Allyn Loyd and Natasa Zupanski

R003

Experiments:

Variables:

Chromosome length is the number of alleles in each organism’s chromosome. Population size is the number of organisms in each generation. Selection is the method used to decide which organisms make up the next generation. Truncation effectively kills of the half of the population with the lowest fitness, duplicating the top half to fill their place. Fitness is the method used to determine how fit an organism is. Number of 1s takes the organism with the most 1s to be the fittest. Elitism is the percentage of the generation, taken from the fittest, that carries over to the next generation unmutated. Mutation rate is the probability that each allele in an organism’s chromosome has of flipping from 1 to 0 or vice versa. Crossover is an option that takes organisms in a generation, taking part of one’s chromosome and adding it to part of another to make a new chromosome. It is close to actual biology in which some of our chromosomes come from one parent while others come from another. Generations is the number of max generations that the population can go through if it does not find the solution or organism of peak fitness before this number of generations is reached.

Experiment 1:

This experiment stands as base case so that differentiations in individual variables than be observed and compared. Additionally, this case serves to explore the usefulness of pure mutation, without crossover, in the evolution of a population.

|  |  |
| --- | --- |
| Chromosome length | 100 |
| Population size | 100 |
| Selection | Truncation Selection |
| Fitness | Number of 1s |
| Elitism | 1%  (i.e. best chromosome cloned) |
| Mutation rate | 1% ( i.e. 1.0/N) |
| Crossover | false |
| Generations | 500 to 1000 (terminate at max fitness) |

Graphical user interface, application

Description automatically generated

Hypothesis: It will take more than 500 generations on average for the simulator to produce a perfect organism of fitness of 100. We estimate the number of generations it takes to be closer to 1000.

Graphical user interface, application

Description automatically generated

Results:

In all 10 runs with the above parameters and a max number of generations to run stretched to 1000, not one reached peak fitness of 100. The variation here, is not much, but that is because the data is outside of our scope of observation. The number of runs it takes to find a solution would likely vary quite a lot because the mutation is entirely random, so it could end early almost never. This is a bit surprising considering that one would hope a solution would be found in 1000 generations, but, given the randomness of mutation, it is understandable.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| run # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| gens | >1000 | >1000 | >1000 | >1000 | >1000 | >1000 | >1000 | >1000 | >1000 | >1000 |

What if anything you can conclude and anything you learned or affirmed by doing so?

Random mutation is a highly ineffective method of reaching any high target value or fitness. Most of the remaining generation, while relatively high in fitness, tends to plateau in its max fitness quite early on. Overall, improvement is quick initially with truncation and low mutation but slows to almost zero by the end.

Experiment 2:

Here, crossover is set to true and thus occurs. All other parameters are the same as in Experiment 1. This experiment’s purpose is to test how crossover affects the number of generations required to reach optimal fitness.

|  |  |
| --- | --- |
| Chromosome length | 100 |
| Population size | 100 |
| Selection | Truncation Selection |
| Fitness | Number of 1s |
| Elitism | 1%  (i.e. best chromosome cloned) |
| Mutation rate | 1% ( i.e. 1.0/N) |
| Crossover | true |
| Generations | 500 (terminate at max fitness) |

Hypothesis: It will take less than 500 generations on average for the simulator to produce a perfect organism of fitness of 100. We estimate the number of generations it takes to be closer to 100. This is because crossover has a higher probability of randomly choosing a crossover point that produces a chromosome of all 1s, it has a range of possible alleles to choose as a point that would produce the solution whereas mutation theoretically has only one point that would produce the solution once an organism is 1 away from its max fitness.

