# Theory of Estimation-of-Distribution Algorithms

PPSN 2024 Tutorial

## 1. Estimation-of-Distribution Algorithms (EDAs)

## 2. Theory

# 1. Estimation-of-Distribution Algorithms (EDAs)

» **What** are they?

# 2. Theory

## 1. Estimation-of-Distribution Algorithms (EDAs)

## 2. Theory

- **What** are they?
- **Focus** on the most commonly theoretically studied ones

## 1. Estimation-of-Distribution Algorithms (EDAs)

» **What** are they?

» **Focus** on the most commonly theoretically studied ones

## 2. Theory

(Mathematical rigor)

## 1. Estimation-of-Distribution Algorithms (EDAs)

» **What** are they?

» **Focus** on the most commonly theoretically studied ones

## 2. Theory

(Mathematical rigor)

» **What** is studied?

## 1. Estimation-of-Distribution Algorithms (EDAs)

» **What** are they?

» **Focus** on the most commonly theoretically studied ones

## 2. Theory

(Mathematical rigor)

» **What** is studied?

» **Selection** of important results

- **Understand** the basic idea

## 1. Estimation-of-Distribution Algorithms (EDAs)

- **What** are they?
- **Focus** on the most commonly theoretically studied ones

## 2. Theory

(Mathematical rigor)

- **What** is studied?
- **Selection** of important results
  - **Understand** the basic idea

» **Depth over breadth**

Martin S. Krejca

martin.krejca@polytechnique.edu


Martin S. Krejca

martin.krejca@polytechnique.edu



Carsten Witt



DOI: **10.1145/3377929.3389888**


(more **breadth**)

Martin S. Krejca

martin.krejca@polytechnique.edu

1806.05392

Carsten Witt

DOI: **10.1145/3377929.3389888**

(more **breadth**)


Martin S. Krejca

martin.krejca@polytechnique.edu

1806.05392


Carsten Witt

GECCO

DOI: **10.1145/3377929.3389888**

(more **breadth**)

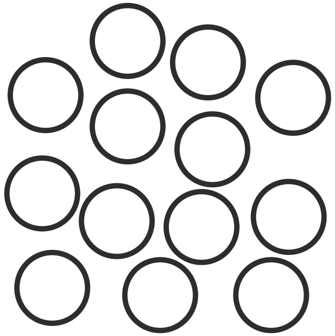
PPSN 2024

**Sunday, Sep. 15**

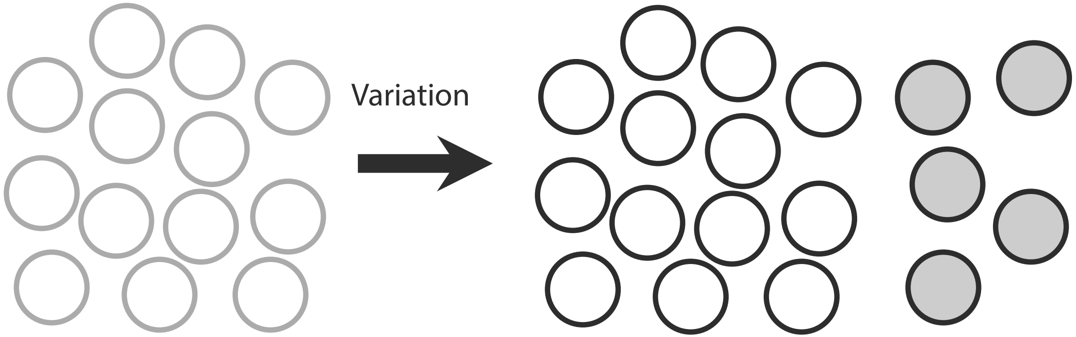
**9:30–12:30**

**30 Years of EDAs**

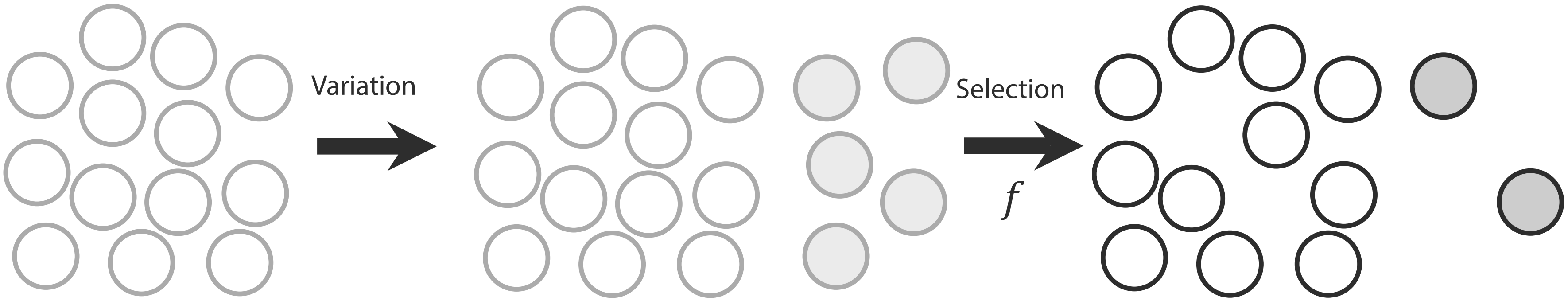
# Evolutionary Algorithms



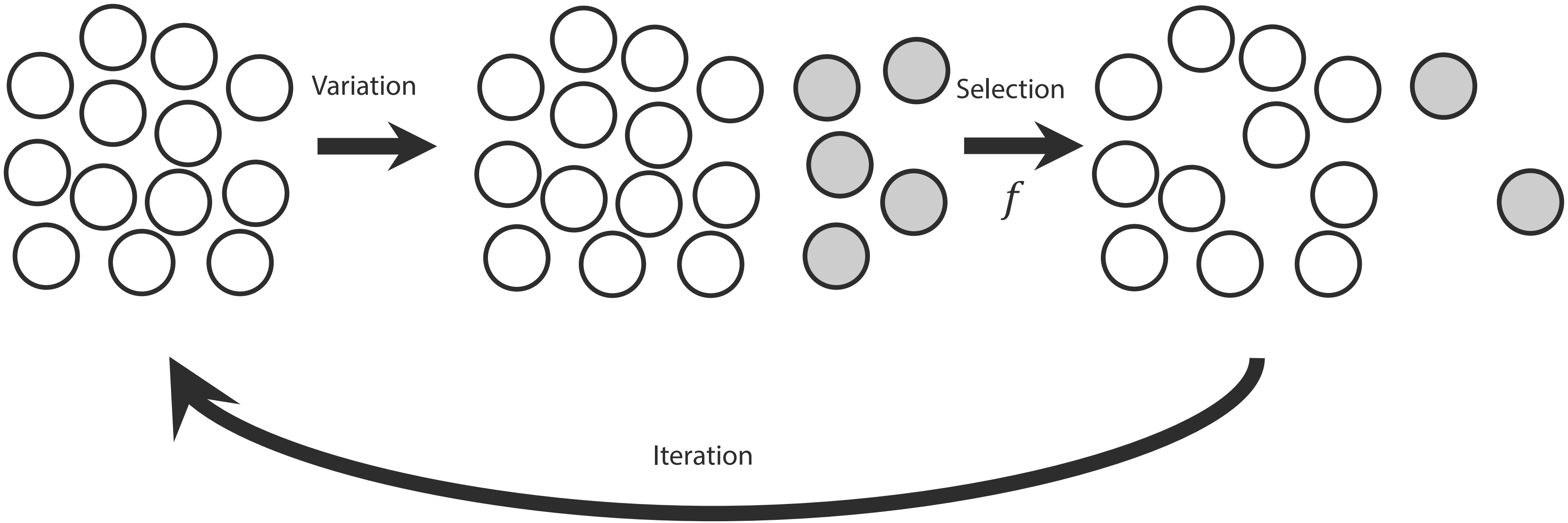
# Evolutionary Algorithms



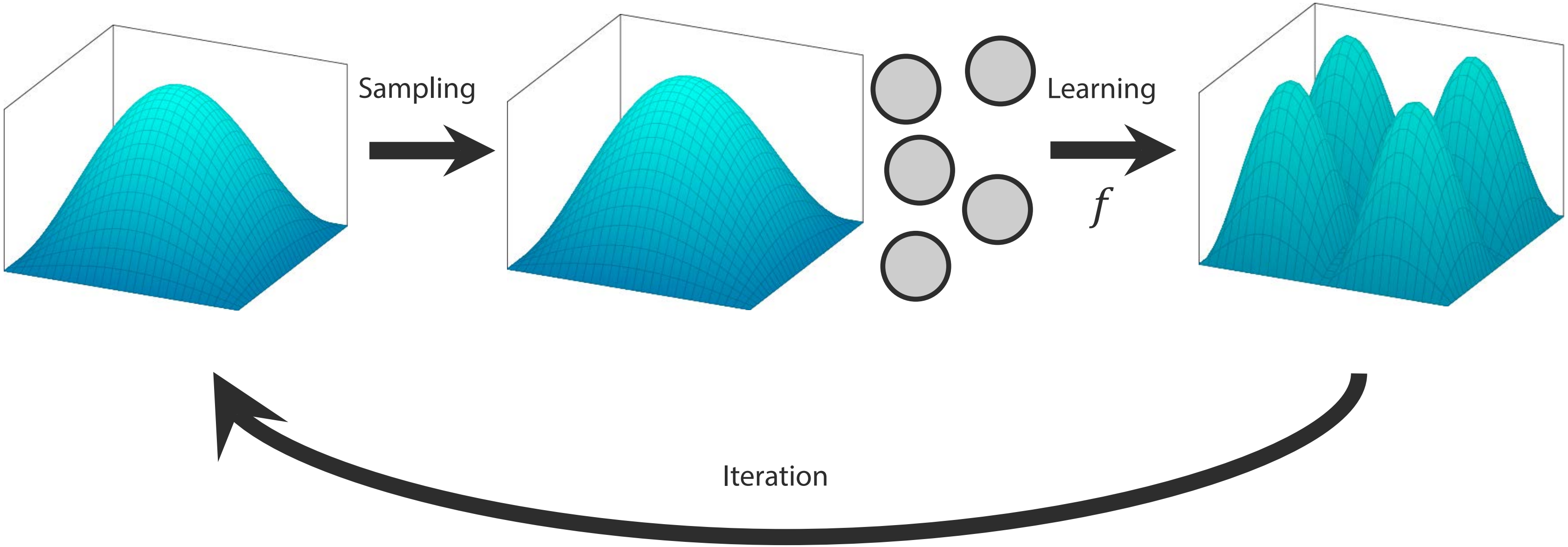
# Evolutionary Algorithms



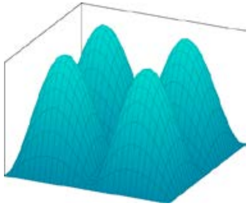
# Evolutionary Algorithms



# Estimation-of-Distribution Algorithms

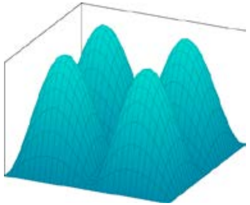


# Probabilistic Models



general

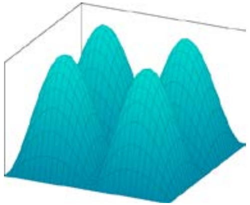
# Probabilistic Models



general

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 2 |
| | | ⋮ | | |

# Probabilistic Models

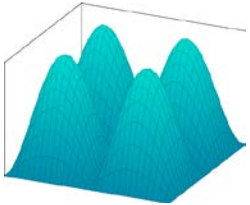


general

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 2 |
| | | ⋮ | | |

$$\prod_{i \in [\#\text{vars}]} \left(\#\text{vals}(x_i)\right) - 1$$

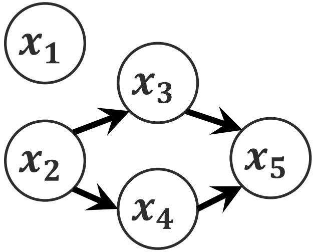
# Probabilistic Models



general

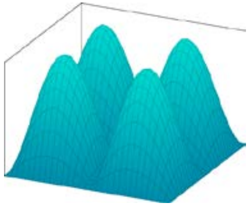
| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 2 |
| | | ⋮ | | |

