**Question 1 – Blackjack**

1. The problem described as a Markov decision process –

Each state consists of two variables where represents the sum of the player’s cards and represents the first dealer’s cards.

The player is initially dealt with two cards therefor the minimal value of is (two s). The maximal sum the player can reach without losing the game is . The dealer’s first card is between and .

Overall the state space is-

In addition, there are three terminal states - therefore - .

The action space is .

The reward for each state is 0 except for the terminal states where –

We do not discount ().

Since the dealer’s policy is given (and independent on the players sum), we can calculate the probability of the final sum of cards for the dealer- marked .

Where B is the value of the card added (generated every time ) and is distributed (independent of ) -

Both and can be initially calculated. The problem’s transition probabilities based on these distributions will be-

All other transitions are with zero probability.

1. As described in section a, initially we shall calculate the probabilities of the total sum reached for the dealer - .

The possible values of are (based on the given policy) –

Where is the case where .

The result after applying the given policy -

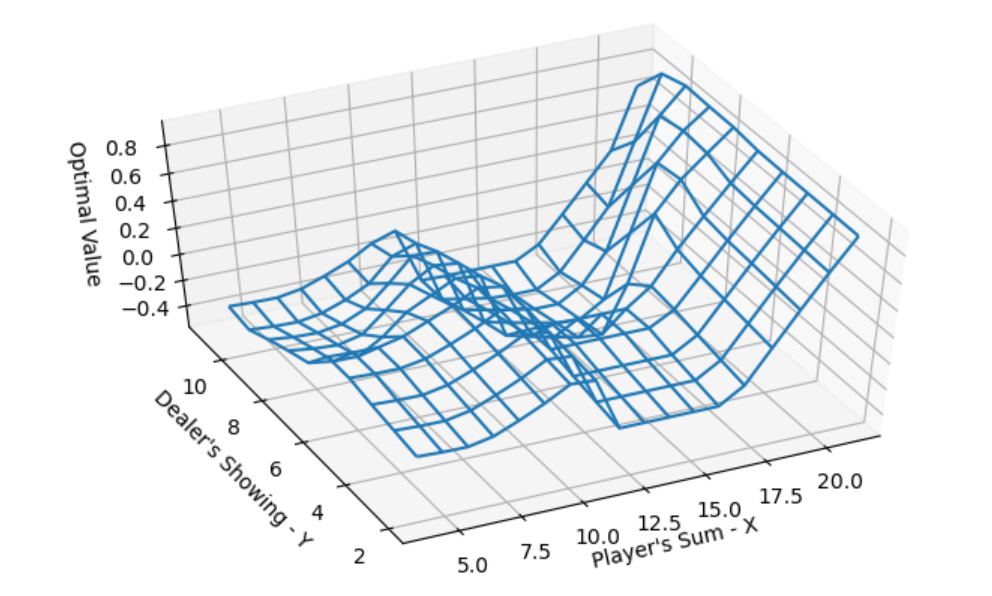
Now we have all the needed possibilities distributions in order to solve the Bellman Equations –

For terminal states –

For non- terminal states -

After running the VI algorithm for iterations we converge to an optimal value .

For example, after iterations –



1. After obtaining the optimal value - we must now derive an optimal policy. By using the Greedy operator-

We receive the following result (plot of the minimal value of X for which the policy choses action ‘stick’):

