Topic 8 Assignment

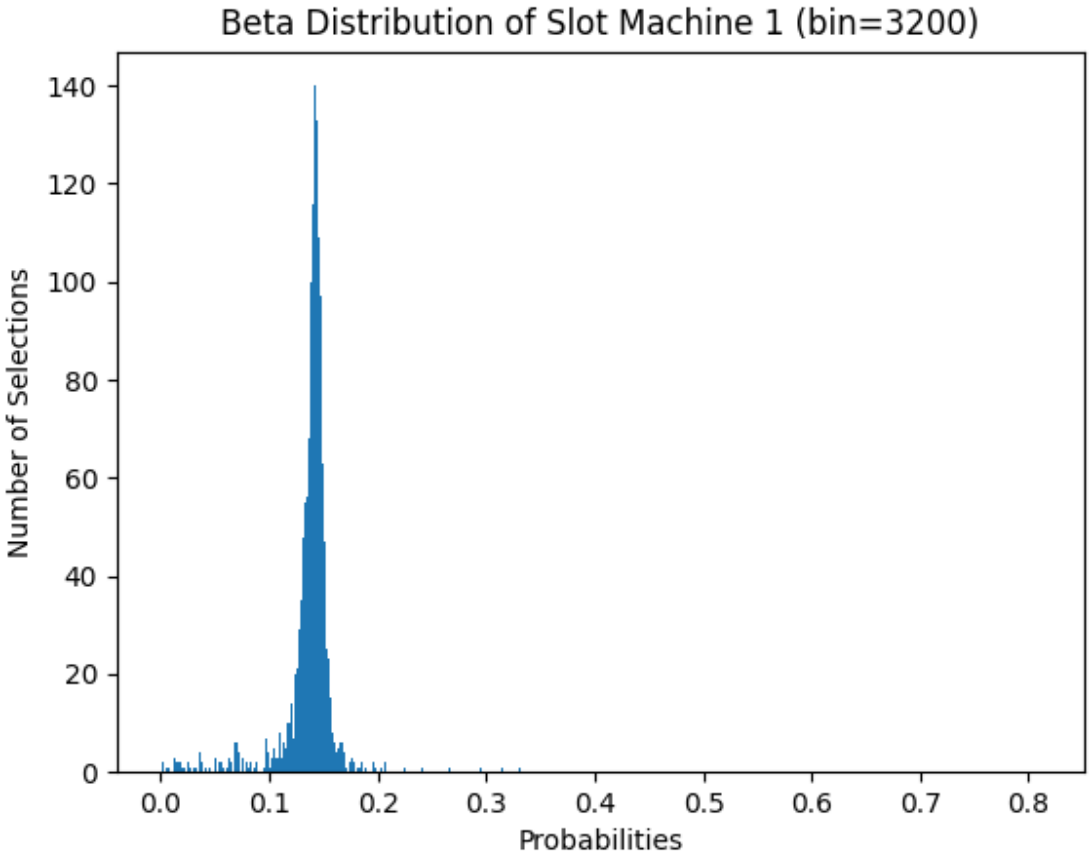
Reinforcement learning is a part of machine learning that focuses on maximizing a reward system. These algorithms are comprised of four main elements:

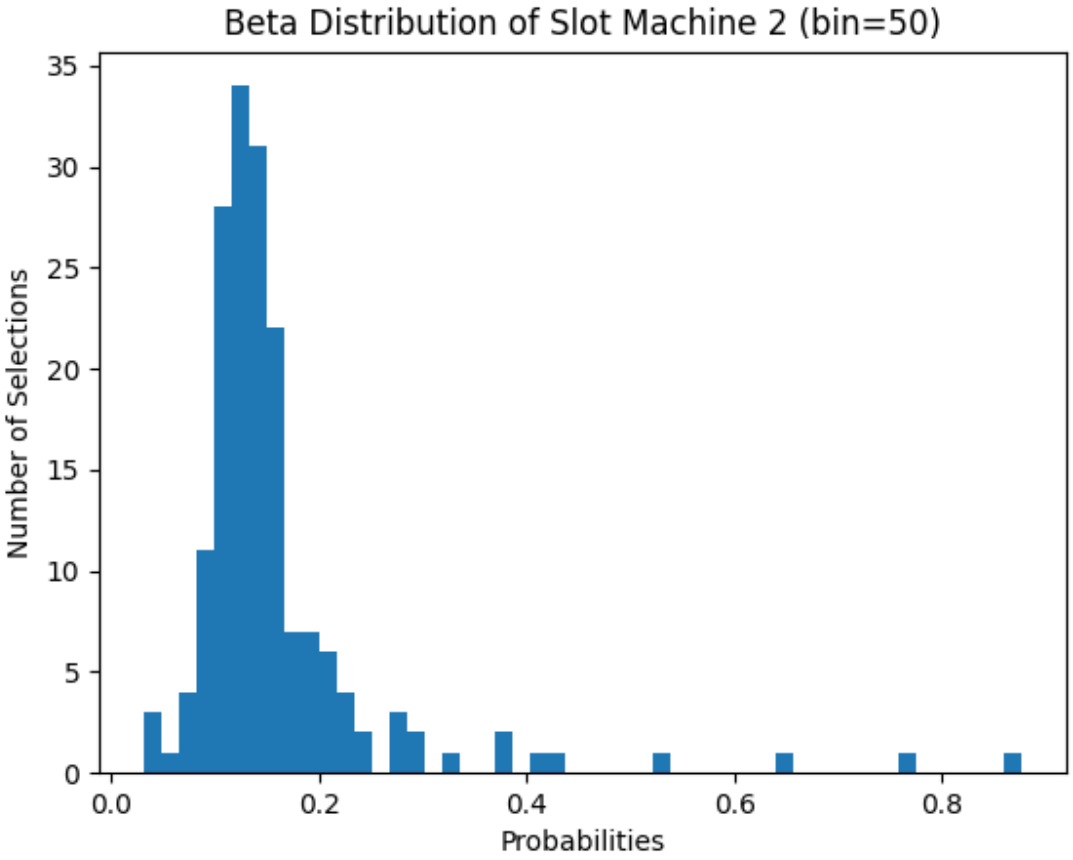
1. *Agent*: The algorithm that interacts with its surrounding. This algorithm receives rewards based on how well it is able to predict the state of its environment.
2. *State*: A snapshot, or piece of the environment that the agent is able to react with. this would surmount to be the data that the agent needs to make its next decision based on the environment.
3. *Reward*: special values that the algorithm tries to maximize over a period of time.
4. *Environment*: This is the overall data or everything outside of the agent. This is where the agent receives the different states based on its actions. The environment also provides a reward together with a new state when the agent takes action (this reward can vary based on whether the action is taken).

