The Hong Kong University of Science and Technology

MSDM 6980 report

Monte Carlo approach to simulating the Monte Carlo "gambling" games

Author:

Manqing Fu

Student ID: 20647103

Supervisor:

Junwei Liu

May 27, 2020

1 Background

Blackjack is a famous gambling game around the world. People bet on the sum of hand cards and the bigger one gain the reward once a time. In this report, it only consider the 1 to 1 game gambling mode from the gamer side to the dealer side and the reward is \$1, \$0 ,\$-1. The detail process is in Figure 1

In this project, I firstly complete the simulation code of Blackjack by using the open-AI package. Then design two gambling strategies for gamer to decide whether take cards and run the simulation 1. According to the result of part 1, I planned to see the performances of strategies in a long term and ran the simulation 2. The last section used two indexes, gambling round time and the expected return, to describe the gamer's happiness during the gambling. More than this, to figure out their relationship with their greedy ratio and risk tolerance ratio. Simulation 3, 4 and 5 show the results.

2 Game strategies

In each gambling round, the gamer can choose whether to take a new card. Suppose the gamer has infinite money and counts the win-loss situation of each round in this section. For both the random strategy and Monte Carlo strategy, the report shows their win-ratio and their average expected return.

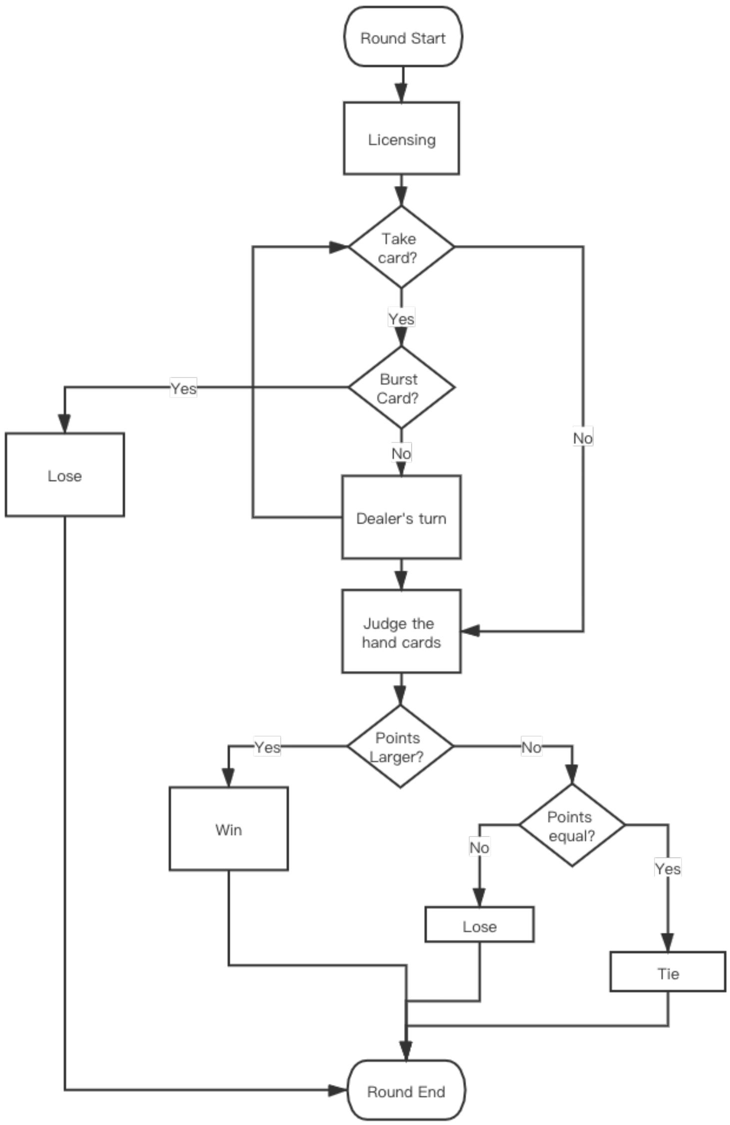

Figure 1: The gambling process and judgement standard of Blackjack.

2.1 Random strategy

With random strategy, gamer will choose to take card with possibility of $P_R = \frac{1}{2}$ and not to take with possibility of $1 - P_R$. After 10^6 simulation times, the gamer's win rate is approximately 30% and the average expected return is \$-0.396. The win-loss possibility distribution of gamer's reward during this simulation is in Figure 2.

