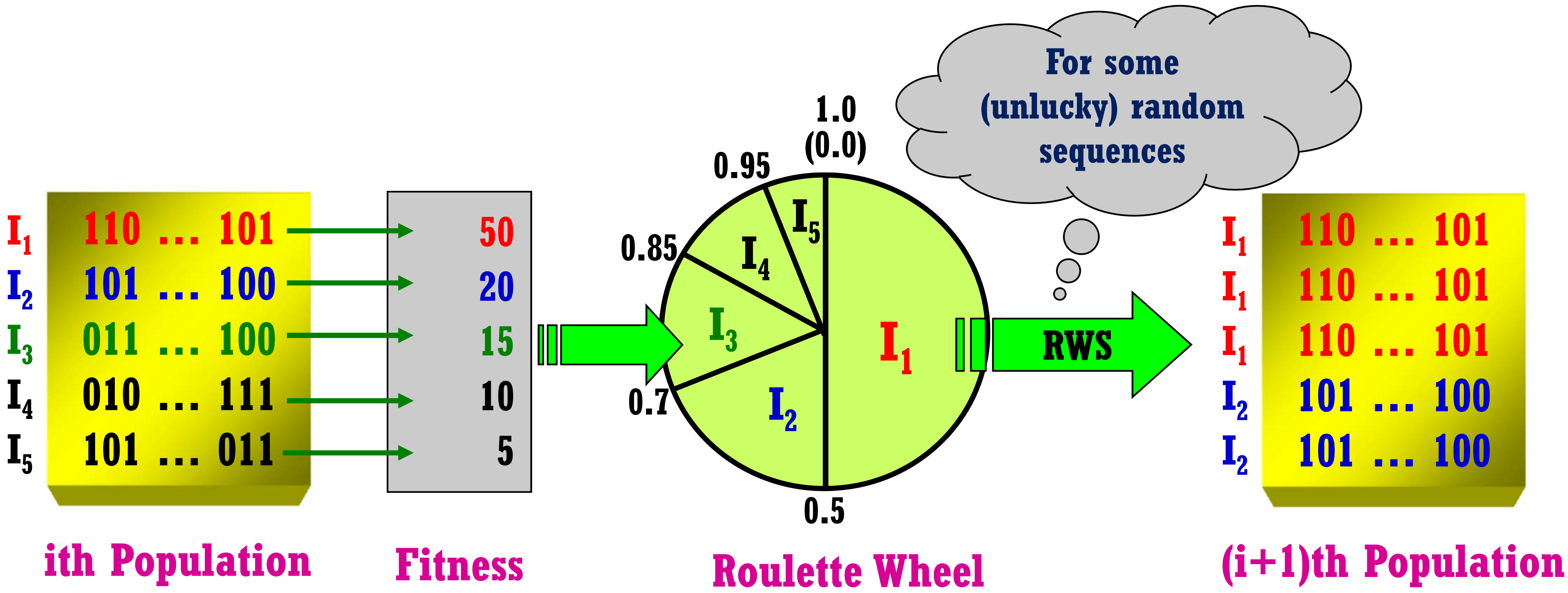
Further Detailed Investigation on Selection

Proportional Selection (1)

Selection Noise

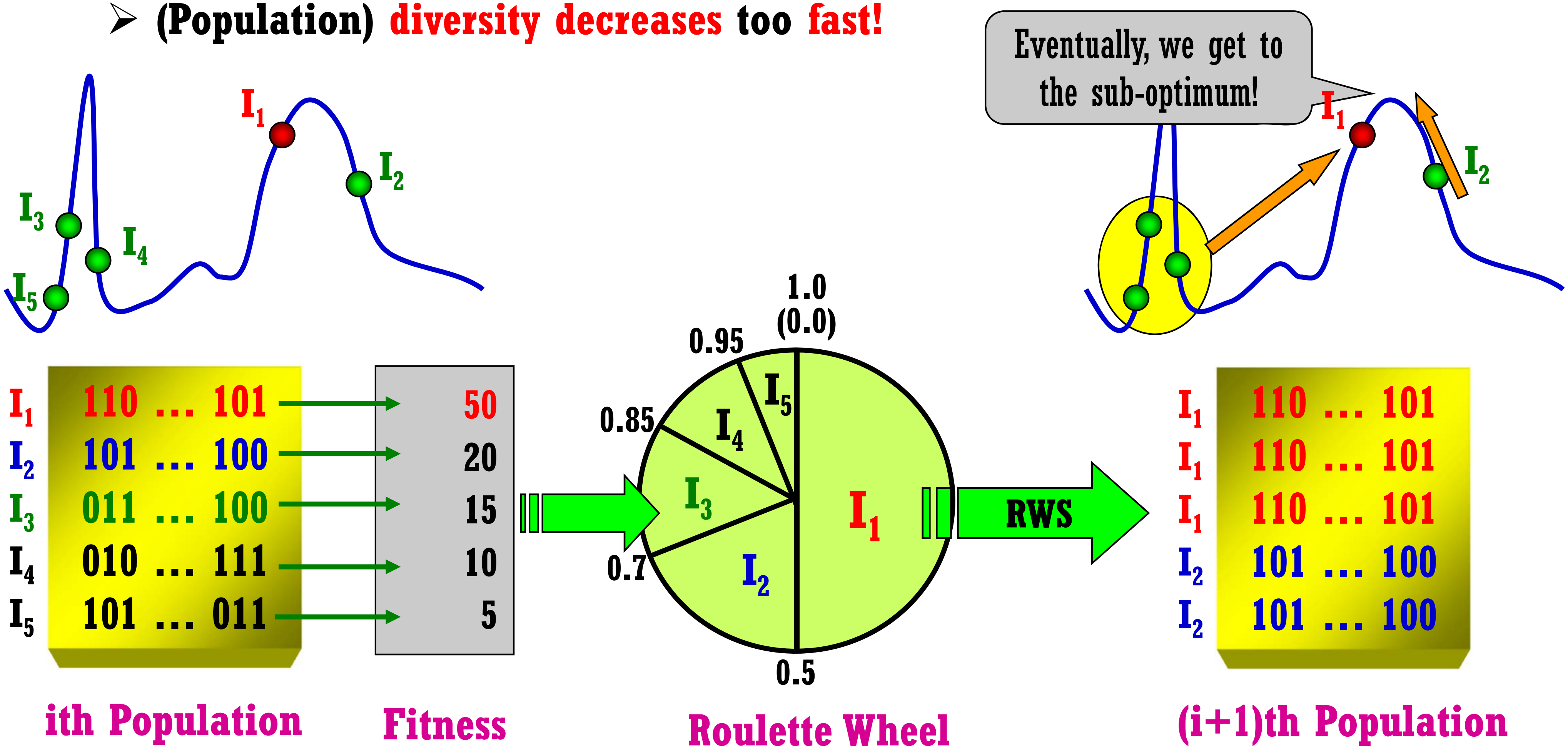
- **What problem exists in the RWS ?**
  - It is prone to be attracted by the selection noise!
  - Thereby, the premature convergence can take place!



