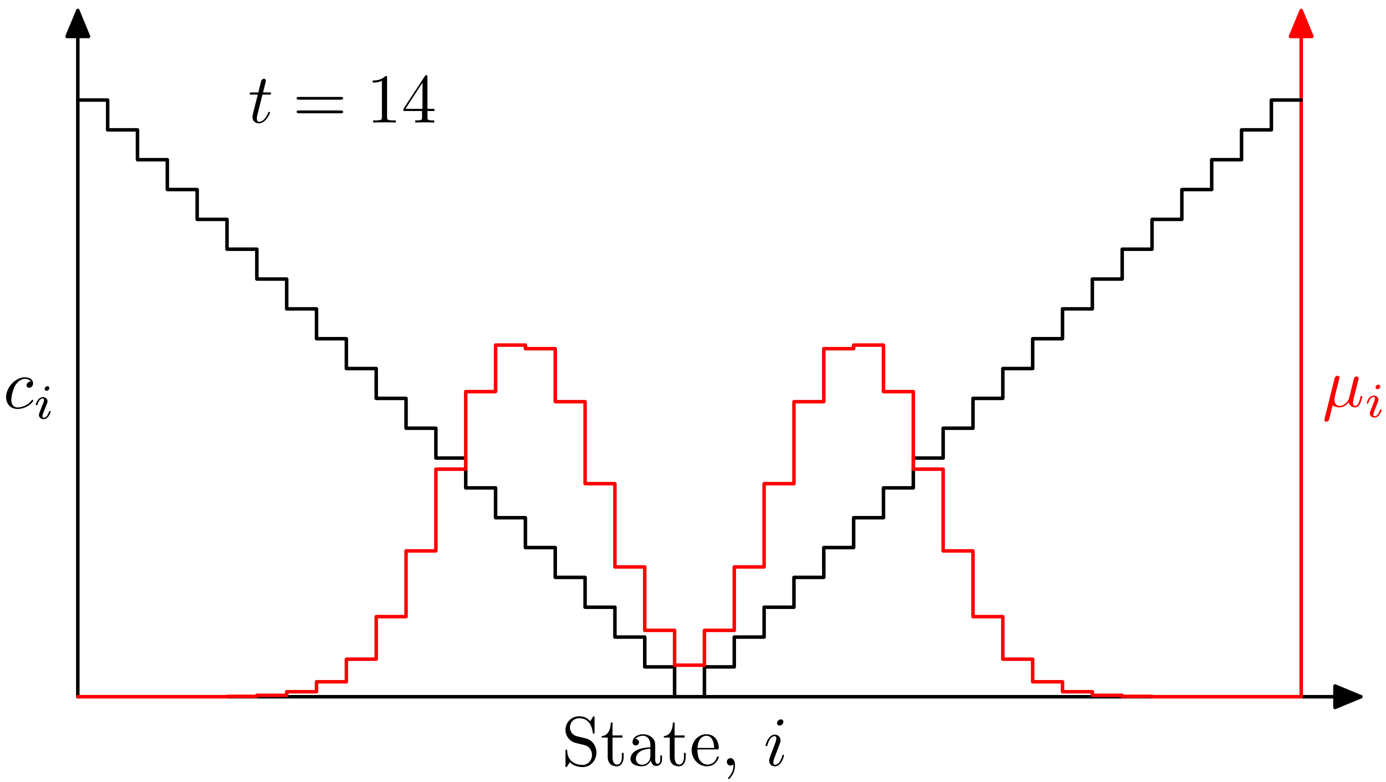
Symmetric Onemax $P = 2$



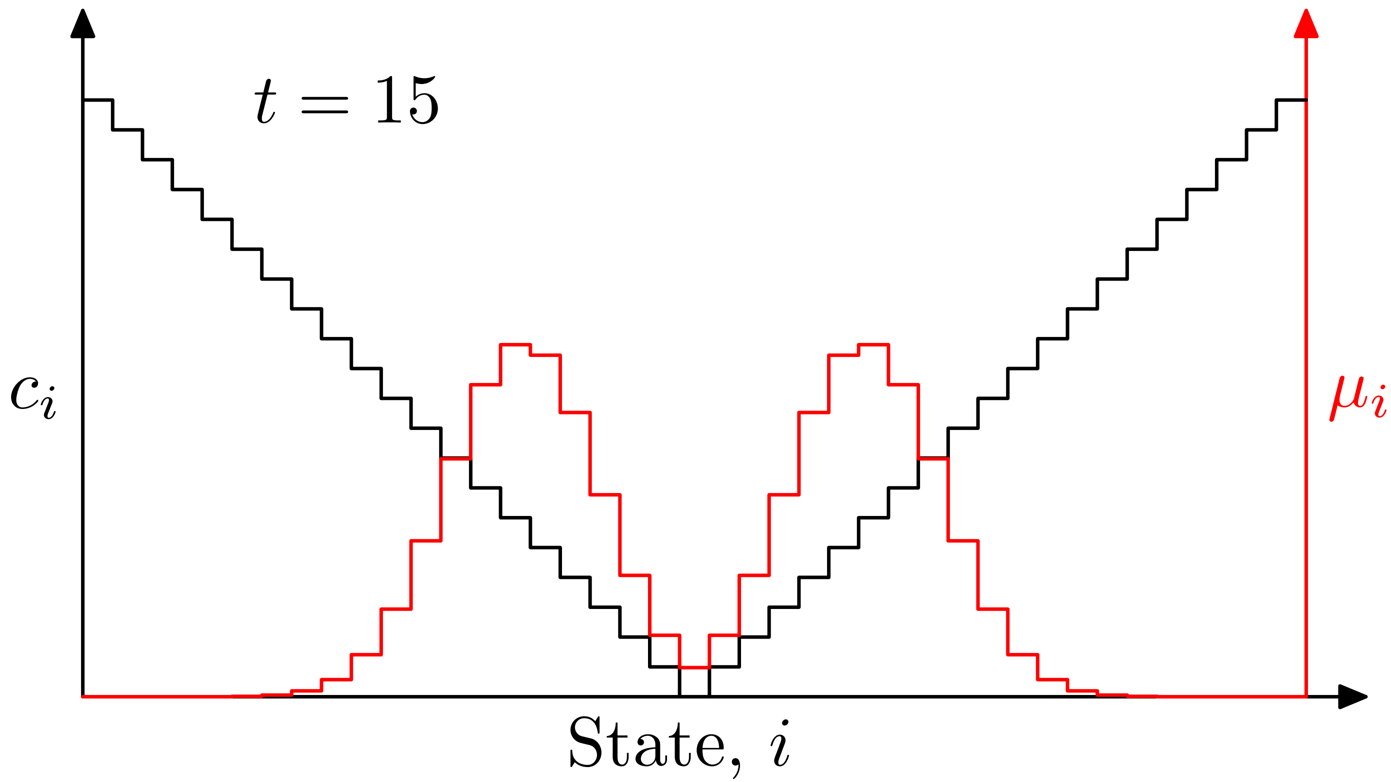
Symmetric Onemax $P = 2$



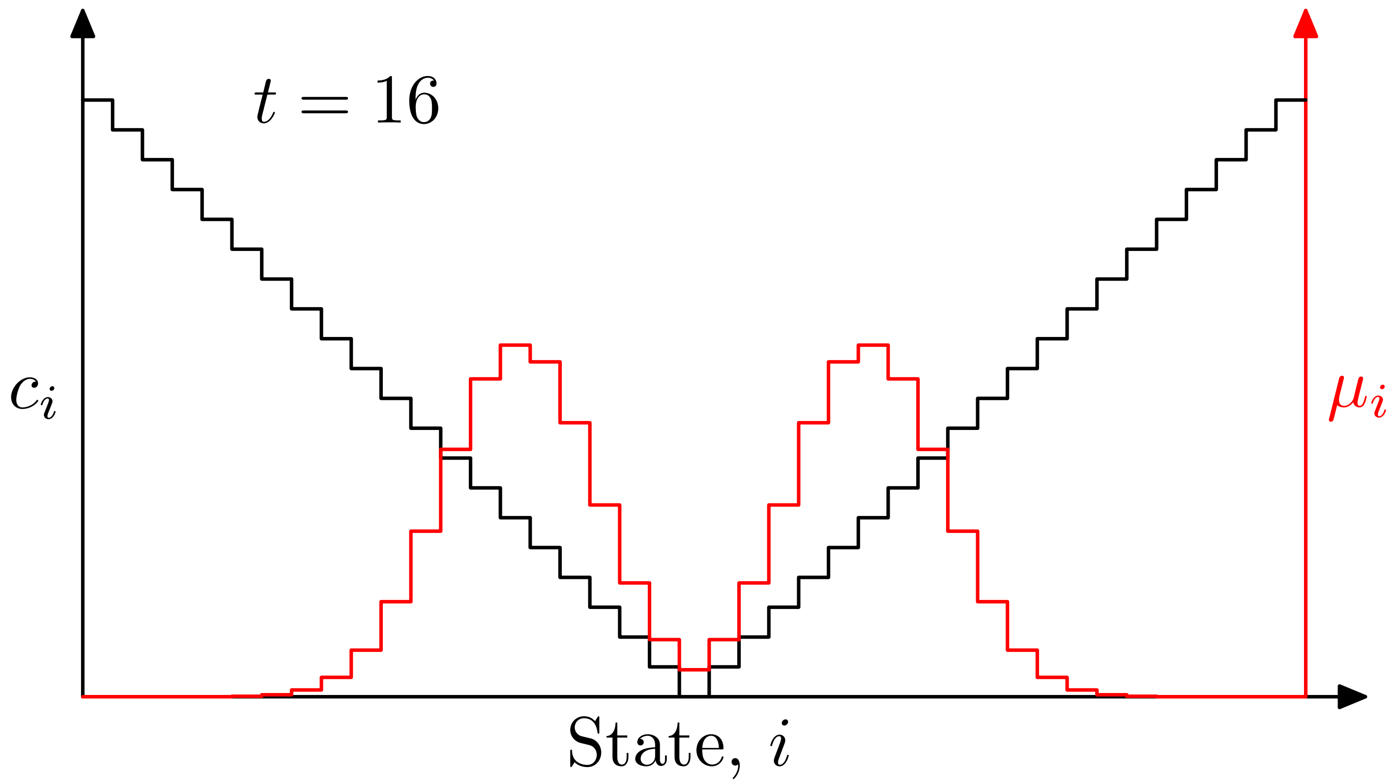
Symmetric Onemax $P = 2$



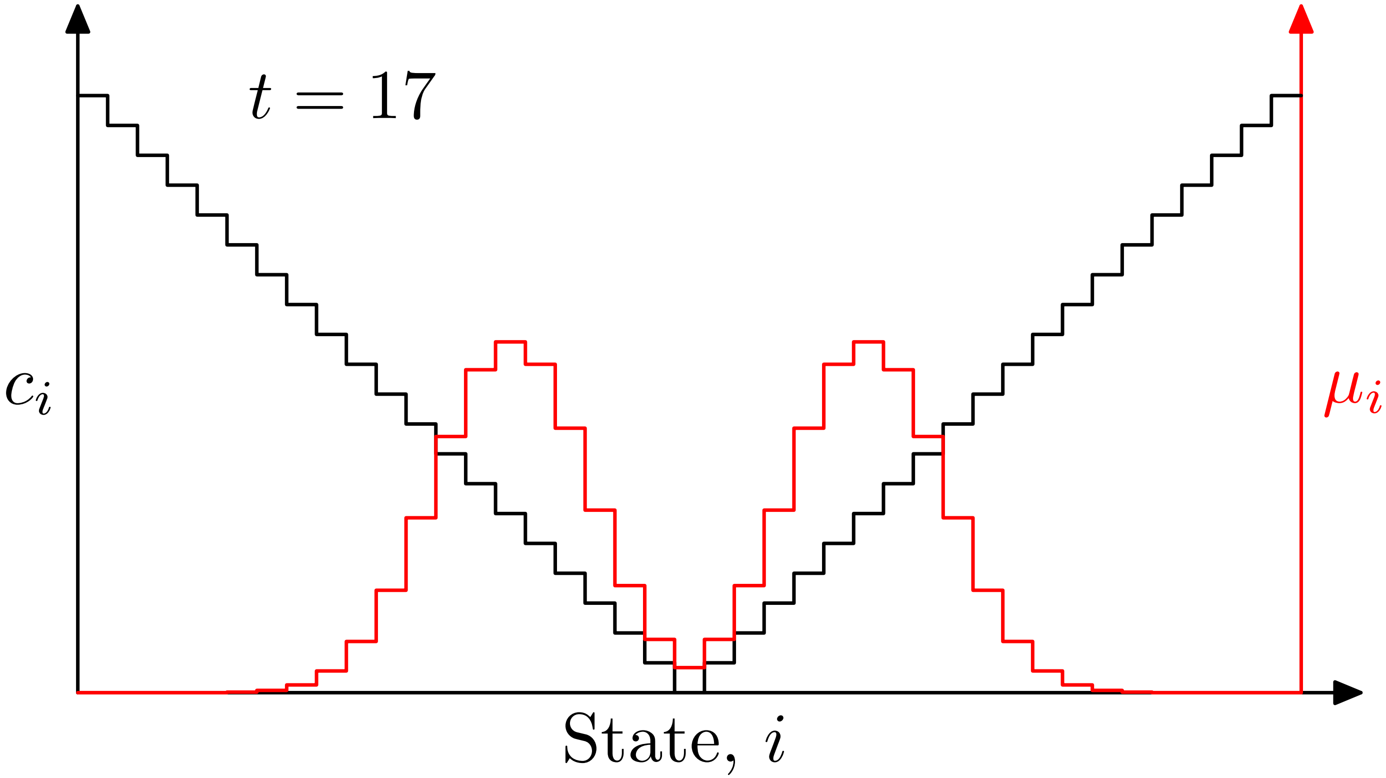
Symmetric Onemax $P = 2$



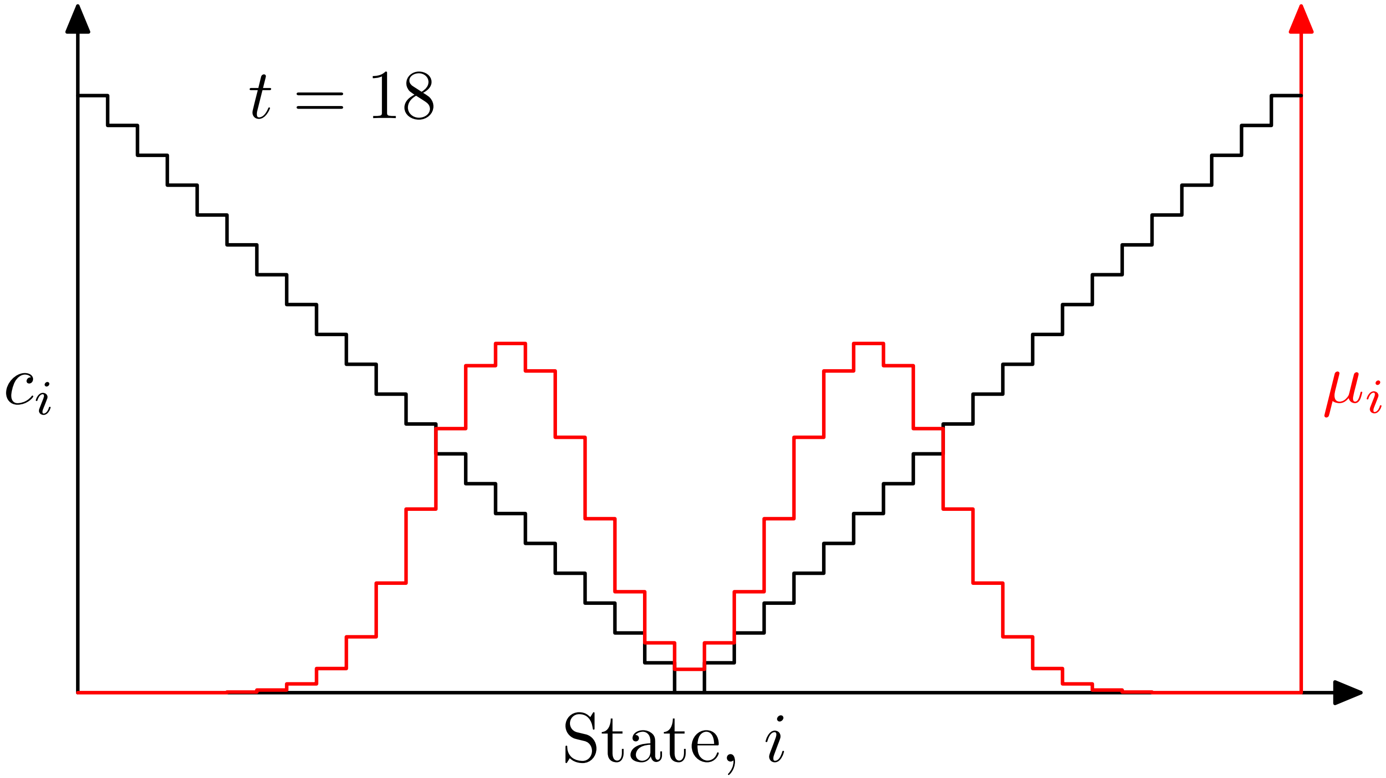
Symmetric Onemax $P = 2$



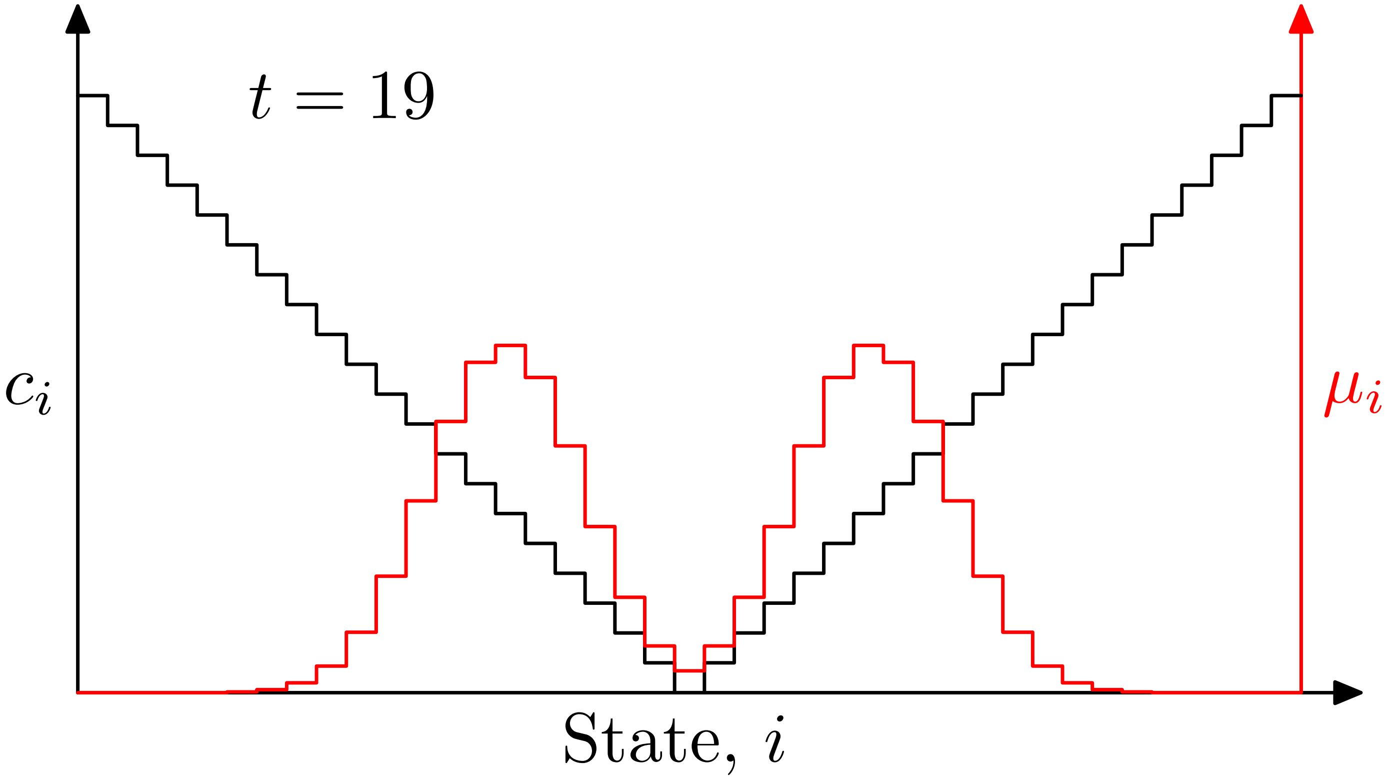
Symmetric Onemax $P = 2$



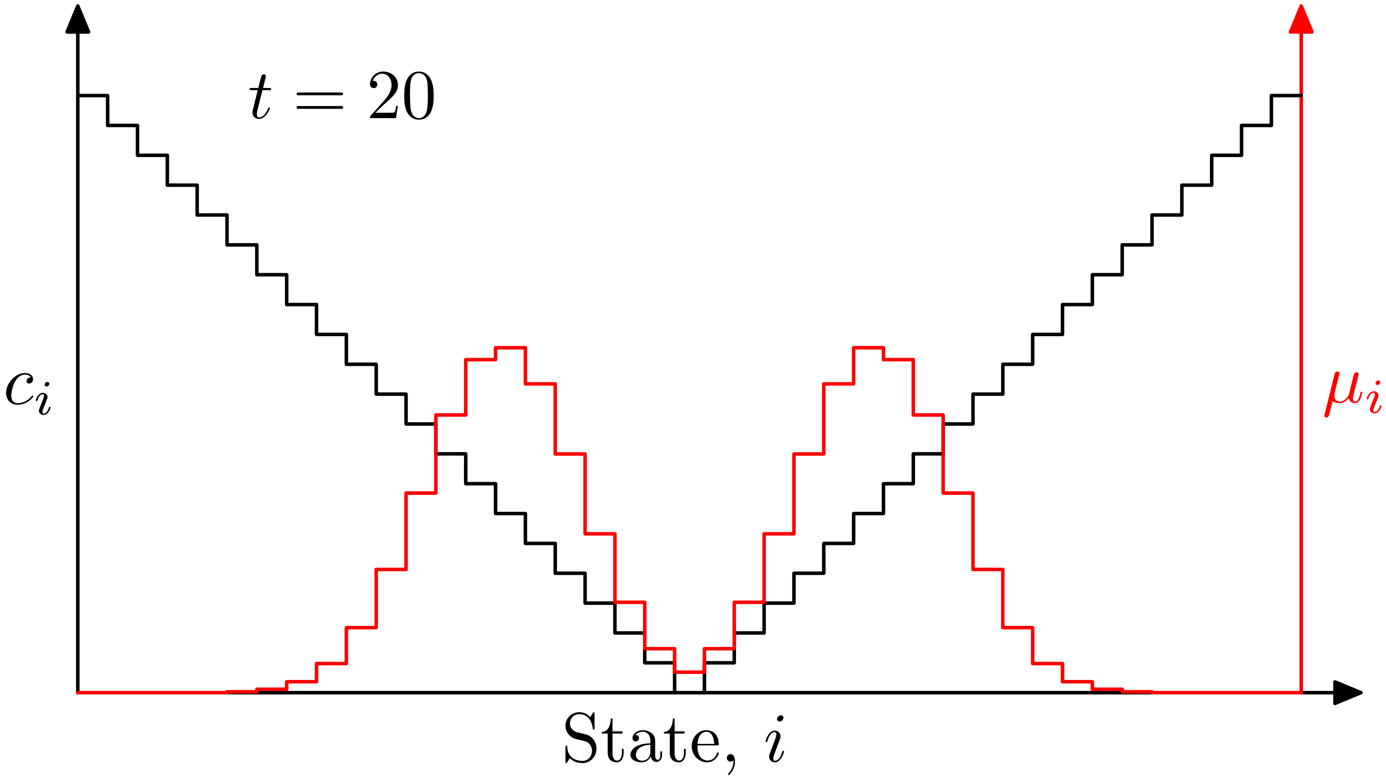
Symmetric Onemax $P = 2$



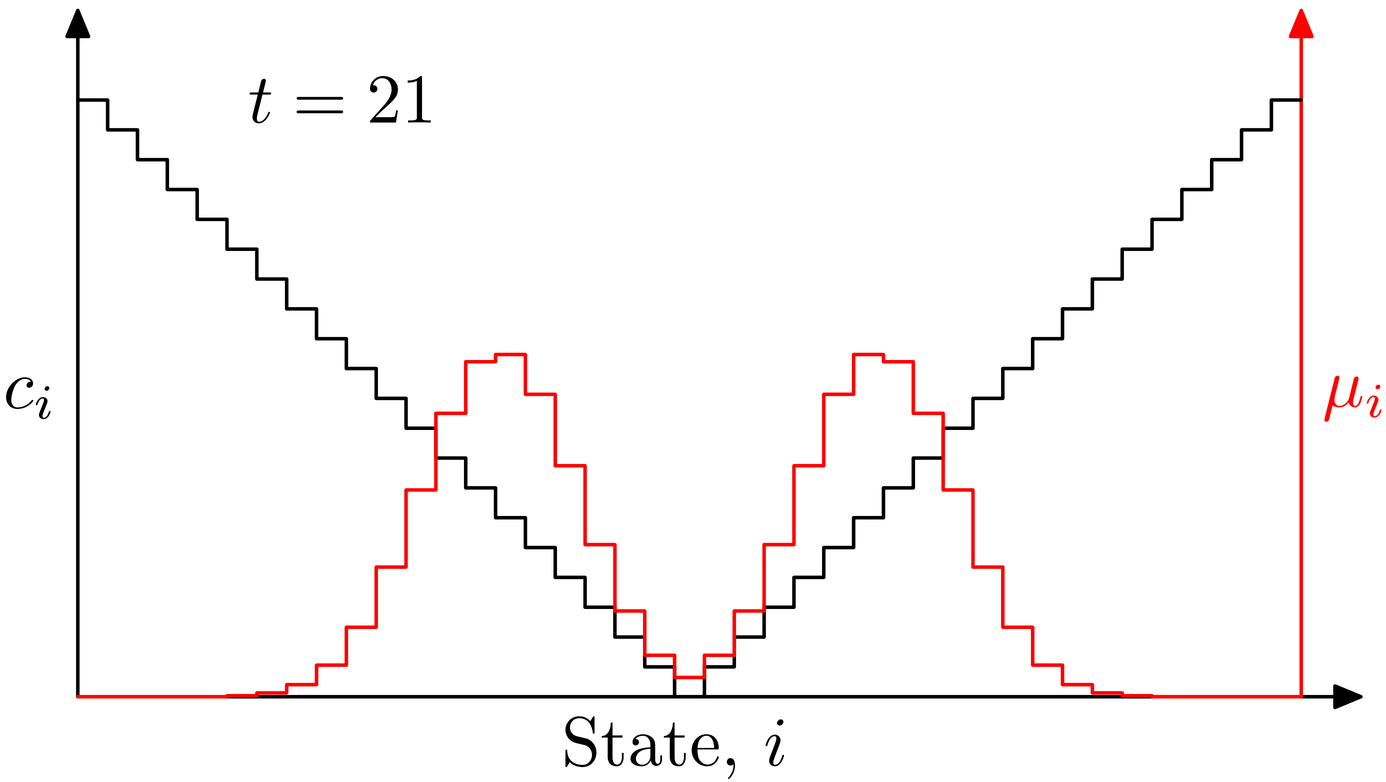
Symmetric Onemax $P = 2$



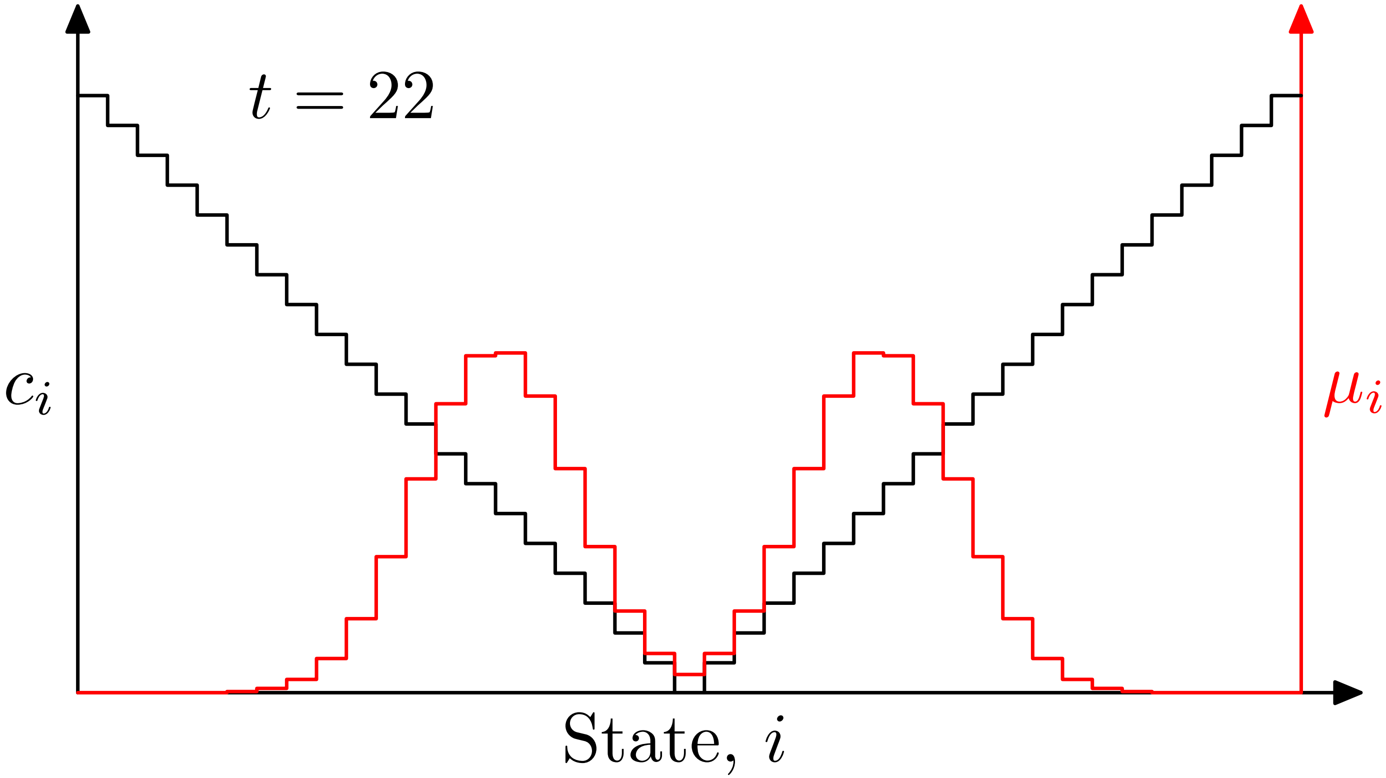
Symmetric Onemax $P = 2$



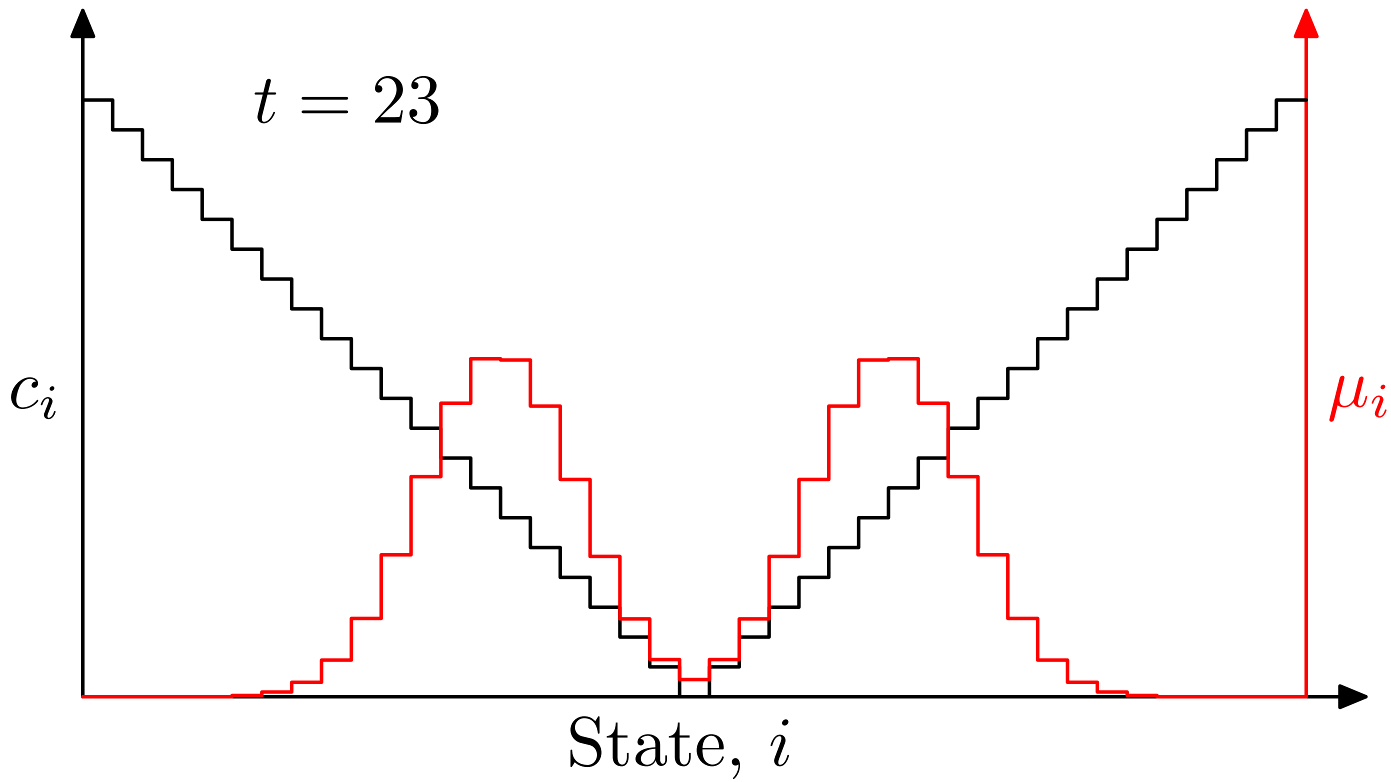
Symmetric Onemax $P = 2$



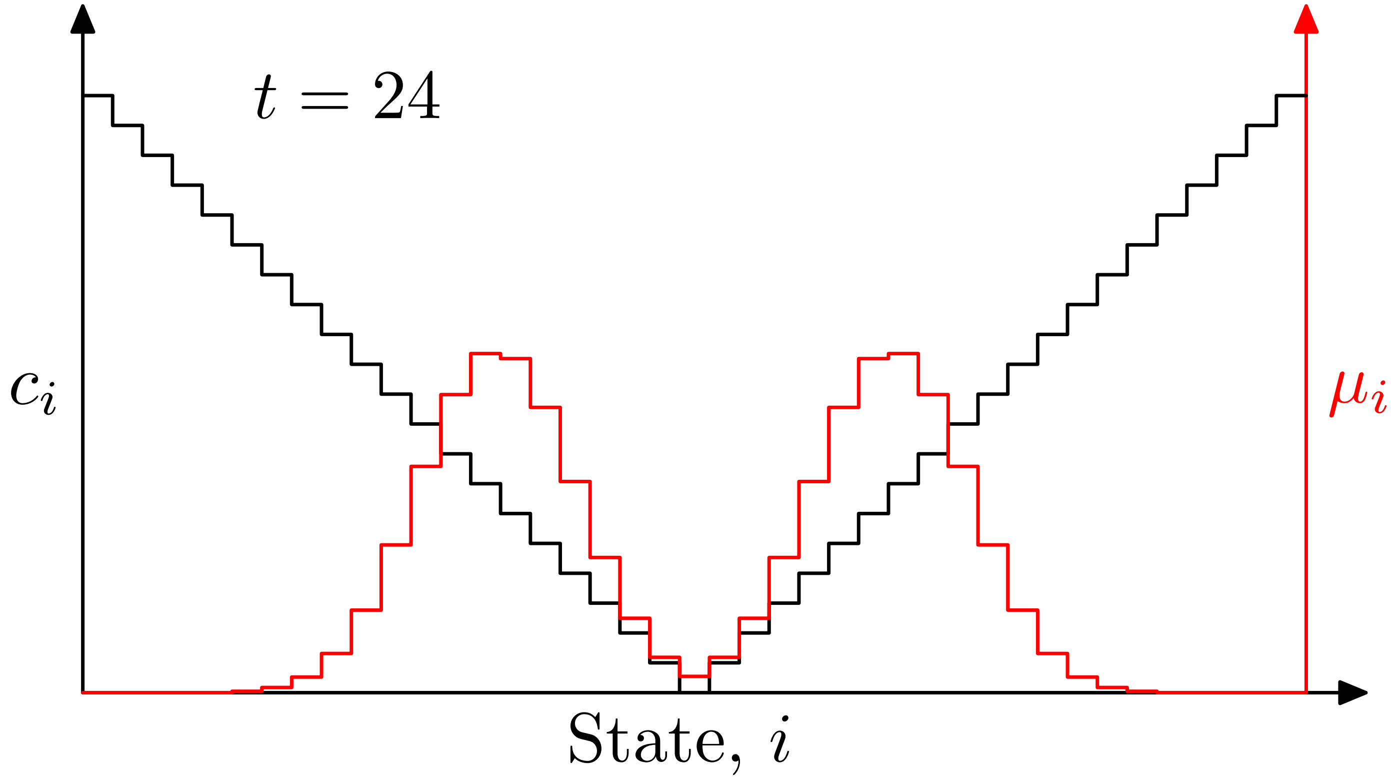
Symmetric Onemax $P = 2$



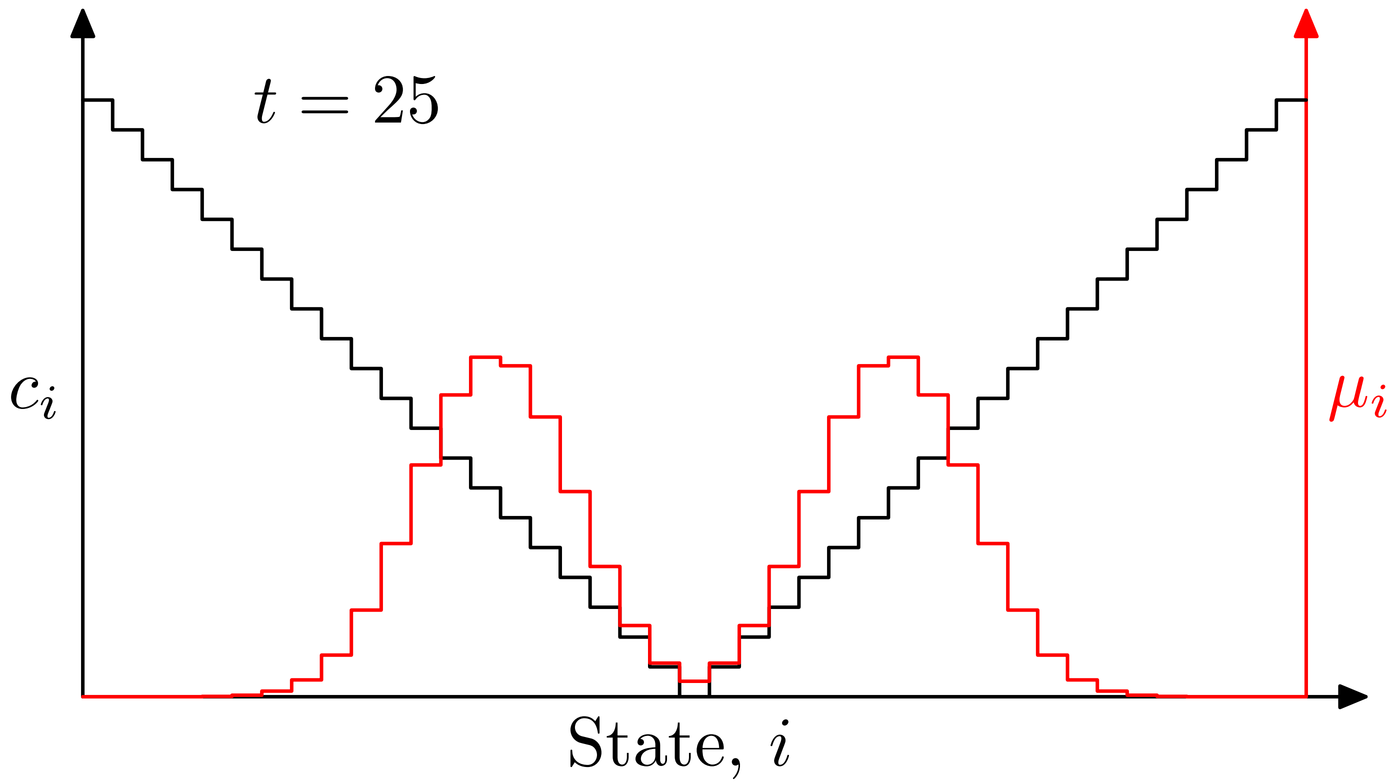
Symmetric Onemax $P = 2$



Symmetric Onemax $P = 2$



Symmetric Onemax $P = 2$



Approximating $g(\boldsymbol{S}, z)$

- For $p > 2$ we don't have enough information to compute $g(\boldsymbol{S}, z)$ exactly
- Many different approaches to approximating $g(\boldsymbol{S}, z)$
- E.g. can use stochastic expansion
- Can assume

$$f_{\boldsymbol{X}}(\boldsymbol{x}, t) = \frac{1}{\sqrt{|2\pi \mathbf{C}|}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

- Correct mean and covariance, but clearly wrong

Approximating $g(\boldsymbol{S}, z)$

- For $p > 2$ we don't have enough information to compute $g(\boldsymbol{S}, z)$ exactly
- Many different approaches to approximating $g(\boldsymbol{S}, z)$
