Homework 3: Solution

Deadline: June 7th, 2019

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.

Answer.

1. The probability the second-highest bid is 2 is $1 - \frac{4}{9} = \frac{5}{9}$. The probability the second-highest bid is 8 is $\frac{1}{9}$. Finally, the probability the second-highest bid is 4 is $1 - \frac{5}{9} - \frac{1}{9} = \frac{1}{3}$. Therefore, the expected revenue is :

$$\frac{5}{9} \times 2 + \frac{1}{9} \times 8 + \frac{1}{3} \times 4 = \frac{10}{3}$$

2. Let's consider a general reserve price r. If $r \leq 2$, then expected revenue remains $\frac{10}{3}$. If $r > 8$, the expected revenue is 0. Now, if $2 < r \leq 4$, the expected revenue is:

$$\frac{4}{9} \times r + 8 \times \frac{1}{9} + 4 \times \frac{1}{3} = \frac{4r + 20}{9}$$

Finally, if $4 < r \leq 8$, the expected revenue is:

$$\frac{4}{9} \times r + \frac{1}{9} \times 8 = \frac{4r+8}{9}.$$

Therefore, if $r = 4$, then the expected revenue is 4 which is higher than the expected utility in part (1). In fact, by setting $r = 4$, in the case the second-highest bid is 2, the seller will obtain a higher expected revenue because the winner has to pay the maximum of r and the second highest-bid (which is $4 > 2$).

3. Yes. According to the analysis provided for part (2), $r = 8$ leads to the highest expected revenue for the seller which is equal to $\frac{40}{9} > 4$.

Part 2. (10 points) Consider an IPV setting where two 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.

Answer. (Let's assume $a \leq \frac{b}{2}$) Note that regarding the uniform distribution over the range $[a, b]$, the probability density function $f(x) = \frac{1}{b-a}$ and cumulative distribution function is $F(x) = \frac{x-a}{b-a}$. For each bidder,

- Virtual valuation function: $\psi(v_i) = v_1 - \frac{1-F(v_i)}{f(v_i)} = 2v_i - b$
- Bidder-specific reserve price: $r_i^* = r^* = \frac{b}{2}$

We divide the problem into three cases:

- $v_1, v_2 \leq r^*$: the good is not sold and the seller revenue is 0. This case happens with a probability of $\left(\frac{r^*-a}{b-a}\right)^2 = \left(\frac{b-2a}{2(b-a)}\right)^2$.
- $v_2 \leq r^* < v_1$ or $v_1 \leq r^* < v_2$. Let's assume $v_2 \leq r^* < v_1$. The winner is agent 1 which will be charged with $\min v_1^*$ where $\psi(v_1^*) \geq 0$ and $\psi(v_1^*) \geq \psi(v_2)$ which implies $v_1^* \geq r^*$. As a result, $\min v_1^* = r^*$. Finally, the expected revenue of the seller is computed as:

$$\int_a^{r^*} \left(\int_{r^*}^{b} r^* \frac{1}{b-a} dv_1 \right) \frac{1}{b-a} dv_2 = \frac{1}{(b-a)^2} \int_{r^*}^{b} r^*(b-r^*) dv_2$$
$$= \frac{1}{(b-a)^2} r^*(b-r^*)^2 = \frac{1}{(b-a)^2} \frac{b^3}{8}.$$

The result is similar to $v_1 \leq r^* < v_2$. Thus, the expected revenue in this second case is $\frac{1}{(b-a)^2} \frac{b^3}{4}$

- $v_1, v_2 > r^*$: the good is sold to the bidder with the highest virtual valuation function. Assume $v_1 > v_2$, then $\psi(v_1) > \psi(v_2)$ which means agent 1 wins. Agent 1 is charged with $\min v_1^*$ where

$\psi(v_1^*) \geq 0$ and $\psi(v_1^*) \geq \psi(v_2)$ which implies $v_1^* \geq v_2$. As a result $\min v_1^* = v_2$. Finally, the expected revenue of the seller is computed as:

$$\int_{r^*}^{b} \left(\int_{v_2}^{b} v_2 \frac{1}{b-a} dv_1 \right) \frac{1}{b-a} dv_2 = \frac{1}{(b-a)^2} \int_{r^*}^{b} v_2(b-v_2) dv_2$$
$$= \frac{1}{(b-a)^2} \frac{b^3}{12}$$

The result is similar to $v_1 < v_2$. Thus, the expected revenue in this second case is $\frac{1}{(b-a)^2}\frac{b^3}{6}$.

By combining all the three cases together, the expected revenue is:

$$0 + \frac{1}{(b-a)^2}\frac{b^3}{4} + \frac{1}{(b-a)^2}\frac{b^3}{6} = \frac{5b^3}{12(b-a)^2}$$

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 $[0, b]$. 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.