Proportional Selection (2)

Premature Convergence

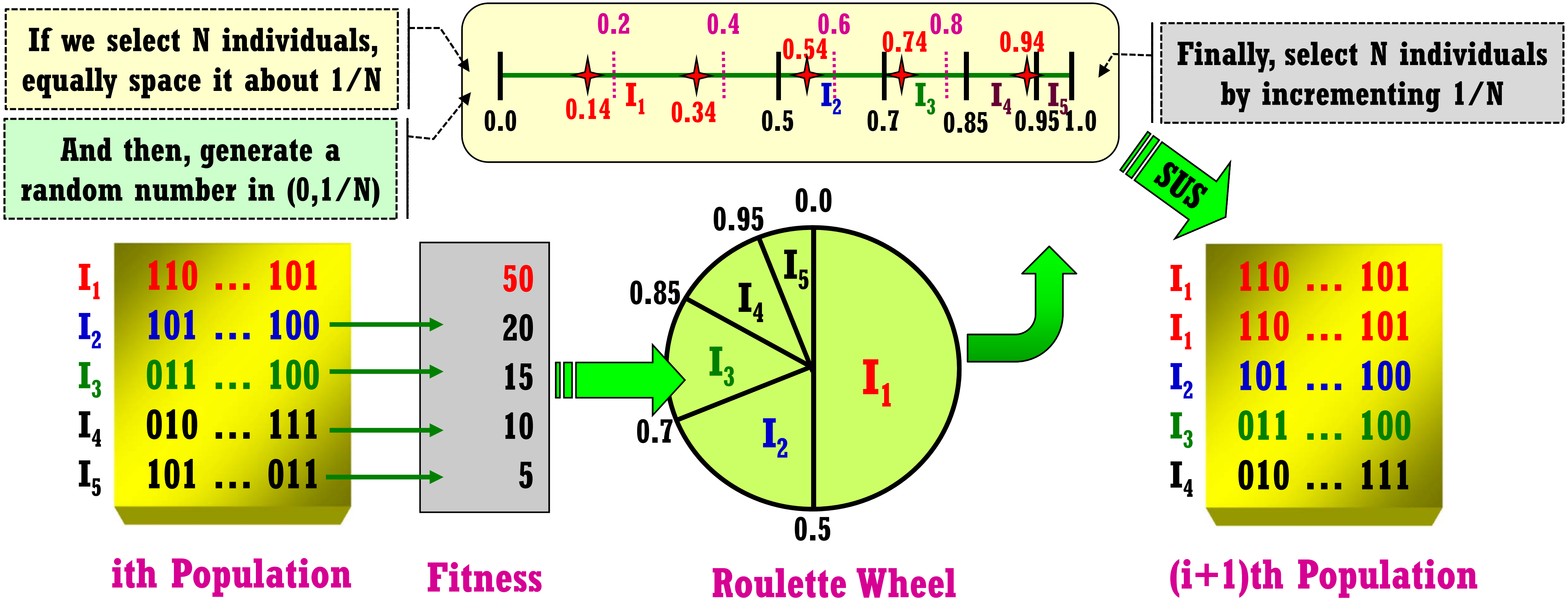
- The whole population is **too early** converged (into **a sub-optimum**)
  - (Population) **diversity decreases** too **fast!**



Proportional Selection (3)

Stochastic Universal Selection (SUS)

