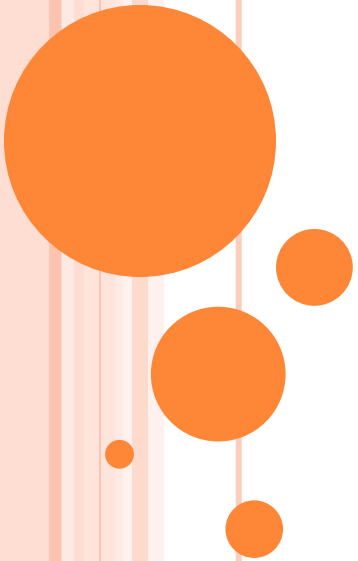
Genetic algorithm based searching



GENETIC ALGORITHMS (GAs)

- This searching technique is based on the mechanics of natural selection and natural genetics.
- It is based on the "*survival of the fittest*" theory i.e the parents (states) which are fittest are allowed to produce children (new states).
- It simulates the process of evolution i.e. it considers evolution as an optimization problem where every successor generation is better.
- It is based on the assumption that if robustness is the objective then nature is the clear winner as it has features of self repair, self guidance and reproduction. So, GAs simulate the process of natural selection and genetics.

Differences between GAs and other Searching Methods

- GAs work with encoding of parameter set and not the parameter themselves.
- GAs search for a population of points, not single point.
- GAs use objective function value (also called fitness function) directly rather than any heuristic function or first or second derivatives, etc.
- GAs use stochastic rules (based on probability or random numbers) rather then deterministic algorithm.

Genetic Algorithm

- Genetic Algorithm works in following phases:

1) Initial Population: It begins with randomly selected population (states) which are satisfactory to the problem. For example, for travelling salesman problem with ten cities, the initial population may contain any sequence of states from start city.

   Encode the population as strings, strings of bits or in octal form or hexadecimal form (if required).

   Usually for n variables, the size of population varies from 2n to 4n.

2) Calculate the value of the fitness function (i.e. objective function) which needs to be maximized or minimized. For objective function which need to minimized can be converted into a maximum problem by taking reciprocal/ negative of the original function.

Genetic Algorithm Contd.....

3) Apply the following operators on the initial population to generate child (offspring's) . These child will act as parents for the next iteration.