- E.g. can use stochastic expansion
- Can assume

$$f_{\boldsymbol{X}}(\boldsymbol{x}, t) = \frac{1}{\sqrt{|2\pi\mathbf{C}|}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

- Correct mean and covariance, but clearly wrong

Approximating $g(\boldsymbol{S}, z)$

- For $p > 2$ we don't have enough information to compute $g(\boldsymbol{S}, z)$ exactly
- Many different approaches to approximating $g(\boldsymbol{S}, z)$
- E.g. can use stochastic expansion
- Can assume

$$f_{\boldsymbol{X}}(\boldsymbol{x}, t) = \frac{1}{\sqrt{|2\pi \mathbf{C}|}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$$

- Correct mean and covariance, but clearly wrong

Approximating $g(\boldsymbol{S}, z)$

- For $p > 2$ we don't have enough information to compute $g(\boldsymbol{S}, z)$ exactly
- Many different approaches to approximating $g(\boldsymbol{S}, z)$
- E.g. can use stochastic expansion
- Can assume

$$f_{\boldsymbol{X}}(\boldsymbol{x}, t) = \frac{1}{\sqrt{|2\pi \mathbf{C}|}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

- Correct mean and covariance, but clearly wrong

Results

- The approximation usually **breaks down!**
- We discuss why latter
- Using the exact sampling approach gives slightly more accurate model than stochastic expansion
- Finding a robust approximation is still unsolved—although I will talk about a possible approach to cure this
- In problems below I use a selection pressure

$$s_i = \mathrm{e}^{\beta c_i}$$

with $\beta \approx 0.1$

Results