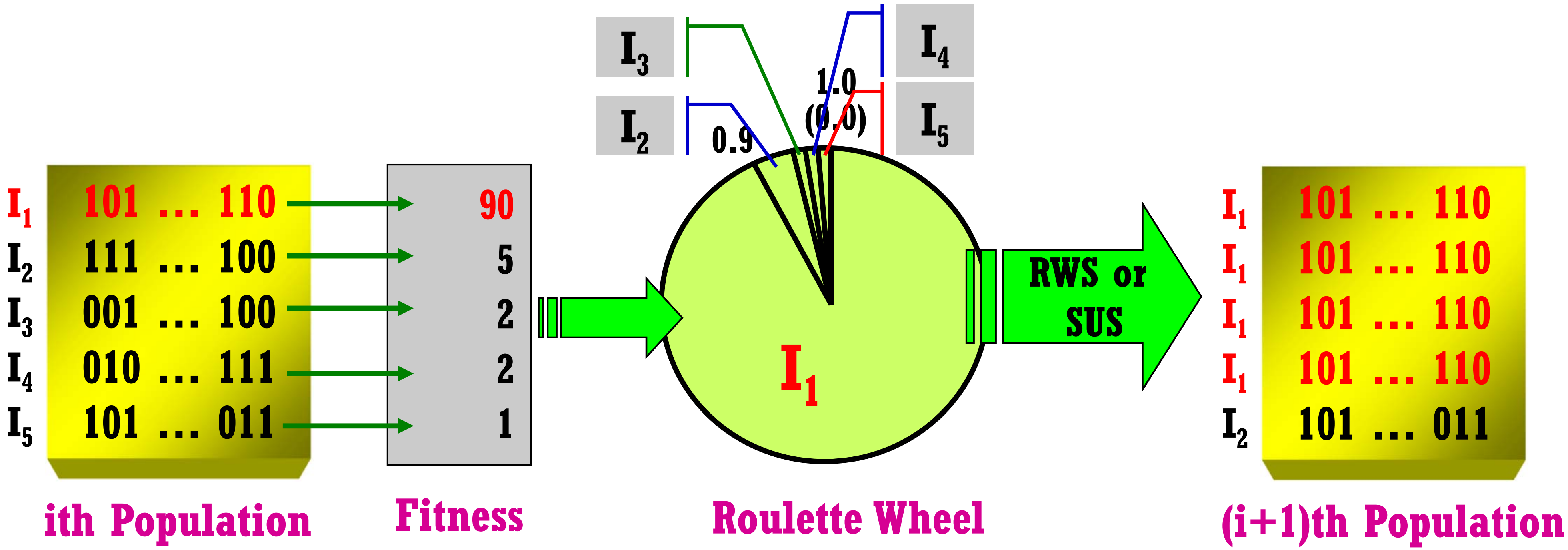
- It is somewhat possible to **alleviate** the problem of RWS
  - A single-phase sampling algorithm with minimum spread and zero bias
  - Guarantee that an individual of τ-portion of the wheel is selected $[\lfloor \tau N \rfloor, \lceil \tau N \rceil]$ times



Proportional Selection (4)

Selection Noise Revisited

- Do you think SUS is Strong enough to endure the selection noise?
  - In fact, SUS is superior to RWS in view of alleviating selection noise!
  - Still, suffer from the premature convergence; see the example below.
  - Premature convergence is an inevitable problem of proportional selection!



Proportional Selection (5)

Linear Scaling

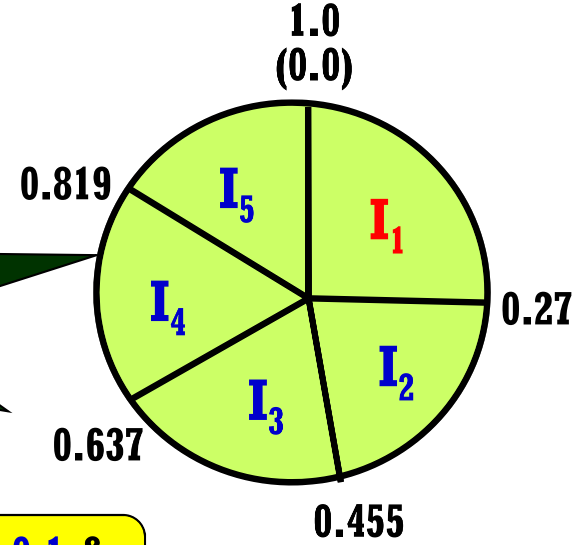
- We can **relax** the weakness of the proportional selection
  - Linear (or non-linear) **scaling** of the fitness

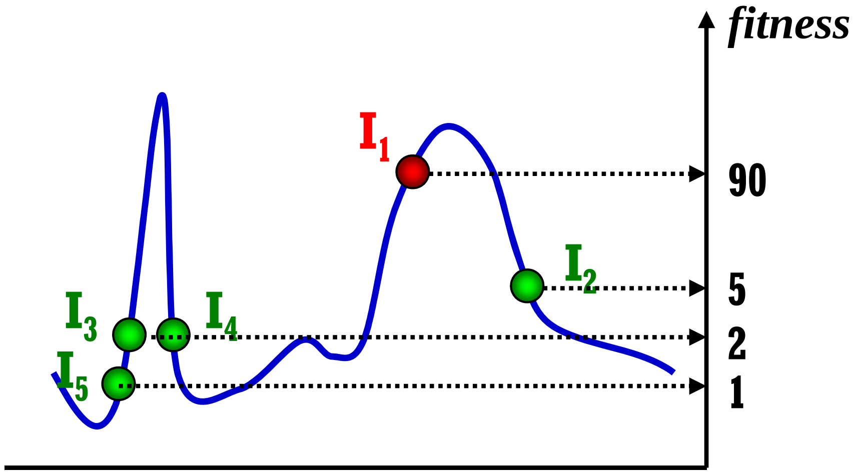
$$f = a \cdot fitness + b$$

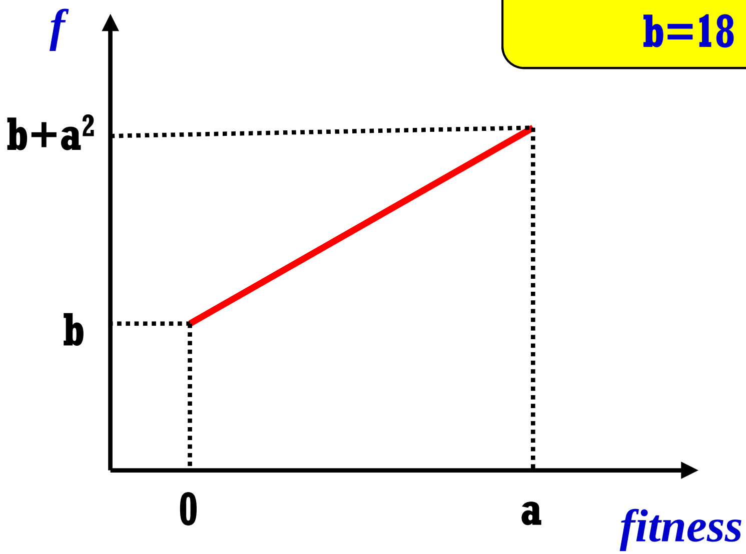
where $\overline{f} = \overline{fitness}$ and $f_{\max} = \varphi \cdot \overline{f}$

But, continual re-scaling is needed. How to set a & b ?