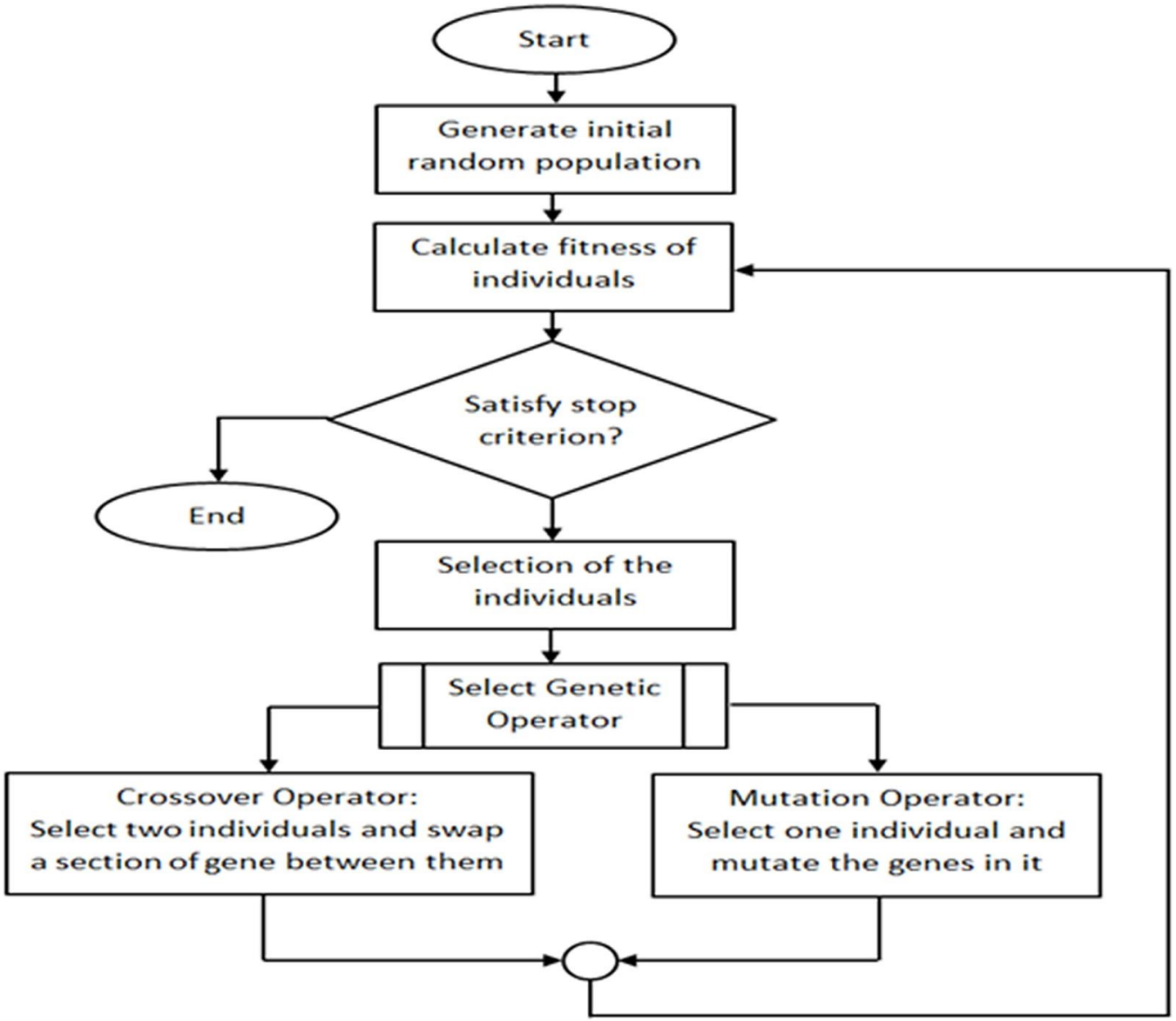
   (a) **Selection**: Parent Selection is the process of selecting parents which mate and recombine to create off-springs for the next generation. Parent selection is very crucial to the convergence rate of the GA as good parents drive individuals to a better and fitter solutions.

   (b)**Crossover:** The crossover operator is analogous to reproduction and biological crossover. In this more than one parent is selected and one or more off-springs are produced using the genetic material of the parents

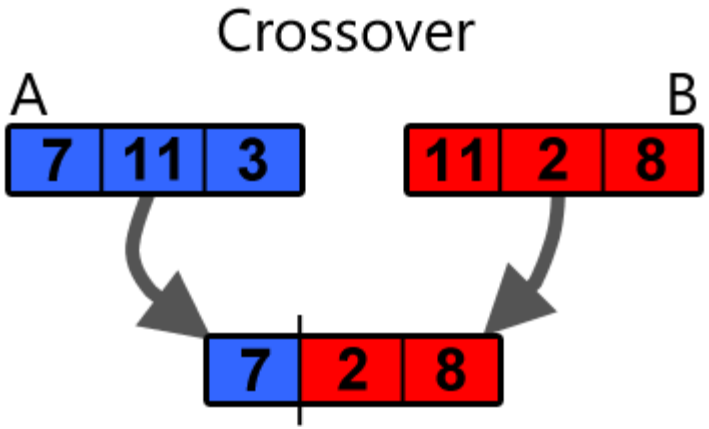
   (c) **Mutation:** Mutation may be defined as a small random change in the chromosome, to get a new solution. It is used to maintain and introduce diversity in the genetic population and is usually applied with a low probability. If the probability is very high, the GA gets reduced to a random search.

