Timothee Guerin
260447866

Comp 424, Assignment 3

Question 1

Question 1.a

We have $p = 0.7$

$$E(X) = p * v_1 + (1 - p) * v_2$$

$$E(X) = 0.7 * (2000 - 1500) + 0.3 * (2000 - 1500 + 700)$$

$$E(X) = 290$$

Question 1.b

Let G be the event where the car is in good shape and let M be the event where he tell it's in good shape. Then we have

- $P(M|G) = 0.8$
- $P(M|\bar{G}) = 0.25$
- $P(G) = 0.7$

Then using baye's rules we have:

$$P(G|M) = \frac{P(M|G)P(G)}{P(M|G)P(G) + P(M|\ \bar{G})P(\bar{G})}$$

$$P(G|M) = \frac{0.8 * 0.7}{0.8 * 0.7 + 0.25 * 0.3}$$

$$P(G|M) = 0.882$$

$$P(\bar{G}|M) = 1 - 0.882 = 0.118$$

$$P(G|\ \bar{M}) = \frac{P(\bar{M}|G)P(G)}{P(\bar{M}|G)P(G) + P(\bar{M}|\ \bar{G})P(\bar{G})}$$

$$P(G|\ \bar{M}) = \frac{0.2 * 0.7}{0.2 * 0.7 + 0.75 * 0.3}$$

$$P(G|\ \bar{M}) = 0.384$$

$$P(\bar{G}|\ \bar{M}) = 1 - 0.384 = 0.616$$

Question 1.c

If we buy the car no matter what the mechanics says then he is of no use and then we don't need to pay him anything.

Timothee Guerin
260447866

Question 1.d

Question 2

Question 2.a

Having optimistic values give an advantage for short run however in long run realistic values will be more accurate and get to the optimal solution. Then if we are running in a long run the optimistic value will not necessarily give the optimal strategy.

Question 2.b

- ϵ-greedy: Will start choosing element randomly between the two arms. But as the number of play increase the mean of the first arm is going to converge to 0 and the mean of the second hand will converge to 0.01. Then if ϵ is not too small at the beginning both arms will be try and it will go for arm 2 in most cases.
- Ucb: This give more chance to less chosen action to be selected. This can allow to make a better choice at the beginning especially where the reward difference are quite small(such as here). However for a long run the ϵ −greedy is more effective.

Question 3

Question 3.a

- At $t = 1$:
  - $P_1(0) = \frac{1}{2}$
  - $P_1(1) = \frac{1}{4}$
  - $P_1(2) = \frac{1}{4}$
- $At\ t = 2$:
  - $P_2(0) = \frac{1}{2}P_1(0) = \frac{1}{4}$
  - $P_2(1) = \frac{1}{2}P_1(1) + \frac{1}{4}P_1(0) + \frac{1}{2}P_1(2) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$
  - $P_2(2) = \frac{1}{2}P_1(2) + \frac{1}{4}P_1(0) + \frac{1}{2}P_1(1) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$

Question 3.b

As we can only go to state 0 from state 0, we need to stay at state 0. Then

$$P_5(0) = \frac{1^5}{2}$$

Question 3.c

We start at 0 then we have $\frac{1}{4}$ chance to go to 1 in one step

If we went to 2 then we have $\frac{1}{2} * \frac{1}{4}$ to do it in 2 step.

If we stayed at 0 then we have one of the probability above time ¼ of doing it in +1 step and 1/4+2

Timothee Guerin
260447866

$$E(1) = 1 * \frac{1}{4} + 2 * \frac{1}{2}\left(\frac{1}{2} * \frac{1}{4} + \frac{1}{4} * \frac{1}{2}\right) + 3 * \frac{1^2}{2}\left(\frac{1^2}{2} * \frac{1}{4} + \frac{1}{4} * \frac{1^2}{2} + \frac{1}{4} * \frac{1}{2}\right) + 4 * \frac{1^3}{2}\left(\frac{1^3}{2} * \frac{1}{4} + \frac{1}{4} * \frac{1^3}{2} + \frac{1}{4} * \frac{1^2}{2} + \frac{1}{4} * \frac{1}{2}\right)$$

$$E(1) = 1 * \frac{1}{4} + 2 * \left(\frac{1}{8}\right) + 3 * \left(\frac{1}{16}\right) + 4 * \left(\frac{1}{32}\right) + \cdots + n\left(\frac{1}{2^{n+1}}\right)$$

Question 3.d

If we run it for infinite t then we will have $P_\infty(0) \to 0$ and then

$$P_\infty(1) \to \frac{1}{2}$$

$$P_\infty(2) \to \frac{1}{2}$$

Question 4

Question 4.a

As we giving 0 rewards to the wrong room and 1 reward to the good room then whatever the path taken the reward will be the same. We need to give some punishment for going the wrong way. To correct that, we could add a small punishment if it's not the destination. Then to maximize the reward we will need to do the smallest path as possible.

Question 4.b

Adding a constant C will break the algorithm as we are going to give reward for some action that needed to be punished then the optimal path is not likely to be found and the algorithm might run indefinitely.

Question 4.c

Adding $\gamma < 1$ make sure the result is finite then as we added a C we are likely to find the wrong solution

Question 5

- States: Position and energy of the rover
- Action: Move, Rough test, pickup
- Reward:
  - Rough test: small positive value
  - Pickup: Normal positive value
  - Move: Small Negative value
- γ can express the amount of energy remaining. As energy get smaller the rover has more chance of stopping

Timothee Guerin
260447866

Question 6

Question 6.a

$$Q^{\pi}(s,a) = E_{\pi}\{r_{t+1} + \gamma r_{t+1} + \cdots | s_t = s, a_t = a\}$$

$$Q^{\pi}(s,a) = r_a(s) + \gamma E_{\pi}\{R_{t+1} | s_{t+1} = s', a_{t+1} = a'\}$$

$$Q^{\pi}(s,a) = r_a(s) + \gamma \sum_{s'} T_a(s,s')\, Q^{\pi}(s',a)$$

then

$$Q^{\pi}(s,a) = r_a(s) + \gamma \sum_{s'} T_a(s,s') \sum_{a'} \pi(s',a') Q^{\pi}(s',a')$$

Question 6.b

Let's do a while where we set the current value of $Q_k(s,a)$ with this formula

$$Q_{k+1}(s,a) \leftarrow r_a(s) + \gamma \sum_{s'} T_a(s,s') \sum_{a'} \pi(s',a') Q_k(s',a')$$

If $|Q_{k+1}(s,a) - Q_k(s,a)| < \epsilon$ for some ϵ small then we stop.