Roulette Wheel







With $a=0.1$ & $b=18$

	Fitness	f
I_1	90	27
I_2	5	18.5
I_3	2	18.2
I_4	2	18.2
I_5	1	18.1

Proportional Selection (6)

Logarithmic Scaling

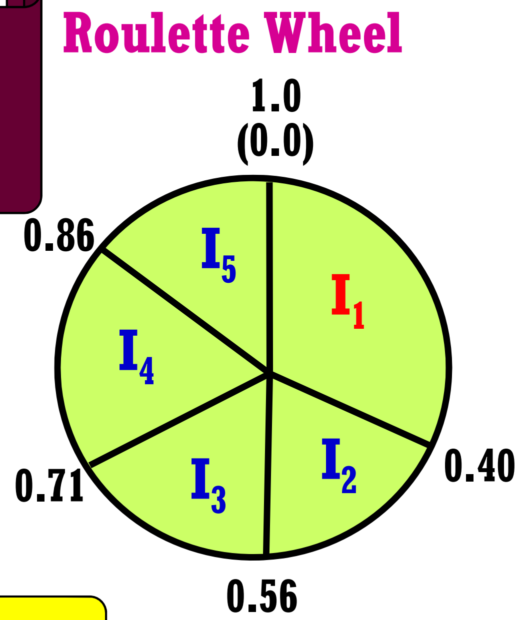
- It is a non-linear scaling of the fitness
  - An example is the *Boltzmann selection*

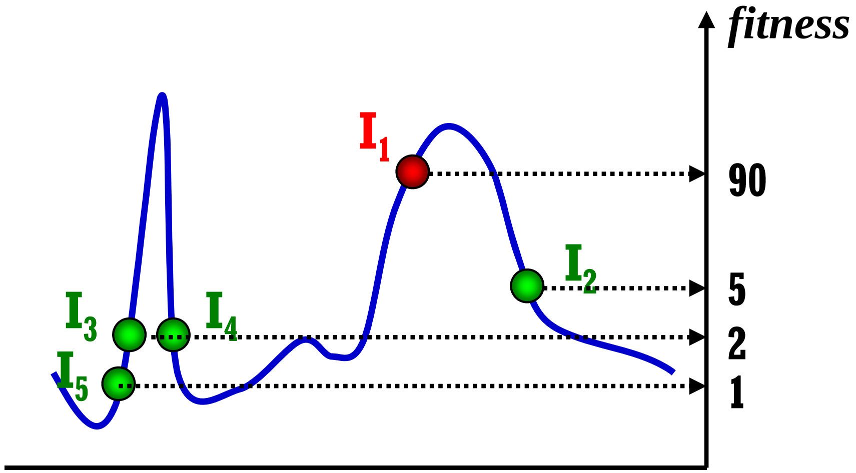
$$f = \exp(fitness/T)$$

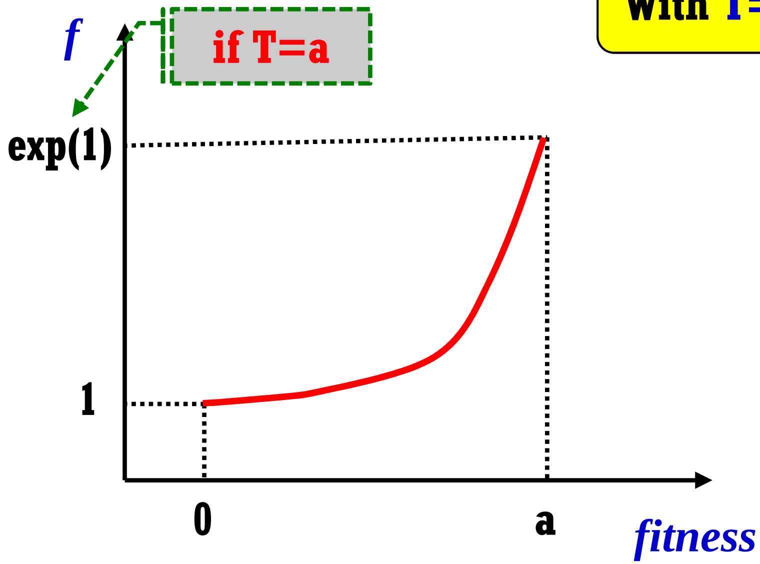
Control the selection pressure!

where the parameter T decreases during the search process.

Also, T should be properly adjusted at every generation!







With T=90

	Fitness	f
I_1	90	2.72
I_2	5	1.06
I_3	2	1.02
I_4	2	1.02
I_5	1	1.01

Ordinal Selection (1)

Ranking Selection

- **Why** do we **resort to fitness** values?
  - The **premature convergence** has been brought forth from the **fitness value** itself
  - A key point in the selection is the **relative dominance** (i.e., **ranking**)!
- **Ranking** may lose some information, but **simpler and more efficient**

	ith Population	Fitness	Rank
I_1	101 … 110	80	5
I_2	111 … 100	10	4
I_3	001 … 100	5	3
I_4	010 … 111	3	2
I_5	101 … 011	2	1

How to select individuals?
What criterion can be used for selection?

- Suppose that the prob. of selecting the kth-rank individual

$$P[k] = \alpha + \beta \cdot k$$

A better individual has a higher rank

- To be a probability distribution

$$\sum_{k=1}^{N} \alpha + \beta \cdot k = N\left(\alpha + \beta \frac{N+1}{2}\right) = 1$$

Assume all individuals are distinct!
N is the population size

Ordinal Selection (2)

Ranking Selection (Cont.)

- Selection pressure is defined by

$$\phi = \frac{P[\textit{selecting the fittest individual}]}{P[\textit{selecting average individual}]}$$

Here, 'average' means 'median'!

$$\Rightarrow \quad \phi = \frac{\alpha + \beta \cdot N}{\alpha + \beta(N+1)/2}$$

$$\Rightarrow \quad \beta = \frac{2(\phi - 1)}{N(N-1)}, \quad \alpha = \frac{2N - \phi(N+1)}{N(N-1)}$$

which implies that $1 \le \phi \le 2N/(N+1) \simeq 2$

- The cumulative prob. distribution can be stated in terms of the sum of an arithmetic progress.

