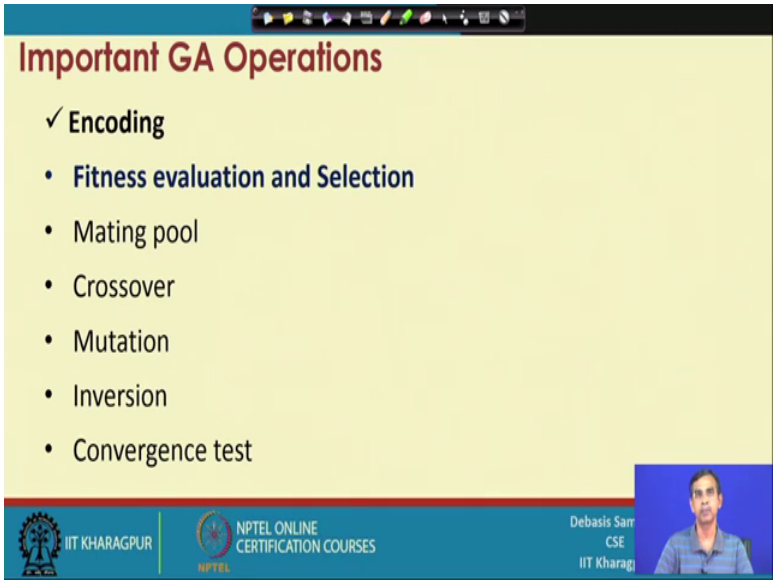
Introduction to Soft Computing
Prof. Debasis Samanta
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture - 18
GA Operator: Selection

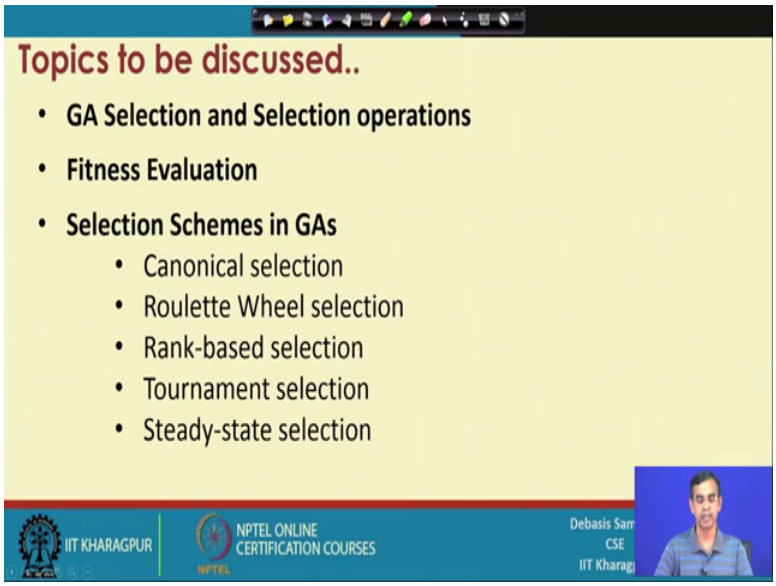
We will learn in this lecture, the another GA operator is the selection operation.

(Refer Slide Time: 00:24)



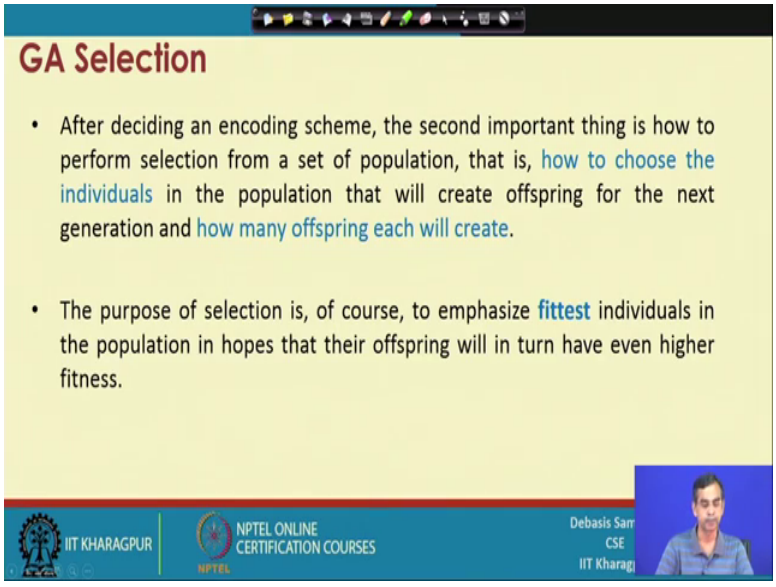
The selection operation is basically the fitness evolution for the solutions that is there in the current population. So, there are different what is called the techniques for the selection mechanism.

(Refer Slide Time: 00:35)



And we will discuss five techniques in this course in this lecture, the five techniques are the canonical selection, then the roulette wheel selection, rank based selection, the tournament selection and steady state selection. So, let us start first with the canonical selection, and then we will discuss one by one.

(Refer Slide Time: 01:02)



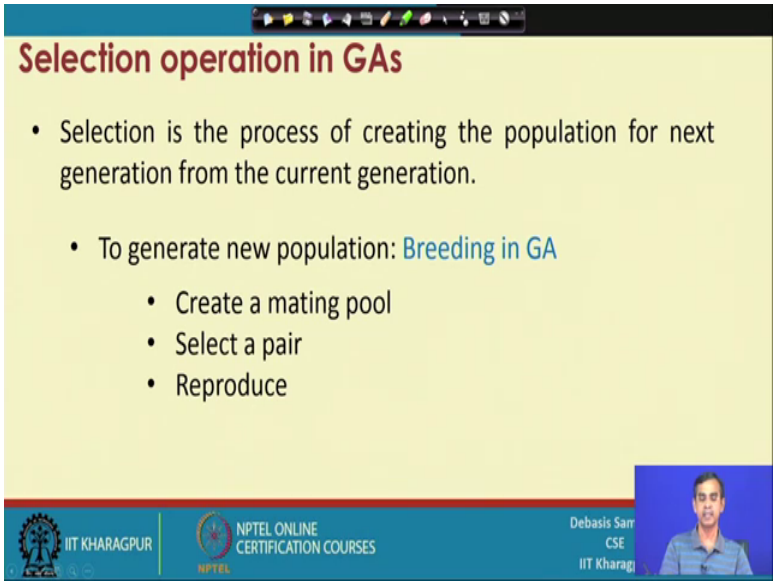
Now, so selection is the one important process before going to the convergence test we have to first evaluate the best solution or the fittest solution. So, basically by means of selection our objective is to how to choose the individuals, those are there in the current

population at any instant, and then if we choose the best solution then the best solution can be passed through unless if it is not converse; that means, the optimal solution is not achieved then we have to go for the next population and.