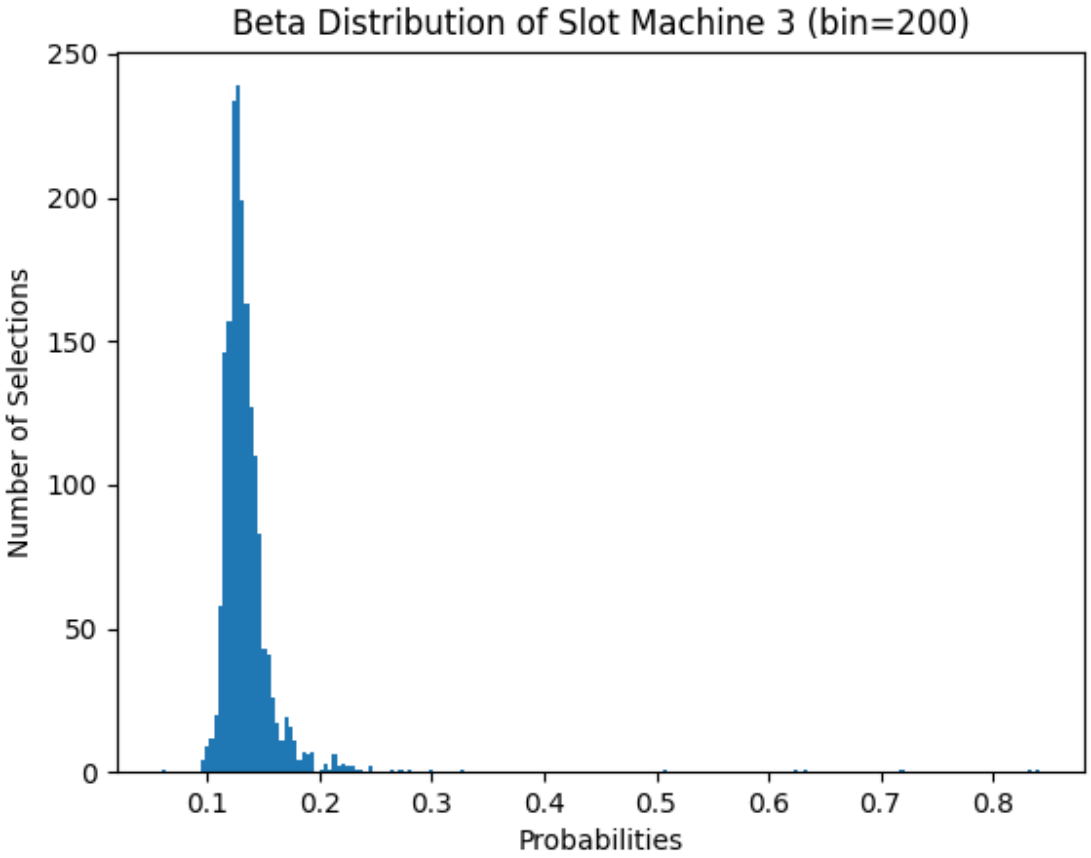
With the aforementioned four elements, one has the basis for a reinforcement learning algorithm. A great example that applies to reinforcement learning is the Multi-Armed Bandit Problem. This problem begins by mentioning that a gambler walks into a casino to gamble his or her money on multiple single-armed slot machines. The gambler then proceeds to explore the slot machines to ascertain which slot machine provides the highest reward, after this, the gambler proceeds to use the slot machines with the highest probability of reward more than others until the gambler obtains that maximum reward. Although this may seem to be a very solid strategy for gambling money, it can have its downside (the best slot machine might not perform best upon initial inspection, and one may be prioritizing a slot machine that is not the best). In this case, the Thompson Sampling algorithm attempts to determine the most rewarding slot machine based on empirical data provided. This algorithm achieves this by using the beta distribution to discover the best-sampled mean between other machines. As the number of actions is provided, the more the beta distribution approaches the empirical mean. Before moving on to explain the Thompson Sampling Algorithm, let's define enumerated terms to apply to the bandit problem:

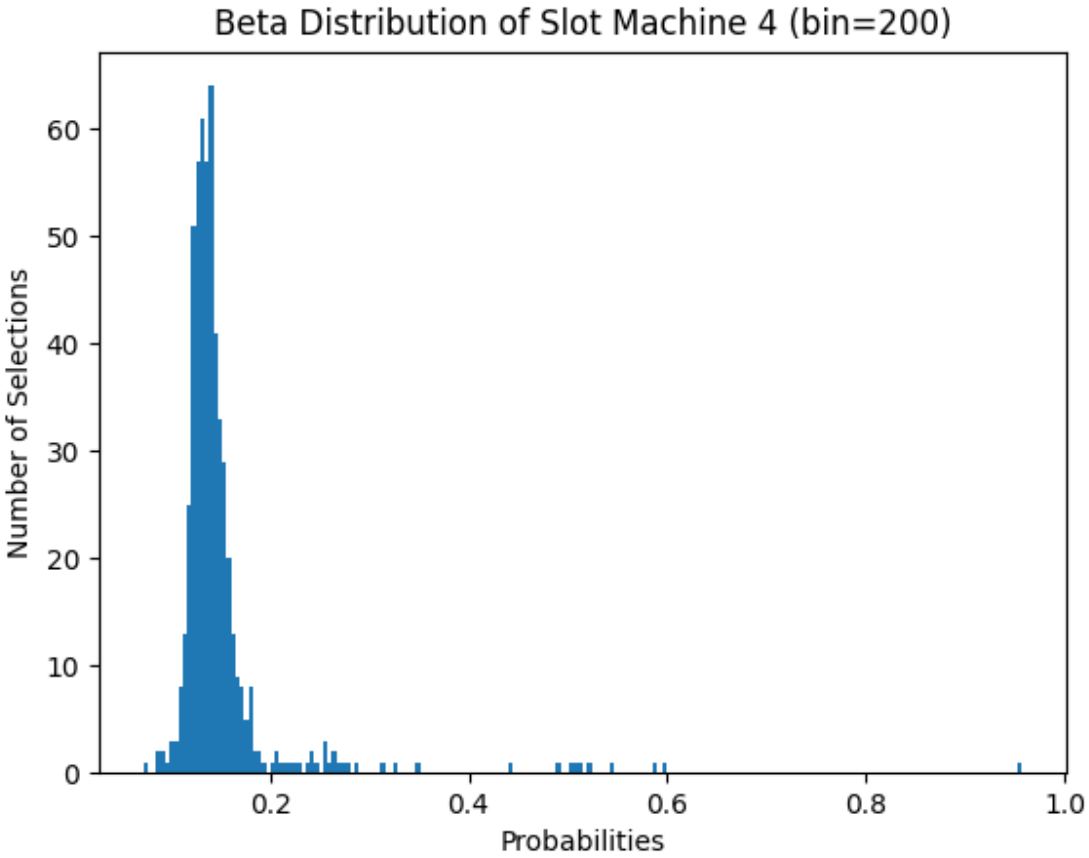
1. *Agent*: The gambler; In this case, the gambler is interacting with everything around himself.
2. *State*: The accumulated number of wins based on the number of trials done.
3. *Reward*: The slot machine's probability of winning.
4. *Environment*: The casino itself is the environment. The agent can react to the result of each chosen slot machine at different iterations (or trials).

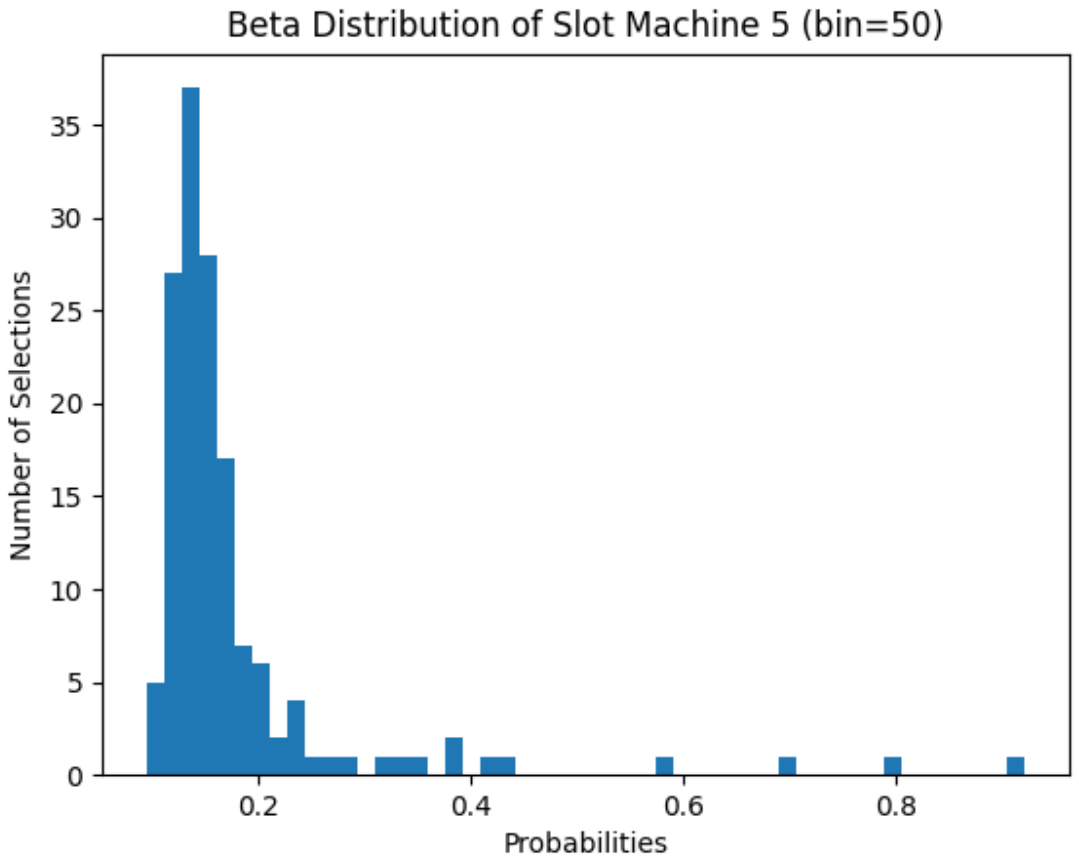
The expected probability for each Slot is 0.15, 0.04, 0.13, 0.11, 0.05. Based on these probabilities and the beta probability distribution figures below, one can state that the first slot perfectly matches the expected probability of 0.15. this may be attributed to the fact that the sample size of the first slot machine is incredibly large (6,541). Slot Machine 2 does not follow the same pattern as it has a sample size of 68 and a probability more closely related to 0.15 (which is more than three times the expected probability). Slot machine 5 follows a similar pattern with a probability closer to 0.15 (sample size=109). This may be due to the extremely small sample size that has not allowed the algorithm to appropriately ascertain the real probability of slot machines 2 and 5. Despite this, the algorithm managed to predict that these two slots contain the worst probabilities as it kept the distribution very low. Slot machines 3 and 4 behaved similarly to slot machine 1, they both represented expected probabilities within their respective distributions.