$$\prod_{i \in [\#\text{vars}]} \left(\#\text{vals}(x_i)\right) - 1$$




compressed, directed

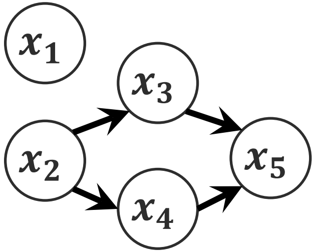
# Probabilistic Models



general

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 2 |
| | | ⋮ | | |

$$\prod_{i\in[\#\text{vars}]}\left(\#\text{vals}(x_i)\right)-1$$




compressed, directed

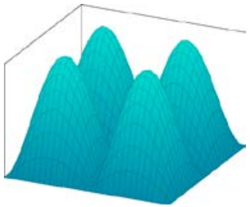
| $x_1$ |
|---|
| 0 |
| 1 |
| 2 |
| ⋮ |

| $x_2$ |
|---|
| 0 |
| 1 |
| 2 |
| ⋮ |

| $x_2$ | $x_3$ |
|---|---|
| 0 | 0 |
| 0 | 1 |
| 0 | 2 |
| ⋮ | |

…

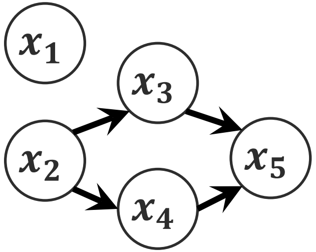
# Probabilistic Models



general

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 2 |
| | | ⋮ | | |

$$\prod_{i\in[\#\text{vars}]}\big(\#\text{vals}(x_i)\big) - 1$$




compressed, directed

| $x_1$ |
|---|
| 0 |
| 1 |
| 2 |
| ⋮ |

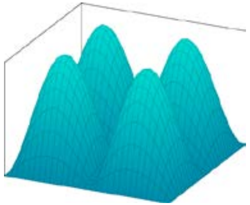
| $x_2$ |
|---|
| 0 |
| 1 |
| 2 |
| ⋮ |

| $x_2$ | $x_3$ |
|---|---|
| 0 | 0 |
| 0 | 1 |
| 0 | 2 |
| ⋮ | |

…

$$\sum_{i\in[\#\text{vars}]}(\#\text{vals}(x_i) - 1)\prod_{y\in\text{pred}(x_i)}\#\text{vals}(y)$$

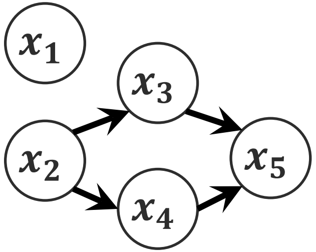
# Probabilistic Models



general

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 2 |
| | | ⋮ | | |

$$\prod_{i\in[\#\mathrm{vars}]}(\#\mathrm{vals}(x_i)) - 1$$




compressed, directed

| $x_1$ |
|---|
| 0 |
| 1 |
| 2 |
| ⋮ |

| $x_2$ |
|---|
| 0 |
| 1 |
| 2 |
| ⋮ |

| $x_2$ | $x_3$ |
|---|---|
| 0 | 0 |
| 0 | 1 |
| 0 | 2 |
| ⋮ | |

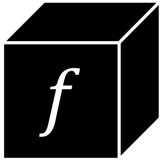
…

$$\sum_{i\in[\#\mathrm{vars}]}(\#\mathrm{vals}(x_i) - 1)\prod_{y\in\mathrm{pred}(x_i)}\#\mathrm{vals}(y)$$

**Theory**


**univariate**

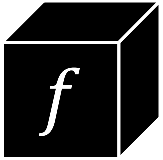
# Theory




$f: \{0,1\}^n \to \mathbf{R}$

» Pseudo-Boolean optimization

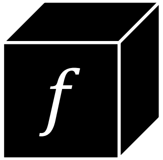
# Theory




$f : \{0,1\}^n \to \mathbf{R}$

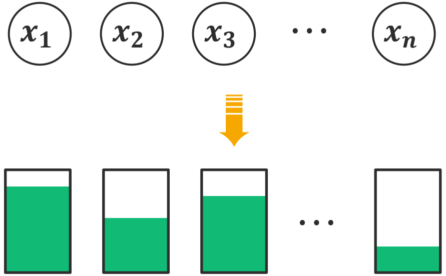
- Pseudo-Boolean optimization
  - **Global optimum** often $1^n$

# Theory

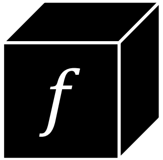



$f: \{0,1\}^n \to \mathbf{R}$

- Pseudo-Boolean optimization
  - **Global optimum** often $1^n$



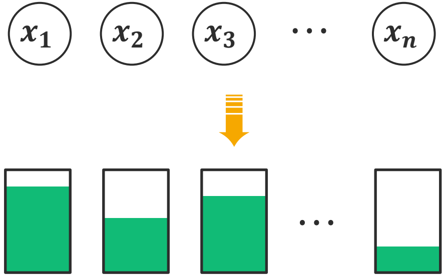
# Theory




$f: \{0,1\}^n \to \mathbf{R}$

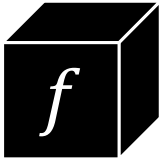
» Pseudo-Boolean optimization

❖ **Global optimum** often $1^n$



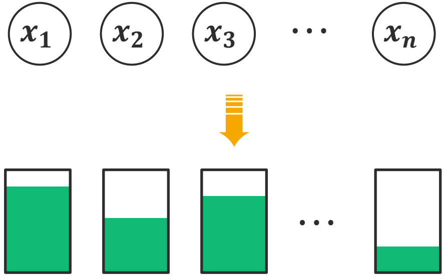

» *Frequency* vector $\boldsymbol{p}$

# Theory



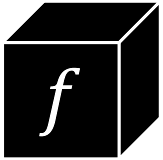

$f : \{0,1\}^n \to \mathbf{R}$

- Pseudo-Boolean optimization
  - **Global optimum** often $1^n$



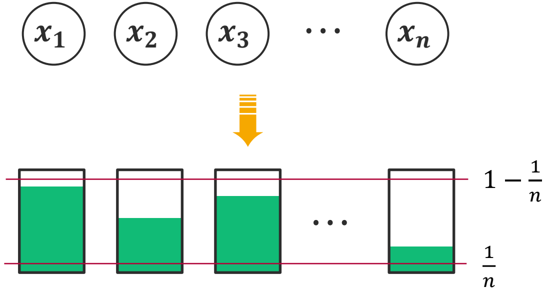

- *Frequency* vector $\boldsymbol{p}$
- $\boldsymbol{p}_i$ … probability to sample a 1 at $i$ (green mass)

# Theory



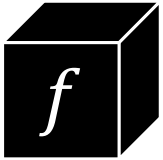

$f: \{0,1\}^n \to \mathbf{R}$

- Pseudo-Boolean optimization
  - **Global optimum** often $1^n$




- *Frequency* vector $\boldsymbol{p}$
- $\boldsymbol{p}_i$ … probability to sample a 1 at $i$ (green mass)

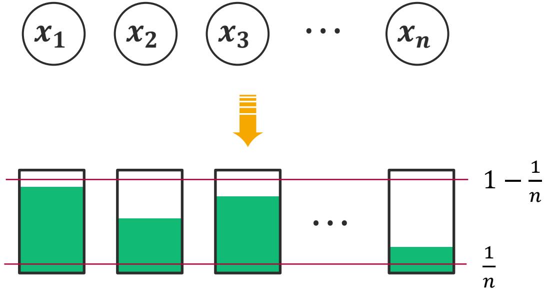
# Theory




$f: \{0,1\}^n \rightarrow \mathbf{R}$

» Pseudo-Boolean optimization

- **Global optimum** often $1^n$

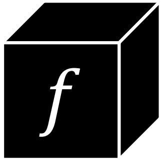



» *Frequency* vector $\boldsymbol{p}$

» $\boldsymbol{p}_i$ … probability to sample a 1 at $i$ (green mass)

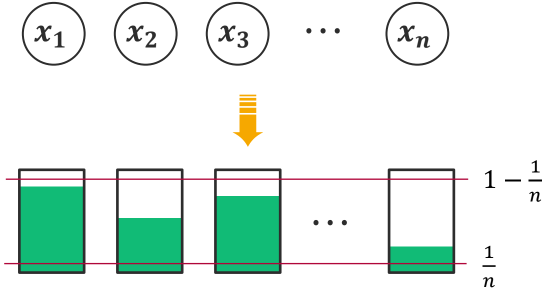
Multi-valued analyses

# Theory




$f: \{0,1\}^n \rightarrow \mathbf{R}$

- Pseudo-Boolean optimization
  - **Global optimum** often $1^n$




- *Frequency vector* $\boldsymbol{p}$
- $\boldsymbol{p}_i$ … probability to sample a 1 at $i$ (green mass)

Multi-valued analyses

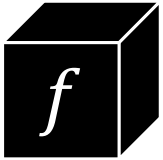
- DOI: **10.1016/j.tcs.2024.114622**

[Ben Jedidia, Doerr, K.'24]

- *Runtime Analysis of a Multi-Valued Compact Genetic Algorithm on Generalized OneMax*

[Adak, Witt'24]

# Theory



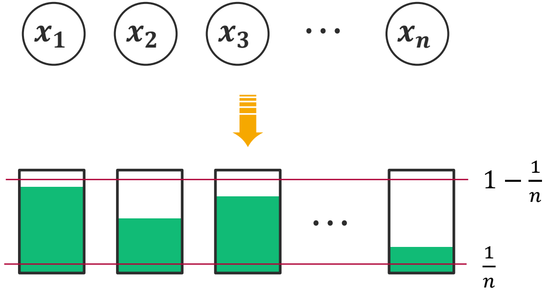

$f: \{0,1\}^n \rightarrow \mathbf{R}$

» Pseudo-Boolean optimization

❖ **Global optimum** often $1^n$

» **Run time** analysis

❖ Number of evaluations of $f$ until a global optimum is sampled




» *Frequency* vector $\boldsymbol{p}$

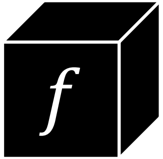
» $\boldsymbol{p}_i$ … probability to sample a 1 at $i$ (green mass)

Multi-valued analyses

» DOI: **10.1016/j.tcs.2024.114622** [Ben Jedidia, Doerr, K.'24]

» *Runtime Analysis of a Multi-Valued Compact Genetic Algorithm on Generalized OneMax* [Adak, Witt'24]

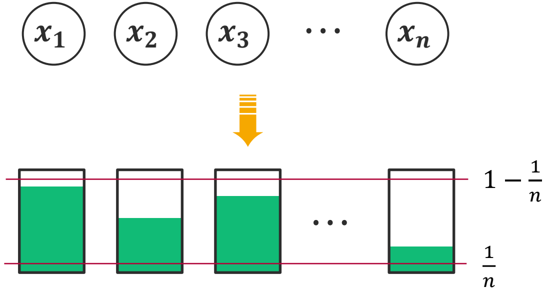
# Theory




$f : \{0,1\}^n \to \mathbf{R}$

» Pseudo-Boolean optimization

❖ **Global optimum** often $1^n$




» *Frequency* vector $\boldsymbol{p}$

» $\boldsymbol{p}_i$ … probability to sample a 1 at $i$ (green mass)