So, for going to the next population our task is to select the mating pool. So, usually the procedure of selecting the mating pool is to select the best individual first and then undergo them to the mating and then reproduction. So, the purpose of the selection is therefore, to I mean ensure that the fittest individuals in the current population is selected to produce the better offspring.

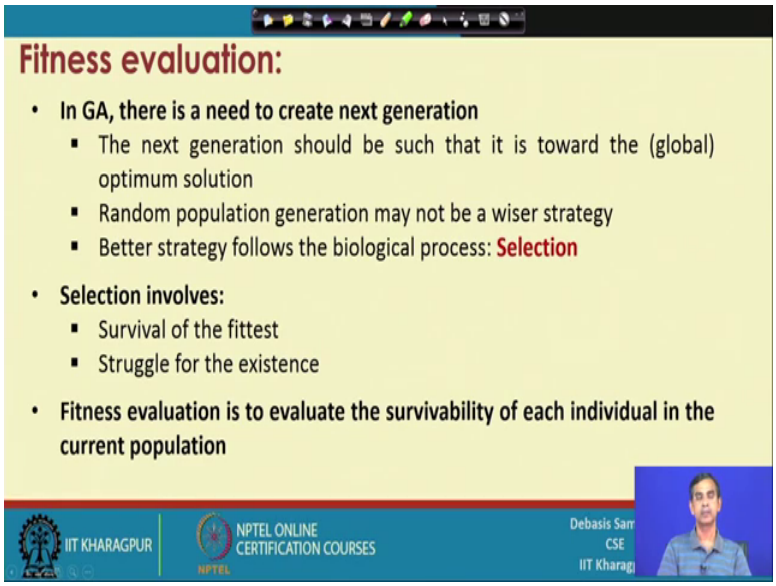
So, this basically necessity the selection procedure is unless the random procedure cannot give the best solution at the quickest way or it will not convert quickly or giving the accurate results. That mean we have to follow certain selection mechanism other than the arbitrary selection or random selection rather.

(Refer Slide Time: 02:27)



Now, so random selection is, so we will this discuss about basically selection. Selection is a prior step of breeding, breeding means reproduction; that means, mating pool creation selecting the mating pair and then reproduction all these things.

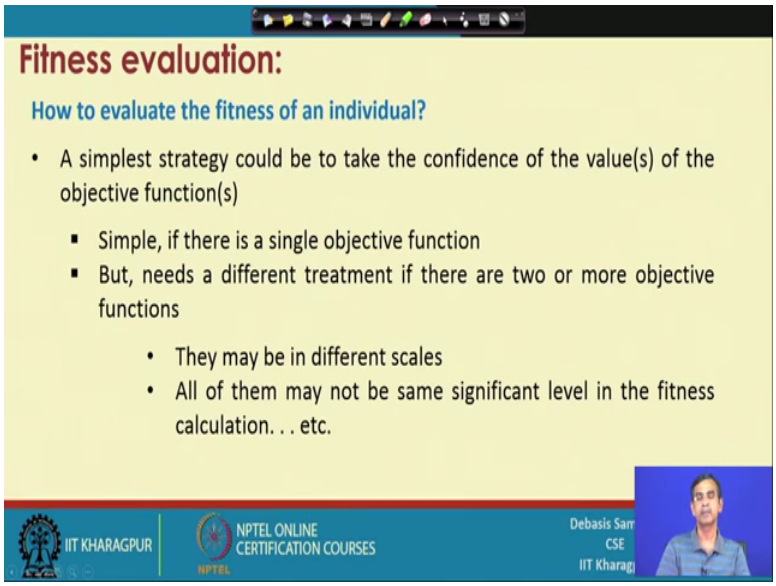
(Refer Slide Time: 02:45)



Now, so far the fitness evolution is concerned. So, as we know so GA, genetic algorithm is a iterative steps; that means, say cyclic process it has to repeat it again and again, or we can say it is basically one population to another population and in each population we have to search for the best solution.

So, best solution is basically ensured by the selection operation in the genetic algorithm and the fitness evolution is the one scheme that will allow us to evaluate the survivability of in the each individual in the current population. Now, let us see what are the different method are there or the selections are there.

(Refer Slide Time: 03:30)

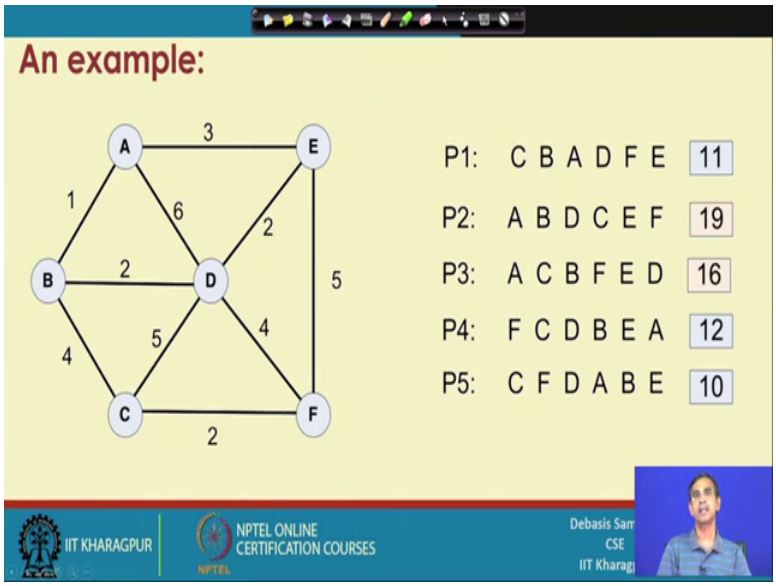


Now, the question is that how to evaluate the fitness of an individual. So, there should be some metric or some policy should be there by which we can apply this one at our hand we have the chromosome or an individual solution and the individual solution is represented by a chromosome we know. So, given the chromosome we have to obtain the fitness value that is the objective actually.