- With a random number r, finding the k is given by

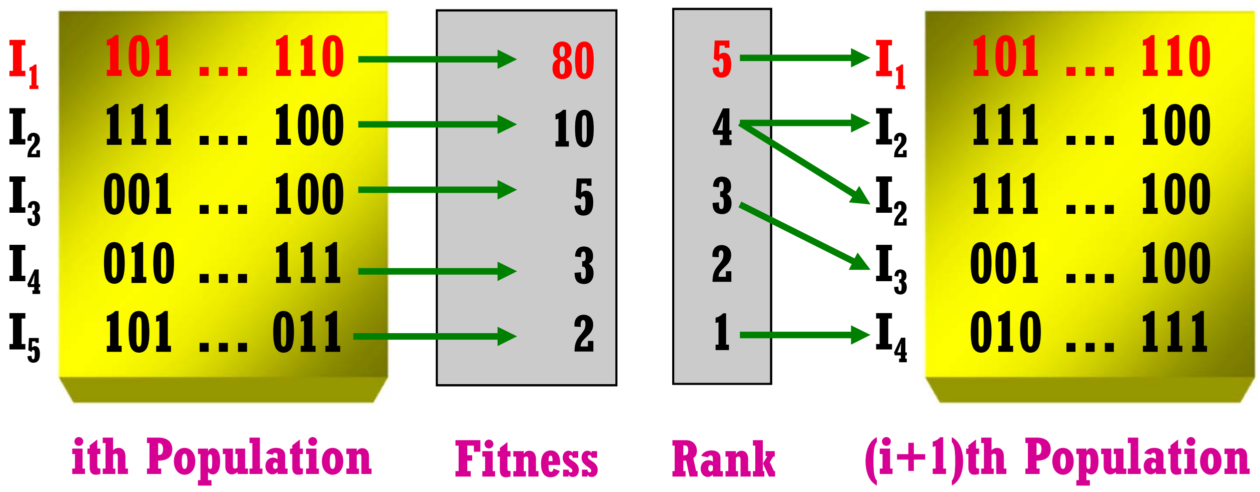
$$\sum_{i=1}^{k} (\alpha + \beta \cdot i) = \alpha \cdot k + \beta \frac{k(k+1)}{2} = r$$

$$\Rightarrow \quad k = \frac{-(2\alpha + \beta) + \sqrt{(2\alpha + \beta)^2 + 8\beta \cdot r}}{2\beta}$$

➔ With a given random number, taking O(1) time, the individual of the $\lceil k \rceil$-th rank can be selected!

Ordinal Selection (3)

Ranking Selection (Cont.)

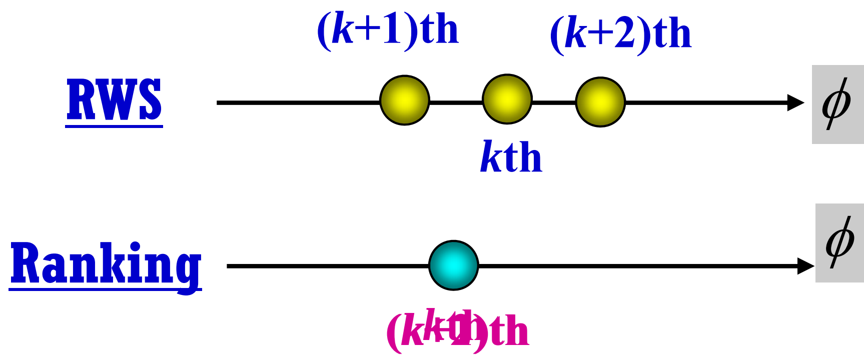


- Finding the proper k for a random number r can be done in $O(1)$
- Searching for k for r using the ordinary proportional selection, needs $O(\log N)$
- But the ranking requires $O(N\log N)$ for a sorting algorithm.
- Nevertheless, the possibility of keeping a constant selection pressure without re-scaling is an attractive one!

- For N=5, ϕ =1.5, we can get $\alpha=\beta=1/20$

- For given random numbers (0.5, 0.9, 0.4, 0.1, 0.7), we have k=(3.22, 4.68, 2.77, 1.0, 4.0)

  ➔ (4, 5, 3, 1, 4) individuals are selected!



Ordinal Selection (4)

Nonlinear Ranking

- **Nonlinear ranking** assigns a **selection prob.** that is a **nonlinear function of rank**
  - **Assumption:** population size N, higher ranks are better, all distinct individuals
  - e.g.) Exponential, Geometric, Biased exponential ranking

	ith Population	Fitness	Rank
I_1	101 … 110	80	5
I_2	111 … 100	10	4
I_3	001 … 100	5	3
I_4	010 … 111	3	2
I_5	101 … 011	2	1