- The approximation usually **breaks down!**
- We discuss why latter
- Using the exact sampling approach gives slightly more accurate model than stochastic expansion
- Finding a robust approximation is still unsolved—although I will talk about a possible approach to cure this
- In problems below I use a selection pressure

$$s_i = \mathrm{e}^{\beta c_i}$$

with $\beta \approx 0.1$

Results

- The approximation usually **breaks down!**
- We discuss why latter
- Using the exact sampling approach gives slightly more accurate model than stochastic expansion
- Finding a robust approximation is still unsolved—although I will talk about a possible approach to cure this
- In problems below I use a selection pressure

$$s_i = \mathrm{e}^{\beta c_i}$$

with $\beta \approx 0.1$

Results

- The approximation usually **breaks down!**
- We discuss why latter
- Using the exact sampling approach gives slightly more accurate model than stochastic expansion
- Finding a robust approximation is still unsolved—although I will talk about a possible approach to cure this
- In problems below I use a selection pressure

$$s_i = \mathrm{e}^{\beta c_i}$$

with $\beta \approx 0.1$

Results

- The approximation usually **breaks down!**
- We discuss why latter
- Using the exact sampling approach gives slightly more accurate model than stochastic expansion
- Finding a robust approximation is still unsolved—although I will talk about a possible approach to cure this
- In problems below I use a selection pressure