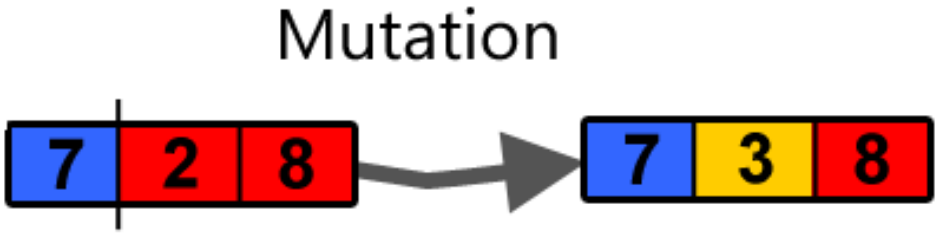
4) Repeat step 3 on the new offspring's produced to produce offspring's for next iteration.

Genetic Algorithms Flow CHART



GA: Crossover/Mutation





GA: Crossover/Mutation

Crossover → for Exploration

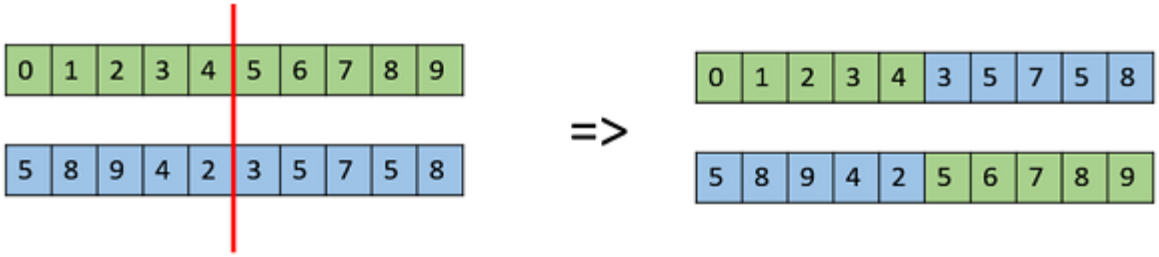
Mutation → for Exploitation

Crossover

- The crossover operator is analogous to reproduction and biological crossover. In this more than one parent is selected and one or more off-springs are produced using the genetic material of the parents. Crossover is usually applied in a GA with a high probability

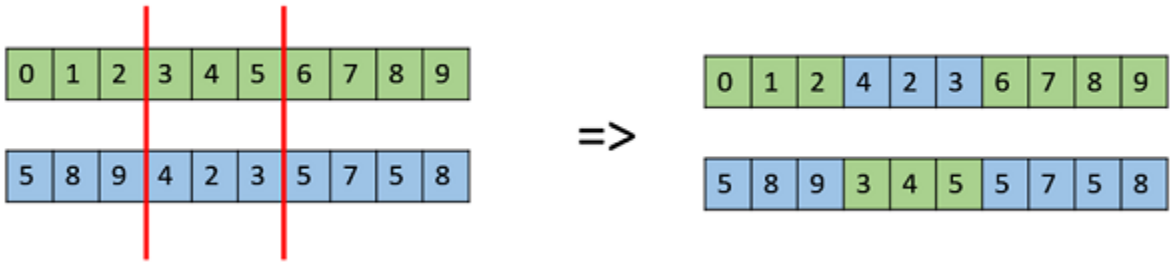
One point Crossover

- In this one-point crossover, a random crossover point is selected and the tails of its two parents are swapped to get new off-springs.