So, idea is basically you know, so at any instant one individual basically represents what are the different values of the design parameter at that instant; that means, if the objective function is f defined in terms of n design parameter, then its phenotype represents that at any, at that instant the different values of this design parameter

Now, so this basically can give us quickly to calculate the objective function. So, one way evaluating the fitness value is basically same way, let us see lets calculate the objective function for the current values, of the gene values, and then this can be a fitness value. So, this is basically broad idea that is followed the f.

(Refer Slide Time: 04:48)



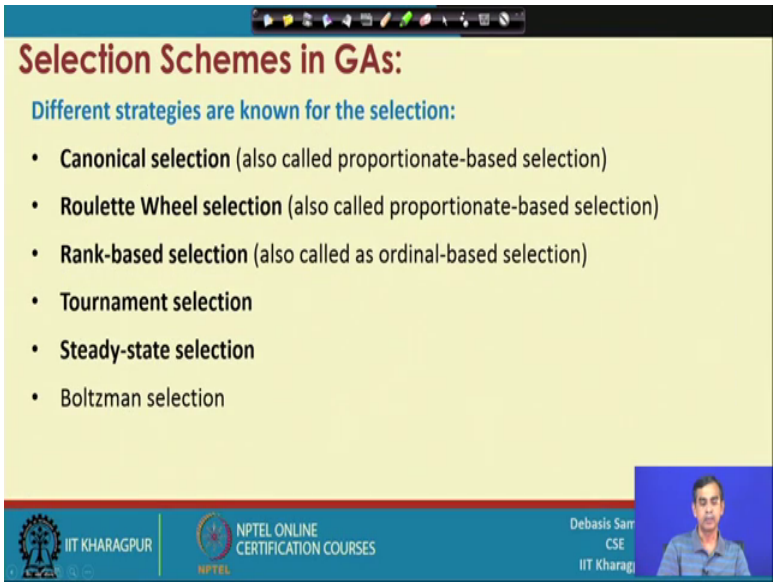
Now, the idea it is again we can represent it like this. So, this is an example to explain how the fitness evolution will work for us. So, this is in the so we can this in the refer with reference to the travelling salesman problem like.

So, here basically five cities are given there A, B, C, D, E 6 cities, of course, 6 cities problem A, B, C, D, E, F and these are the different solutions. So, for example, P 1 is one path and these are the different solutions are there and the cost of this solutions as we know exactly the cost means; what is the cost of C to B, then B to A, then A to D, then D to E, and D F to F to E, like the cost is basically here in 11.

So, at any instance suppose this is the population, population includes five different individuals or five solutions, and then we can apply the cost function. So, applying the cost function for P 1, it gives this is the cost value and similarly this one

Now, out of this different cost value that is obtained in this case, we can say this 19 is the best so; that means, out of the solution if it is a minimum cost of course, is not the maximum right. So, if it is a minimum cost then out of the solution 10 is the best solution the next best is 11 and then 12 and like this one. So, 19 and 16 are the worst, two worst solutions here. So, this basically gives an idea about how the fitness is, fitness can be collect can be calculated with the help of the objective function.

(Refer Slide Time: 06:41)

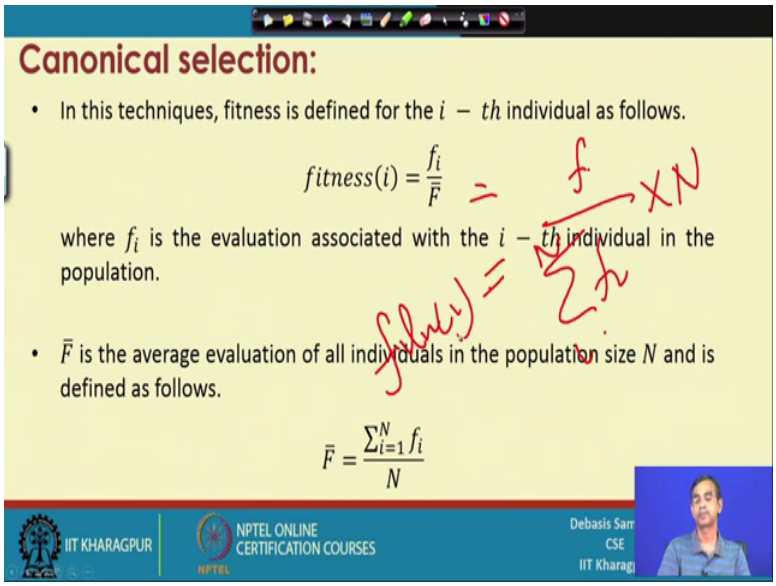


Now, so we can say in other words that fitness value for the measurement of fitness value one metric that can be considered, is a objective function usually it is followed in many of the cases, otherwise it is some other strategies or policies followed.

Now we will discuss about different selection; that means, that is depending on the survivability or some other way or some more procedure how from a set of I mean from a given population, how the best some solutions can be selected for the mating or for the mating pool actually. Now, we will first discuss about the canonical selection the canonical selection. So, first we will decide canonical selection, the in GA theory this canonical selection is also called proportionate based selection.

The like canonical selection the roulette wheel selection is also another version of this it is also called proportionate based selection. The both are the proportional based selection the rank based selection, again it is called the ordinal based selection, and then there are some other selection strategy also known it is called the tournament selection steady state selection anyway we will discuss one by one then, First let us discuss about canonical selection.