$$s_i = \mathrm{e}^{\beta c_i}$$

with $\beta \approx 0.1$

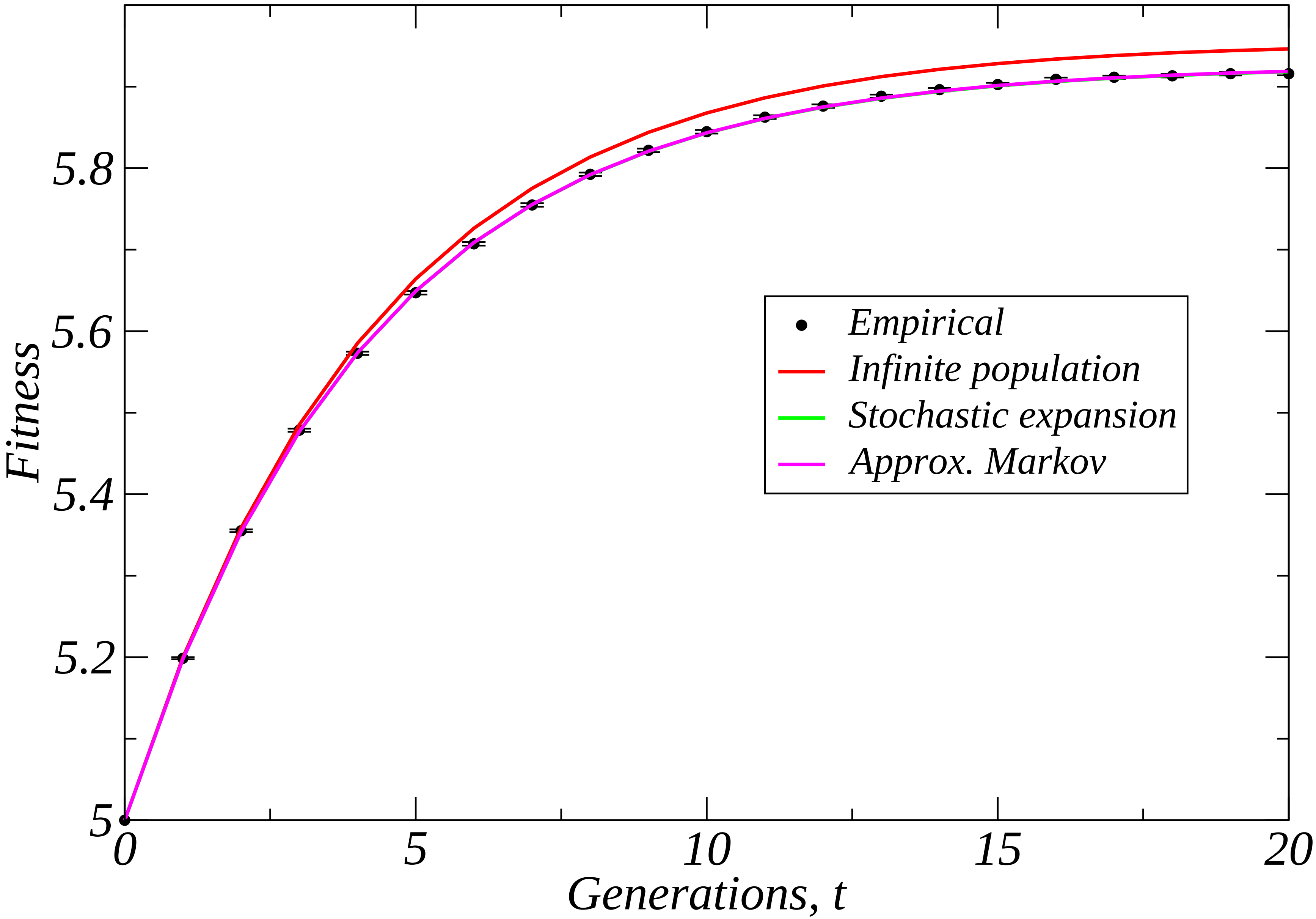
Results

- The approximation usually **breaks down!**
- We discuss why latter
- Using the exact sampling approach gives slightly more accurate model than stochastic expansion
- Finding a robust approximation is still unsolved—although I will talk about a possible approach to cure this
- In problems below I use a selection pressure