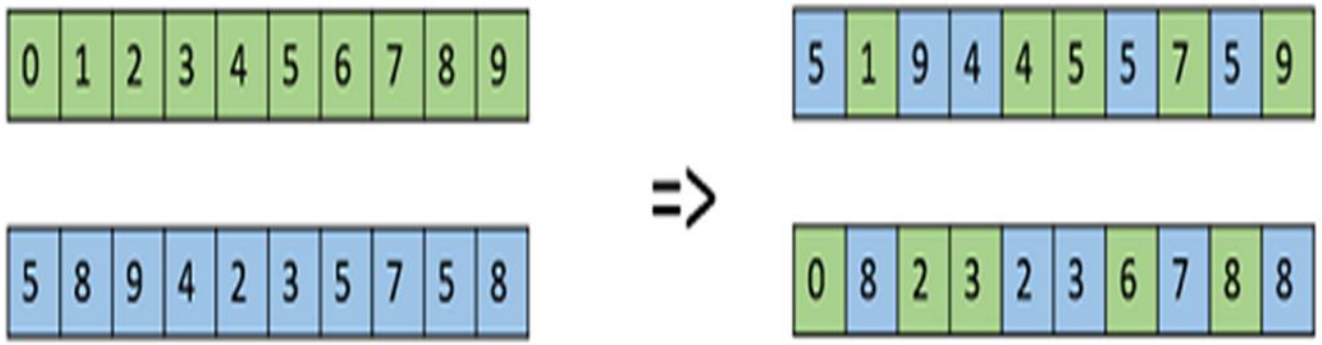
Multi point Crossover

- Multi point crossover is a generalization of the one-point crossover wherein alternating segments are swapped to get new off-springs.



Uniform Crossover

- In a uniform crossover, we don't divide the chromosome into segments, rather we treat each gene separately. In this, we essentially flip a coin for each chromosome to decide whether or not it'll be included in the off-spring. We can also bias the coin to one parent, to have more genetic material in the child from that parent.



Whole Arithmetic Recombination

- This is commonly used for integer representations and works by taking the weighted average of the two parents by using the following formulae −

  $Child1 = \alpha.x + (1-\alpha).y$

  $Child2 = (1-\alpha).x + \alpha.y$

Obviously, if $\alpha = 0.5$, then both the children will be identical.

GA: Tuning Parameters

- Population Size: 50, 100
- Crossover Rate: 0.5 to 0.9
- Mutation Rate: 0.01
- Fitness Function

Selection Methods in GAs

- **Roulette Wheel Selection (example explained at the end)**

1) Compute the probability of selection for each population and convert into percentage as:

$$\text{Probi} = \frac{f(x)_i}{\sum_{i=1}^{n} f(x)_i} \quad \text{X } 100\%$$

where

n = size of population

f(x)=fitness value corresponding to a particular individual in the population

2) Represent these probabilities on the roulette wheel according to the percentage of their probabilities. A fixed point is chosen on the wheel circumference as shown and the wheel is rotated. The region of the wheel which comes in front of the fixed point is chosen as the parent. For the second parent, the same process is repeated.

1. Fitness Proportionate Selection

In this selection, every individual can become a parent with a probability which is proportional to its fitness value. Therefore, fitter individuals have a higher chance of mating and propagating their features to the next generation.

