Monte Carlo Methods

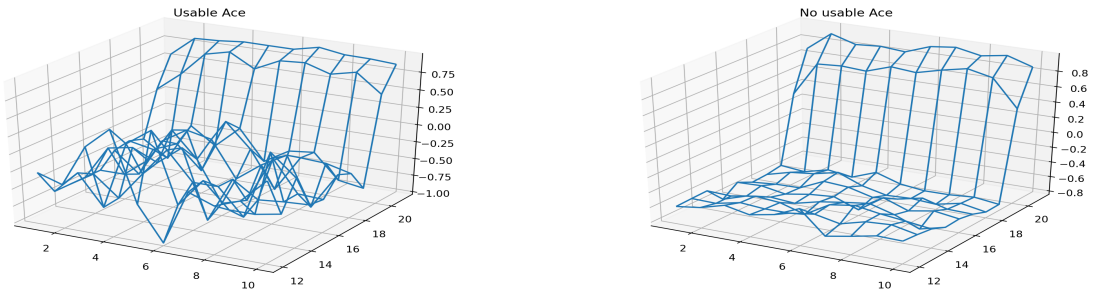
Exercise 1

(a) The most important difference is that we don't need to know complete probability distributions of all possible transitions. Also, and following from the first one, we can learn from experience which is much easier to be obtained in real world scenarios than a complete knowledge of the environment. We naturally pay more attention on more important transitions.

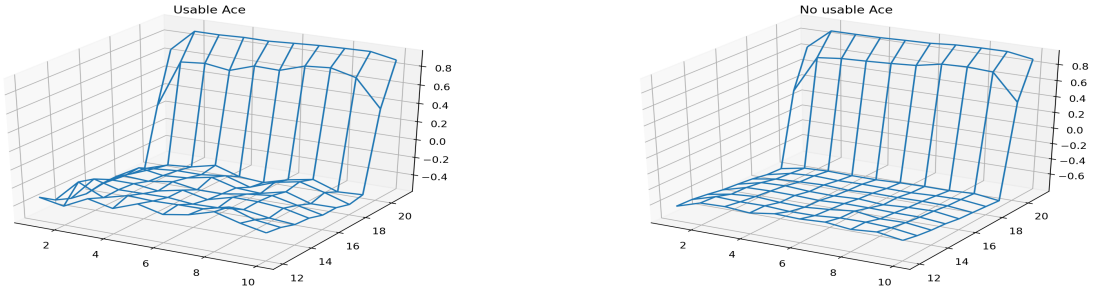
(b) For problems with a lot of possible states, like a game of chess, it is not feasible to maintain a complete dynamic of the environment. That's why a MC method based on actual or simulated games is more realistic and still converging to an optimal policy.

Exercise 2

(a)



(a) after 10000 iterations



(b) after 500000 iterations

Figure 1: Approximate state-value functions for the blackjack policy that sticks only on 20 or 21, computed by Monte Carlo policy evaluation

(b)

```
After 500000 iterations and with gamma=1:
No usable Ace:
[[1 0 1 1 0 0 1 1 1 1]
 [1 1 0 1 0 0 1 1 1 1]
 [1 0 0 0 0 0 1 1 1 1]
 [1 0 0 0 0 0 1 1 0 1]
 [1 0 0 0 0 0 1 1 1 0]
 [0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0]]
Usable ace:
[[1 1 1 1 1 1 1 1 1 1]
 [1 1 1 1 1 1 1 1 1 1]
 [1 1 1 1 1 1 1 1 1 1]
 [1 1 0 1 1 1 1 1 1 1]
 [1 0 1 1 1 1 1 1 1 1]
 [1 1 1 1 1 1 1 1 1 1]
 [1 0 0 0 0 0 0 0 1 1]
 [0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0]]
```

(Rows are players sum, columns are dealer card)