Beta Distribution of Slot Machine 1 (bin=3200)
Number of Selections
Probabilities
0
20
40
60
80
100
120
140
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


Beta Distribution of Slot Machine 2 (bin=50)
Number of Selections
Probabilities
0
5
10
15
20
25
30
35
0.0
0.2
0.4
0.6
0.8


Beta Distribution of Slot Machine 3 (bin=200)
Number of Selections
250
200
150
100
50
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Probabilities


Beta Distribution of Slot Machine 4 (bin=200)
Number of Selections
Probabilities
0
10
20
30
40
50
60
0.2
0.4
0.6
0.8
1.0



Below is a count of the times each slot machine won and the times that each slot machine did not win for all of the trials that the slot machines were used:

$N_1^1(6,541) = 948$

$N_1^0(6,541) = 5593.0$

$N_2^1(68) = 2$

$N_2^0(68) = 66$

$N_3^1(2,276) = 301$

$N_3^0(2,276) = 1975$

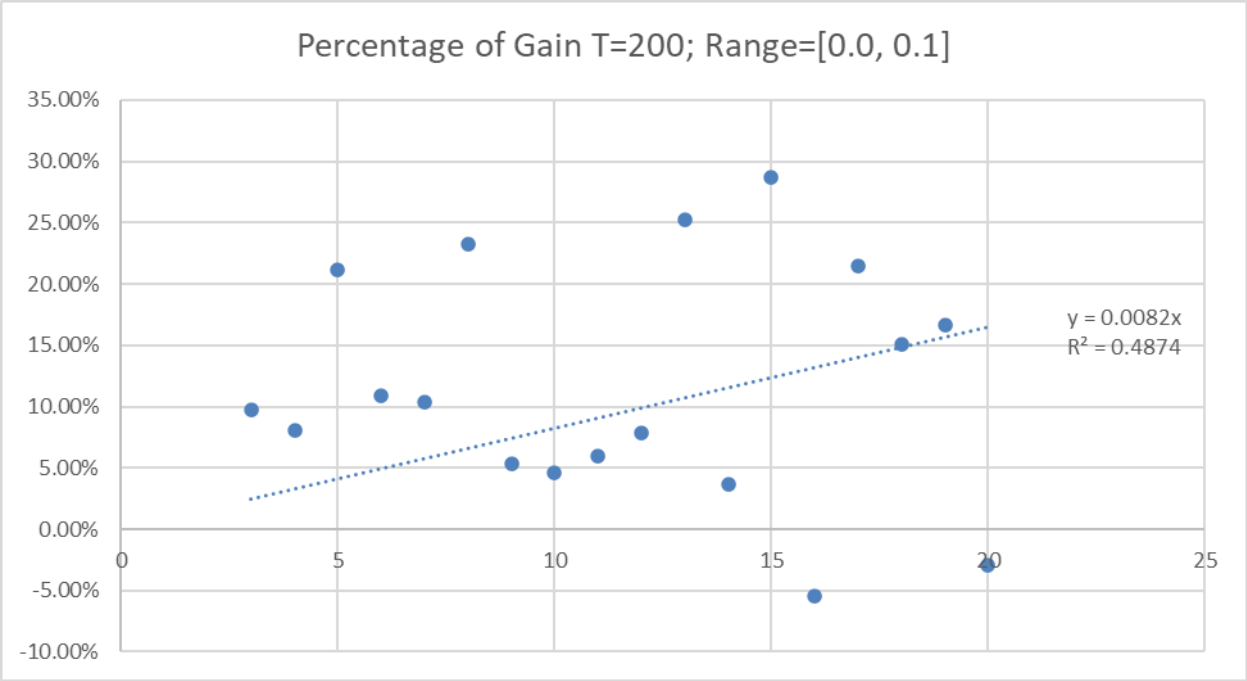
$N_4^1(1,006) = 124$

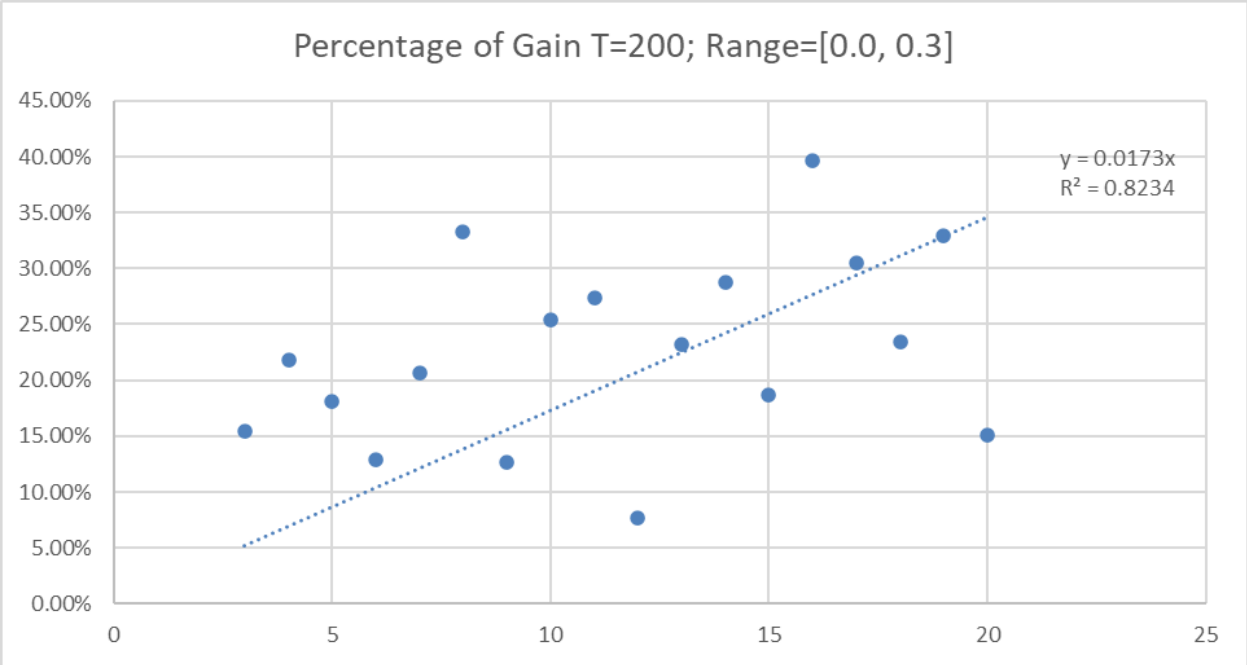
$N_4^0(1,006) = 882$

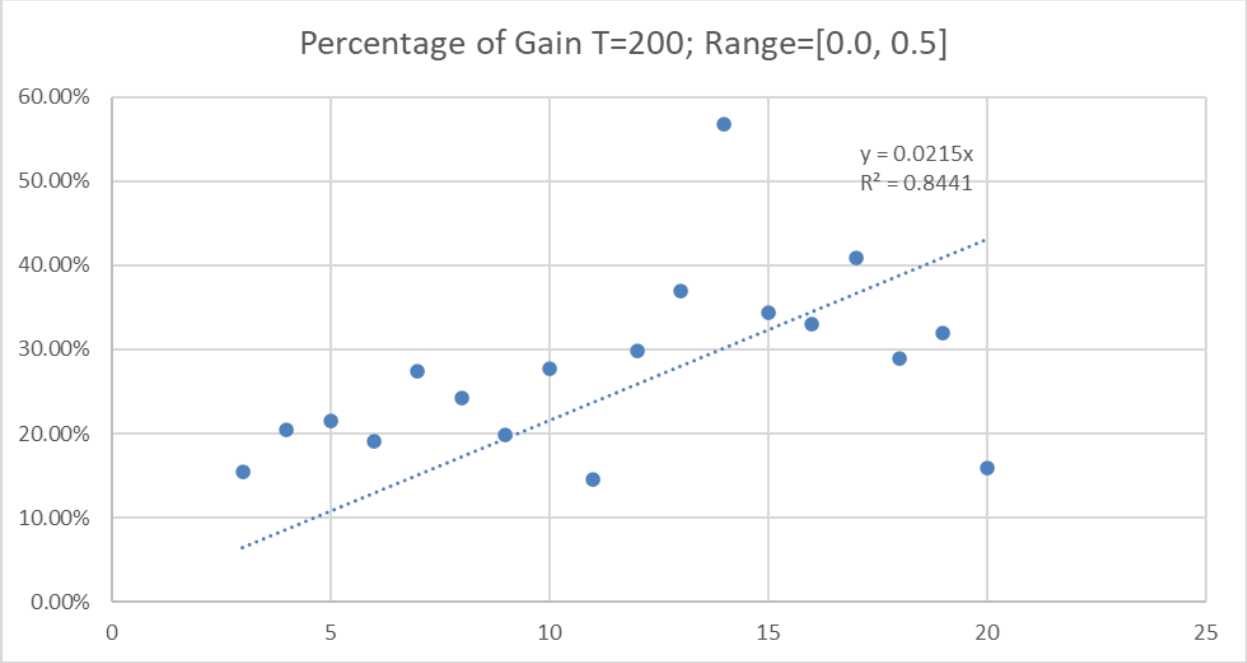
$N_5^1(109) = 5$

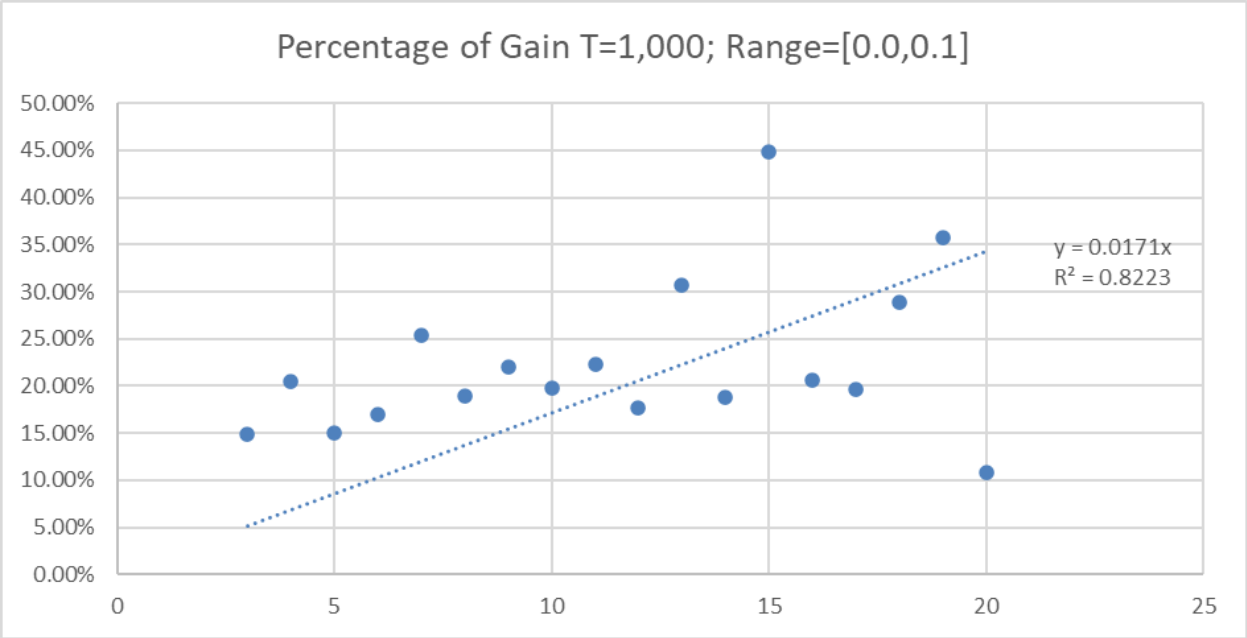
$$N_5^0(109) = 104$$

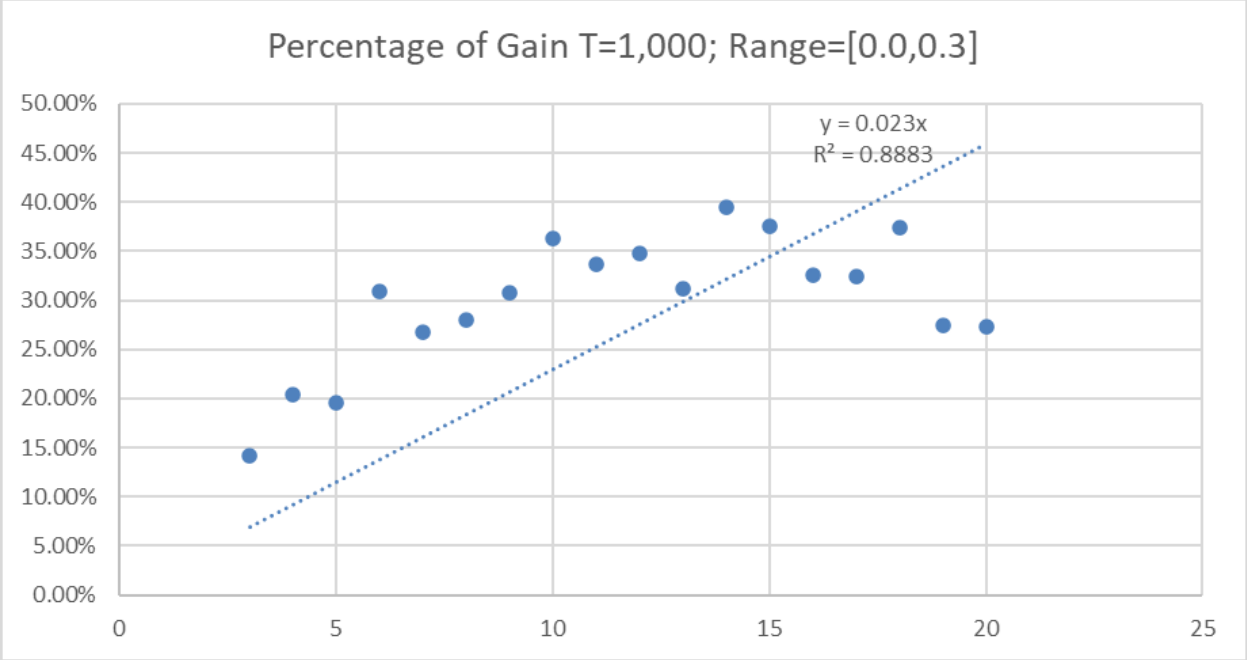
As can be observed from the above, the slot machine with the lowest use was slot machine two (with an expected probability of 0.04), followed by slot machine number 5 with an expected probability of 0.05. The undisputed best slot machine was slot machine one, which had 948 wins out of 6,541 uses (this slot machine also contains the highest expected probability of 0.15). Now that one is able to tell that the Thompson Sampling Algorithm is able to accurately determine the best slot machine, let's compare this algorithm with the standard sampling algorithm.