(Refer Slide Time: 08:12)

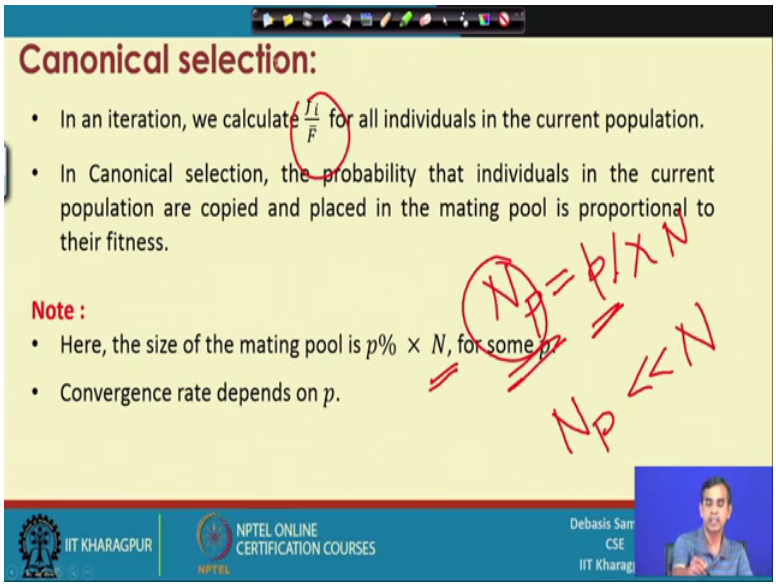


Now, canonical selection idea is very simple, it basically calculates two halloos for each individual. So, f ith fitness value of the ith solution and this f bar capital F bar, this represents the it is represents the average evolution of all individuals in the current population, as in that population of size N then the f bar can be calculated using this formula.

If we know the fitness values of each individuals, in the population then it is a summation of the fitness value of each follow divided by the total number of size of the population so this gives the f bar. So, essentially, so this is basically it is same also fi summation of i equals to 1 to N fi into N, this is the formula that is can be followed to calculate this one.

So, these basically calculate the fitness of any ith individual. Now, the canonical selection follows this formula and we will see exactly how this formula can be applied to I mean to you use the selection of mating pool.

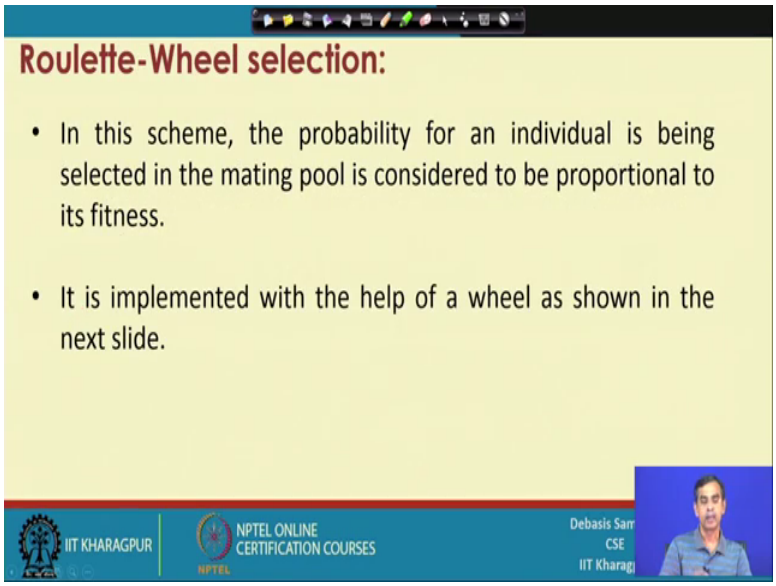
(Refer Slide Time: 09:34)



So, in any instance of the current population, we calculate this fitness value, this fitness value for all individuals and according to this canonical selection, the probability that an individual in the current population will be selected for the mating pool is proportional to this fitness value; that means, the individual which has the highest value of these value will be selected first then the next value will be selected and this will continue to select Np number of solutions.

So, Np is basically some p% of N. So, p maybe 25 or maybe 30% or like this one; that means, this way out of the N. Where N is the current population size we will select Np number of individuals, where Np is less than N and then they will be selected based on this fitness value that we have calculated for each that mean the higher the fitness value will be selected first and so on, so on, this way we will select NP. So, this is the concept of canonical selection here.

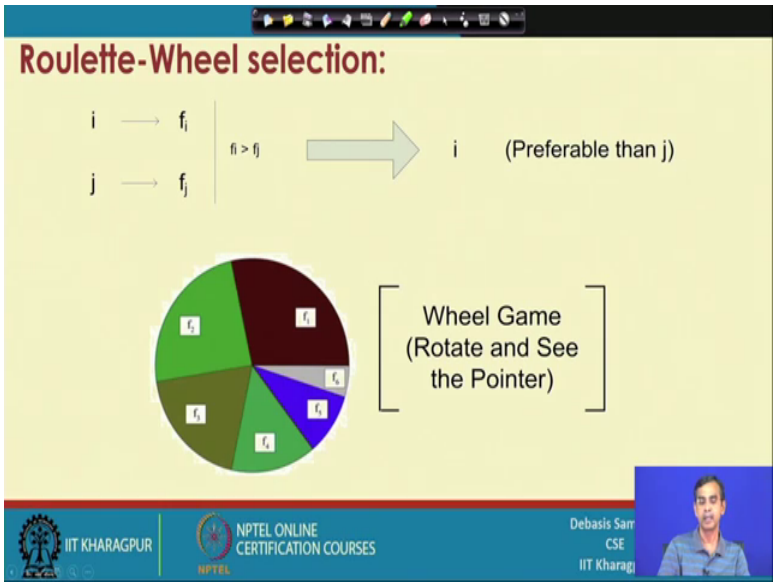
(Refer Slide Time: 11:03)



Now, after knowing the canonical selection, we will learn about the next selection strategy it is called the roulette wheel selection strategy. In this scheme the probability for an individual is being selected for the mating pool is considered to be the proportional to its fitness that is basically the concept of canonical same as this one that is why both the technique is called proportionate based, but it has little bit different the idea.

The idea is basically, if we know one wheel let see exactly what is the idea that is followed there in roulette wheel selection it is there.

(Refer Slide Time: 11:44)



So, roulette wheel selection can be better understood by a wheel game. So, wheel game is basically is a wheel like this and the different wheel is marked with the different colour or different symbols and then the different area, rather or different regions, the different regions is proportional the different fitness value.