» **Run time** analysis

❖ Number of evaluations of $f$ until a global optimum is sampled

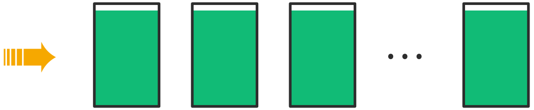
Multi-valued analyses

» DOI: **10.1016/j.tcs.2024.114622**

[Ben Jedidia, Doerr, K.'24]

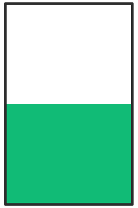
» *Runtime Analysis of a Multi-Valued Compact Genetic Algorithm on Generalized OneMax*

[Adak, Witt'24]



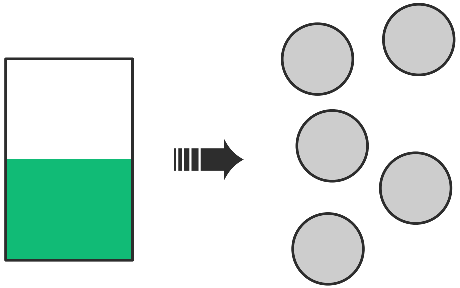
» Favorable albeit not necessary model

# Updating a Frequency (commonly)



$\boldsymbol{p}_i^{(t)}$

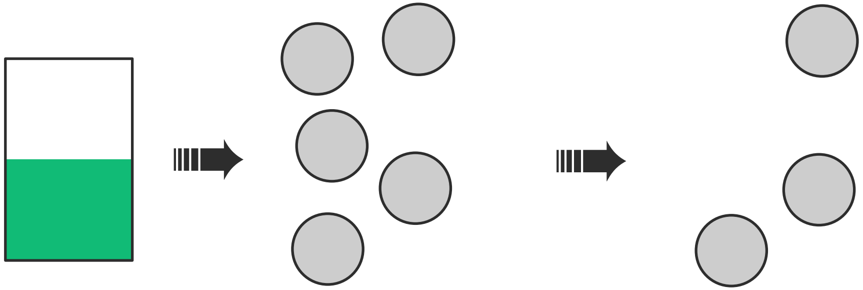
# Updating a Frequency (commonly)



$\boldsymbol{p}_i^{(t)}$

samples $\left\{\boldsymbol{x}^{(j)}\right\}_{j \in [\lambda]}$

# Updating a Frequency (commonly)



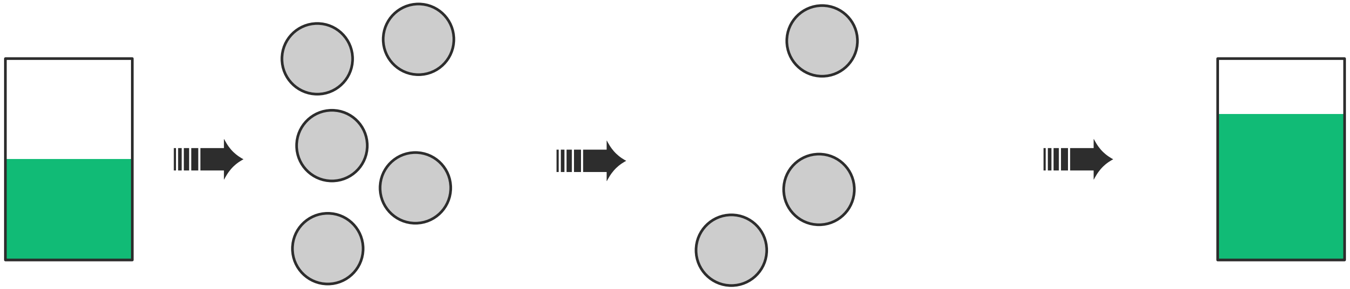
$\boldsymbol{p}_i^{(t)}$

samples $\left\{\boldsymbol{x}^{(j)}\right\}_{j\in[\lambda]}$

**selected** samples $\left\{\hat{\boldsymbol{x}}^{(k)}\right\}_{k\in[\mu]}$

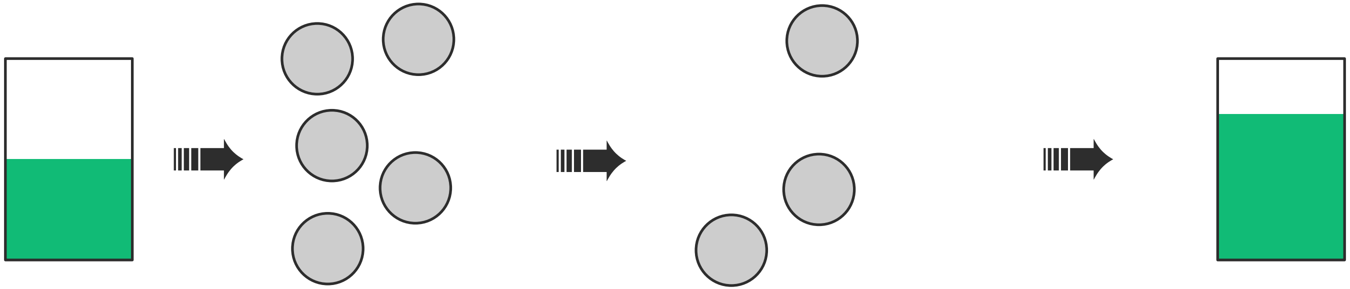
(ordered by decreasing fitness)

# Updating a Frequency (commonly)



$\boldsymbol{p}_i^{(t)}$ samples $\left\{\boldsymbol{x}^{(j)}\right\}_{j\in[\lambda]}$ **selected** samples $\left\{\hat{\boldsymbol{x}}^{(k)}\right\}_{k\in[\mu]}$ (ordered by decreasing fitness) $\boldsymbol{p}_i^{(t+1)}$

# Updating a Frequency (commonly)

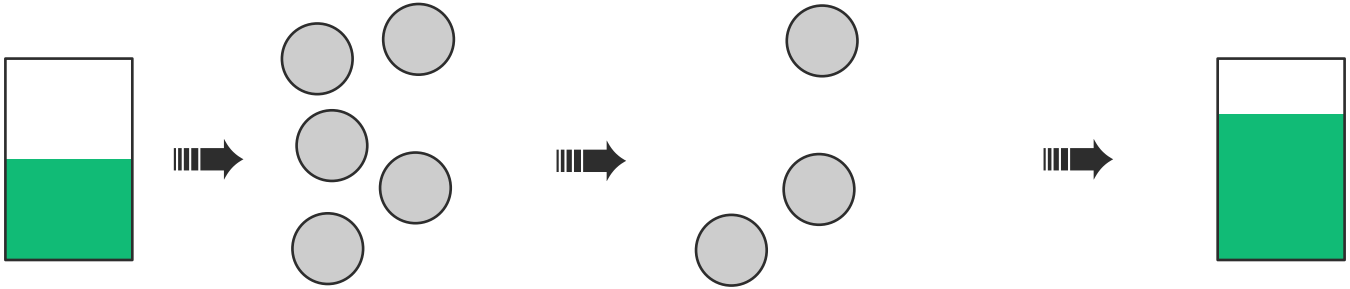


$\boldsymbol{p}_i^{(t)}$ samples $\left\{\boldsymbol{x}^{(j)}\right\}_{j\in[\lambda]}$ **selected** samples $\left\{\hat{\boldsymbol{x}}^{(k)}\right\}_{k\in[\mu]}$ (ordered by decreasing fitness) $\boldsymbol{p}_i^{(t+1)}$

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

drift

# Updating a Frequency (commonly)



$\boldsymbol{p}_i^{(t)}$

samples $\left\{\boldsymbol{x}^{(j)}\right\}_{j\in[\lambda]}$

**selected** samples $\left\{\hat{\boldsymbol{x}}^{(k)}\right\}_{k\in[\mu]}$

(ordered by decreasing fitness)

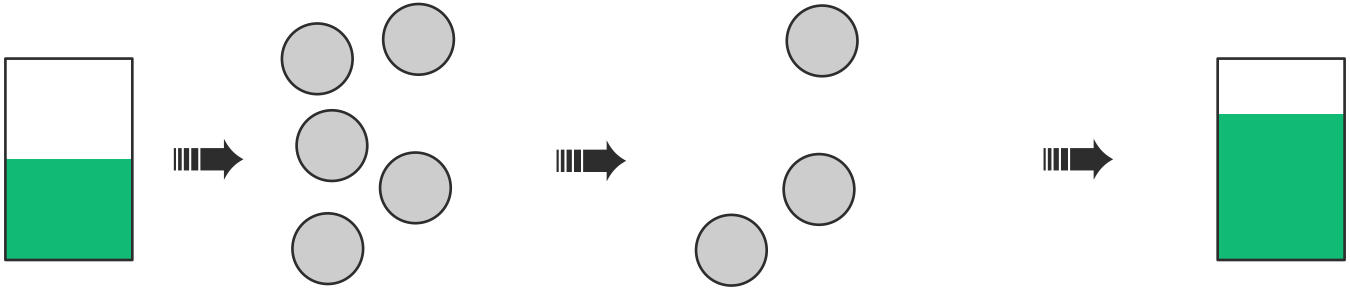
$\boldsymbol{p}_i^{(t+1)}$

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

drift

» Estimating the drift, estimates the expected time that the frequency passes a specific value

# Updating a Frequency (commonly)



$\boldsymbol{p}_i^{(t)}$ samples $\left\{\boldsymbol{x}^{(j)}\right\}_{j\in[\lambda]}$ **selected** samples $\left\{\hat{\boldsymbol{x}}^{(k)}\right\}_{k\in[\mu]}$ (ordered by decreasing fitness) $\boldsymbol{p}_i^{(t+1)}$

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

drift

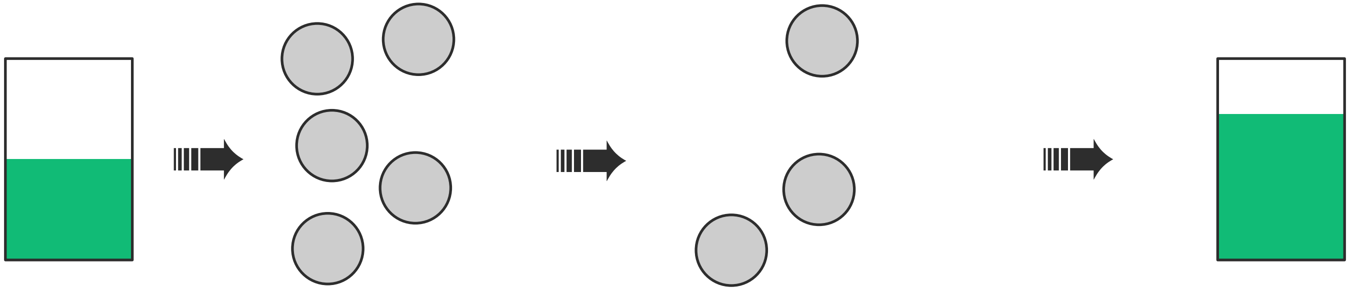
» Estimating the drift, estimates the expected time that the frequency passes a specific value

❖ **Drift analysis**

- 1712.00964 [Lengler'18]
- 2406.14589 [Kötzing'24]

# Updating a Frequency (commonly)

**Very tricky!**



$\boldsymbol{p}_i^{(t)}$ samples $\{\boldsymbol{x}^{(j)}\}_{j\in[\lambda]}$ **selected** samples $\{\hat{\boldsymbol{x}}^{(k)}\}_{k\in[\mu]}$ (ordered by decreasing fitness) $\boldsymbol{p}_i^{(t+1)}$

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

drift

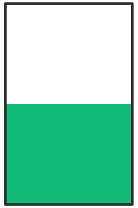
» Estimating the drift, estimates the expected time that the frequency passes a specific value

❖ **Drift analysis** 1712.00964 [Lengler'18]

2406.14589 [Kötzing'24]

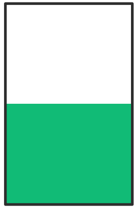
# Updating a Frequency (Genetic drift)

» **What** happens if the samples are **not** informative?