- Suppose that the prob. of selecting the kth-rank individual

Exponential Ranking

$$P[k] = c^{-1}(1 - e^{1-k})$$

where c is a normalization factor

Geometric Distribution Ranking

$$P[k] = a(1-a)^{N-k}$$

where $a \in (0,1)$

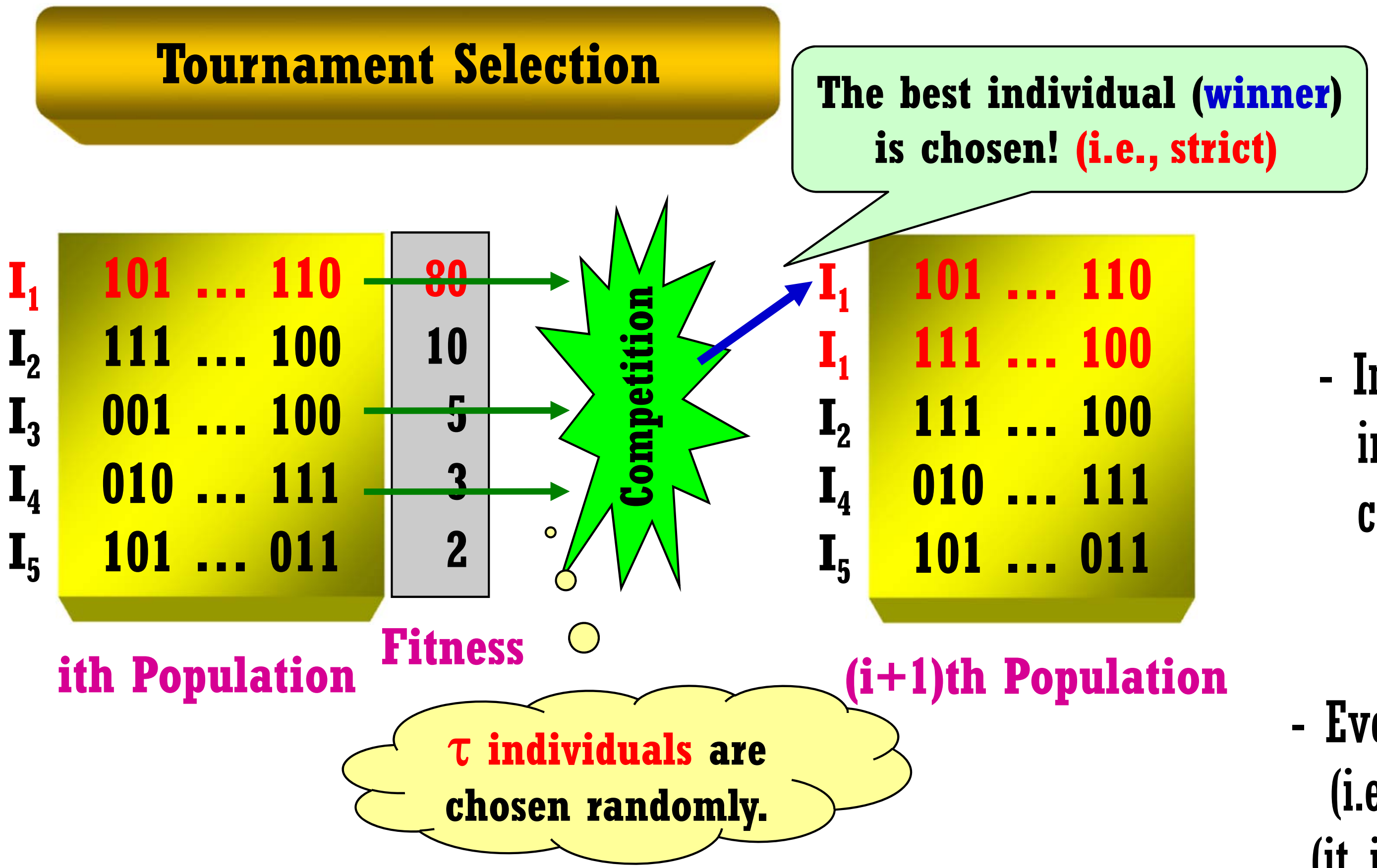
Biased Exponential Ranking

$$P[k] = \frac{1-a}{1-a^N} a^{N-k}$$

where $a \in (0,1)$

Remaining procedures are the same as those done in Linear Ranking!

Ordinal Selection (5)



- In a complete cycle (i.e., generating N individuals), each individual will be compared τ times on average

- Every time it is compared, the best one (i.e., winner) is selected all the time. (it is called 'strict' tournament.)

$$P[selecting\ I_{best} \mid I_{best}\ is\ chosen] = 1$$

$$\phi = \frac{P[selecting\ I_{best}]}{P[selecting\ I_{avg}]}$$

$$= \frac{P[selecting\ I_{best} \mid I_{best}\ is\ chosen]P[I_{best}\ is\ chosen]}{P[selecting\ I_{avg} \mid I_{avg}\ is\ chosen]P[I_{avg}\ is\ chosen]}$$

Ordinal Selection (6)

Tournament Selection (Cont.)

I_1	101 … 110
…	… … …
I_{avg}	001 … 100
…	… … …
I_N	101 … 011

ith Population

Probability of choosing any individual below the median is 1/2

- The chance of the median individual being surviving is the prob. that the remaining (τ-1) ones are all worse: $(½)^{\tau-1}$

$$P[\text{selecting } I_{avg} \mid I_{avg} \text{ is chosen}] = (1/2)^{\tau-1}$$

- Since the probability of selecting each individual is equally likely as 1/N, we get the following result regardless of τ

$$P[I_{best} \text{ is chosen}] = P[I_{avg} \text{ is chosen}]$$

- Thus, the selection pressure is $\phi = 2^{\tau-1}$

- To obtain a selection pressures below 2, soft tournament can be used such that the chance of winning of the best is $p<1$