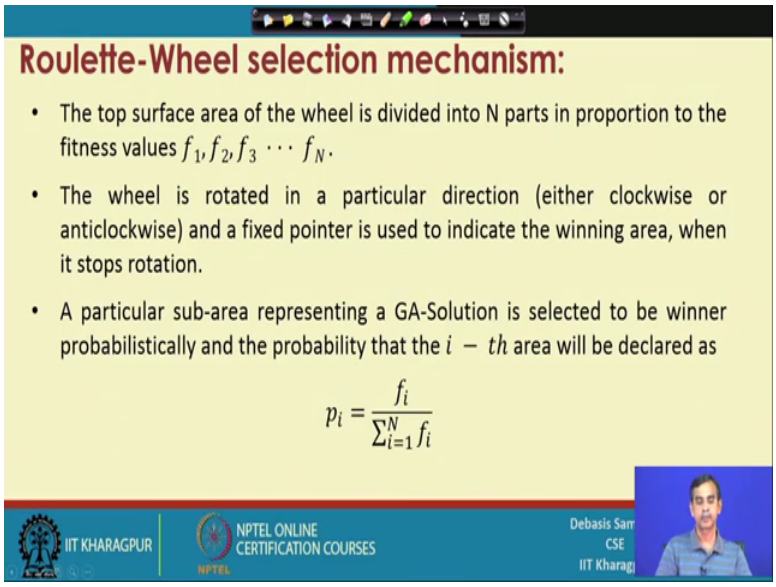
For example, see suppose one solution having the fitness value f 1. And it if it is area that is can be cover proportionally so this is, and similarly, f 2 is this one proportion and this one. Say here basically f 6 is the lowest fitness value for the solution. So, the wheel can be calibrated based on the different fitness values and it is like this one. Now having this is the wheel and supposes it is rotated in this direction and there is a pointer like.

And let it be rotated and then when the wheel stop it, it will point to some point it is there. So, if it points to this one then the f one will be selected. So, let the wheel be rotate for NP times, where NP number of individuals to be selected for the mating pool. So, for each time we will select the individual which basically this one.

So, the probability that an individual will be selected for this, in fact, proportional to their fitness value become. So, this has the least chance to stop it there because it is the wheel game is like that. So, the idea it is followed there the same idea is basically followed in the wheel scheme is there and then.

So, it is basically the, it can be; obviously, that wheel on that if the fitness value is greater than, then that greater fitness value the individual with greater fitness value will be selected for the mating pool.

(Refer Slide Time: 13:41)

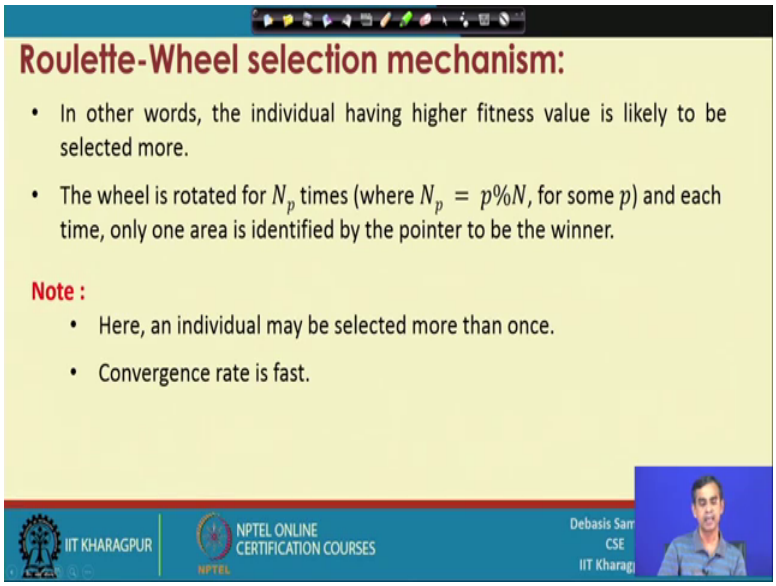


Now, this idea is basically followed there in the roulette wheel scheme, now the mechanism is therefore, the top surface area of the wheel is divided into N parts in proportion to the fitness values of each individuals. The wheel is rotated in a particular direction may be clockwise or anticlockwise and a fixed pointer is used to indicate the winning area when its stops rotation.

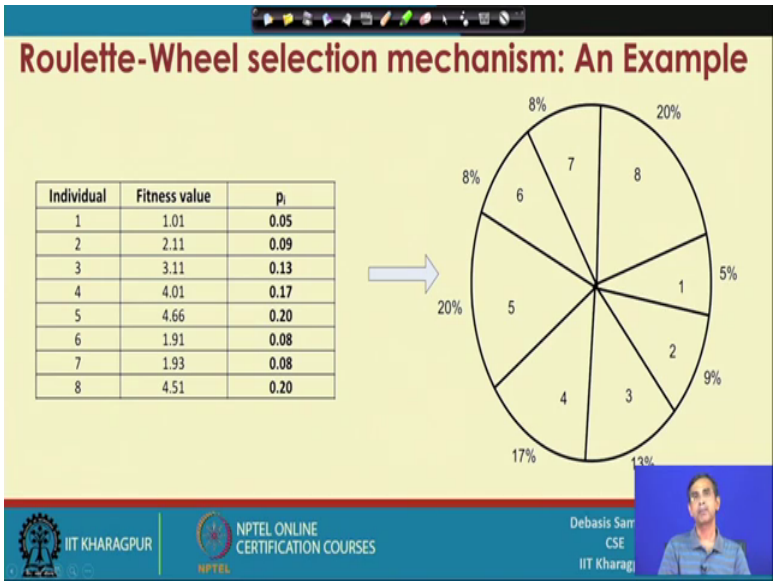
A particular sub-area representing, a solution is selected to be winner probabilistically and probability that the ith area will be declared as the winner is using this formula. So, this is basically the formula that basically Pi, that this particular area will be selected.

Now you can note that this formula pi is basically same as the fitness value, that we have calculated in case of the canonical wheel only the product n is there, but here it is not there and otherwise it is same so, but in this roulette wheel scheme this basically gives the calculation of the probabilistic value, that at any ith rotation of the wheel, the ith individual with the fitness value fi will be selected for the mating pool. Now, having this is the understanding; then let us see how it will work? It is like this.