$\boldsymbol{p}_i^{(t)}$

# Updating a Frequency (Genetic drift)



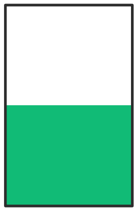
$\boldsymbol{p}_i^{(t)}$

» **What** happens if the samples are **not** informative?

- Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j\in[\lambda]}$ and $\left\{\widehat{\boldsymbol{x}}_i^{(k)}\right\}_{k\in[\mu]}$ follows the same distribution

» Can happen regularly during a run

# Updating a Frequency (Genetic drift)



$\boldsymbol{p}_i^{(t)}$

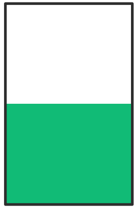
- **What** happens if the samples are **not** informative?
  - Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j\in[\lambda]}$ and $\left\{\widehat{\boldsymbol{x}}_i^{(k)}\right\}_{k\in[\mu]}$ follows the same distribution
  - ➡ **Removes** the tricky situation

- Can happen regularly during a run

# Updating a Frequency (Genetic drift)

- **What** happens if the samples are **not** informative?
  - Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j\in[\lambda]}$ and $\left\{\widehat{\boldsymbol{x}}_i^{(k)}\right\}_{k\in[\mu]}$ follows the same distribution
  - **Removes** the tricky situation

- Can happen regularly during a run



$\boldsymbol{p}_i^{(t)}$

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \boldsymbol{p}_i^{(t)}$$

balanced

- Typical property of univariate EDAs

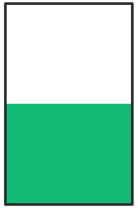
# Updating a Frequency (Genetic drift)

» **What** happens if the samples are **not** informative?

❖ Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j\in[\lambda]}$ and $\left\{\hat{\boldsymbol{x}}_i^{(k)}\right\}_{k\in[\mu]}$ follows the same distribution

» Can happen regularly during a run

➡ **Removes** the tricky situation



$\boldsymbol{p}_i^{(t)}$

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \boldsymbol{p}_i^{(t)}$$

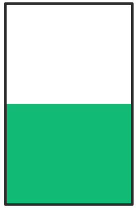
balanced

» Typical property of univariate EDAs

➡ $\boldsymbol{p}_i$ is a martingale and **approaches the borders quickly**

» Due to random fluctuations; despite a clear signal

# Updating a Frequency (Genetic drift)



$\boldsymbol{p}_i^{(t)}$

» **What** happens if the samples are **not** informative?

- Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j\in[\lambda]}$ and $\left\{\hat{\boldsymbol{x}}_i^{(k)}\right\}_{k\in[\mu]}$ follows the same distribution

» Can happen regularly during a run

➡ **Removes** the tricky situation

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \boldsymbol{p}_i^{(t)}$$

balanced

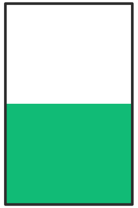
» Typical property of univariate EDAs

➡ $\boldsymbol{p}_i$ is a martingale and **approaches the borders quickly**

» Due to random fluctuations; despite a clear signal

- Genetic drift

# Updating a Frequency (Genetic drift)



$\boldsymbol{p}_i^{(t)}$

- **What** happens if the samples are **not** informative?
  - Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j\in[\lambda]}$ and $\left\{\widehat{\boldsymbol{x}}_i^{(k)}\right\}_{k\in[\mu]}$ follows the same distribution
  - Can happen regularly during a run

➡ **Removes** the tricky situation

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \boldsymbol{p}_i^{(t)}$$

balanced

- Typical property of univariate EDAs

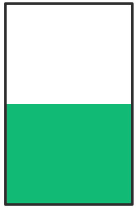
➡ $\boldsymbol{p}_i$ is a martingale and **approaches the borders quickly**

- Due to random fluctuations; despite a clear signal

- Genetic drift

- **How fast** does $\boldsymbol{p}_i$ cover distance $d \in [0,1]$?

# Updating a Frequency (Genetic drift)



$\boldsymbol{p}_i^{(t)}$

- **What** happens if the samples are **not** informative?
  - Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j\in[\lambda]}$ and $\left\{\widehat{\boldsymbol{x}}_i^{(k)}\right\}_{k\in[\mu]}$ follows the same distribution
  - Can happen regularly during a run
  - ➡ **Removes** the tricky situation

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \boldsymbol{p}_i^{(t)}$$

balanced

- Typical property of univariate EDAs

➡ $\boldsymbol{p}_i$ is a martingale and **approaches the borders quickly**

- Due to random fluctuations; despite a clear signal

- Genetic drift

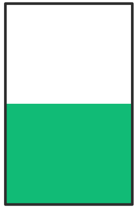
- **How fast** does $\boldsymbol{p}_i$ cover distance $d \in [0,1]$?

drift via ➡ variance

$$\approx \frac{d}{\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right]}$$

[K.'19]

# Updating a Frequency (Genetic drift)



$\boldsymbol{p}_i^{(t)}$

» **What** happens if the samples are **not** informative?

- Each $\left\{\boldsymbol{x}_i^{(j)}\right\}_{j \in [\lambda]}$ and $\left\{\widehat{\boldsymbol{x}}_i^{(k)}\right\}_{k \in [\mu]}$ follows the same distribution

» Can happen regularly during a run

➡ **Removes** the tricky situation

$$\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \boldsymbol{p}_i^{(t)}$$

balanced

» Typical property of univariate EDAs

➡ $\boldsymbol{p}_i$ is a martingale and **approaches the borders quickly**

» Due to random fluctuations; despite a clear signal

- Genetic drift

» **How fast** does $\boldsymbol{p}_i$ cover distance $d \in [0,1]$?

drift via ➡ variance

$$\approx \frac{d}{\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right]}$$

[K.'19]

» This hitting time is **concentrated**

[Doerr, Zheng'20]

# Commonly Studied EDAs

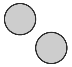
Compact Genetic Algorithm (cGA)
[Harik, Lobo, Goldberg'99]

Univariate Marginal Distribution Algorithm (UMDA)
[Mühlenbein, Paaß'96]

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]



» $\lambda = 2$ samples each iteration

## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]

» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

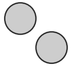
## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]



» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

» **Adjust** in favor of victor

## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]

» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

» **Adjust** in favor of victor

$$p_i^{(t+1)} = p_i^{(t)} + \frac{1}{K}\left(\hat{x}_i^{(1)} - \hat{x}_i^{(2)}\right)$$

hypothetical population size

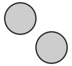
## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]



» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

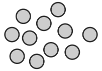
» **Adjust** in favor of victor

$$p_i^{(t+1)} = p_i^{(t)} + \frac{1}{K}\left(\hat{x}_i^{(1)} - \hat{x}_i^{(2)}\right)$$

hypothetical population size

## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]

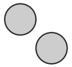


» $\lambda$ samples each iteration

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]



» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

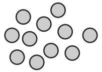
» **Adjust** in favor of victor

$$\boldsymbol{p}_i^{(t+1)} = \boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right)$$

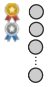
hypothetical population size

## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]



» $\lambda$ samples each iteration

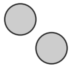


» Select $\mu \leq \lambda$ best samples, ranking them

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]



» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

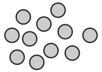
» **Adjust** in favor of victor

$$\boldsymbol{p}_i^{(t+1)} = \boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right)$$

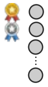
hypothetical population size

## Univariate Marginal Distribution Algorithm (UMDA)

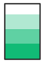
[Mühlenbein, Paaß'96]



» $\lambda$ samples each iteration



» Select $\mu \leq \lambda$ best samples, ranking them

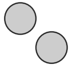


» **Set** to the relative number of 1s

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]



» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

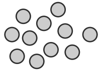
» **Adjust** in favor of victor

$$\boldsymbol{p}_i^{(t+1)} = \boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right)$$

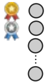
hypothetical population size

## Univariate Marginal Distribution Algorithm (UMDA)

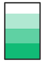
[Mühlenbein, Paaß'96]



» $\lambda$ samples each iteration



» Select $\mu \leq \lambda$ best samples, ranking them



» **Set** to the relative number of 1s

$$\boldsymbol{p}_i^{(t+1)} = \frac{1}{\mu}\sum_{k\in[\mu]} \widehat{\boldsymbol{x}}_i^{(k)}$$

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]

- » $\lambda = 2$ samples each iteration
- » Select $\mu = 2$ best samples, ranking them
- » **Adjust** in favor of victor

$$\boldsymbol{p}_i^{(t+1)} = \boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right)$$

hypothetical population size

## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]

- » $\lambda$ samples each iteration
- » Select $\mu \leq \lambda$ best samples, ranking them
- » **Set** to the relative number of 1s

$$\boldsymbol{p}_i^{(t+1)} = \frac{1}{\mu}\sum_{k\in[\mu]} \widehat{\boldsymbol{x}}_i^{(k)}$$

- » Population-based incremental learning [Baluja'**94**]
- » 2-Max-Min ant system with iteration-best update [Neumann, Sudholt, Witt'10]

# Commonly Studied EDAs

## Compact Genetic Algorithm (cGA)

[Harik, Lobo, Goldberg'99]

» $\lambda = 2$ samples each iteration

» Select $\mu = 2$ best samples, ranking them

» **Adjust** in favor of victor

$$\boldsymbol{p}_i^{(t+1)} = \boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right)$$

hypothetical population size ($K$)

## Univariate Marginal Distribution Algorithm (UMDA)

[Mühlenbein, Paaß'96]

» $\lambda$ samples each iteration

» Select $\mu \leq \lambda$ best samples, ranking them

» **Set** to the relative number of 1s

$$\boldsymbol{p}_i^{(t+1)} = \frac{1}{\mu}\sum_{k\in[\mu]} \widehat{\boldsymbol{x}}_i^{(k)}$$

» Population-based incremental learning [Baluja'**94**]

» 2-Max-Min ant system with iteration-best update [Neumann, Sudholt, Witt'10]

All are **balanced!** (starting with the uniform distribution)

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

## Univariate Marginal Distribution Algorithm (UMDA)

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

## Univariate Marginal Distribution Algorithm (UMDA)

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

$$= \frac{2}{K^2} \boldsymbol{p}_i{}^{(t)}\left(1 - \boldsymbol{p}_i{}^{(t)}\right)$$

## Univariate Marginal Distribution Algorithm (UMDA)

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

$$= \frac{2}{K^2}\, \boldsymbol{p}_i{}^{(t)}\left(1 - \boldsymbol{p}_i{}^{(t)}\right)$$

➡ $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with constant probability within $\Theta(K^2)$ iterations

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

$$= \frac{2}{K^2}\, \boldsymbol{p}_i{}^{(t)}\left(1 - \boldsymbol{p}_i{}^{(t)}\right)$$

➡ $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with constant probability within $\Theta(K^2)$ iterations

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\frac{1}{\mu}\sum_{k\in[\mu]} \widehat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

$$= \frac{2}{K^2} \boldsymbol{p}_i{}^{(t)}\left(1 - \boldsymbol{p}_i{}^{(t)}\right)$$