A picture containing graphical user interface

Description automatically generated

Results:

All 10 runs stayed in the range of 40 to 60 generations, with the median being 52.5. This was lower than our hypothesis predicted, but our hypothesis was close to the truth in assuming crossover to lower the number of generations it took to get a solution drastically. The variation here is quite small, but this makes sense. Crossover reduces from the element of randomness, making a solution far more likely quite quickly, which confers to a smaller range of generations needed.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| run # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| gens | 58 | 53 | 44 | 57 | 59 | 49 | 52 | 50 | 49 | 57 |

What if anything you can conclude and anything you learned or affirmed by doing so?

Crossover overwhelmingly decreases the amount of generations needed to reach peak fitness.

Experiment 3:

This experiment sets the mutation rate to 0 percent. All other parameters are the same as in Experiment 2. The aim of this experiment is to test the effectiveness of mutation in the number of generations required to reach optimal fitness when crossover is used.

|  |  |
| --- | --- |
| Chromosome length | 100 |
| Population size | 100 |
| Selection | Truncation Selection |
| Fitness | Number of 1s |
| Elitism | 1%  (i.e. best chromosome cloned) |
| Mutation rate | 0% ( i.e. 1.0/N) |
| Crossover | true |
| Generations | 500 (terminate at max fitness) |

Timeline

Description automatically generated

Hypothesis: It will take less than 500 generations on average for the simulator to produce a perfect organism of fitness of 100. We estimate the number of generations it takes to be closer to 250. Crossing over the initial random population seems to lead to a no less random population, thus it will take longer than if there was initial mutation; however, this would also seem to prevent the worst organism from going down.

Results:

All 10 runs went of 500 in the generations it took to find the solution. Indeed, none of them could have encountered the solution because the worst solution rose to meet the best one so that all organisms had the same imperfect chromosome that, without mutation, could never change. Our hypothesis did not foresee this, but we did predict that no mutation would take longer to find a solution and were correct in that respect.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| run # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| gens | >500 | >500 | >500 | >500 | >500 | >500 | >500 | >500 | >500 | >500 |

What if anything you can conclude and anything you learned or affirmed by doing so?

Crossover provides an immense, initial rise in fitness for a population, but it cannot be relied on to get that population to the greatest fitness. Once the worst solution rises to meet the best one, with no mutation the population will never reach the solution. It is simply not possible. Mutation is therefore required for any population to continue to rise in its fitness.

Experiment 4:

The population in this case is 1000. The rest of the parameters are the same as in Experiment 2. This experiment aims to understand the impact of increasing the population size on the number of generations it takes to reach optimal fitness.

|  |  |
| --- | --- |
| Chromosome length | 100 |
| Population size | 1000 |
| Selection | Truncation Selection |
| Fitness | Number of 1s |
| Elitism | 1%  (i.e. best chromosome cloned) |
| Mutation rate | 1% ( i.e. 1.0/N) |
| Crossover | true |
| Generations | 500 (terminate at max fitness) |

Chart

Description automatically generated with medium confidence

Hypothesis: It will take less than 50 generations on average for the simulator to produce a perfect organism of fitness of 100. We estimate the number of generations it takes to be closer to 40 on average. A higher population means more chances for one to reach peak fitness, so it will take fewer generations to reach this.

Results:

All 10 runs went below 50 generations as expected, with the median being 38.9 with the mode being 38. It was quite surprising to get three runs of 38 in a row, but random chance does sometimes work out like that. Our hypothesis was correct in guessing lower than 50 and around 40 generations needed to find the solution.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| run # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| gens | 41 | 43 | 40 | 38 | 38 | 38 | 37 | 39 | 33 | 42 |

What if anything you can conclude and anything you learned or affirmed by doing so?

Increasing the population or generation size decreases the number of generations needed to find a solution; however, our increase from 100 to 1000, a factor of 10, only brought this number down by around 10 generations. It takes a large increase in the generation size to make an impact. Additionally, the variance in this sample is smaller than in the one with the population size of 100, which had a range of about 20 in its generation values. The range here is about 10, going from around 30 to around 40.

Experiment 5:

With the same parameters of Experiment 2, change the mutation rate to 5. We use this experiment to examine the influence of higher mutation rates in the number of generations required to reach peak fitness.