(Refer Slide Time: 14:59)



(Refer Slide Time: 15:06)



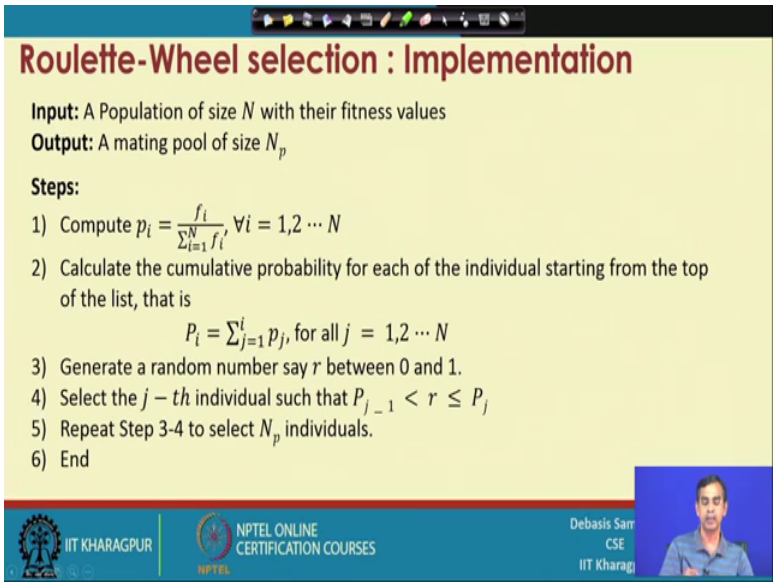
I can give an example, so that you can understand it. Say suppose at any instant there are 8 individuals, and each individual has their fitness core which is mentioned here right. These are the fitness core for 8 individuals is calculated by sum means and then based on this fitness core and using the probabilistic calculation, we can calculate these are the probabilistic value for each individual.

So, for the 8 individuals their fitness can be calculated and finally, their probability of selection can be calculated. Now the same thing it can be represented here, so these are

the different fitness value of the different solution, and then these are the probabilistic value that is there for each.

So, this basically the, so if this is the roulette wheel alternatively this is also the roulette wheel in the tabular form this is the pictorial form of the roulette wheel and this is the tabular form of the roulette wheel. Now, I will see exactly how roulette wheel mechanism can be followed to select the individual.

(Refer Slide Time: 16:16)

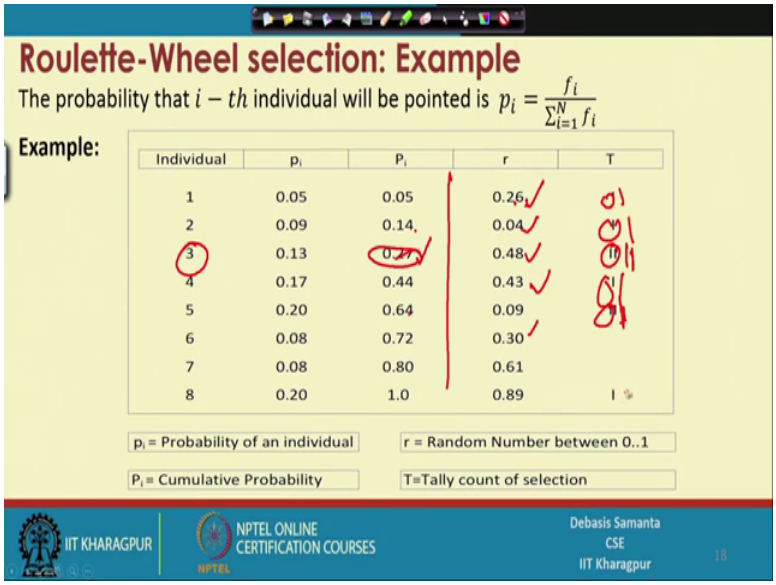


So, there are few steps involved, so as we have said already input is the N number of individuals in a current population, and output is basically we have to select NP number of population out of N, the for the mating. So, the task, task in this method is compute pi for each, even the f i the fitness value for each. Then next step is basically calculate the cumulative probability for each.

So, its cumulative probability value can be calculate in this formula. So, j equals to 1, 2 i, pj for the ith individual and it is denoted by capital P i. So, it is basically called cumulative probability. Now then we have to generate a random number between 0 and r, let this random number be 0 let this random number be denoted as r. Then if the r is in between this cumulative value Pj and Pj minus 1, then select P jth individual as the winner, and we have to repeat these steps right for N times to select N P number of individuals.

So, the idea it is like this we have to calculate f i, the fitness core for each individual and then we will calculate the probability of selection pi, for each individual and then cumulative probability and then generate a random number r and based on the random number r in between 0 and 1, we have to decide the jth individual for the selection. Now, we can give an example, so that we can understand it better.

(Refer Slide Time: 18:00)



Now, here is an one example like this earlier. So, this 8 individuals these are the pi value the probability of searching is here and then these are the cumulative probability. Now, you can see what is the cumulative probability is basically, for the first it gives 0.1 then this is added to give this one, then this is added to this one it is give this one.

So, this way so it is basically the so 27 means, it will add this one, then this 27. Similarly this one means add up to this one this is the cumulative probability this one. So, this way the cumulative probability for each will be calculated. Now, if 8 individuals are there for their pi, small pi and then capital pi can be calculated. So, this is the step next is the selection according the roulette will.

So, idea is that we have to select a random number in between 0 and 1. So, let us see this is the current instant the first toss and the random number is 0.26. Now if so then we have to select the winner out of which how you can select it. So, if the 0.26 is less than to the highest cumulative probability and then. So, 0.26 is basically lie in this one, so 0.26 I

mean; that means, it will. So, it is basically this value should be less than the cumulative value greater than this one.

So, it is basically this one is the selected that means; we will select three. So, if 0.26 is tossed first then the third individual that mean it will be selected. So, it give a tile mark here 1 so this is selected.