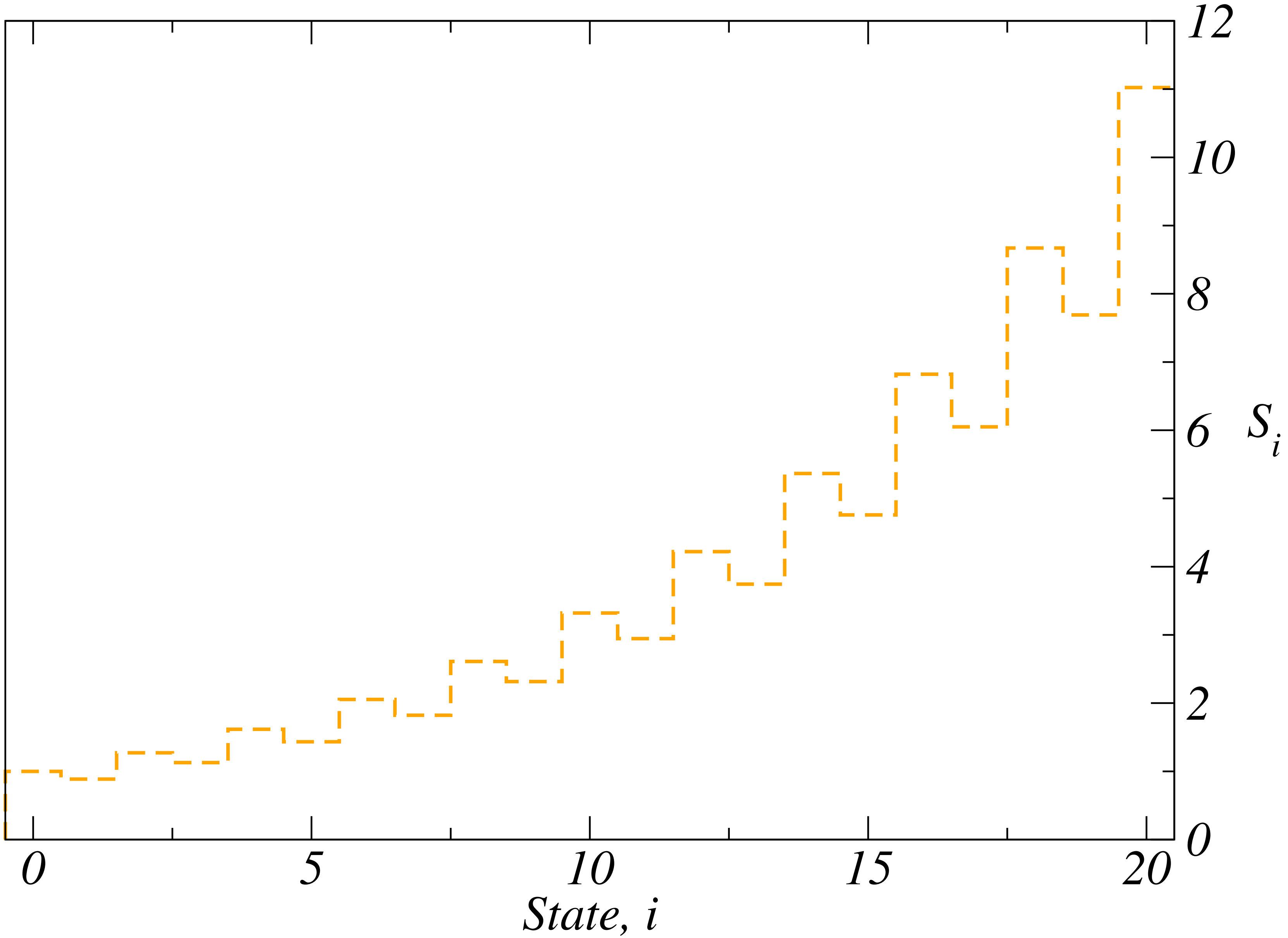
$$s_i = \mathrm{e}^{\beta c_i}$$

with $\beta \approx 0.1$

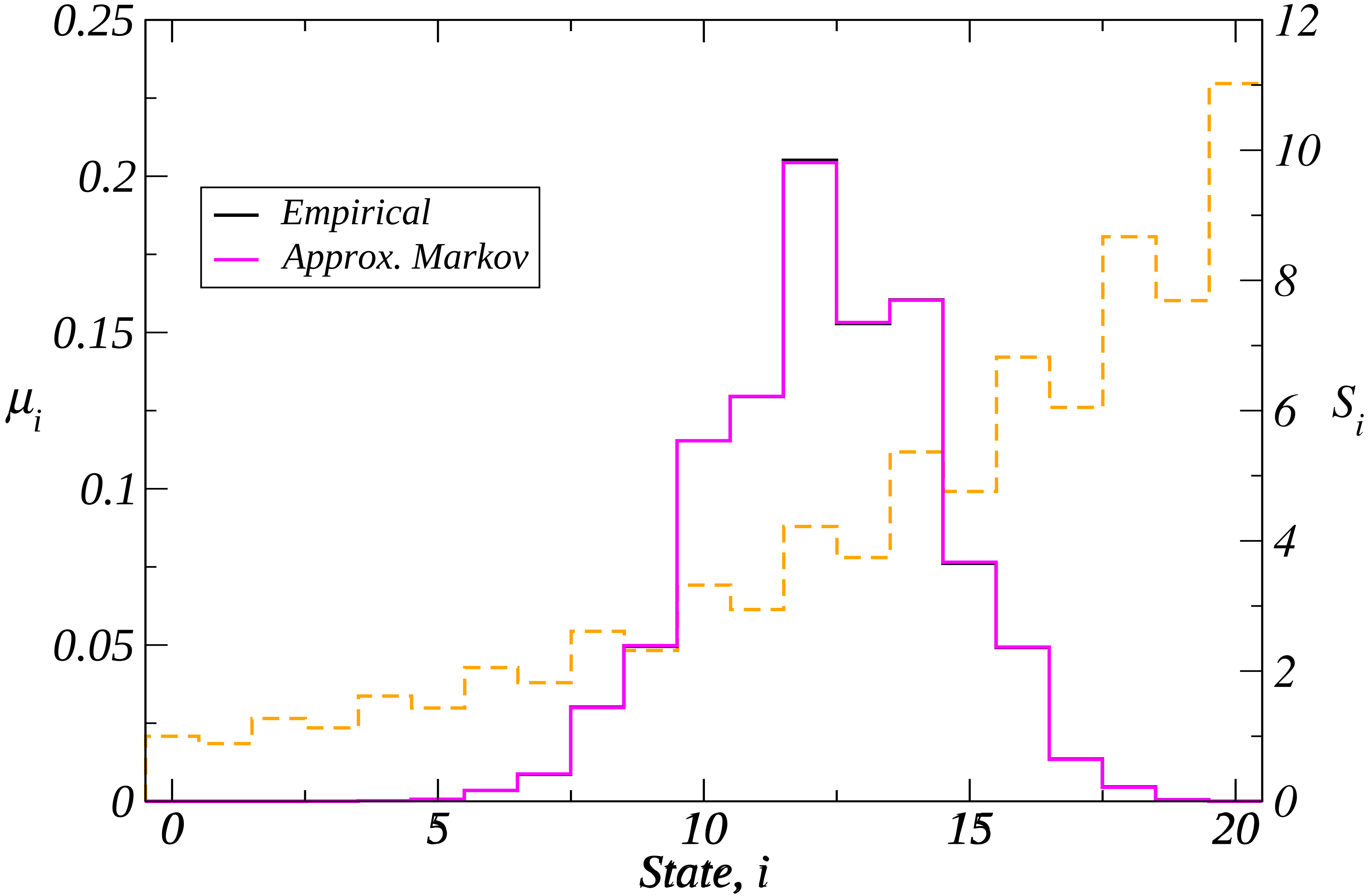
Ones-max Dynamics



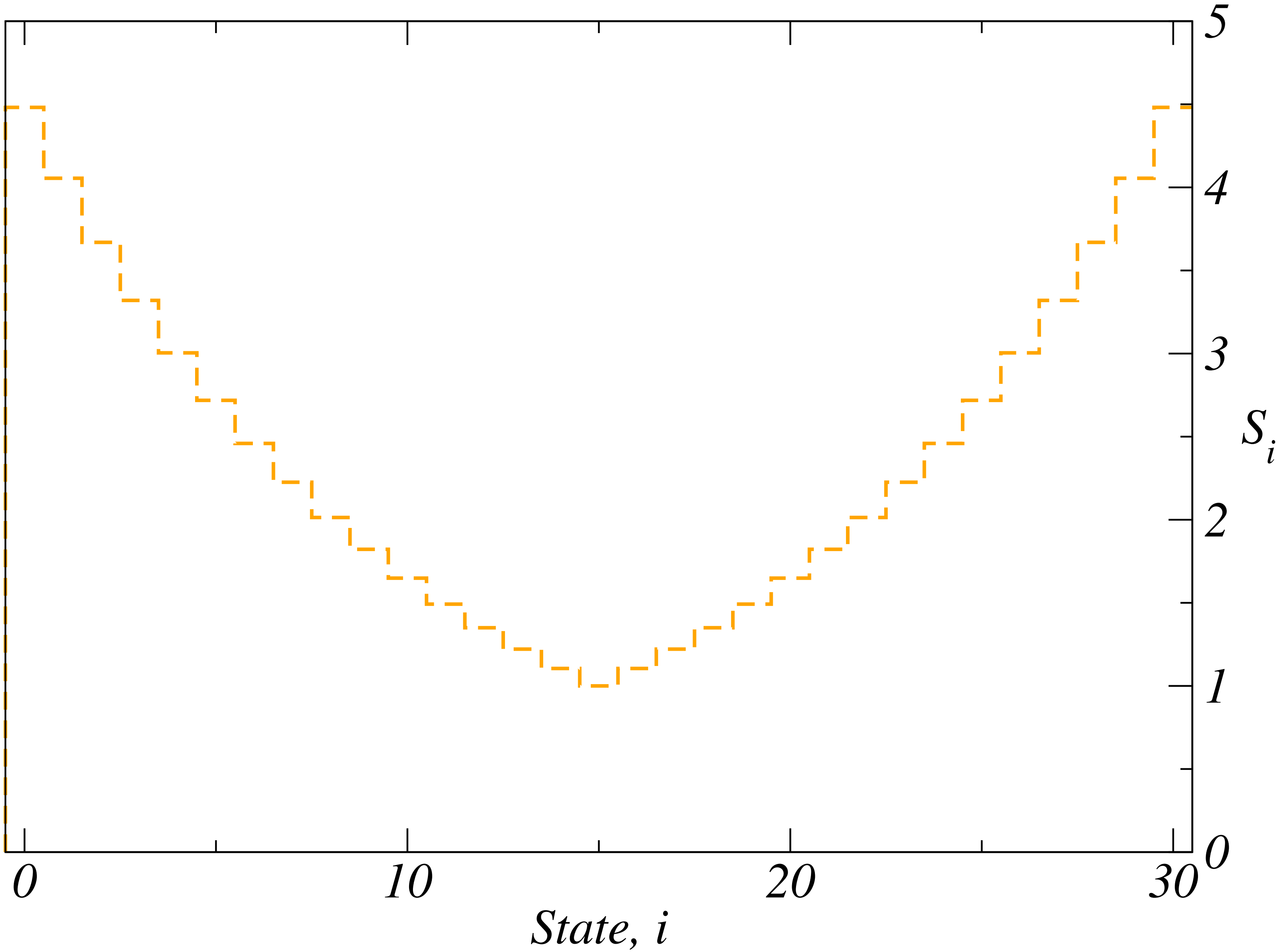
Hurdle Problem After 5 Generations



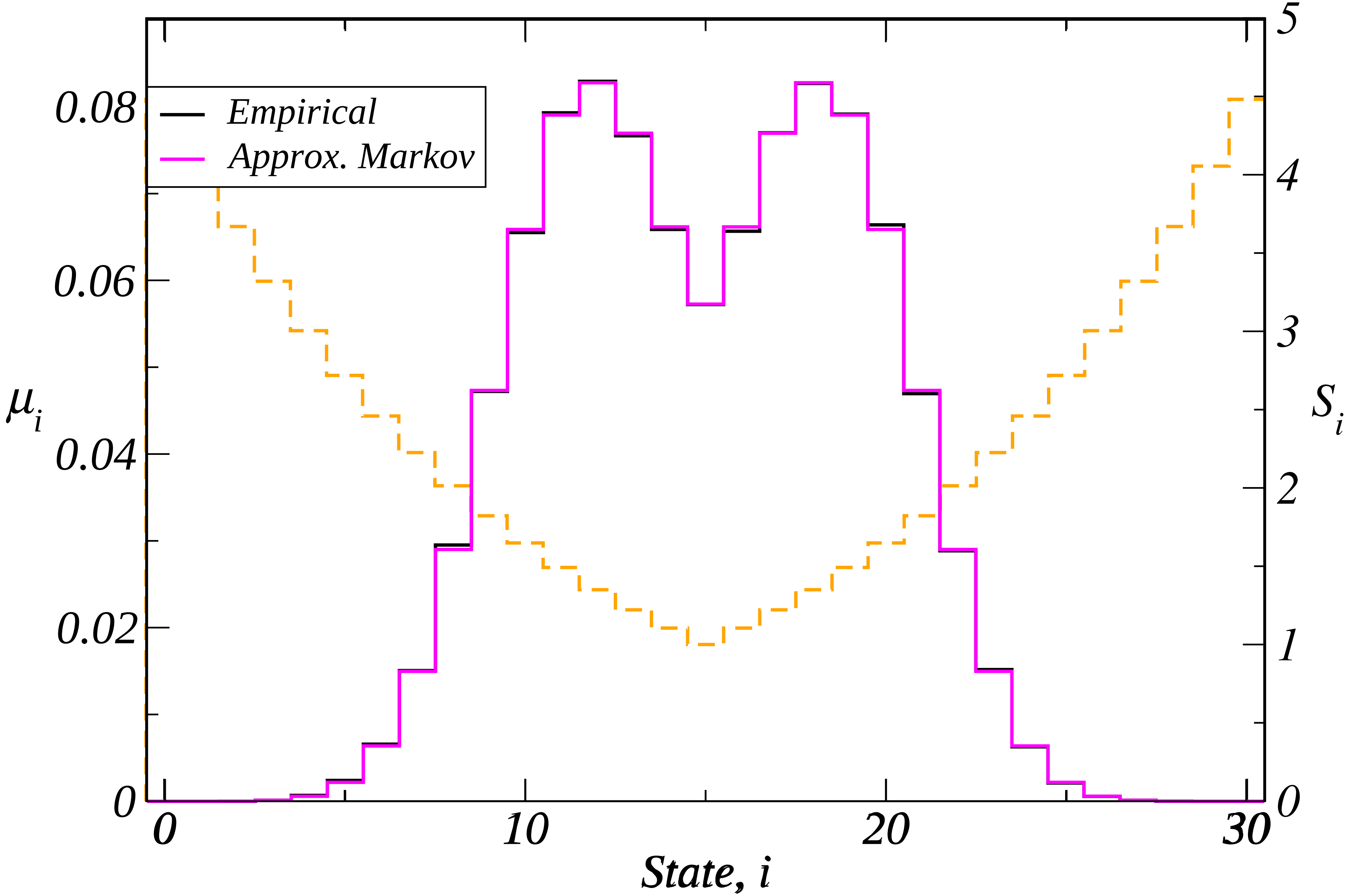
Hurdle Problem After 5 Generations