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\frac{1}{\mu}\sum_{k\in[\mu]} \hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu^2}\mathrm{Var}\left[\sum_{k\in[\mu]} \hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

➡ $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with constant probability within $\Theta(K^2)$ iterations

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

$$= \frac{2}{K^2}\, \boldsymbol{p}_i{}^{(t)}\left(1 - \boldsymbol{p}_i{}^{(t)}\right)$$

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\frac{1}{\mu}\sum_{k\in[\mu]} \hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu^2}\mathrm{Var}\left[\sum_{k\in[\mu]} \hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu^2}\sum_{k\in[\mu]} \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

➠ $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with constant probability within $\Theta(K^2)$ iterations

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} - \hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

$$= \frac{2}{K^2} \boldsymbol{p}_i{}^{(t)}\left(1 - \boldsymbol{p}_i{}^{(t)}\right)$$

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\frac{1}{\mu} \textstyle\sum_{k \in [\mu]} \hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu^2} \mathrm{Var}\left[\textstyle\sum_{k \in [\mu]} \hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu^2} \textstyle\sum_{k \in [\mu]} \mathrm{Var}\left[\hat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu} \boldsymbol{p}_i{}^{(t)}\left(1 - \boldsymbol{p}_i{}^{(t)}\right)$$

➡ $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with constant probability within $\Theta(K^2)$ iterations

# Commonly Studied EDAs (variances)

## Compact Genetic Algorithm (cGA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\boldsymbol{p}_i^{(t)} + \frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{K}\left(\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)}\right) \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} - \widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{K^2} \cdot \left(\mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(1)} \mid \boldsymbol{p}^{(t)}\right] + \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(2)} \mid \boldsymbol{p}^{(t)}\right]\right)$$

$$= \frac{2}{K^2}\, {\boldsymbol{p}_i}^{(t)}\left(1 - {\boldsymbol{p}_i}^{(t)}\right)$$

➡ $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with constant probability within $\Theta(K^2)$ iterations

## Univariate Marginal Distribution Algorithm (UMDA)

$$\mathrm{Var}\left[\boldsymbol{p}_i^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \mathrm{Var}\left[\frac{1}{\mu}\sum_{k\in[\mu]} \widehat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu^2}\mathrm{Var}\left[\sum_{k\in[\mu]} \widehat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu^2}\sum_{k\in[\mu]} \mathrm{Var}\left[\widehat{\boldsymbol{x}}_i^{(k)} \mid \boldsymbol{p}^{(t)}\right]$$

$$= \frac{1}{\mu}\, {\boldsymbol{p}_i}^{(t)}\left(1 - {\boldsymbol{p}_i}^{(t)}\right)$$

➡ $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with constant probability within $\Theta(\mu)$ iterations

# Theoretician's Trick

[Sudholt, Witt'16; Friedrich, Kötzing, K.'16; Doerr, Zheng'20]

- We wish genetic drift to be low for at least $\Theta(T)$ iterations

# Theoretician's Trick

[Sudholt, Witt'16; Friedrich, Kötzing, K.'16; Doerr, Zheng'20]

- We wish genetic drift to be low for at least $\Theta(T)$ iterations
- $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with probability at least $1 - \exp\left(-\Theta\left(\frac{1}{\mathrm{Var} \cdot T}\right)\right)$ [Doerr, Zheng'20]

# Theoretician's Trick

[Sudholt, Witt'16; Friedrich, Kötzing, K.'16; Doerr, Zheng'20]

- We wish genetic drift to be low for at least $\Theta(T)$ iterations
- $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with probability at least $1 - \exp\left(-\Theta\left(\frac{1}{\mathrm{Var} \cdot T}\right)\right)$ [Doerr, Zheng'20]

**Choose** algorithm parameters such that $\mathrm{Var} \in \Theta\left(\frac{1}{T \log n}\right)$

- Guarantee with high probability

# Theoretician's Trick

[Sudholt, Witt'16; Friedrich, Kötzing, K.'16; Doerr, Zheng'20]

- We wish genetic drift to be low for at least $\Theta(T)$ iterations
- $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with probability at least $1 - \exp\left(-\Theta\left(\frac{1}{\mathrm{Var} \cdot T}\right)\right)$ [Doerr, Zheng'20]

➡ **Choose** algorithm parameters such that $\mathrm{Var} \in \Theta\left(\frac{1}{T \log n}\right)$

- Guarantee with high probability

- Several times derived before, not fully rigorously [Thierens, Goldberg, Pereira'98; Lobo,Goldberg, Pelikan'00; Shapiro'05]

# Theoretician's Trick

[Sudholt, Witt'16; Friedrich, Kötzing, K.'16; Doerr, Zheng'20]

- We wish genetic drift to be low for at least $\Theta(T)$ iterations
- $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with probability at least $1 - \exp\left(-\Theta\left(\frac{1}{\mathrm{Var}\cdot T}\right)\right)$ [Doerr, Zheng'20]

➡ **Choose** algorithm parameters such that $\mathrm{Var} \in \Theta\left(\frac{1}{T \log n}\right)$

- Guarantee with high probability

- Several times derived before, not fully rigorously [Thierens, Goldberg, Pereira'98; Lobo,Goldberg, Pelikan'00; Shapiro'05]

- **Smart-restart strategy** for EDAs [Doerr, Zheng'20]
  - Choose some variance and run the EDA for $\Theta\left(\frac{1}{\mathrm{Var}}\right)$ iterations

# Theoretician's Trick

[Sudholt, Witt'16; Friedrich, Kötzing, K.'16; Doerr, Zheng'20]

- We wish genetic drift to be low for at least $\Theta(T)$ iterations
- $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with probability at least $1 - \exp\left(-\Theta\left(\frac{1}{\text{Var}\cdot T}\right)\right)$ [Doerr, Zheng'20]

➡ **Choose** algorithm parameters such that $\text{Var} \in \Theta\left(\frac{1}{T \log n}\right)$

- Guarantee with high probability

- Several times derived before, not fully rigorously [Thierens, Goldberg, Pereira'98; Lobo,Goldberg, Pelikan'00; Shapiro'05]

- **Smart-restart strategy** for EDAs [Doerr, Zheng'20]
  - Choose some variance and run the EDA for $\Theta\left(\frac{1}{\text{Var}}\right)$ iterations
  - If not happy, **halve the variance**, repeat the algorithm

# Theoretician's Trick

[Sudholt, Witt'16; Friedrich, Kötzing, K.'16; Doerr, Zheng'20]

» We wish genetic drift to be low for at least $\Theta(T)$ iterations

» $\boldsymbol{p}_i$ covers a distance of at most $\frac{1}{4}$ with probability at least $1 - \exp\left(-\Theta\left(\frac{1}{\mathrm{Var}\cdot T}\right)\right)$ [Doerr, Zheng'20]

➡ **Choose** algorithm parameters such that $\mathrm{Var} \in \Theta\left(\frac{1}{T \log n}\right)$

» Guarantee with high probability

» Several times derived before, not fully rigorously [Thierens, Goldberg, Pereira'98; Lobo,Goldberg, Pelikan'00; Shapiro'05]

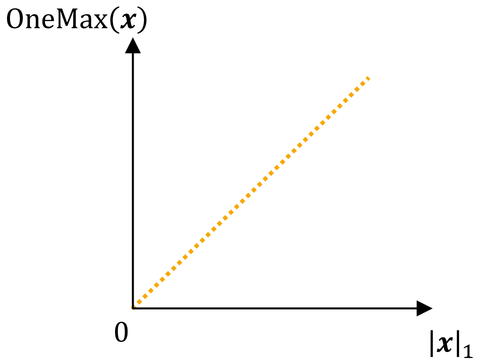
» **Smart-restart strategy** for EDAs [Doerr, Zheng'20]

- Choose some variance and run the EDA for $\Theta\left(\frac{1}{\mathrm{Var}}\right)$ iterations
- If not happy, **halve the variance**, repeat the algorithm

» The trick carries over to weak preferences of bit values [Doerr, Zheng'20]

# Commonly Studied Benchmarks

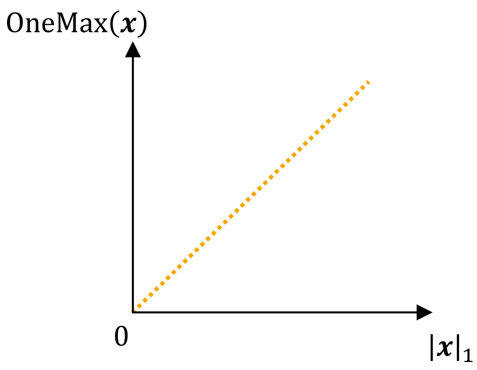
OneMax




$$\boldsymbol{x} \mapsto \sum_{i \in [n]} \boldsymbol{x}_i =: |\boldsymbol{x}|_1$$

# Commonly Studied Benchmarks

OneMax




LeadingOnes

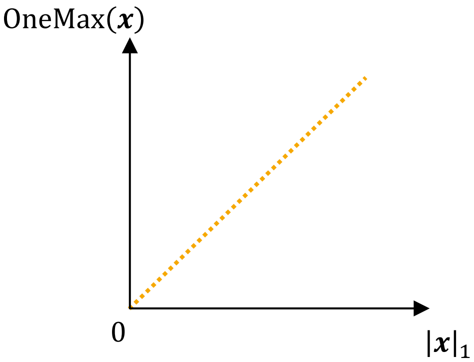
110000011010

111110000100

$\boldsymbol{x} \mapsto \sum_{i\in[n]} \boldsymbol{x}_i =: |\boldsymbol{x}|_1$ $\quad$ $\boldsymbol{x} \mapsto \max\{i \in [0..n] \mid \forall j \in [i]: \boldsymbol{x}_j = 1\}$

# Commonly Studied Benchmarks

## OneMax




$$\boldsymbol{x} \mapsto \sum_{i \in [n]} \boldsymbol{x}_i =: |\boldsymbol{x}|_1$$

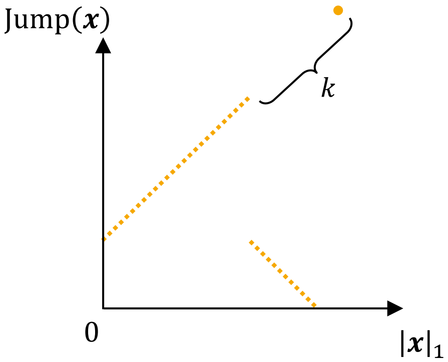
## LeadingOnes

110000011010

111110000100

$$\boldsymbol{x} \mapsto \max\{i \in [0..n] \mid \forall j \in [i]: \boldsymbol{x}_j = 1\}$$

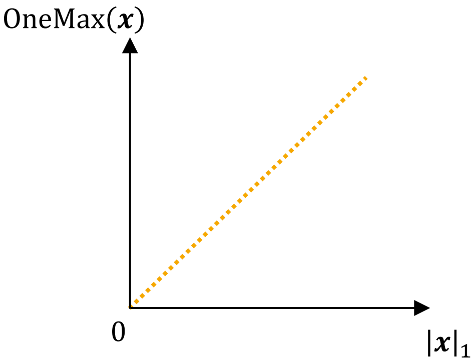
## Jump




$$\boldsymbol{x} \mapsto \begin{cases} k + |\boldsymbol{x}|_1 & \text{if } |\boldsymbol{x}|_1 \in [n-k] \cup \{n\} \\ n - |\boldsymbol{x}|_1 & \text{else} \end{cases}$$