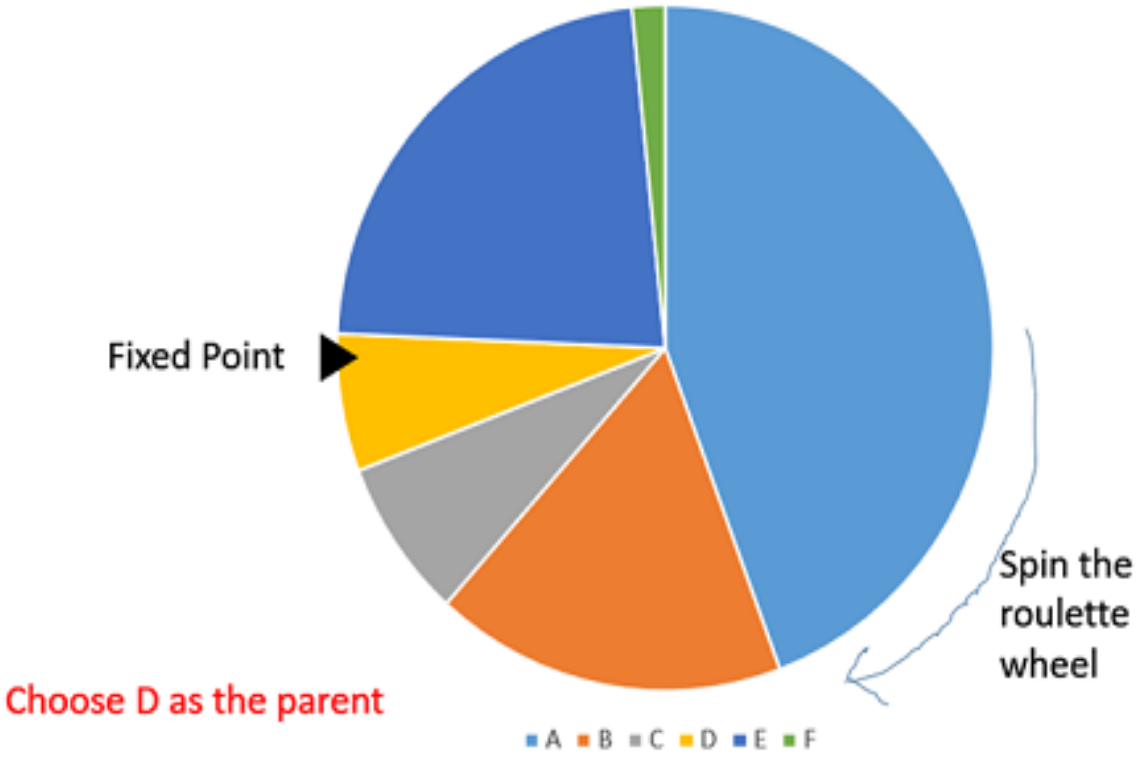
Two implementations of fitness proportionate selection are possible –

- Roulette Wheel Selection
- Stochastic Universal Sampling (SUS)

Note: Fitness proportionate selection methods don't work for cases where the fitness can take a negative value.

Roulette Wheel Selection

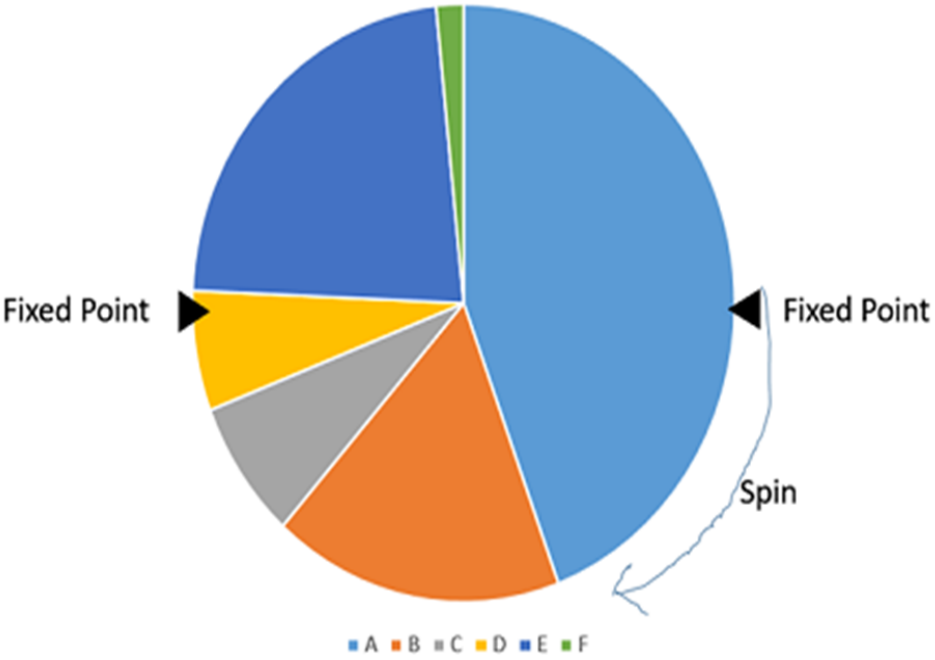
- In this, the circular wheel is divided into **n pies**, where n is the size of the population.
- Each individual solution gets a portion of the circle which is proportional to its fitness value.
- A fixed point is chosen on the wheel circumference and the wheel is rotated. The region of the wheel which comes in front of the fixed point is chosen as the parent.
- For the second parent, the same process is repeated.