Answer. Let's suppose that all agents $i \neq 1$ bids $\frac{n-1}{n}v_i$. We are going to prove that the best response of agent 1 is to bid $\frac{n-1}{n}v_1$. Denote by s_1 the bid of agent 1. Note that agent 1 only wins if he bids the highest value, which means $s_1 \geq \frac{n-1}{n}v_i \iff v_i \leq \frac{n}{n-1}s_1$, for all agents $i \neq 1$. In this case, agent 1 obtains a utility of $v_1 - s_1$. Otherwise, agent 1 will lose and receive a utility of zero. Therefore, agent 1's expected utility can be computed as follows (assuming $\frac{n}{n-1}s_1 \leq b$):

$$E[u_1] \propto \int_0^{\frac{n}{n-1}s_1} \int_0^{\frac{n}{n-1}s_1} \cdots \int_0^{\frac{n}{n-1}s_1} (v_1 - s_1) dv_2 dv_3 \ldots dv_n$$
$$= (v_1 - s_1) \left(\frac{n}{n-1} s_1 \right)^{n-1}$$

$E[u_1]$ is maximized when:

$$\frac{\partial E[u_1]}{\partial s_1} = 0$$
$$\implies -\left(\frac{n}{n-1} s_1 \right)^{n-1} + (v_1 - s_1)(n-1)\frac{n}{n-1}\left(\frac{n}{n-1} s_1 \right)^{n-2} = 0$$
$$\implies s_1 = \frac{n-1}{n} v_1$$

Note: In fact, for a complete proof, you have to analyze the case when $\frac{n}{n-1}s_1 \geq b$ as well. For the grading, you will get a full grade for providing only part of the proof as above.

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

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:

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?

Answer. Note that Q-learning: $Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha\left(r_t + \gamma \max_{a'} Q(s_{t+1}, a')\right)$

$$Q(A, Down) = (1-0.5) \times 0 + 0.5 \times (2 + 0.5 \times \max_{a' \in \{Down, Up\}} Q(B, a')) = 1.0$$

$$Q(B, Down) = (1-0.5) \times 0 + 0.5 \times (3 + 0.5 \times \max_{a' \in \{Down, Up\}} Q(C, a')) = 1.5$$

$$Q(C, Up) = (1-0.5) \times 0 + 0.5 \times (-2 + 0.5 \times \max_{a' \in \{Down, Up\}} Q(B, a')) = -0.625$$

$$Q(B, Down) = (1-0.5) \times 1.5 + 0.5 \times (0 + 0.5 \times \max_{a' \in \{Down, Up\}} Q(B, a')) = 1.125$$

and so on.

(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.

Answer. The estimates of transition probabilities:
$\hat{T}(A, Down, B) = \hat{T}(A, Down, C) = \frac{1}{2}$, $\hat{T}(A, Up, A) = 0$
$\hat{T}(A, Up, B) = \hat{T}(A, Up, C) = \frac{1}{2}, \hat{T}(A, Up, A) = 0$
$\hat{T}(B, Down, A) = \hat{T}(B, Down, C) = \hat{T}(B, Down, B) = \frac{1}{3}$
$\hat{T}(B, Up, A) = 1, \hat{T}(B, Up, B) = \hat{T}(B, Up, C) = 0$
$\hat{T}(C, Down, A) = \hat{T}(C, Down, B) = \frac{1}{2}, \hat{T}(C, Down, C) = 0.$

$\hat{R}(A, Down, B) = 2$, $\hat{R}(A, Down, C) = -3$, $\hat{R}(A, Down, A) = na$
$\hat{R}(A, Up, B) = 3$, $\hat{R}(A, Up, C) = 1$, $\hat{R}(A, Up, A) = na$
$\hat{R}(B, Down, A) = 2$, $\hat{R}(B, Down, C) = 3$, $\hat{R}(B, Down, B) = 0$
$\hat{R}(B, Up, A) = 1$, $\hat{R}(B, Up, C) = \hat{R}(B, Up, B) = na$
$\hat{R}(C, Down, A) = 2$, $\hat{R}(C, Down, B) = 2$, $\hat{R}(C, Down, C) = na$
$\hat{R}(C, Up, A) = na$, $\hat{R}(C, Up, B) = -2$, $\hat{R}(C, Up, C) = na$

(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 $x \in \mathbb{R}$, $|x| < 1$, we have $1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}$.

Answer. Follow the policy iteration approach:
First, initialize a policy: $\pi_0(B) = Up$, $\pi_0(C) = Down$, $\pi_0(A) = Up$

Second, perform policy evaluation:
$\hat{V}^{\pi_0}(B) = \hat{T}(B, \pi_0(B), B)[8 + 0.5 \times \hat{V}^{\pi_0}(B)] = 8 + 0.5\hat{V}^{\pi_0}(B) = \cdots = 8 \times (1 + 0.5 + 0.5^2 + \dots) = \frac{8}{1-0.5} = 16$
Similarly, $\hat{V}^{\pi_0}(C) = 24$, $\hat{V}^{\pi_0}(A) = 24$.

Third, for fixed values, update policy, we still obtain $\pi_1 \equiv \pi_0$. This means the policy converges and thus $\hat{\pi}^*(s) = \pi_0(s)$ and $\hat{V}^*(s) = \hat{V}^{\pi_0}(s)$ with $s \in \{A, B, C\}$.

(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.

Answer. The Q-learning algorithm will not converge to the optimal values V^* for the MDP because the experience sequence and transition frequencies replayed are not necessarily representative of the underlying MDP. However, for the MDP with transition function $\hat{T}$ and reward function $\hat{R}$, replaying this experience repeatedly will result in Q-learning converging to its optimal values $\hat{V}^*$ computed from (i).