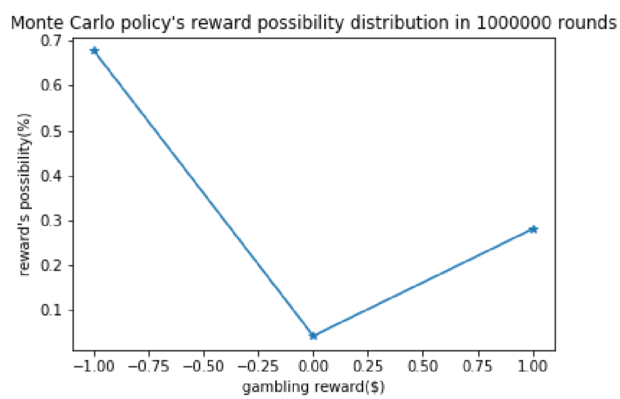

Figure 2: Using the Random Strategy simulate for 10^6 times.

2.2 Monte Carlo strategy

Monte Carlo Strategy consider the Blackjack gambling as a Markov Decision Process issue. In Monte Carlo theory, the process is composed by a time series list which contains the *states*, *actions* and *rewards*. In $t = 0, 1, 2, 3, ...$, the Monte Carlo sequence is like:

$$S_0 \rightarrow A_0 \rightarrow R_0 \rightarrow S_1 \rightarrow A_1 \rightarrow R_1 \rightarrow ...$$

The Monte Carlo strategy is decided by the formula Equation (1). When the gamer facing with a state (gamer's hand cards and dealer's hand cards) and get the corresponding rewards. Gamer's action decision is the action with the maximum expected return according to in the matrix Q.

$$\max Q(s,a) \tag{1}$$

In this project, Q is a $19 \times 11 \times 2$ matrix. [1] Before the official simulation, the state-action value matrix needed to calculated out in advance, in which the Q is a converged solution among the per simulation. The pre-simulation process is in Figure 3. The value α is the minimum converged circle time. In Figure 4 and Figure 5, typically discussed about the value of α. This simulation represents the Q with different training gambling round times and their performances scores with different round-time-training Q. The score is calculated by the expected return of another 10^3 times of simulation using the corresponding Q. From Figure 5 we can see that approximately after 2.5×10^4 round times, the expected return of Q wanders around \$-0.09, so that $\alpha = 2.5 \times 10^4$.

After 10^6 rounds gambling times using Monte Carlo strategy, the gamer's win rate is 44% and the expected return is \$-0.082. The win-loss possibility distribution of gamer's reward during this simulation is in Figure 6.

[1] According to the Open-AI environments observation structure. Related source code: https://github.com/openai/gym

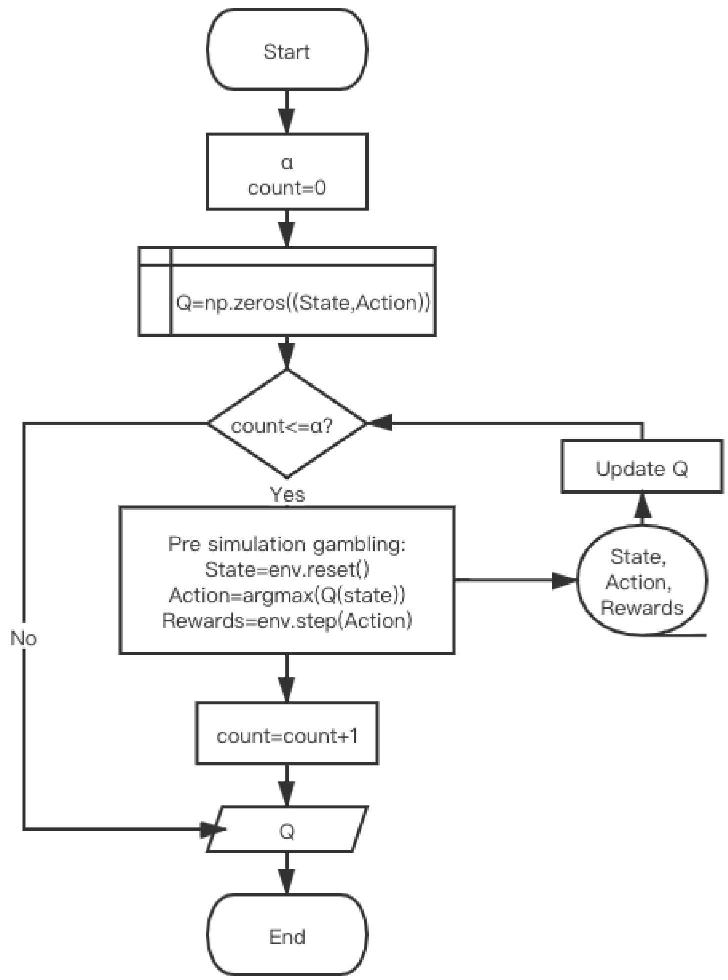

Figure 3: The flow chart for pre-simulation to calculate the matrix Q.