Hypothesis: It will take longer to find the solution.

Results:

Within the 500-generation period, none of the organisms reached peak fitness, and the best fitness plateaued below a fitness of 90.

Conclusion:

A seemingly small increase in mutation rate can lead to an immense increase in the number of generations required to find the solution.

Experiment 6:

With a population size of 1000 and using Roulette Wheel as a method of selection, we take the same parameters as in Experiment 2 to understand how this selection method impacts the number of generations required to reach peak fitness.

Results:

Within the 500-generation period, this population did not reach peak fitness, but got to a best fitness of around 85.

Conclusion:

Roulette Wheel is extremely ineffective as a selection method and this cannot be easily made up by altering other variables, such as the population size.

Experiment 7:

Starting with the initial parameters of Experiment 1, the fitness method will be changed in the middle of the evolutionary process. The fitness method with change from Number of 1s to Target Organism.

Hypothesis: Without crossover, it will be similarly hard for the population to reach peak fitness, even after it switches fitness method. This switch would likely only add time needed to reach peak fitness as the fitness would essentially drop as soon as the fitness method was switched.

A picture containing graphical user interface

Description automatically generated

Results:

The fitness dropped as soon as the fitness method was changed, but it improved as almost the same speed as it was with the previous fitness method. While peak fitness was not reached within the 500 generations, we can postulate that this setback in fitness would make the number of generations needed to reach peak fitness larger than had the fitness method not been changed.

Conclusion:

Changing the fitness method like this is not productive to reaching peak fitness; however, that does not mean something like this could not happen in nature. It is possible that what an organism needs to be to be the fittest could change rapidly if its environment and general situation change rapidly as may be the case in certain natural disasters. This shows how such events can set back entire populations, potentially even wiping them out if they can’t adapt quickly enough. There is no current relationship between diversity and these results as the diversity stays nearly constant; however, in a real-life situation where natural disasters change the environment and criteria for survival so quickly, there would be a decrease in diversity. A potentially better way to model such a situation would be to use a version of ranked selection that would evolve the population at a slower rate while allowing the initial fitness method to remain for a longer period. This would sufficiently evolve the population away from the second fitness ideal, introducing a more severe method of selection to reflect the high-danger environment. This would likely drastically decrease the diversity.

Research Replication:

|  |  |
| --- | --- |
| Chromosome length | 20 |
| Population size | 1000 |
| Selection | Learning Chance |
| Fitness | N/A |
| Elitism | 0%  (i.e. best chromosome cloned) |
| Mutation rate | 0% ( i.e. 1.0/N) |
| Crossover | True |
| Generations | 100 (terminate at max fitness) |

This hopes to replicate the findings of the research paper that started this project. The results of this paper show the undesirable alleles, the zeros, quite quickly reaching zero in the population while the desirable alleles, the 1s, take over as much of the space as they can. Meanwhile, the number of ? alleles in the population says almost constant.

A picture containing graphical user interface

Description automatically generated

The blue line in the graph above shows the number of 1s in the population, they cyan line the number of zeros. The grey line shows the number of ?s in the population. As one can see, the results of the research paper are replicated. The blue line goes up suddenly as the cyan line goes down suddenly, indicating an increase in 1s and a decrease in 0s that reaches 0.

Interestingly, this process even works with a mutation rate of 1 per the length of the chromosome, or, in this instance, 5%. Higher levels of mutation have it take longer before the distinct split in the line of the 1s and 0s can be seen.

A picture containing graphical user interface

Description automatically generated

It would be interesting to see how such a simulation would play out if the alleles of the chromosome were slowly set to be constant. This would decrease the amount of ? alleles in the population over time as a function and perhaps hinder the progress of the population if alleles were set as zeros, while helping the population if alleles were set as 1s. It would also be interesting to run an experiment where the number of generations until there were no zero alleles was determined.