Chromosome	Fitness Value
A	8.2
B	3.2
C	1.4
D	1.2
E	4.2
F	0.3

Stochastic Universal Sampling (SUS)

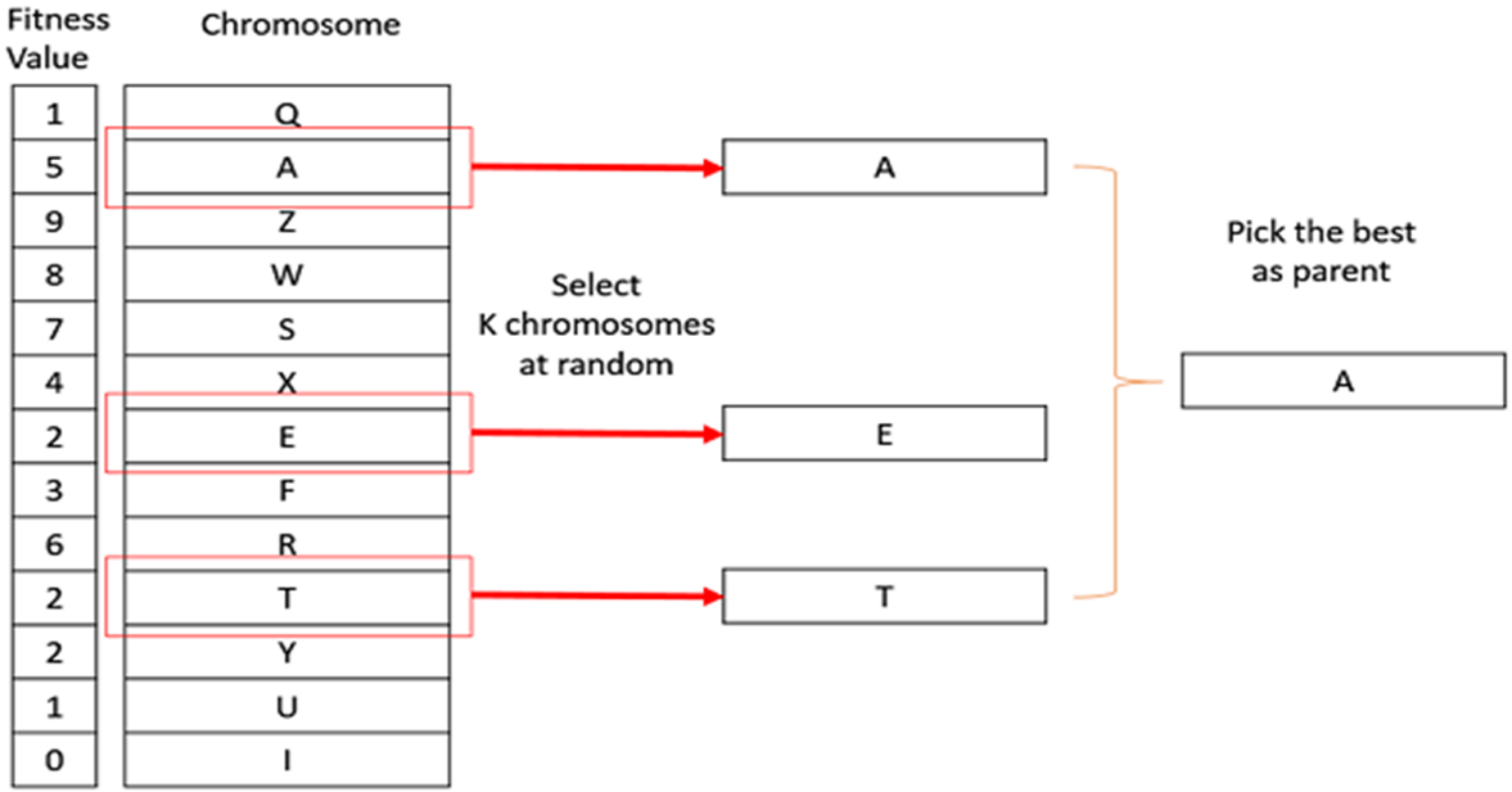
It is quite similar to Roulette wheel selection, however instead of having just one fixed point, we have multiple fixed points. Therefore, all the parents are chosen in just one spin of the wheel. Also, such a setup encourages the highly fit individuals to be chosen at least once.