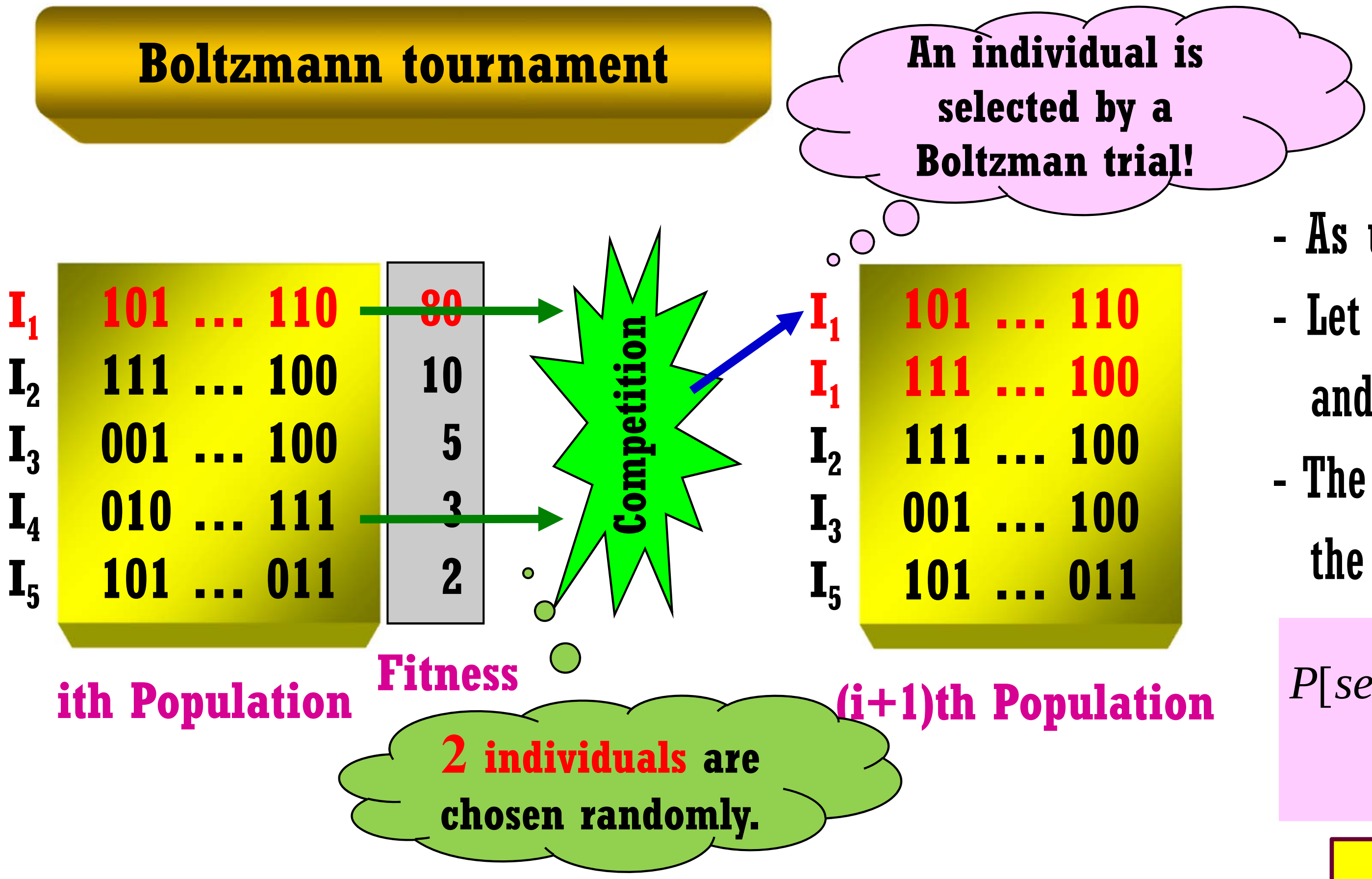
- We can get $\phi = 2^{\tau-1} p$

- The pair-wise soft tournament selection can produce the selection pressure: $\phi = 2p$ that exists [0, 2]

- If the selection pressure when $p \geq 0.5$ is the same as that of the ranking selection: $\phi = [1, 2]$

Ordinal Selection (7)

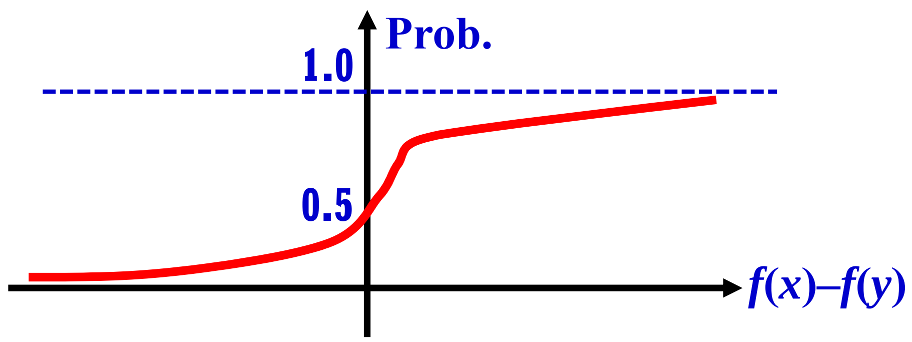
Boltzmann tournament

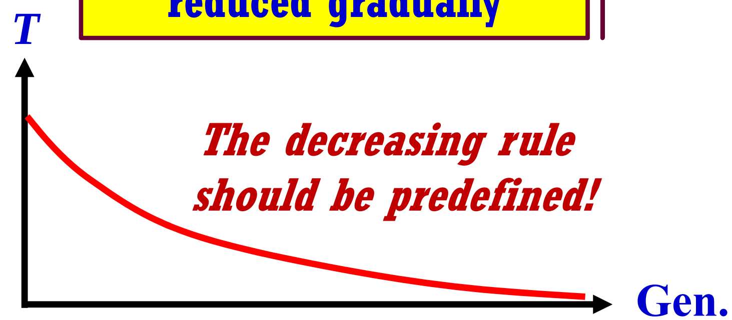


- As usual, consider a maximization problem.
- Let x be the current solution, and y an alternative solution.
- The individual x wins (is selected) with the following probability:

$$P[selecting\ \ x] = \frac{1}{1+\exp\left(-\dfrac{f(x)-f(y)}{T}\right)}$$

T is the temperature reduced gradually

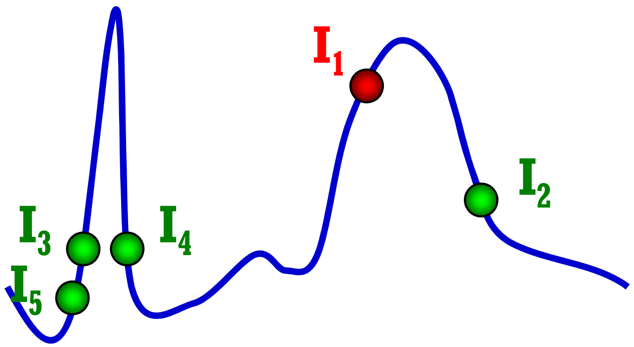


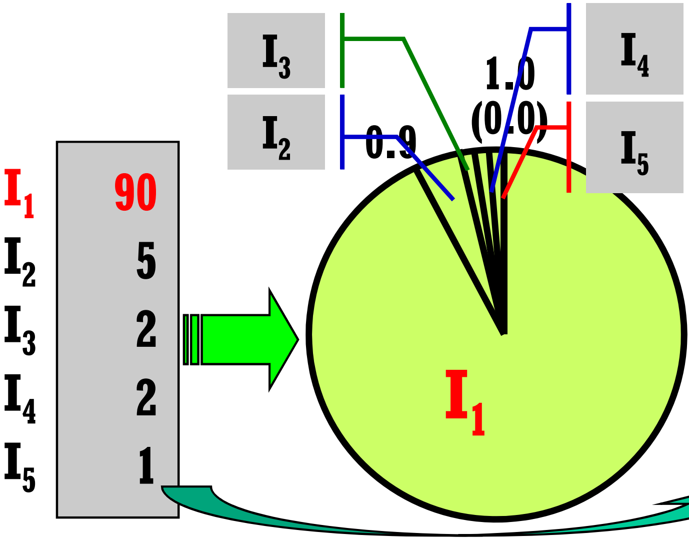


Comparison I

Proportional Selection

- Premature convergence may happen





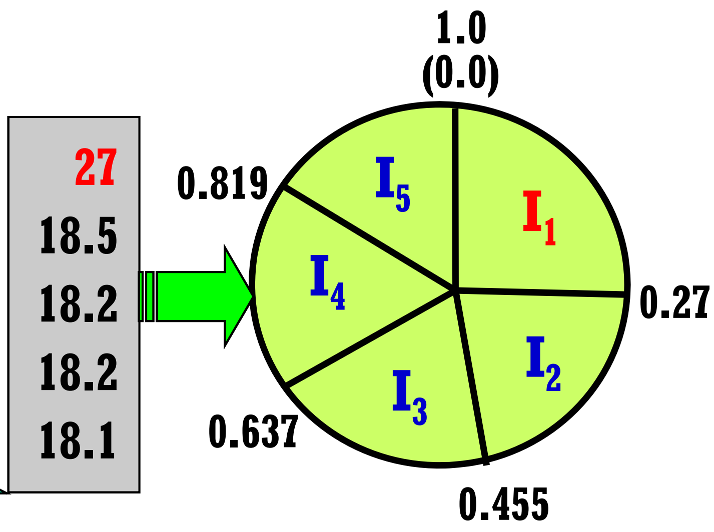
Fitness Roulette Wheel

Scaling

- Relax the premature convergence
- Continual re-scaling is needed

$$f = a \cdot fitness + b$$

$$\bar{f} = \overline{fitness} \qquad f_{\max} = \phi \cdot \bar{f}$$



f Roulette Wheel

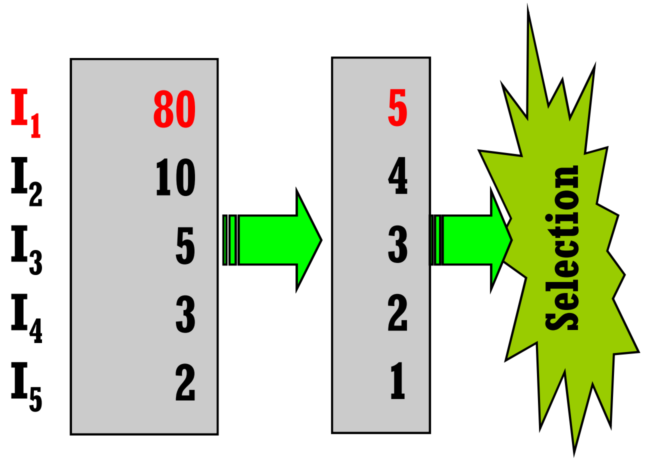
Ranking Selection

- Ranks of individuals are used
- Constant selection pressure

$$P[k] = \alpha + \beta \cdot k$$

With random number r

$$k = \frac{-(2\alpha+\beta)+\sqrt{(2\alpha+\beta)^2+8\beta\cdot r}}{2\beta}$$



Fitness Rank

Comparison II

Proportional Selection

RWS: $O(N\log N)$

SUS: $O(N)$

$$\phi = \frac{P[selecting\ I_{best}]}{P[selecting\ I_{avg}]} = \frac{fitness^{(best)}}{\frac{1}{N}\sum_{k=1}^{N} fitness^{(k)}}$$

- The selection pressure varies each generation!
- We never control the selection pressure.

Scaling

$$f_{\max} = \phi \cdot \overline{f} \quad \Rightarrow \quad \phi = f_{\max} / \overline{f}$$

$$\phi = f_{\max} \Big/ \frac{1}{N}\sum_{k=1}^{N}(a \cdot fitness^{(k)} + b)$$

- Also, the selection pressure varies each generation!
- But, it can somehow control the selection pressure.
- To do this, re-scaling is needed every generation.

Ranking Selection

$O(N\log N)$

$$\phi = \frac{\alpha + \beta \cdot N}{\alpha + \beta(N+1)/2}$$

$$\beta = 2(\phi - 1)/N(N-1), \quad \alpha = \frac{2N - \phi(N+1)}{N(N-1)}$$

- The selection pressure is not altered!
- It controls the pressure without any re-scaling.
- But, the selection pressure only exists in [1, 2]

Tournament Selection

$O(N)$

I. Strict Tournament

$$\phi = 2^{\tau - 1}$$

II. Soft Tournament

$$\phi = 2^{\tau - 1} p$$

- The selection pressure does not vary!
- It controls the pressure without any re-scaling.
- But, the selection pressure can have any value.