Master MOSIG – Mathematical Foundations of Machine Learning – Dec 10, 2024.

Firstname : Possible Correction.
Lastname :

Please be **concise** and **precise**. All documents are allowed. Electronic devices are not allowed.
Duration : 30min.

We want to compute the optimal strategy of a gambler. Initially, the gambler starts with $S_0 \in \{1\dots,4\}$ euros. At each time step, he/she can bet up to S_t euros (bets are integer between 0 and S_t).

If the gambler bets A_t, then with probability $p \in (0,1)$, he/she wins $\cancel{2A_t}$ A_t and with probability $q = 1-p$, the gambler looses A_t:

$$S_{t+1} = \begin{cases} S_t + 2A_t & \text{with probability } p \\ S_t - A_t & \text{with probability } q = 1-p \end{cases}$$

1. We assume that the gambler wants to maximize the probability of reaching $S_t \geq 5$ euros within T bets. Explain why this can be modeled as a MDP and provide the state space, action space, reward and transition probabilities. Shall we use a finite-horizon or a discounted reward criterion?

We consider a MDP with state space $S = \{0, 1, 2, 3, 4, \geq 5\}$. The state represents the amount of money that we have. Actions allowed in state s are $\{0, 1 \dots, s\}$. The reward is 1 for the state "≥ 5" at time T.

The transitions probabilities are:

$P(s \mid s, 0) = 1$ (if we bid 0)

$\begin{cases} P(s+2a \mid s, a) = p \\ P(s-a \mid s, a) = 1-p \end{cases}$ if we bid $a > 0$

We consider a finite horizon problem.

2. Let $V_t(s)$ be the probability that the gambler ends with 5 or more euros given that he/she started with s euros after t bets. We set $V_T(5) = 1$ and $V_T(s) = 0$ for $s < 5$. Write below a set of equations that link the values of $V_t(s)$, $V_{t+1}(s')$, p and q (for $t < T$):

- $V_t(0) =$ 0
- $V_t(1) =$ $\max(V_t(1), p V_{t+1}(3) + 1-p\, V_{t+1}(0)) = p V_{t+1}(3)$
- $V_t(2) =$ $\max(V_{t+1}(2); p V_{t+1}(4) + (1-p) V_{t+1}(1), p V_{t+1}(\geq 5) + \cancel{(1-p) V_t(0)} = 0$
- $V_t(3) =$ $\max(V_{t+1}(3); p V_t(\geq 5) + (1-p) V_{t+1}(2); p V_{t+1}(\geq 5) + (1-p) V_{t+1}(1); p V_{t+1}(\geq 5))$
- $V_t(4) =$ $\max(V_{t+1}(4); \underbrace{p V_t(\geq 5) + (1-p) V_{t+1}(3)}_{\text{this is larger than this except maybe for "2"}}; p V_t(\geq 5) + (1-p) V_{t+1}(2) \dots)$
- $V_t(5) =$ 1

the max is never this one

this is larger than this except maybe for "2".

3. Assume that $T = 1$. Compute the optimal strategy and its expected performance.

We apply the above equation and we get:

$V_0(0) = V_0(1) = 0$; $V_0(2) = V_0(3) = V_0(4) = p$; $V_0(5) = 1$

The optimal strategy is to bid "2" in state 2, 3, 4 and 0 in state $\{0; 1; 5\}$. There are other optimal strategies.

4. Assume that $T = 4$ and $p = 1/2$. Complete the following table of $V_t(s)$ (justify below).

V	$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = T = 4$
$s = 0$	$V_0(0) = 0$	0	0	0	0
$s = 1$	3/8	3/8	1/4	0	0
$s = 2$	5/8	1/2	1/2	1/2	0
$s = 3$	3/4	3/4	3/4	1/2	0
$s = 4$	7/8	7/8	3/4	1/2	0
$s = 5$	1	1	1	1	1

$V_1(2) = \max\left(\frac{1}{2} \times 1, \frac{1}{2} \times \left(\frac{3}{4} + \frac{1}{4}\right)\right) = \frac{1}{2}$

$V_1(3) = \frac{1}{2}\left(1 + \frac{1}{2}\right)$; $V_1(4) = \frac{1}{2}\left(1 + \frac{3}{4}\right)$

$V_0(2) = \max\left(\frac{1}{2} ; \frac{1}{2}\left(\frac{7}{8} + \frac{3}{8}\right)\right) = \frac{5}{8}$

$V_0(3) = \frac{1}{2}\left(1 + \frac{1}{2}\right)$; $V_0(4) = \frac{1}{2}\left(1 + \frac{3}{4}\right) = 7/8$

5. Describe the optimal strategy (for $p = 1/2$).

The optimal strategy is to bid "1" in almost all cases except for $t = \{2, 3\}$ and $s = 2$ where it is optimal to bid "2".

For $t = 1$ and $s = 2$, actions bid = 1 and bid = 2 are both optimal.