# D3) Dynamic Programming

Jiin Sang

2021-01-15

## Contents

## Exercise 1

**How would you generalize this game with arbitrary value of $m_1$ (minimum increment), $m_2$ (maximum increment), and $N$ (the winning number)?**

| |
|---|
| A and B are to play a game. They take turn to call out integers.<br>1. The serving player must call out an integer between 1 or 2.<br>2. The opponent player 1) takes the other player's number and 2) increments it by 1 or 2, then 3) call out the number.<br>3. Keep playing back and forth until someone calling out the number 31. The person calling out 31 is winner. |

---

A and B are to play a game. They take turn to call out integers.

1. The serving player must call out an integer between 1 or 2.
2. The opponent player 1) takes the other player's number and 2) increments it by the range between $m_1$ and $m_2$ (inclusive), then 3) call out the number.
3. Keep playing back and forth until someone calling out the number $N$. The person calling out $N$ is winner.

## Exercise 2

**Two players are to play a game. The two players take turns to call out integers. The rules are as follows. Describe A's winning strategy.**

- A must call out an integer between 4 and 8, inclusive.
- B must call out a number by adding A's last number and an integer between 5 and 9, inclusive.
- A must call out a number by adding B's last number and an integer between 2 and 6, inclusive.
- Keep playing until the number larger than or equal to 100 is called by the winner of this game.

---

## Exercise 3

**There is only finite number of deterministic stationary policy. How many is it?**

---

$A$ is an action space, $S$ is a state space.

Since policy function $\Pi : S \to A$, number of deterministic stationary policy equals $|\Pi| = |A||S|$.

$\therefore |\Pi| = |A||S|$

## Exercise 4

**Formulate the first example in this lecture note using the terminology including state, action, reward, policy, transition. \ Describe the optimal policy using the terminology as well.**

| |
|---|
| A and B are to play a game. They take turn to call out integers.<br>1. The serving player must call out an integer between 1 or 2.<br>2. The opponent player 1) takes the other player's number and 2) increments it by 1 or 2, then 3) call out the number.<br>3. Keep playing back and forth until someone calling out the number 31. The person calling out 31 is winner. |

---

State space $S = \{1, 2, 3, ...30, 31\}$.

Action space $A = \{a_1, a_2\}$ ($a_1$ represents incrementing by 1, and $a_2$ represents incrementing by 2).

Reward function $R(s, a) = E[r_t | s_t = s, A_t = s]$, where $R(30, a_1) = R(29, a_2) = 1$, and $R(s, a) = 0$ for all other states.

The player who calls out 31, 28, 25, … ,4, 1 becomes the winner.

Optimal policy $ ^{*}(1) = $

## Exercise 5

From the first example,

- Assume that your opponent increments by 1 with prob. 0.5 and by 2 with prob. 0.5.
- Assume that the winning number is 10 instead of 31.
- Your opponent played first and she called out 1.
- Your current policy $\pi_0$ is that
  - If the current state $s \leq 5$ then increment by 2.
  - If the current state $s > 5$ then increment by 1.
- Evaluate $V^{\pi_0}(1)$.

---