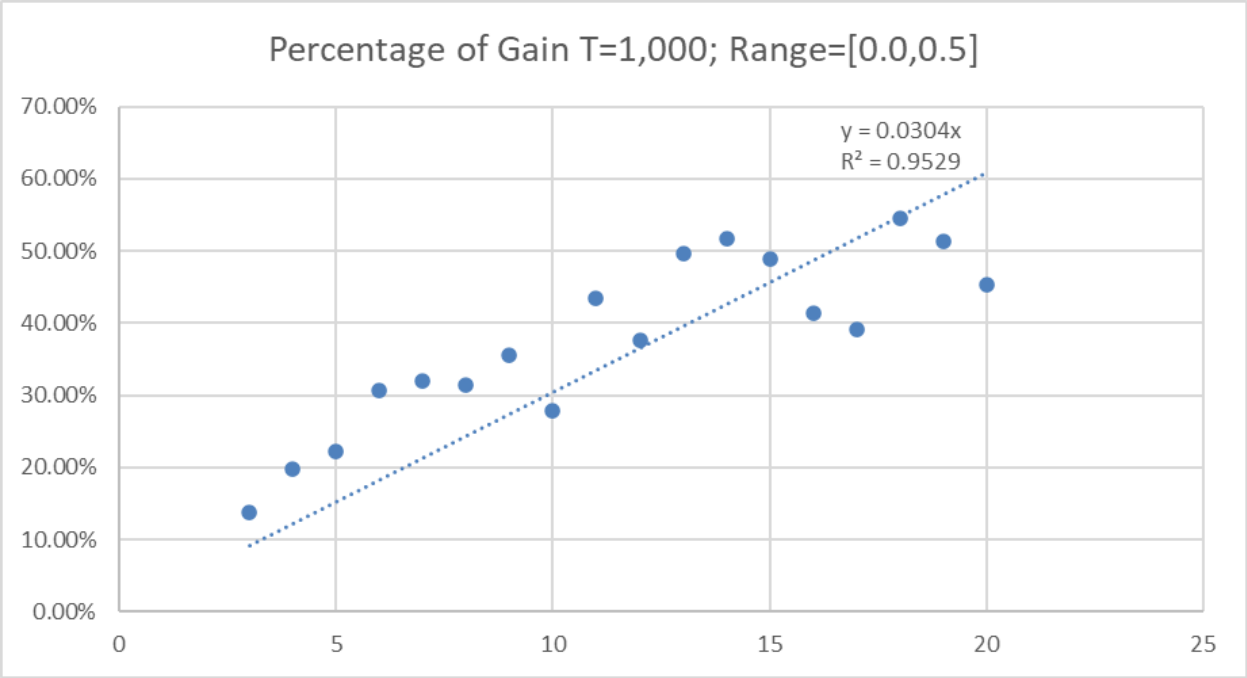


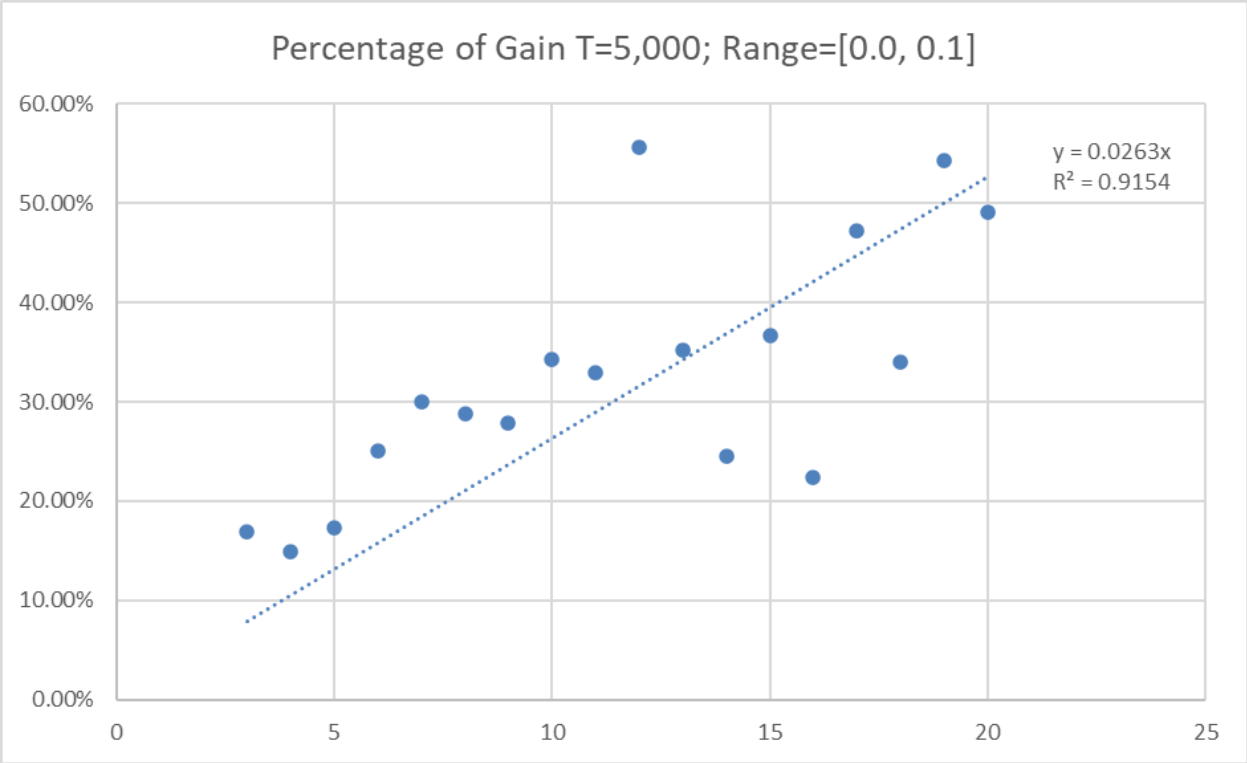

Percentage of Gain T=200; Range=[0.0, 0.3]
45.00%
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
0
5
10
15
20
25
