Homework 3

Deadline: June 7th, 2019

Instruction: You may discuss these problems with classmates, but please complete the write-ups individually. Your answers must be **typewritten**, except for figures, which may be hand-drawn. Please submit your answers (pdf format only for non-programming assignments) on **Canvas**.

P1. Programming [35 points]

Undergrads: Implement the GUARD algorithm to address the attacker observational uncertainty.

Grads: Implement the COBRA-C algorithm which addresses both bounded rationality and observational uncertainty of the attacker.

Instruction: The **input** of your program is two .csv files: *param.csv* and *payoff.csv*. The *param.csv* file has four numbers *#targets,#resources,*α*,*ϵ in which *#targets* is the number of targets, *resources* is the number of the defender's resources, α is the parameter of the observational uncertainty, and ϵ is the parameter of the attacker's bounded rationality. For undergrads, you can skip the ϵ parameter for your program.
Output: The output of your program is a CSV file, named *solution.csv*. Each line of the output file is in the format of *target id,defender coverage probability*. A sample of the three files are provided.

Your **submission** must include: (i) source codes; (ii) documentary including description of your program and instruction to run it. Your program will be tested based on different games.

P2. Problem Solving

Q1. Auctions

Part 1. (25 points) Suppose a seller runs a second-price, sealed-bid auction for a painting. There are two bidders with independent, private values. The seller does not know their precise valuations, but knows: (a) each bidder i has one of three values, $v_i = 2$, $v_i = 4$ or $v_i = 8$; and (b) each of these values is equally likely (i.e., occurs with probability $\frac{1}{3}$). When running the auction, if the two bids are tied (say, at x), the winner is chosen at random (and pays x).

1. Assume both bidders use their dominant strategies for bidding in a second-price auction. What is the seller's expected revenue in this auction? Please explain your answer.

2. Now the seller decides to set a reserve price of r. This means that if the highest bid is below r, the seller will not sell the item. If the highest bid is at least r, then the painting will go to the highest bidder, and the winner will pay the maximum of r and the second-highest bid. Suppose the reserve price is set to $r = 4$. Assume both bidders use their dominant strategies. What is the seller's expected revenue in this auction? Please explain your answer. If the expected revenue increases or decreases relative to your answer in part (1), give a qualitative explanation for why this change occurs.

3. Is there a better reserve price than $r = 4$ (i.e., that will provide more revenue for the seller)? Give a brief justification for your response.

Part 2. (10 points) Consider an IPV setting where three bidders are risk-neutral and each bidder's valuation is drawn from a uniform distribution over the range $[a, b]$. Describe the optimal auction using virtual valuation and bidder-specific reserve price for this setting. What is the seller's expected revenue using this optimal auction mechanism, given bidders play the dominant strategy.

Part 3. [Graduates only] (10 points) Consider a first price sealed-bid auction with n risk-neutral agents whose valuations $v_1, v_2, \ldots, v_n$ are independently drawn from a uniform distribution on the interval $[a, b]$. Prove that $(\frac{n-1}{n}v_1, \frac{n-1}{n}v_2, \ldots, \frac{n-1}{n}v_n)$ is a Bayes-Nash equilibrium.

Q2. Reinforcement Learning [30 points]

Imagine an unknown game which has three states $\{A, B, C\}$ and in each state the agent has two actions to choose from $\{Up, Down\}$. Suppose a game agent chooses actions according to some policy π and generates the following sequence of actions and rewards in the unknown game:

t	s_t	a_t	s_{t+1}	r_t
0	A	Down	B	2
1	B	Down	C	3
2	C	Up	B	-2
3	B	Down	B	0
4	B	Up	A	1
5	A	Down	C	-3
6	C	Down	A	2
7	A	Up	C	1
8	C	Down	B	2
9	B	Down	A	2
10	A	Up	B	3

Unless specified otherwise, assume a discount factor $\gamma = 0.5$ and a learning rate $\alpha = 0.5$.

(a) (10 pts) Assume that all Q-values are initialized as 0. What are the Q-values learned by running Q-learning with the above experience sequence?

(b) (8 pts) In model-based reinforcement learning, we first estimate the transition function $T(a, s, a')$ and the reward function $R(s, a, s')$. Write down the estimates of T and R, estimated from the experience above. Write "n/a" if not applicable or undefined.

(c) (12 pts) Assume we had *a different experience* and ended up with the following estimates of the transition and reward functions:

s	a	s'	$\hat{T}(s,a,s')$	$\hat{R}(s,a,s')$
A	Up	A	1	12
A	Down	B	0.5	2
A	Down	C	0.5	-3
B	Up	B	1	8
B	Down	C	1	-6
C	Down	C	1	12
C	Up	C	0.5	2
C	Up	B	0.5	-2

(i) Give the optimal policy $\hat{\pi}^*(s)$ and $\hat{V}^*(s)$ for the MDP with transition function $\hat{T}$ and reward function $\hat{R}$. Explain your answers.
Hint: for any $x \in \mathbb{R}$, $|x|$, *we have* $1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}$.

(ii) If we repeatedly feed this new experience sequence through our Q-learning algorithm, what values will it converge to? Assume that the convergence is guaranteed.