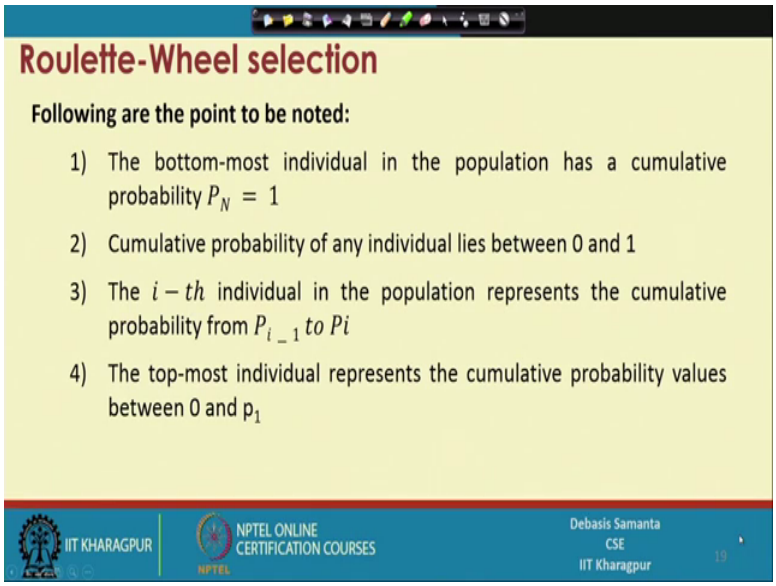
Now, next is 0.4, so 0.4 is basically selecting this one because, it is within less than this one. So, one will be selected, so one will be tiled here, then 0.48 the next toss now 0.48 is basically this one. So, it means select 5, so we can be tile one.

Then 0.43, 0.43 means four. So, the 4 element will be selected; that means, this will be selected, then 0.09, 0.09 will be this one. So, 2 will be selected, then 0.30, 0.30 is this one, so 4 will be selected. So, 0.30, so 0.30 under this category, so 3 will be selected, 3 1 selected first then next 3 again selected, and then 0.61, 0.61 is basically. So, five will be selected, so 5 is selected earlier 5 selected next 5 are selected second times, and 0.89 it basically 8 will be selected so 8 will be selected.

So, this tally mark basically shows the how many individuals are selected and if an individual selected more than once or not it is like this. So, in this example we see the individual one is selected once two is selected once; however, these selected three I mean the individual 3 selected twice, fourth is 1 and then individual 5 is selected twice; however, individual 6 and 7 never selected at all and then individual 8 is selected ones. So, these are way so 8 here basically, 8 round we have carried out an in eighth round how the 8 individual will be selected there.

So, we need not to carry the eight rounds. In fact, we have to carried out NP rounds depending on how many individuals you have to select for the mating pool actually. So, this procedure needs to be repeated for N P times and this time we have to have a random number in between 0 and 1 and accordingly we have to select it. So, this basically the idea about the roulette wheel scheme and then we will follow it there.

(Refer Slide Time: 21:58)

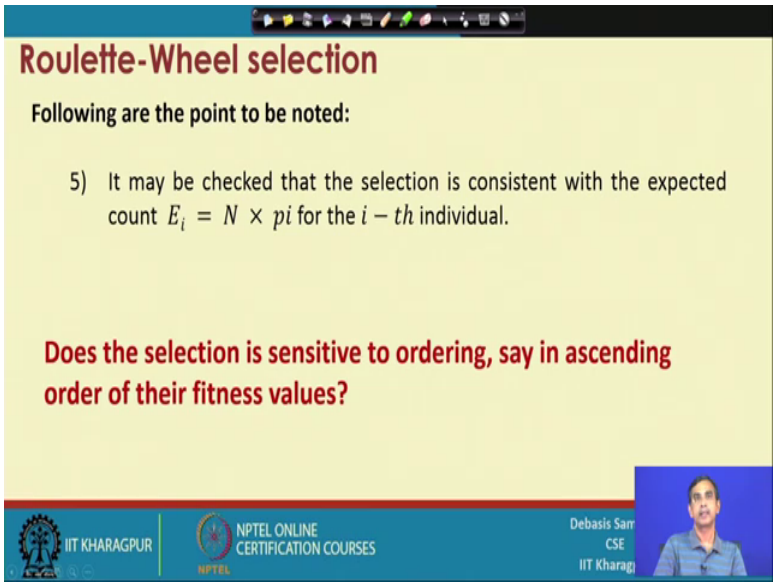


Now, following at the important features in the roulette wheel scheme we can understand it the bottommost individual in the population has a cumulative probability Pn equals to 1. So, this can be example like this so it is the bottommost means this is the bottommost individual and it has the cumulative probability 1.

The cumulative probability of any individual lies in between 0 and 1. So, we can see again the cumulative probability always in between 0 and 1. The ith individual in the population represents the cumulative probability from pi minus 1 to Pi. So, it can be like this, so the ith individual if suppose the fourth individual as the cumulative value between pi to pi minus 1 to pi; that means, third to fourth cumulative probability that is the problem that it will be selected depending on the random number that can be.

And then the topmost individuals, in this representation having the cumulative probability values 0 and then first Pi. So, again this is the topmost one in this top most the cumulative probability in between 0 and 0.5 that is basically the first one, so these are this one. So, these are the property that the cumulative probability holds in this case.

(Refer Slide Time: 23:35)



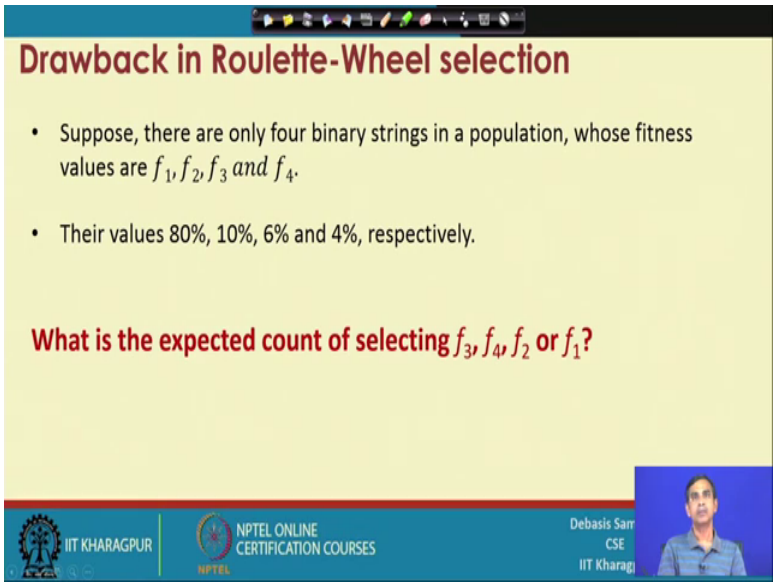
And one idea it is just I have to say it that the expected count that can be obtained that can be also proportional to the probability that the ith element will be selected. So, it basically hold good this formula if Ei denote, that the how many times that Ei will be selected expected for the selection that can be calculated.