y = 0.0173x
R² = 0.8234

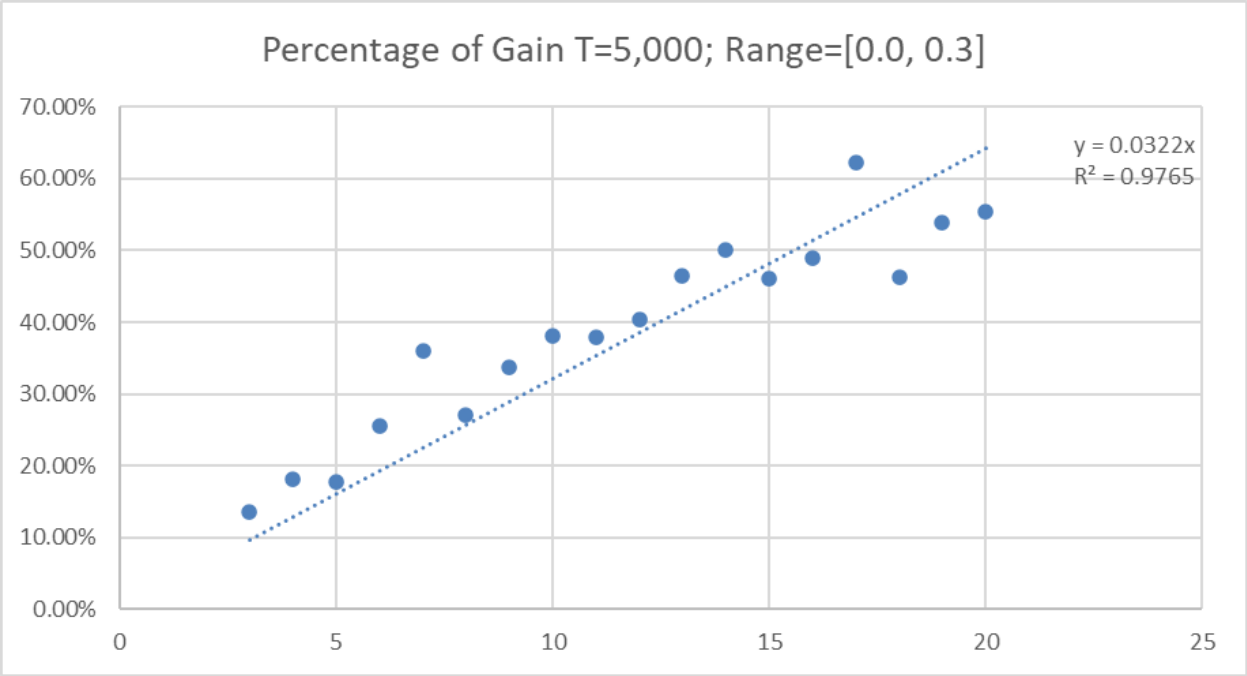

Percentage of Gain T=200; Range=[0.0, 0.5]
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
0
5
10
15
20
25
y = 0.0215x
R² = 0.8441