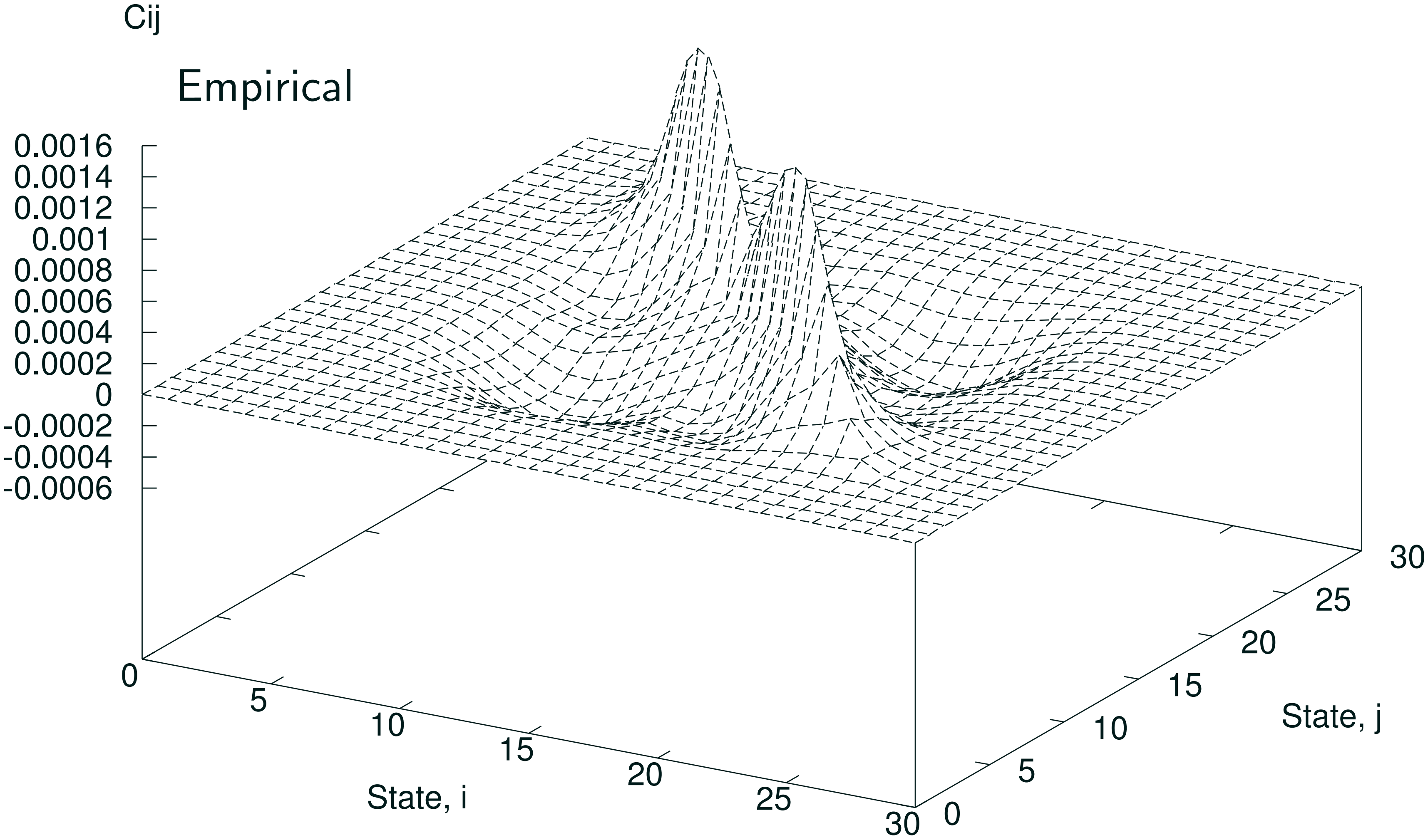
Symmetry Breaking After 5 Generations



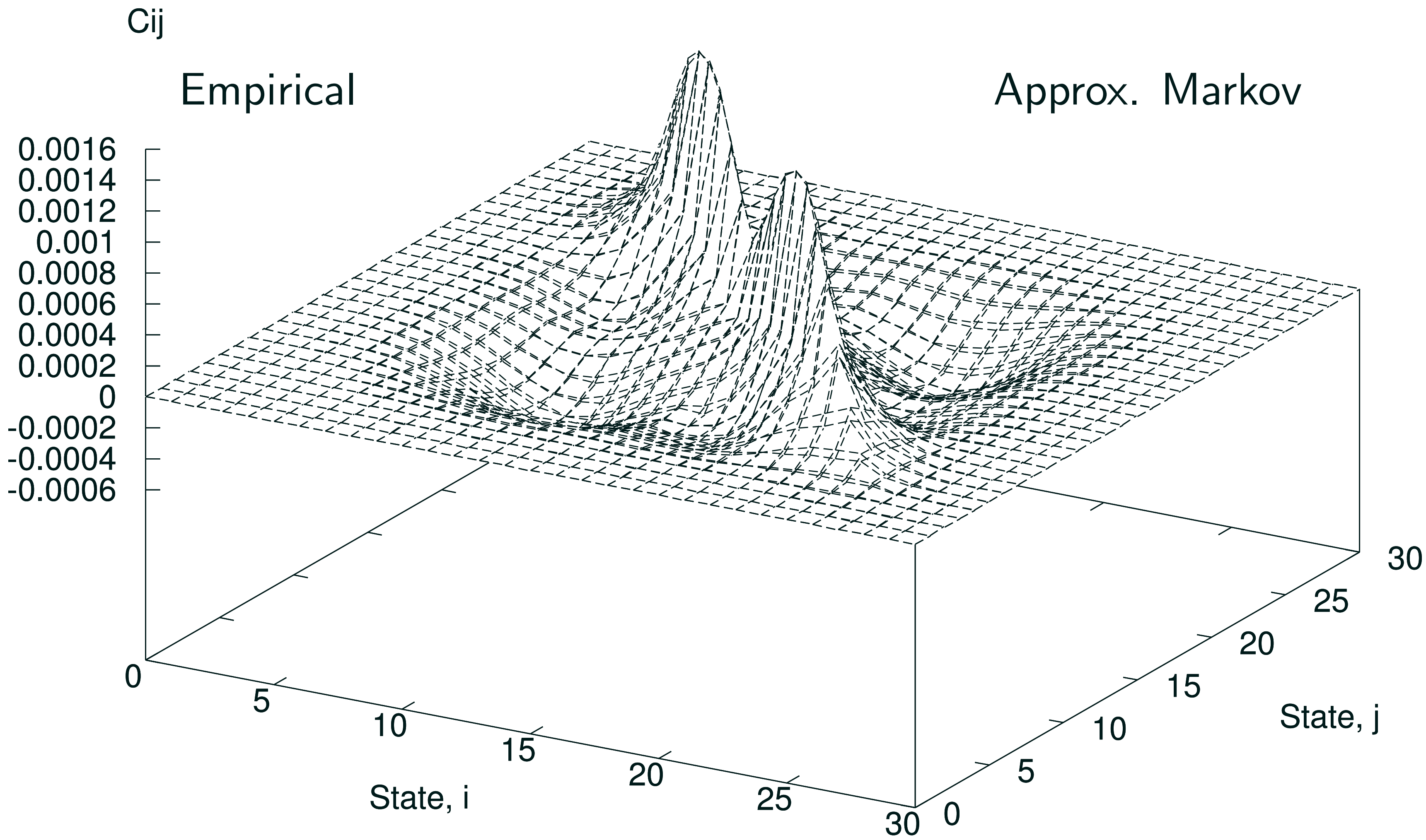
Symmetry Breaking After 5 Generations



Covariance


Cij
Empirical
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
-0.0002
-0.0004
-0.0006
0
5
10
15
20
25
30
State, i
0
5
10
15
20
25
30
State, j