Chromosome	Fitness Value
A	8.2
B	3.2
C	1.4
D	1.2
E	4.2
F	0.3

2. Tournament Selection

In K-Way tournament selection, we select K individuals from the population at random and select the best out of these to become a parent. The same process is repeated for selecting the next parent. Tournament Selection is also extremely popular in literature as it can even work with negative fitness values.



3. Rank Selection

Rank Selection also works with negative fitness values and is mostly used when the individuals in the population have very close fitness values (this happens usually at the end of the run). This leads to each individual having an almost equal share of the pie and hence each individual no matter how fit relative to each other has an approximately same probability of getting selected as a parent. This in turn may lead to a loss in the selection procedure towards fitter individuals, making the GA to make poor parent selections in such situations.

In this, we remove the concept of a fitness value while selecting a parent. However, every individual in the population is ranked according to their fitness. The selection of the parents depends on the rank of each individual and not the fitness. The higher ranked individuals are preferred more than the lower ranked ones.

Chromosome	Fitness Value	Rank
A	8.1	1
B	8.0	4
C	8.05	2
D	7.95	6
E	8.02	3
F	7.99	5

EXAMPLE 1

- Consider the following initial population and fitness function values.

Population	Chromosome	Fitness
P1	10001	50
P2	10010	6
P3	00100	36
P4	11001	30
P5	11111	36
P6	00110	28

- Generate mating pool for crossover operation using Roulette wheel and ranking selection method.
- Apply crossover operator using single point

SOLUTION

Roulette Wheel Selection:

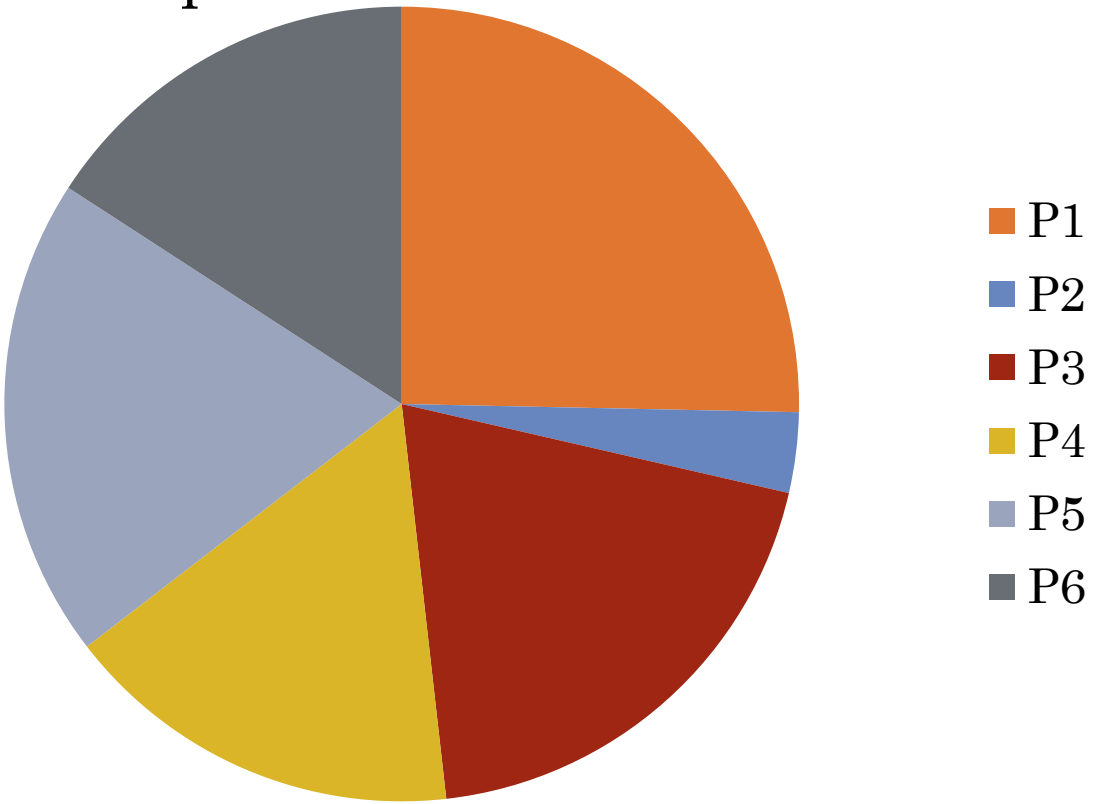
First calculate probability of selection of each population as