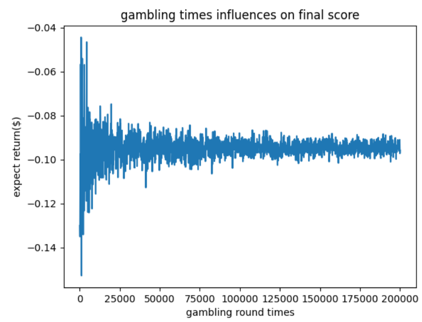

Figure 4: Q's expected return in different training times.

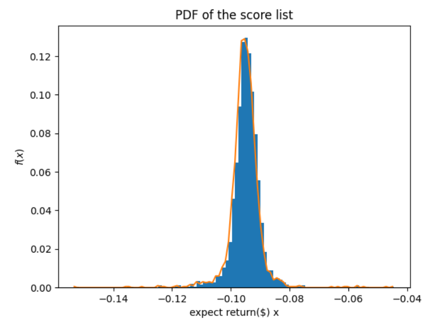

Figure 5: Mean=-0.095, variance=0.006

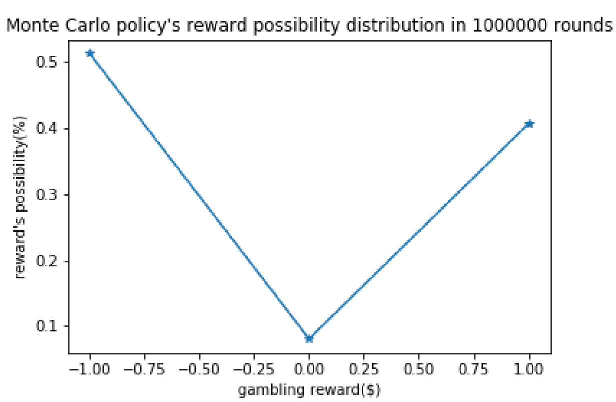

Figure 6: Using the Monte Carlo Strategy simulate for 10^6 times.

3 Long term simulation

After obtaining the win rate and expected return of these two strategy, this section discusses their performance in a long term simulation. Suppose 100 people has infinite money and bet on \$1 once a round. Figure 7 records their 2×10^5 gambling rounds return(\$) and the y-axis is these 100 gamers' average return(\$). From the Figure 7, the average return(\$) against the round times are similar to linear functions. Then Figure 8 reflects their average returns in each rounds and shows that the slopes of these two functions are approximately -0.04 and -0.01, which are coincide with the expected returns of \$-0.396 and \$-0.082 in section 2.1 and 2.2. [2]

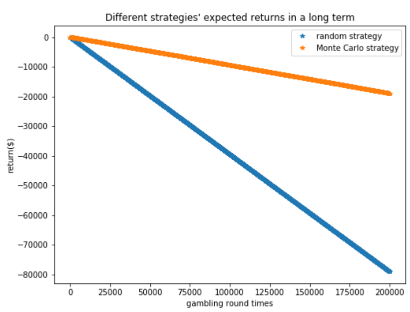

Figure 7: Expected return.

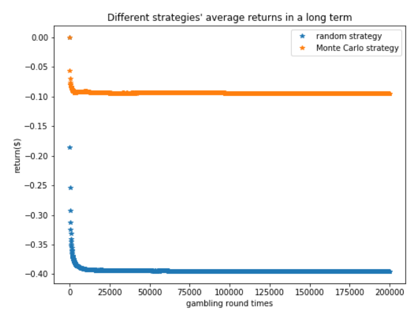

Figure 8: Average round return.

[2] Because of its better performance, the gamer in following simulations will use the Monte Carlo strategy.

4 Be a happy gamer!

4.1 Greedy rewards

This part simulates the influence of gamer's greedy in gambling games. Firstly, it gives 100 gamer $100 each as a startup. Then, it supposes each gamer will set a target profit and the gamer will not quit the gambling process as long as he has money and doesn't reach his target. In the pre-simulation, it has been found that when the gamer's greedy ratio (target profit ratio) bigger than 30%, the gamer's average expect return comes to 0 with no growth trend. So in the following simulation, the greedy ratio has been set from 0% to 30%, each gaps 1%. In Figure 9, the result shows that the higher greedy ratio will bring the lower expected returns and this return becomes to 0 when the greedy ratio set larger than 30%.