Covariance


Cij
Empirical
Approx. Markov
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
-0.0002
-0.0004
-0.0006
0
5
10
15
20
25
30
State, i
0
5
10
15
20
25
30
State, j

Break down

- Both the stochastic expansion and exact sampling approach break down for large selection rates or large problems
- The problem occurs for the fit states that have a very small probability of being visited
- The fluctuations for these states grow larger in the theory than in practice
- Eventually, this drives some of the probabilities negative—everything then falls apart

Break down

- Both the stochastic expansion and exact sampling approach break down for large selection rates or large problems
- The problem occurs for the fit states that have a very small probability of being visited
- The fluctuations for these states grow larger in the theory than in practice
- Eventually, this drives some of the probabilities negative—everything then falls apart

Break down

- Both the stochastic expansion and exact sampling approach break down for large selection rates or large problems

- The problem occurs for the fit states that have a very small probability of being visited

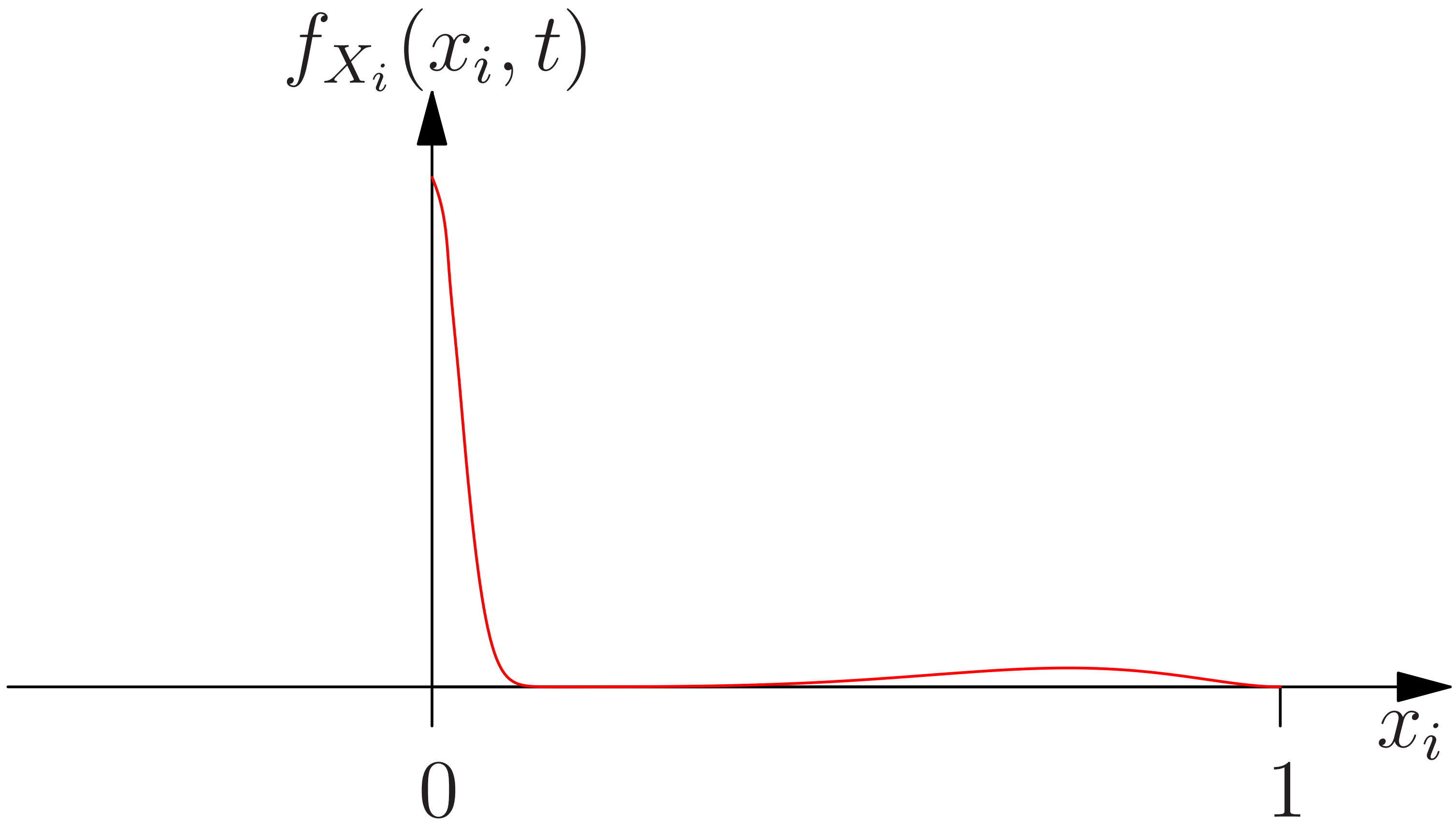
- The fluctuations for these states grow larger in the theory than in practice

- Eventually, this drives some of the probabilities negative—everything then falls apart

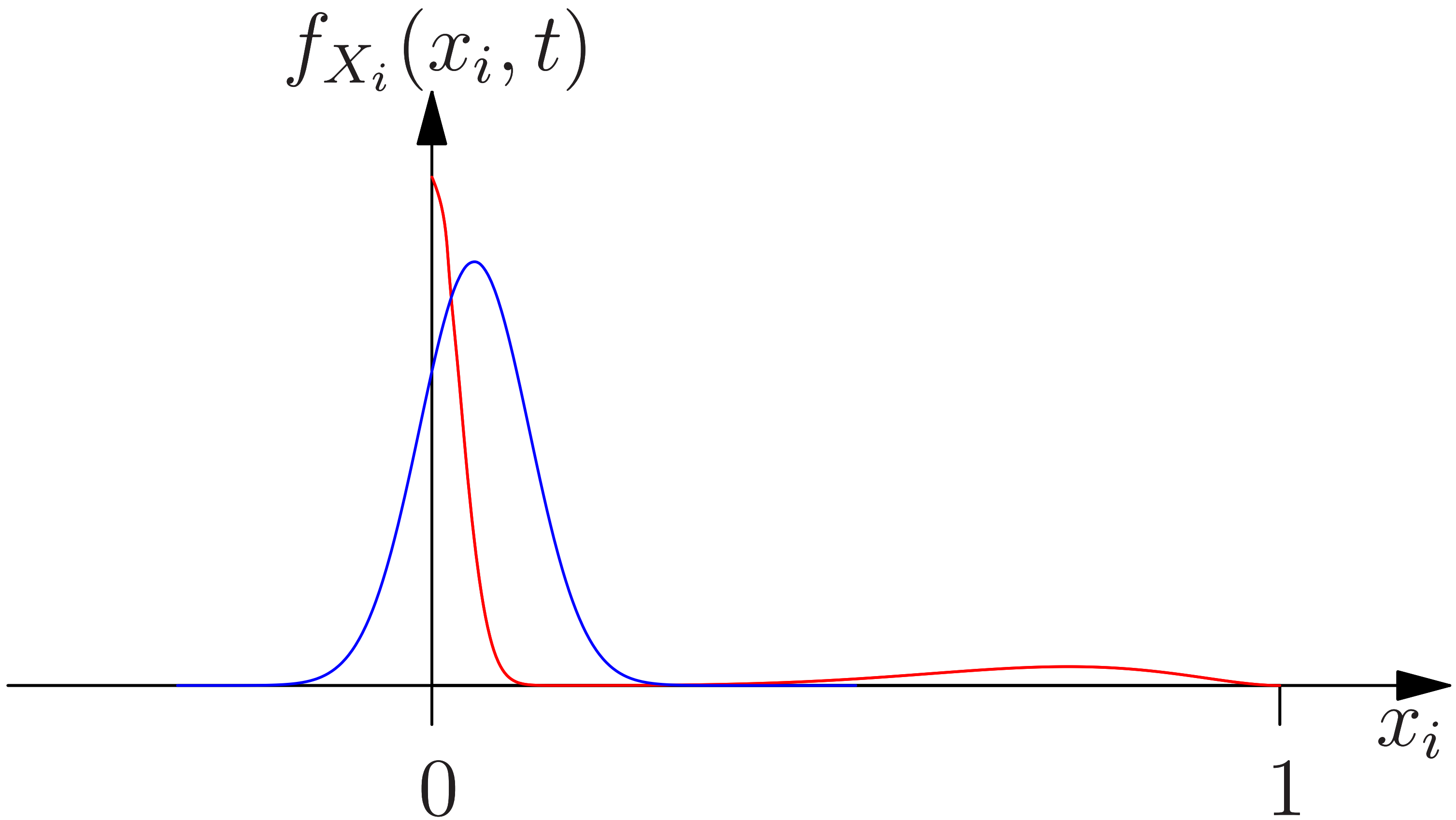
Break down

- Both the stochastic expansion and exact sampling approach break down for large selection rates or large problems
- The problem occurs for the fit states that have a very small probability of being visited
- The fluctuations for these states grow larger in the theory than in practice
- Eventually, this drives some of the probabilities negative—everything then falls apart