$$\text{Probi} = \frac{f(x)_i}{\sum_{i=1}^{n} f(x)_i} \quad X\ 100$$

Population	Chromosome	Fitness	Probability	Probability %
P1	10001	50	50 / 186 = 0.2688	26.88
P2	10010	6	6 /186 = 0.347	3.47
P3	00100	36	36/ 186 = 0.2081	20.81
P4	11001	30	30/186= 0.1734	17.34
P5	11111	36	36/186 = 0.2081	20.81
P6	00110	28	28/186 = 0.1681	16.18
	Total	186		

Solution Contd….

- Represent these probabilities in Roulette Wheel



Solution Contd….

- Now, the spin is selected and the mating parents are selected.
- Let us decide that one with probability more than 25% is selected twice and one with probability less than 5% is not selected. So the new mating pool for crossover is:

Population	Chromosome
P1	10001
P1	10001
P3	00100
P4	11001
P5	11111
P6	00110

EXAMPLE II

- Consider the following travelling salesman problem which uses following distance matrix and starting city A.

	A	B	C	D	E	F	G	H
A	0	1	4	2	9	8	3	2
B	1	0	5	3	7	2	5	1
C	2	5	0	6	1	4	7	7
D	4	3	6	0	5	2	1	6
E	9	7	1	5	0	9	1	1
F	8	2	4	2	9	0	3	5
G	3	5	7	1	1	3	0	2
H	2	1	7	6	1	5	2	0

- GENERATE INITIAL POPULATION AND SELECT PARENTS FOR FIRST ITERATION USING ROULETTE WHEEL METHOD? (Take population size 4)

Solution

- Let initial Population be:

Initial Population	Chromosome	Distance value	Fitness = 1/D (Done to convert into maximization)
P1	ABCDEFGHA	1+5+6+5+9+3+2+2=33	1/33 = 0.0303
P2	ACBDEFGHA	4+5+3+5+9+3+2+2=33	1/33 = 0.0303
P3	ADBCFEHGA	2+3+5+4+9+1+2+3=29	1/29 = 0.0345
P4	AHGFEDBCA	2+2+3+9+5+3+5+2= 31	1/31= 0.0322

Solution Contd…..

- Roulette Wheel Selection

Initial Population	Chromosome	Fitness	Prob	Prob %
P1	ABCDEFGHA	0.0303	0.2380	23.80%
P2	ACBDEFGHA	0.0303	0.2380	23.80%
P3	ADBCFEHGA	0.0345	0.2710	27.10%
P4	AHGFEDBCA	0.0322	0.2529	25.29%
	Total	0.1273		

- Let us keep two samples each with probability greater than 25%.
- So mating population for cross over is

P3,P3, P4,P4