# Commonly Studied Benchmarks

## OneMax




$\boldsymbol{x} \mapsto \sum_{i \in [n]} \boldsymbol{x}_i =: |\boldsymbol{x}|_1$

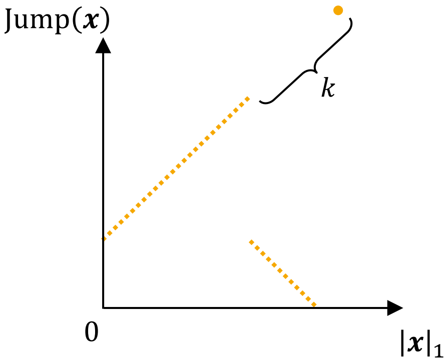
## LeadingOnes

110000011010

111110000100

$\boldsymbol{x} \mapsto \max\{i \in [0..n] \mid \forall j \in [i]: \boldsymbol{x}_j = 1\}$

## Jump




$\boldsymbol{x} \mapsto \begin{cases} k + |\boldsymbol{x}|_1 & \text{if } |\boldsymbol{x}|_1 \in [n-k] \cup \{n\} \\ n - |\boldsymbol{x}|_1 & \text{else} \end{cases}$

## BinVal

$[110000011010]_2$

$[111110000100]_2$

$\boldsymbol{x} \mapsto \sum_{i \in [n]} 2^{n-i} \cdot \boldsymbol{x}_i$

# Commonly Studied Benchmarks

| | OneMax | LeadingOnes | Jump | BinVal |
|---|---|---|---|---|
| cGA | ✔ [Droste'06; Lengler, Sudholt, Witt'18; Sudholt, Witt'19] | ✖ | ✔ [Hasenöhrl, Sutton'18; Doerr'21; Witt'23] | ✔ [Droste'06; Witt'18] |
| UMDA | ✔ [Dang, Lehre, Nguyen'18; Witt'19; K., Witt'20] | ✔ [Dang, Lehre, Nguyen'18; Doerr, K.'21] | ✖ | ✔ [Dang, Lehre, Nguyen'18] |

# Commonly Studied Benchmarks

| | OneMax | LeadingOnes | Jump | BinVal |
|---|---|---|---|---|
| cGA | ✓ [Droste'06; Lengler, Sudholt, Witt'18; Sudholt, Witt'19] | ✗ | ✓ [Hasenöhrl, Sutton'18; Doerr'21; Witt'23] | ✓ [Droste'06; Witt'18] |
| UMDA | ✓ [Dang, Lehre, Nguyen'18; Witt'19; K., Witt'20] | ✓ [Dang, Lehre, Nguyen'18; Doerr, K.'21] | ✗ | ✓ [Dang, Lehre, Nguyen'18] |

**Also:** (non-exhaustive)

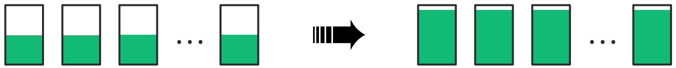
» Analyses on noise [Friedrich, Kötzing, K., Witt'17; Lehre, Nguyen'19; Lehre, Nguyen'21; Kötzing, Radhakrishnan'22]

» Analyses on deception [Lehre, Nguyen'19; Doerr, K.'21]

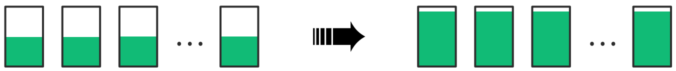
» New EDAs [Doerr, K.'20; Ajimakin, Devi'23]

» Multi-valued EDAs [Ben Jedidia, Doerr, K.'24; Adak, Witt'24]

# cGA on OneMax

Our aim: 

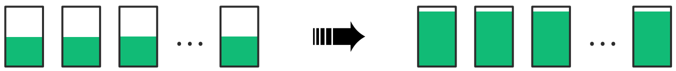
# cGA on OneMax

Our aim: 

» Choose $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$

» And other mild assumptions, such as that $K$ is polynomial

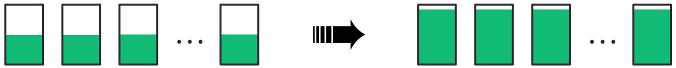
# cGA on OneMax

Our aim: 

» Choose $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$ » And other mild assumptions, such as that $K$ is polynomial

**All** frequencies are with high probability at least $\frac{1}{4}$ for at least $T$ iterations » We determine $T$ later

# cGA on OneMax

Our aim: 

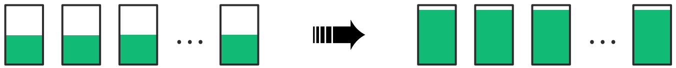
» Choose $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$ » And other mild assumptions, such as that $K$ is polynomial

➡ **All** frequencies are with high probability at least $\frac{1}{4}$ for at least $T$ iterations » We determine $T$ later

» Consider the potential $\varphi^{(t)} \coloneqq \sum_{i \in [n]} \left(1 - \boldsymbol{p}_i^{(t)}\right)$, which we aim to **minimize**

❖ $\mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \sum_{i \in [n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$

# cGA on OneMax

Our aim: 

» Choose $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$

» And other mild assumptions, such as that $K$ is polynomial

➡ **All** frequencies are with high probability at least $\frac{1}{4}$ for at least $T$ iterations

» We determine $T$ later

» Consider the potential $\varphi^{(t)} \coloneqq \sum_{i \in [n]} \left(1 - \boldsymbol{p}_i^{(t)}\right)$, which we aim to **minimize**

» This potential holds a secret we will uncover soon

❖ $\mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \sum_{i \in [n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$

# cGA on OneMax (informative bits)

- We bound $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}_i^{(t)}\right]$

# cGA on OneMax (informative bits)

» We bound $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}_i^{(t)}\right]$

- The bits at position $i$ are not always **informative**

# cGA on OneMax (informative bits)

- We bound $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}_i^{(t)}\right]$
  - The bits at position $i$ are not always **informative**

➡ When they are not informative, the expected change is $0$

- For simplicity, assume no bounds

# cGA on OneMax (informative bits)

- We bound $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}_i^{(t)}\right]$
  - The bits at position $i$ are not always **informative**

➡ When they are not informative, the expected change is 0

- For simplicity, assume no bounds

- **When** are bits informative?

$\dot{\boldsymbol{x}}^{(1)}$ 1100 ? 11110

$\dot{\boldsymbol{x}}^{(2)}$ 0011 ? 00011

$i$

# cGA on OneMax (informative bits)

- We bound $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}_i^{(t)}\right]$
  - The bits at position $i$ are not always **informative**

➡ When they are not informative, the expected change is 0

- For simplicity, assume no bounds

- **When** are bits informative?

$\dot{\boldsymbol{x}}^{(1)}$ 1100 ? 11110
$\dot{\boldsymbol{x}}^{(2)}$ 0011 ? 00011
$i$

➡ Since $\underbrace{\left|\left|\dot{\boldsymbol{x}}^{(1)}\right|_1 - \left|\dot{\boldsymbol{x}}^{(2)}\right|_1\right|}_{=:D} \geq 2$, position $i$ is not informative

# cGA on OneMax (informative bits)

» We bound $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}_i^{(t)}\right]$

❖ The bits at position $i$ are not always **informative**

➡ When they are not informative, the expected change is 0

» For simplicity, assume no bounds

» **When** are bits informative?

$\dot{\boldsymbol{x}}^{(1)}$ 1100 ? 11110

$\dot{\boldsymbol{x}}^{(2)}$ 0011 ? 00011

$i$

➡ Since $\underbrace{\left|\left|\dot{\boldsymbol{x}}^{(1)}\right|_1 - \left|\dot{\boldsymbol{x}}^{(2)}\right|_1\right|}_{=:\,D} \geq 2$, position $i$ is not informative

➡ $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \geq \frac{2}{K}\boldsymbol{p}_i^{(t)}\left(1 - \boldsymbol{p}_i^{(t)}\right) \cdot \Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$

» Ignoring the case $D = 1$ does not change the result asymptotically here

# cGA on OneMax (the sampling variance)

» **How** to bound $\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$?

# cGA on OneMax (the sampling variance)

- **How** to bound $\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$?
  - Let $X = \sum_{j \in [n] \setminus \{i\}} \boldsymbol{x}_i^{(1)}$ and $\mathrm{Y} = \sum_{j \in [n] \setminus \{i\}} \boldsymbol{x}_i^{(2)}$

# cGA on OneMax (the sampling variance)

» **How** to bound $\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$?

❖ Let $X = \sum_{j\in[n]\setminus\{i\}} \boldsymbol{x}_i^{(1)}$ and $\mathrm{Y} = \sum_{j\in[n]\setminus\{i\}} \boldsymbol{x}_i^{(2)}$

➡ $X$ and $Y$ follow independent Poisson-binomial distributions

## cGA on OneMax (the sampling variance)

- **How** to bound $\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$?
  - Let $X = \sum_{j\in[n]\setminus\{i\}} \boldsymbol{x}_i^{(1)}$ and $\mathrm{Y} = \sum_{j\in[n]\setminus\{i\}} \boldsymbol{x}_i^{(2)}$

➡ $X$ and $Y$ follow independent Poisson-binomial distributions

- These are concentrated around their expectation within their standard deviation $\sigma$

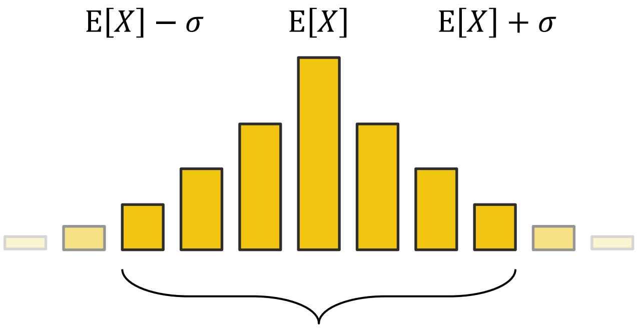
# cGA on OneMax (the sampling variance)

» **How** to bound $\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$?

❖ Let $X = \sum_{j \in [n] \setminus \{i\}} \boldsymbol{x}_i^{(1)}$ and $\mathrm{Y} = \sum_{j \in [n] \setminus \{i\}} \boldsymbol{x}_i^{(2)}$

⇒ $X$ and $Y$ follow independent Poisson-binomial distributions

» These are concentrated around their expectation within their standard deviation $\sigma$




Constant fraction of the probability mass
in interval $I \coloneqq [\mathrm{E}[X] - \sigma \,..\, \mathrm{E}[X] + \sigma\,]$

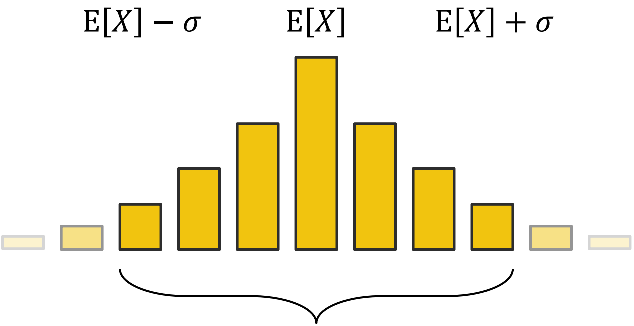
# cGA on OneMax (the sampling variance)

» **How** to bound $\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$?