High Fitness State

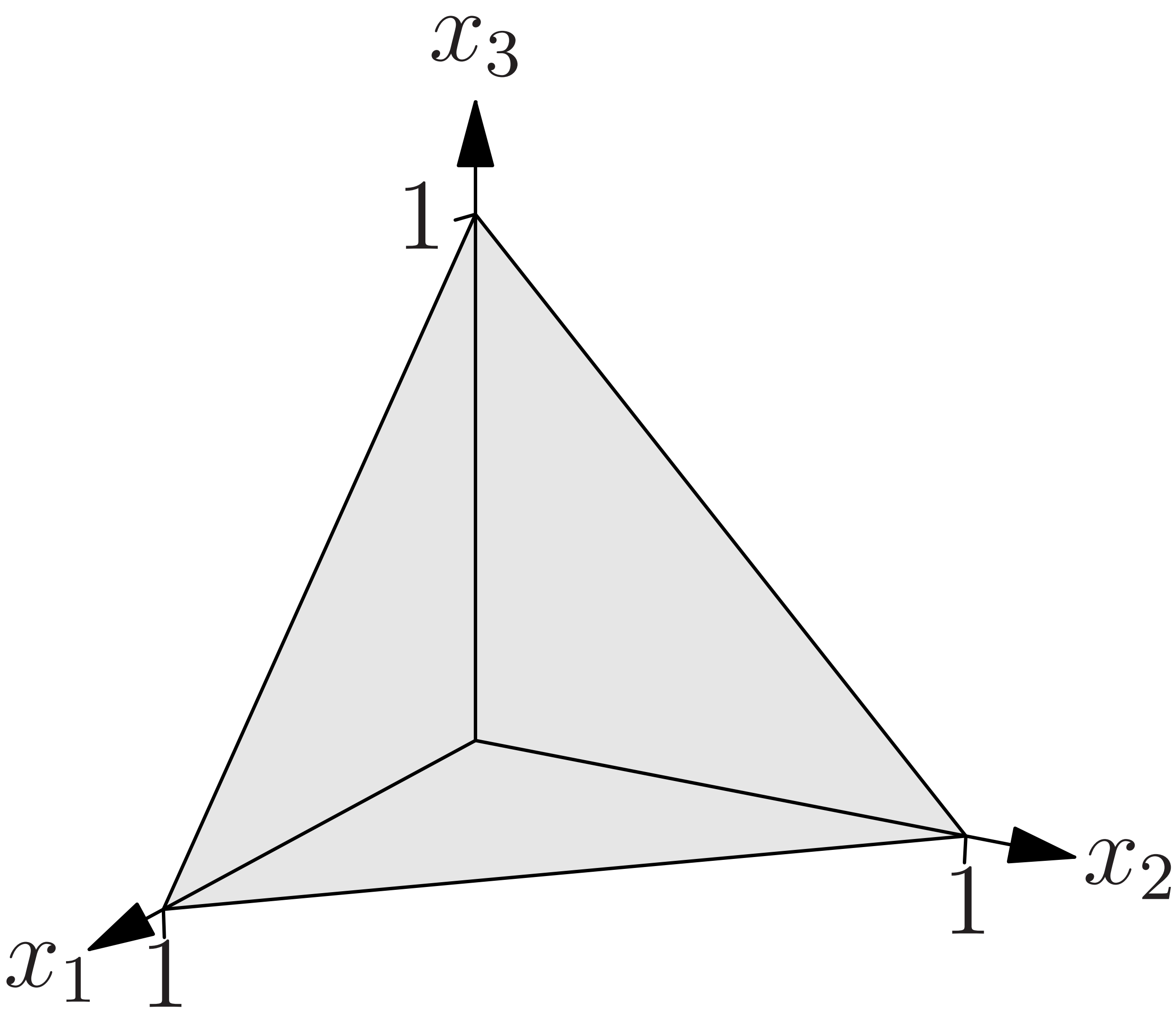


High Fitness State

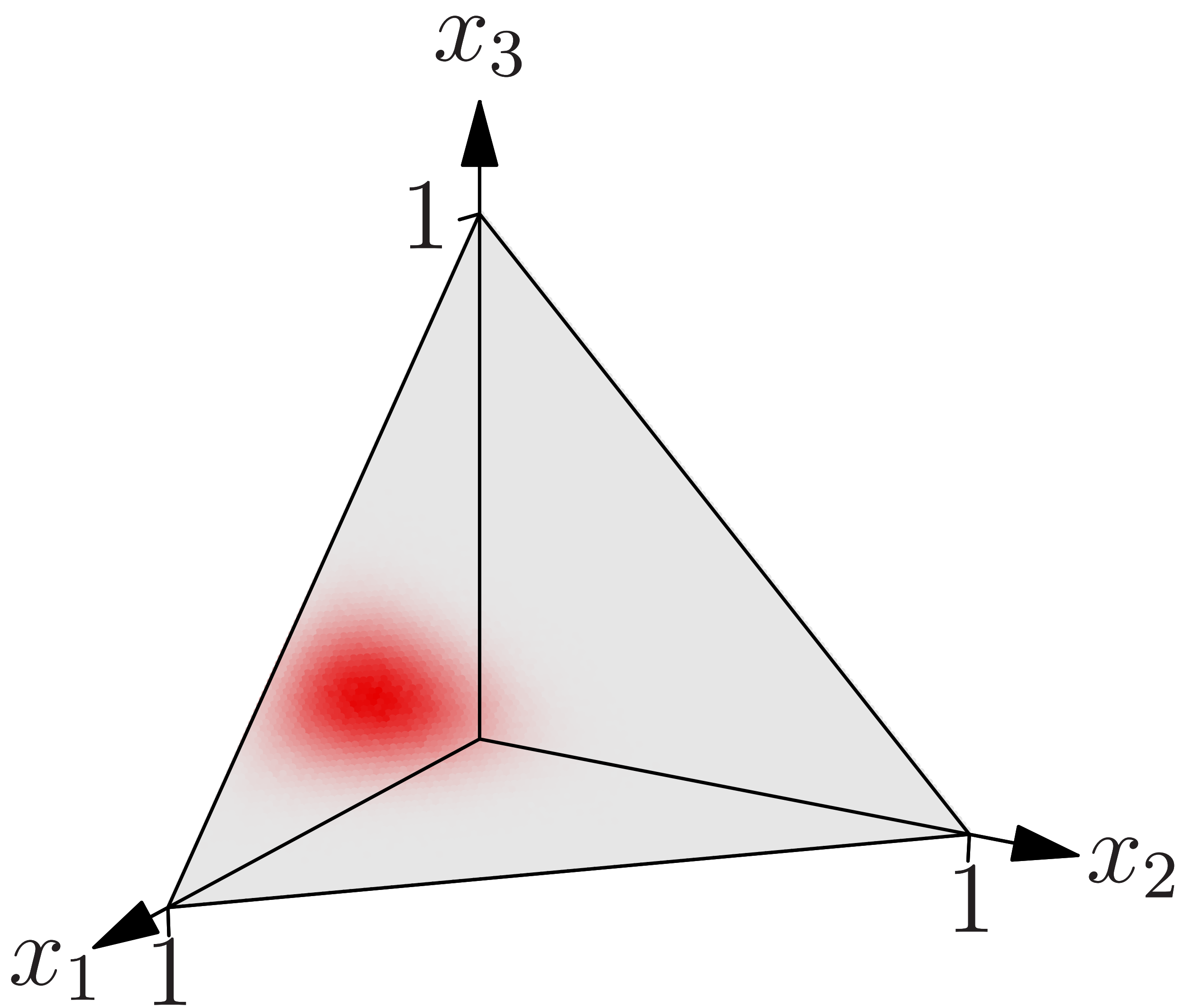


c.f. Foga 2002

Distribution on the Simplex



Distribution on the Simplex



Multivariate Distribution on Simplex

- Require a multivariate distribution defined on the unit simplex
- No standard distribution
- Easy to define

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu}, \mathbf{L}) = \int_{-\infty}^{\infty} N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu}, \mathbf{L}) \prod_{i=1}^{n} \delta\left(x_i - \frac{y_i^2}{\sum_j y_j^2}\right) \mathrm{d}\boldsymbol{y}$$

$$N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu}, \mathbf{L}) = \frac{1}{|\sqrt{2\pi}\mathbf{L}|} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \mathbf{L}^{-1} \mathbf{L}^{T-1} (\boldsymbol{y}-\boldsymbol{\mu})}$$

where $\boldsymbol{\mu}$ is a vector and $\mathbf{L}$ a lower diagonal matrix

Multivariate Distribution on Simplex

- Require a multivariate distribution defined on the unit simplex
- No standard distribution
- Easy to define

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu}, \mathbf{L}) = \int_{-\infty}^{\infty} N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu}, \mathbf{L}) \prod_{i=1}^{n} \delta\left(x_i - \frac{y_i^2}{\sum_j y_j^2} \right) \mathrm{d}\boldsymbol{y}$$

$$N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu}, \mathbf{L}) = \frac{1}{|\sqrt{2\pi}\mathbf{L}|} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \mathbf{L}^{-1} \mathbf{L}^{T-1} (\boldsymbol{y}-\boldsymbol{\mu})}$$

where $\boldsymbol{\mu}$ is a vector and $\mathbf{L}$ a lower diagonal matrix

Multivariate Distribution on Simplex

- Require a multivariate distribution defined on the unit simplex
- No standard distribution
- Easy to define

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu},\mathbf{L}) = \int_{-\infty}^{\infty} N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu},\mathbf{L}) \prod_{i=1}^{n} \delta\left(x_i - \frac{y_i^2}{\sum_j y_j^2}\right) \mathrm{d}\boldsymbol{y}$$

$$N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu},\mathbf{L}) = \frac{1}{|\sqrt{2\pi}\mathbf{L}|} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \mathbf{L}^{-1} \mathbf{L}^{T-1}(\boldsymbol{y}-\boldsymbol{\mu})}$$

where $\boldsymbol{\mu}$ is a vector and $\mathbf{L}$ a lower diagonal matrix

Averages Over Simplex Distribution

- This distribution is not unique, e.g.

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu}, \mathbf{L}) = \int_{-\infty}^{\infty} N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu}, \mathbf{L})\, \delta\left(\sum_i \mathrm{e}^{y_i} - 1\right) \prod_{i=1}^{n} \delta(x_i - \mathrm{e}^{y_i})\, \mathrm{d}\boldsymbol{y}$$

- However, to be practical it must be possible to compute average quantities

$$\langle g(\boldsymbol{X}) \rangle = \int g(\boldsymbol{X}) f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu}, \mathbf{L}) \mathrm{d}\boldsymbol{x}$$

- This is possible for previous distribution

Averages Over Simplex Distribution

- This distribution is not unique, e.g.

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu},\mathbf{L}) = \int_{-\infty}^{\infty} N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu},\mathbf{L})\,\delta\!\left(\sum_i \mathrm{e}^{\,y_i} - 1\right)\prod_{i=1}^{n}\delta(x_i - \mathrm{e}^{\,y_i})\;\mathrm{d}\boldsymbol{y}$$

- However, to be practical it must be possible to compute average quantities

$$\langle g(\boldsymbol{X})\rangle = \int g(\boldsymbol{X})\,f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu},\mathbf{L})\mathrm{d}\boldsymbol{x}$$

- This is possible for previous distribution

Averages Over Simplex Distribution

- This distribution is not unique, e.g.

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu},\mathbf{L}) = \int_{-\infty}^{\infty} N_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{\mu},\mathbf{L})\,\delta\left(\sum_i \mathrm{e}^{\,y_i} - 1\right)\prod_{i=1}^{n}\delta(x_i - \mathrm{e}^{\,y_i})\;\mathrm{d}\boldsymbol{y}$$

- However, to be practical it must be possible to compute average quantities

$$\langle g(\boldsymbol{X})\rangle = \int g(\boldsymbol{X})\,f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu},\mathbf{L})\mathrm{d}\boldsymbol{x}$$

- This is possible for previous distribution

Conclusions

- An approximate Markov model would be desirable for studying GAs
- Need to go beyond infinite population model
- Plausible that using the means and covariances an approximate Markov model can be made
- However, there remains a problem with modelling the above fitness states

Conclusions

- An approximate Markov model would be desirable for studying GAs
- Need to go beyond infinite population model
- Plausible that using the means and covariances an approximate Markov model can be made
- However, there remains a problem with modelling the above fitness states

Conclusions

- An approximate Markov model would be desirable for studying GAs
- Need to go beyond infinite population model
- Plausible that using the means and covariances an approximate Markov model can be made
- However, there remains a problem with modelling the above fitness states

Conclusions

- An approximate Markov model would be desirable for studying GAs
- Need to go beyond infinite population model
- Plausible that using the means and covariances an approximate Markov model can be made
- However, there remains a problem with modelling the above fitness states

Conclusions

- An approximate Markov model would be desirable for studying GAs
- Need to go beyond infinite population model
- Plausible that using the means and covariances an approximate Markov model can be made
- However, there remains a problem with modelling the above fitness states—This work is still in progress

Conclusions

- An approximate Markov model would be desirable for studying GAs
- Need to go beyond infinite population model
- Plausible that using the means and covariances an approximate Markov model can be made
- However, there remains a problem with modelling the above fitness states—This work is still in progress

Questions?