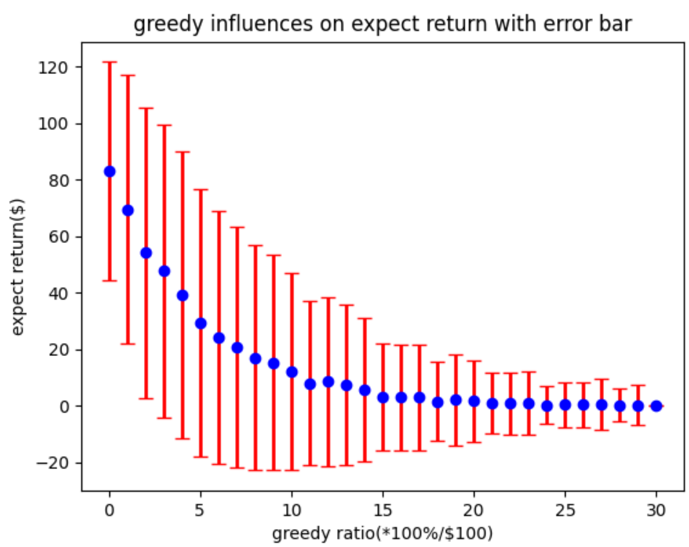

Figure 9: An error bar plot, while red bar represents the variance.

4.2 Risk tolerance

Some gamer want to gain more money, while the other gamer want to play more with the same \$100 as a startup. Hence, here use risk tolerance ratio to define the maximum money can be lose is tolerated by the gamer. The Figure 10 shows a trend that the higher risk tolerance can bring more game round times. By the way, as the conclusion from section 3, gamer approximately will lose \$0.1 each round. The average game time points can be connected in a straight line with the slope of $\frac{1}{0.1} = 10$, which meets the statistic guess.

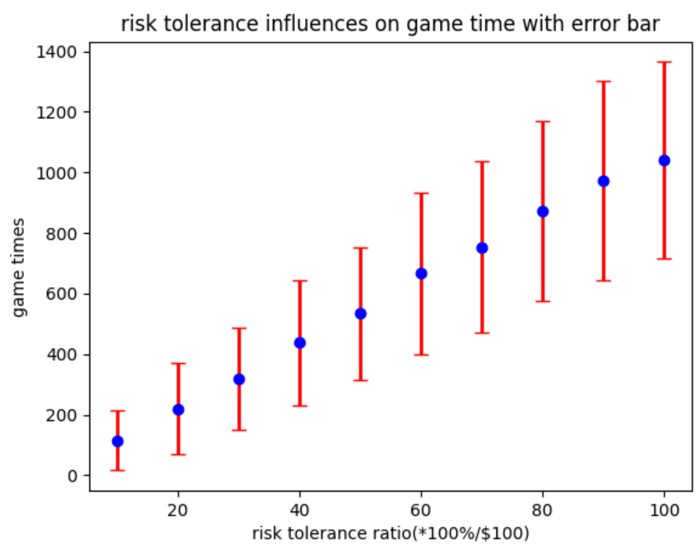

Figure 10: An error bar plot, while red bar represents the variance.

4.3 Play long, gain more

Gamer walking into the casino usually doesn't have the simple limit of greedy ratio or risk tolerance. They sometimes want to gain some money, but don't want to lose too much in a bad luck. Suppose 30×10 kinds of situation that gamer will end the gambling when they arrive their target profit or lose the maximum they can bear. The corresponding expected returns and average round times are shown in Figure 11 and Figure 12 among 100 people with the same \$100 startup.

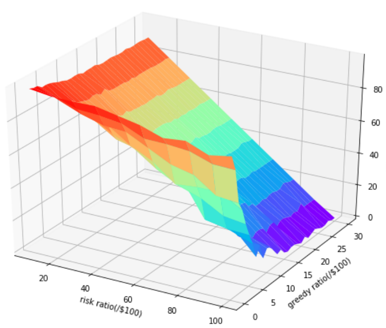

Figure 11: average expect return

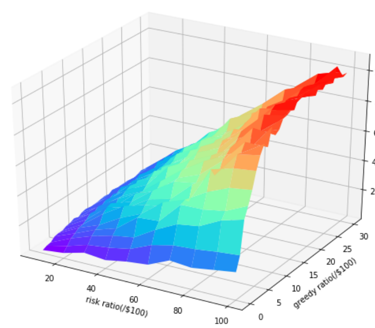

Figure 12: average round times

Than I rank the average expect return and gambling round time in descending order, for I suppose most of the people wants to play more and gain more. According to the rank list, the best performance psychological expectation is to take 20% greedy ratio and 40% risk tolerance ratio.

More than these, if we keep the expected return($) and the average round times as two random variables and the greedy ratio and risk tolerance ratio as two absorption walls(any variable will stop moving as soon as touch the wall), the interval between the walls makes different effects on the return variable and the time variable. As for the expected return variable, wider interval reduce both the variance and the mean value. As for the round time variable, wider interval will rise the variance and the mean value. It's easy to think that, as for the single expected return of Monte Carlo strategy in Blackjack, wider interval will give it more movement steps and higher possibility to bankrupt, vice versa.