❖ Let $X = \sum_{j\in[n]\setminus\{i\}} \boldsymbol{x}_i^{(1)}$ and $\mathrm{Y} = \sum_{j\in[n]\setminus\{i\}} \boldsymbol{x}_i^{(2)}$

➡ $X$ and $Y$ follow independent Poisson-binomial distributions

» These are concentrated around their expectation within their standard deviation $\sigma$




Constant fraction of the probability mass in interval $I \coloneqq [\mathrm{E}[X] - \sigma \,..\, \mathrm{E}[X] + \sigma\,]$

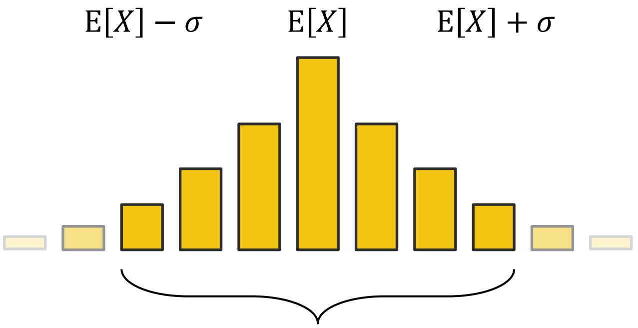
➡ The probability of each outcome in $I$ is very roughly the same, that is, $\frac{1}{|I|} \approx \frac{1}{\sigma}$

# cGA on OneMax (the sampling variance)

- **How** to bound $\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right]$?
  - Let $X = \sum_{j \in [n] \setminus \{i\}} \boldsymbol{x}_i^{(1)}$ and $Y = \sum_{j \in [n] \setminus \{i\}} \boldsymbol{x}_i^{(2)}$

➡ $X$ and $Y$ follow independent Poisson-binomial distributions

- These are concentrated around their expectation within their standard deviation $\sigma$




Constant fraction of the probability mass in interval $I \coloneqq [\mathrm{E}[X] - \sigma \,..\, \mathrm{E}[X] + \sigma]$

➡ The probability of each outcome in $I$ is very roughly the same, that is, $\frac{1}{|I|} \approx \frac{1}{\sigma}$

- Note that $\sigma = \sqrt{\mathrm{Var}\left[X \mid \boldsymbol{p}^{(t)}\right]} = \sqrt{\sum_{j \in [n] \setminus \{i\}} \boldsymbol{p}_j^{(t)} \left(1 - \boldsymbol{p}_j^{(t)}\right)}$

sampling variance

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$

$$= \sum_{j \in [n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} q_j^2$$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$

$$= \sum_{j \in [n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$

$$= \sum_{j \in [n-1]} q_j^2$$

$$\geq \sum_{j \in I} q_j^2$$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} q_j^2$$
$$\geq \sum_{j\in I} q_j^2$$

Jensen's inequality: $\frac{\sum_{k\in L} a_k^2}{|L|} \geq \left(\frac{\sum_{k\in L} a_k}{|L|}\right)^2$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} q_j^2$$
$$\geq \sum_{j\in I} q_j^2$$

Jensen's inequality: $\frac{\sum_{k\in L} a_k^2}{|L|} \geq \left(\frac{\sum_{k\in L} a_k}{|L|}\right)^2$

$$\Longrightarrow \sum_{k\in L} a_k^2 \geq \frac{\left(\sum_{k\in L} a_k\right)^2}{|L|}$$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j \in [n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j \in [n-1]} q_j^2$$
$$\geq \sum_{j \in I} q_j^2$$

(Jensen)
$$\geq \frac{\left(\sum_{j \in I} q_j\right)^2}{|I|}$$

Jensen's inequality: $\frac{\sum_{k \in L} a_k^2}{|L|} \geq \left(\frac{\sum_{k \in L} a_k}{|L|}\right)^2$

$\Rightarrow$ $\sum_{k \in L} a_k^2 \geq \frac{\left(\sum_{k \in L} a_k\right)^2}{|L|}$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$
$$= \sum_{j\in[n-1]} q_j^2$$
$$\geq \sum_{j\in I} q_j^2$$

(Jensen)

$$\geq \frac{\left(\sum_{j\in I} q_j\right)^2}{|I|}$$
$$\geq \frac{\left(\Pr\left[|X-\mathrm{E}[X]|<\sigma+1 \mid \boldsymbol{p}^{(t)}\right]\right)^2}{|I|}$$

Jensen's inequality: $\frac{\sum_{k\in L} a_k^2}{|L|} \geq \left(\frac{\sum_{k\in L} a_k}{|L|}\right)^2$

$\Rightarrow \sum_{k\in L} a_k^2 \geq \frac{\left(\sum_{k\in L} a_k\right)^2}{|L|}$

## cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] = \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right]$$

$$= \sum_{j \in [n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right]$$

$$= \sum_{j \in [n-1]} q_j^2$$

$$\geq \sum_{j \in I} q_j^2$$

(Jensen)

$$\geq \frac{\left(\sum_{j \in I} q_j\right)^2}{|I|}$$

(Chebyshev)

$$\geq \frac{\left(\Pr\left[|X - \mathrm{E}[X]| < \sigma + 1 \mid \boldsymbol{p}^{(t)}\right]\right)^2}{|I|}$$

$$\gtrsim \frac{1}{|I|}$$

Jensen's inequality:

$$\frac{\sum_{k \in L} a_k^2}{|L|} \geq \left(\frac{\sum_{k \in L} a_k}{|L|}\right)^2$$

$$\Rightarrow \sum_{k \in L} a_k^2 \geq \frac{\left(\sum_{k \in L} a_k\right)^2}{|L|}$$

# cGA on OneMax (the sampling variance) [Neumann, Sudholt, Witt'10; Alistair, Lehre'24]

$$
\begin{aligned}
\Pr\left[D = 0 \mid \boldsymbol{p}^{(t)}\right] &= \Pr\left[X = Y \mid \boldsymbol{p}^{(t)}\right] \\
&= \sum_{j\in[n-1]} \Pr\left[X = j \wedge Y = j \mid \boldsymbol{p}^{(t)}\right] \\
&= \sum_{j\in[n-1]} q_j^2 \\
&\geq \sum_{j\in I} q_j^2 \\
\text{(Jensen)} \quad &\geq \frac{\left(\sum_{j\in I} q_j\right)^2}{|I|} \\
&\geq \frac{\left(\Pr\left[|X-\mathrm{E}[X]|<\sigma+1 \mid \boldsymbol{p}^{(t)}\right]\right)^2}{|I|} \\
\text{(Chebyshev)} \quad &\gtrsim \frac{1}{|I|} \\
&\approx \frac{1}{\sigma}
\end{aligned}
$$

Jensen's inequality: $\frac{\sum_{k\in L} a_k^2}{|L|} \geq \left(\frac{\sum_{k\in L} a_k}{|L|}\right)^2$

$\Rightarrow \quad \sum_{k\in L} a_k^2 \geq \frac{\left(\sum_{k\in L} a_k\right)^2}{|L|}$

# cGA on OneMax

- Let $\sigma_i = \sqrt{\sum_{j \in [n] \setminus \{i\}} \boldsymbol{p}_j^{(t)} \left(1 - \boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)}$

# cGA on OneMax

- Let $\sigma_i = \sqrt{\sum_{j \in [n] \setminus \{i\}} \boldsymbol{p}_j^{(t)} \left(1 - \boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)}$
- We have $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{2}{K\sigma_i} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right) \geq \frac{2}{K\sigma} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)$

# cGA on OneMax

- Let $\sigma_i = \sqrt{\sum_{j\in[n]\setminus\{i\}} \boldsymbol{p}_j^{(t)}\left(1-\boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)}$
- We have $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{2}{K\sigma_i}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right) \geq \frac{2}{K\sigma}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)$

⇒ $\mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \sum_{i\in[n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$

## cGA on OneMax

- Let $\sigma_i = \sqrt{\sum_{j\in[n]\setminus\{i\}} \boldsymbol{p}_j^{(t)}\left(1-\boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)}$
- We have $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{2}{K\sigma_i}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right) \geq \frac{2}{K\sigma}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)$

$$\Rightarrow \quad \mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \sum_{i\in[n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

$$\gtrsim \frac{2}{K\sigma}\sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)$$

# cGA on OneMax

» Let $\sigma_i = \sqrt{\sum_{j\in[n]\setminus\{i\}} \boldsymbol{p}_j^{(t)}\left(1-\boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)}$

» We have $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{2}{K\sigma_i}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right) \geq \frac{2}{K\sigma}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)$

$$\begin{aligned}
\Rightarrow \quad \mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] &= \sum_{i\in[n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \\
&\gtrsim \frac{2}{K\sigma}\sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right) \\
&= \frac{2}{K}\sigma
\end{aligned}$$

# cGA on OneMax

» Let $\sigma_i = \sqrt{\sum_{j \in [n] \setminus \{i\}} \boldsymbol{p}_j^{(t)} \left(1 - \boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)}$

» We have $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{2}{K\sigma_i} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right) \geq \frac{2}{K\sigma} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)$

$$\Rightarrow \quad \mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \sum_{i \in [n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

$$\gtrsim \frac{2}{K\sigma} \sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)$$

$$= \frac{2}{K} \sigma$$

» Note that since $\boldsymbol{p}_i^{(t)} \geq \frac{1}{4}$, we have $\sigma^2 = \sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right) \geq \frac{1}{4} \varphi^{(t)}$

» The secret property of $\varphi$!
For one-sided bounded frequencies, sampling variance and distance to the optimal model are roughly the same!

# cGA on OneMax

- Let $\sigma_i = \sqrt{\sum_{j\in[n]\setminus\{i\}} \boldsymbol{p}_j^{(t)}\left(1-\boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)}$
- We have $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{2}{K\sigma_i}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right) \ge \frac{2}{K\sigma}\boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)$

$$\Rightarrow \quad \mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \sum_{i\in[n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

$$\gtrsim \frac{2}{K\sigma}\sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)$$

$$= \frac{2}{K}\sigma$$

- Note that since $\boldsymbol{p}_i^{(t)} \ge \frac{1}{4}$, we have $\sigma^2 = \sum_{i\in[n]} \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right) \ge \frac{1}{4}\varphi^{(t)}$

$$\Rightarrow \quad \mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{1}{K}\sqrt{\varphi^{(t)}}$$

- The secret property of $\varphi$!
  For one-sided bounded frequencies, sampling variance and distance to the optimal model are roughly the same!

# cGA on OneMax

- Let $\sigma_i = \sqrt{\sum_{j \in [n] \setminus \{i\}} \boldsymbol{p}_j^{(t)} \left(1 - \boldsymbol{p}_j^{(t)}\right)}$ and $\sigma = \sqrt{\sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)}$
- We have $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{2}{K\sigma_i} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right) \geq \frac{2}{K\sigma} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)$

$$\Rightarrow \quad \mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] = \sum_{i \in [n]} \mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right]$$

$$\gtrsim \frac{2}{K\sigma} \sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right)$$

$$= \frac{2}{K} \sigma$$

- Note that since $\boldsymbol{p}_i^{(t)} \geq \frac{1}{4}$, we have $\sigma^2 = \sum_{i \in [n]} \boldsymbol{p}_i^{(t)} \left(1 - \boldsymbol{p}_i^{(t)}\right) \geq \frac{1}{4} \varphi^{(t)}$

- The secret property of $\varphi$!
  For one-sided bounded frequencies, sampling variance and distance to the optimal model are roughly the same!

$$\Rightarrow \quad \mathrm{E}\left[\varphi^{(t)} - \varphi^{(t+1)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \frac{1}{K} \sqrt{\varphi^{(t)}} \quad \xrightarrow[\text{drift}]{\text{variable}} \quad \mathrm{E}[T] \lesssim K\sqrt{n}$$

# cGA on OneMax

- We have $\mathrm{E}[T] \lesssim K\sqrt{n}$ under the assumption $K \in \Omega\left(\sqrt{\mathrm{E}[T]\log n}\right)$

# cGA on OneMax

- We have $\mathrm{E}[T] \lesssim K\sqrt{n}$ under the assumption $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$
  - Solving for $K$ yields $K \in \Omega(\sqrt{n} \log n)$

# cGA on OneMax

- We have $\mathrm{E}[T] \lesssim K\sqrt{n}$ under the assumption $K \in \Omega\left(\sqrt{\mathrm{E}[T]\log n}\right)$
  - Solving for $K$ yields $K \in \Omega(\sqrt{n}\log n)$

➡ The cGA with $K \in \Omega(\sqrt{n}\log n)$ maximizes OneMax in $\mathrm{O}(K\sqrt{n})$ iterations in expectation, which is $\mathrm{O}(n\log n)$ for minimal $K$.