If N is the size of the population and pi denotes the probability that ith individual will be calculated then it gives the expected selection now. So, this idea is basically, this basically the idea about the roulette wheel scheme rather. So, in the roulette wheel scheme ok.

So, that tally mark that we have used is basically counting the expected count for each individual actually. So, it is basically N into the multiplication of the pi now here; obviously, one question that arises is that whether this selection is sensitive to ordering say in ascending order of their fitness values if we do of this one.

What I want to say that it is independent of the sensitivity of the ordering, if you take the individuals in any order it will give the results in a probabilistic manner of course, but it is independent of the any ordering scheme that we follow. So, whether all the individuals are order according the fitness value and then their pi value is calculated and then cumulative probability it hardly matters.

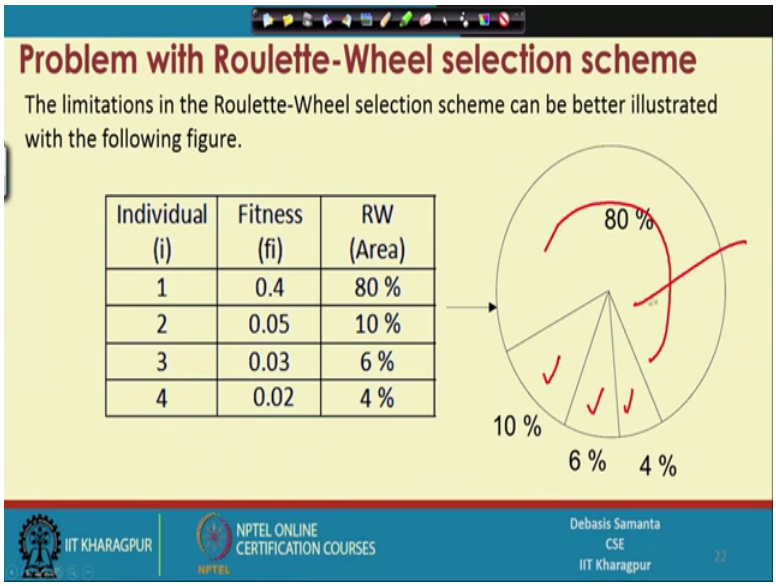
(Refer Slide Time: 25:14)



Now, so roulette wheel selection is more, I mean a better approach than canonical selection because canonical selection is basically very naive approach, but roulette selection gives a favour; however, it suffers from one limitation which I want to discuss it as an example the limitation of the roulette wheel can be understood suppose at any instant there are only four individuals and they are denoted by f 1, f 2, f 3 and f 4, and their fitness value is here.

So, 80, 10, 6, and 4, these are the fitness score in percentage we have represented it anyway. So, it is there or we can say 0.8, 0.1, 0.6 and 0.4 are the fitness core of all these things on there. Then if we apply the roulette wheel scheme, then let us see what will happen to these kind of selection and these basically gives an idea about what is the drawback of the method.

(Refer Slide Time: 26:08)



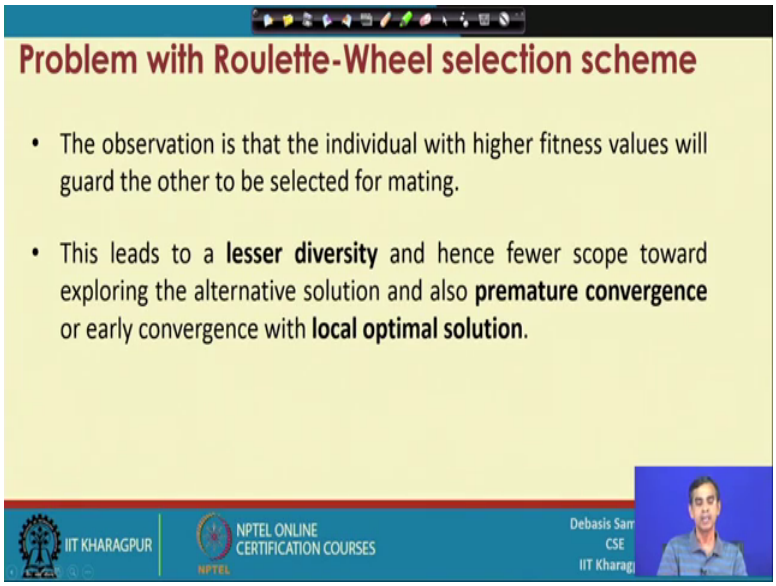
So, it is basically the fitness value of the four individual that we have used it, and then these are fitness core we have represented it in the point formula it is like this basically is the same 80% to this one also same actually the representation.

So, if we see the roulette marking wheel game then the individuals covered with the highest one. And then this is the second individual, third individual and fourth individual, and if we play the game then the chance that the maximum time it will point is the individual one. So, if we run it for the four times the probability that the first individual will be selected; that means, the wheels will favour that individual which having the fitness value it is desirable.

But sometimes it is also not, so desirable because it deprives or it basically ignore the other to be selected, and in the GA strategy the idea is that you have to give a fair chance to all other also to be selected; obviously, the best individual will be selected most of the time, but sometime the other individuals if it is selected it can gives the better diversity in the problem solution.

So, we should not ignore the other individuals also or other that always favouring to a particular individual also not a good what is called the strategy to follow.

(Refer Slide Time: 27:36)



So, that is why this is idea is that the higher individuals with, the individuals with the higher fitness value will favour according to the roulette wheels mechanism and these become a problem or it creates a lesser diversity and hence there is a chance that the genetic algorithm will terminate with a local optimum solution or the premature convergence will resolved.

So, this is the one limitation there in this roulette wheel selection scheme and the same limitations can be overcome using some other approach and this approach is called the roulette selection scheme, and the roulette selection scheme will be discussed in our next slides next lectures.

Thank you.