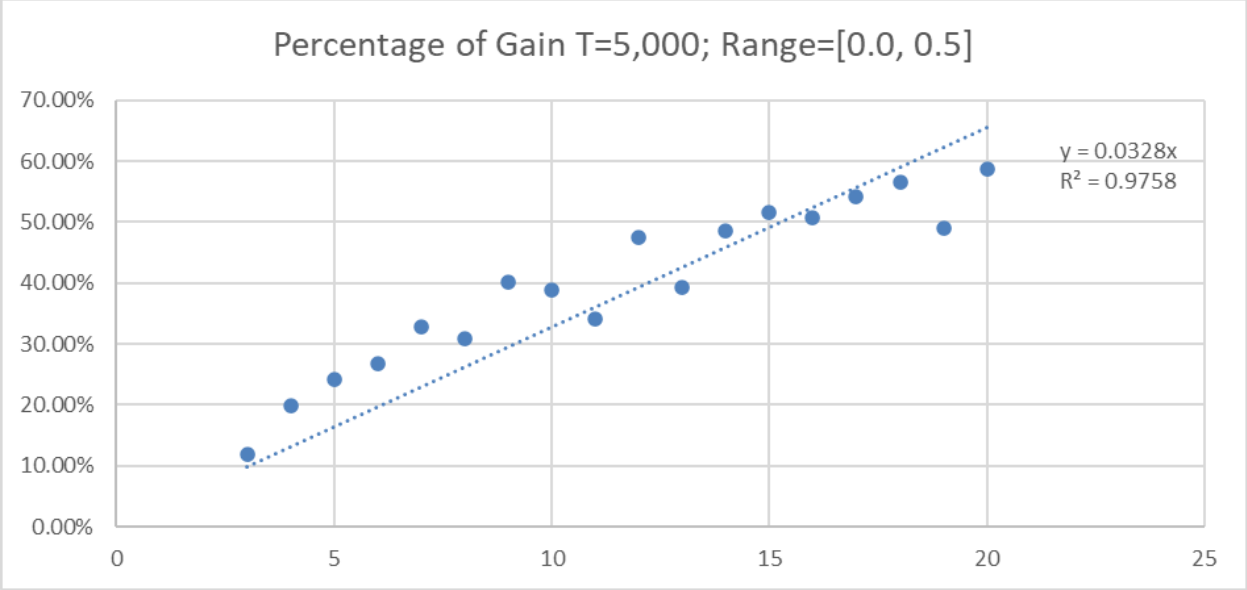

Percentage of Gain T=1,000; Range=[0.0,0.1]
50.00%
45.00%
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
0
5
10
15
20
25
y = 0.0171x
R² = 0.8223


Percentage of Gain T=1,000; Range=[0.0,0.3]
50.00%
45.00%
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
0
5
10
15
20
25
y = 0.023x
R² = 0.8883


Percentage of Gain T=1,000; Range=[0.0,0.5]
y = 0.0304x
R² = 0.9529
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
0
5
10
15
20
25


Percentage of Gain T=5,000; Range=[0.0, 0.1]
y = 0.0263x
R² = 0.9154
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
0
5
10
15
20
25





Scatter plots were made from the results of comparing the standard sampling to the Thompson Sampling algorithm at varying sampling sizes (200, 1000, and 5000). The percentage of gain from the Thompson Algorithm was then calculated using the equation $Gain = (ThompsonSampling - StandardSampling)/StandardSampling$. These gain plots show that as the range increases for the different slots, the Thompson sampling algorithm increases the gap between itself and the standard sampling method increases. Similarly, as the

sample size increases, so does the gap between the two sampling methods as well (this is all shown by the linear correlation between the gain and the number of slots). This coordinates with the standard Thompson Sampling method which is able to discover the ideal way to maximize rewards via its fast sampling capabilities.

Ethical Design Specification

Ethical design is a methodology that focuses on attributing design and specifications to one's ethical values. The first train of thought and the current main topic about ethical algorithms is regarding Amazon's algorithms to efficiently assign tasks to workers at the warehouse. As the warehouse workers have unionized, they have voted toward a law that enforces the algorithms to follow certain ethical rules for a more humane and less dangerous work environment. This law will enforce the responsibility of software developers, data scientists, and executives to develop and plan algorithms that provide tasks that are possible in a comfortable and able work environment. Amazon is infamous for a harsh work environment where warehouse workers often did not have the opportunity to use the bathroom throughout the day or even eat due to the demand of the job.

One way to verify whether the algorithm has changed would be to sample whether workers at different warehouses had the opportunity to eat comfortably for each day of the week, or how many times each warehouse worker sampled has gone to the bathroom throughout the day. One can then observe past statistics and compare the past statistics before the law takes effect and the statistics after to see if there is a difference. One may also observe whether the number of tasks assigned per warehouse worker assigned has changed, or even observe the tasks completed vs the number of total tasks assigned to each worker of the warehouse and observe if the rate of tasks completed a day per total has increased.

Many statistical methods could be used to verify whether the law enforcing ethics in technology would provide. Although Amazon shopping has become a common commodity throughout the United States, the majority of the Amazon business is

focused on Amazon Web Services which causes the headquarters executives to neglect the work environment in warehouses as it is not as significant. This is where having ethical principles predetermined can become a great advantage for managing large companies such as amazon. From a Christian perspective, the status of amazon's abuse towards its warehouse workers is something that should not be allowed. Jesus had once stated to treat others you would like to be treated to summarize how to accomplish all of the law as one who treats their fellow man as themselves would not abuse their fellow man in this world. The fact of the matter that the status of amazon warehouses is not related to greed or money, but rather a disregard of one's fellow man simply because it is possible. Although they have been working under terrible working environments, this does not mean that it should become the norm, God created work for Adam by having him name all animals in his kingdom and made him and Eve as representatives of God for the sake of exploring all of God's creation and find the good in creation. Due to the sin of man, their nature becomes one where work is a weight, and the norm of the work environment is one where competition is king and everyone tries to upstart the other as opposed to working together. Not all environments are like this, there are workplaces where cooperation and enjoyment of work is encouraged, but ones where this is perpetual are rare.

Source(s)

1. Thompson Sampling for Multi-Armed Bandit Problem in Python (codespeedy.com)
2. Thompson Sampling for Multi-Armed Bandit Problem | by Amit Ranjan | Analytics Vidhya | Medium
3. The Intuition Behind Thompson Sampling Explained With Python Code (analyticsindiamag.com)
4. Basic Understanding of Environment and its Types in Reinforcement Learning - MLK - Machine Learning Knowledge
5. Reinforcement Learning: An Introduction to the Concepts, Applications and Code | by Ryan Wong | Towards Data Science
6. 3.1 The Agent-Environment Interface (incompleteideas.net)
7. States, Observation and Action Spaces in Reinforcement Learning | by #Cban2020 | The Startup | Medium

8. Multi-Armed Bandit Analysis of Thompson Sampling Algorithm | by Kenneth Foo | Analytics Vidhya | Medium
9. The principles of ethical design (and how to use them) - 99designs
10. Exploratory reflection on design ethics | by Rauno Pello | Medium