- And other mild assumptions, such as that $K$ is polynomial

# cGA on OneMax

- We have $\mathrm{E}[T] \lesssim K\sqrt{n}$ under the assumption $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$
  - Solving for $K$ yields $K \in \Omega(\sqrt{n} \log n)$

➡ The cGA with $K \in \Omega(\sqrt{n} \log n)$ maximizes OneMax in $\mathrm{O}(K\sqrt{n})$ iterations in expectation, which is $\mathrm{O}(n \log n)$ for minimal $K$.

- And other mild assumptions, such as that $K$ is polynomial

! The constraint on $K$ is due to the variance of a single frequency, and the factor $\sqrt{n}$ can be traced back to the sampling variance!

# cGA on OneMax

- We have $\mathrm{E}[T] \lesssim K\sqrt{n}$ under the assumption $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$
  - Solving for $K$ yields $K \in \Omega(\sqrt{n} \log n)$

➡ The cGA with $K \in \Omega(\sqrt{n} \log n)$ maximizes OneMax in $\mathrm{O}(K\sqrt{n})$ iterations in expectation, which is $\mathrm{O}(n \log n)$ for minimal $K$.

- And other mild assumptions, such as that $K$ is polynomial

The constraint on $K$ is due to the variance of a single frequency, and the factor $\sqrt{n}$ can be traced back to the sampling variance!

- Almost matching lower bounds [Droste'06] and **matching lower bounds** [Sudholt, Witt'19] exist

# cGA on OneMax

- We have $\mathrm{E}[T] \lesssim K\sqrt{n}$ under the assumption $K \in \Omega\left(\sqrt{\mathrm{E}[T]\log n}\right)$
  - Solving for $K$ yields $K \in \Omega(\sqrt{n}\log n)$

➡ The cGA with $K \in \Omega(\sqrt{n}\log n)$ maximizes OneMax in $O(K\sqrt{n})$ iterations in expectation, which is $O(n\log n)$ for minimal $K$.

- And other mild assumptions, such as that $K$ is polynomial

The constraint on $K$ is due to the variance of a single frequency, and the factor $\sqrt{n}$ can be traced back to the sampling variance!

- Almost matching lower bounds [Droste'06] and **matching lower bounds** [Sudholt, Witt'19] exist
  - Sudholt and Witt approximate the potential of the sum of frequencies via a normal distribution

# cGA on OneMax

- We have $\mathrm{E}[T] \lesssim K\sqrt{n}$ under the assumption $K \in \Omega\left(\sqrt{\mathrm{E}[T] \log n}\right)$
  - Solving for $K$ yields $K \in \Omega(\sqrt{n} \log n)$

➡ The cGA with $K \in \Omega(\sqrt{n} \log n)$ maximizes OneMax in $O(K\sqrt{n})$ iterations in expectation, which is $O(n \log n)$ for minimal $K$.

- And other mild assumptions, such as that $K$ is polynomial

The constraint on $K$ is due to the variance of a single frequency, and the factor $\sqrt{n}$ can be traced back to the sampling variance!

- Almost matching lower bounds [Droste'06] and **matching lower bounds** [Sudholt, Witt'19] exist
  - Sudholt and Witt approximate the potential of the sum of frequencies via a normal distribution
- For $K \in O\left(\frac{\sqrt{n}}{\log(n) \log \log n}\right)$ the run time is in $\Omega\left(K^{1/3}n + n \log n\right)$ [Lengler, Sudholt, Witt'18]
  - Covers **high** genetic drift

# UMDA on OneMax (with borders)

[Witt'18] 1. The UMDA with $\mu \in \Omega(\sqrt{n} \log n)$ and $\lambda = (1 + \Theta(1))\mu$ optimizes OneMax in $O(\lambda\sqrt{n})$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.

- **Low** genetic drift.
- Holds with high probability also without borders.

# UMDA on OneMax (with borders)

[Witt'18]

1. The UMDA with $\mu \in \Omega(\sqrt{n}\log n)$ and $\lambda = \big(1 + \Theta(1)\big)\mu$ optimizes OneMax in $O(\lambda\sqrt{n})$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.

2. The UMDA with $\mu \in \Omega(\log n) \cap O(n^{1-\varepsilon})$ and $\lambda = \big(1 + \Theta(1)\big)\mu$ optimizes OneMax in $O(\lambda n)$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.

» **Low** genetic drift.

» Holds with high probability also without borders.

» **High** genetic drift.

» Only holds **with** borders, otherwise infinite even whp.

# UMDA on OneMax (with borders)

[Witt'18]

1. The UMDA with $\mu \in \Omega(\sqrt{n}\log n)$ and $\lambda = (1 + \Theta(1))\mu$ optimizes OneMax in $O(\lambda\sqrt{n})$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.
   - » **Low** genetic drift.
   - » Holds with high probability also without borders.

2. The UMDA with $\mu \in \Omega(\log n) \cap O(n^{1-\varepsilon})$ and $\lambda = (1 + \Theta(1))\mu$ optimizes OneMax in $O(\lambda n)$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.
   - » **High** genetic drift.
   - » Only holds **with** borders, otherwise infinite even whp.

[Lehre, Nguyen'18]

The UMDA with $\mu \in \Omega(\log n) \cap O(\sqrt{n})$ and $\lambda = (1 + \Omega(1))\mu$ optimizes OneMax in $O(\lambda n)$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.

- » **High** genetic drift.

# UMDA on OneMax (with borders)

[Witt'18]

1. The UMDA with $\mu \in \Omega(\sqrt{n}\log n)$ and $\lambda = \big(1 + \Theta(1)\big)\mu$ optimizes OneMax in $O(\lambda\sqrt{n})$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.

   » **Low** genetic drift.

   » Holds with high probability also without borders.

2. The UMDA with $\mu \in \Omega(\log n) \cap O(n^{1-\varepsilon})$ and $\lambda = \big(1 + \Theta(1)\big)\mu$ optimizes OneMax in $O(\lambda n)$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.

   » **High** genetic drift.

   » Only holds **with** borders, otherwise infinite even whp.

[Lehre, Nguyen'18]

The UMDA with $\mu \in \Omega(\log n) \cap O(\sqrt{n})$ and $\lambda = \big(1 + \Omega(1)\big)\mu$ optimizes OneMax in $O(\lambda n)$ function evaluations in expectation. This is $O(n \log n)$ for minimal $\lambda$.

» **High** genetic drift.

» Analysis by Witt very similar to the cGA, but some steps are vastly different

- It analogously holds that $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \boldsymbol{p}_i^{(t)}\left(1 - \boldsymbol{p}_i^{(t)}\right)$, but for different reasons, involving 1st- and 2nd-class individuals

# UMDA on OneMax (with borders)

[Witt'18] 1. The UMDA with $\mu \in \Omega(\sqrt{n}\log n)$ and $\lambda = (1+\Theta(1))\mu$ optimizes OneMax in $O(\lambda\sqrt{n})$ function evaluations in expectation. This is $O(n\log n)$ for minimal $\lambda$.

- **Low** genetic drift.
- Holds with high probability also without borders.

2. The UMDA with $\mu \in \Omega(\log n) \cap O(n^{1-\varepsilon})$ and $\lambda = (1+\Theta(1))\mu$ optimizes OneMax in $O(\lambda n)$ function evaluations in expectation. This is $O(n\log n)$ for minimal $\lambda$.

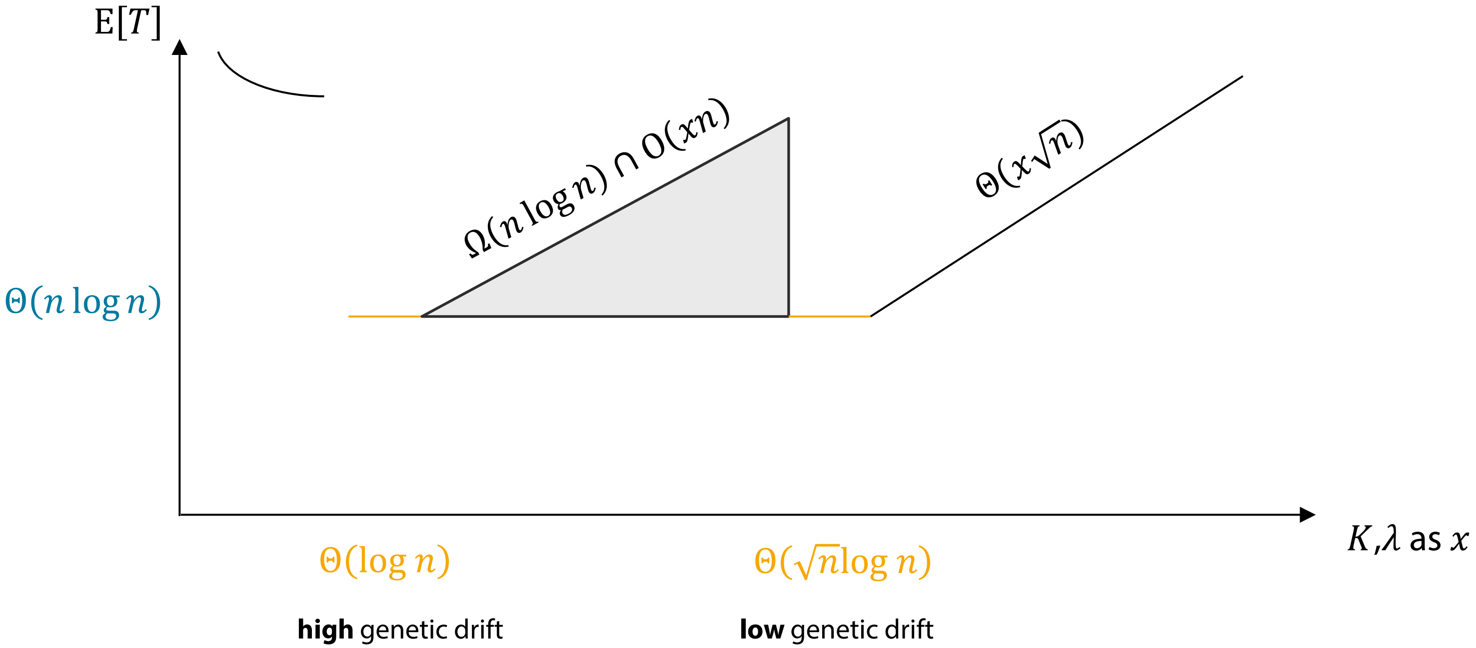
- **High** genetic drift.
- Only holds **with** borders, otherwise infinite even whp.

[Lehre, Nguyen'18] The UMDA with $\mu \in \Omega(\log n) \cap O(\sqrt{n})$ and $\lambda = (1+\Omega(1))\mu$ optimizes OneMax in $O(\lambda n)$ function evaluations in expectation. This is $O(n\log n)$ for minimal $\lambda$.

- **High** genetic drift.

- Analysis by Witt very similar to the cGA, but some steps are vastly different
  - It analogously holds that $\mathrm{E}\left[\boldsymbol{p}_i^{(t+1)} - \boldsymbol{p}_i^{(t)} \mid \boldsymbol{p}^{(t)}\right] \gtrsim \boldsymbol{p}_i^{(t)}\left(1-\boldsymbol{p}_i^{(t)}\right)$, but for different reasons, involving 1st- and 2nd-class individuals
- Analysis by Lehre and Nguyen simpler, utilizing a **level-based method**

# cGA and UMDA on OneMax



# cGA and UMDA on OneMax