

Assignment 7

Task 7.1

1. states:
 x_1 - door1 is reward door, door2 is tiger door,
 x_2 - door1 is tiger door, door2 is reward door;
2. actions:
 u_1 - open door1 and finish game, terminal action,
 u_2 - open door2 and finish game, terminal action,
 u_3 - listen, sensing action, gives measurement $\{z_1, z_2\}$ as result;
3. measurement space:
 z_1 - noise behind the door1,
 z_2 - noise behind the door2;
4. cost function (expected rewards):
 $r(b, u_1) = p_1 * r(x_1, u_1) + p_2 * r(x_2, u_1) = p_1 * (+200) + (1 - p_1) * (-1000),$
 $r(b, u_2) = p_1 * r(x_1, u_2) + p_2 * r(x_2, u_2) = p_1 * (-1000) + (1 - p_1) * (+200),$
 $r(b, u_3) = -50;$
5. associated probabilities:
in state x_1 person gets z_1 with probability 0.2 and z_2 with probability 0.8, thus with probability 0.2 he thinks, that tiger is behind the door1 and changes state to x_2 ,
in state x_2 person gets z_1 with probability 0.8 and z_2 with probability 0.2, thus with probability 0.2 he thinks, that tiger is behind the door2 and changes state to x_1 .

Whole scheme is depicted in the Figure 1.

Task 7.2

Cumulative reward (cost) of the sequence "Listen, listen, open door1" is:

$$\begin{aligned} R &= r(b, u_3) + r(b, u_3) + r(b, u_1) = \\ &= -50 - 50 + (+200) * p_1 + (-1000) * (1 - p_1) = -1100 - 800p_1 \end{aligned}$$

where p_1 is probability of being in x_1

Person choose action u_1 anyway, independently of the measurement from u_3 , thus we just sum up doubled cost of doing u_3 and expected reward after doing u_1 .

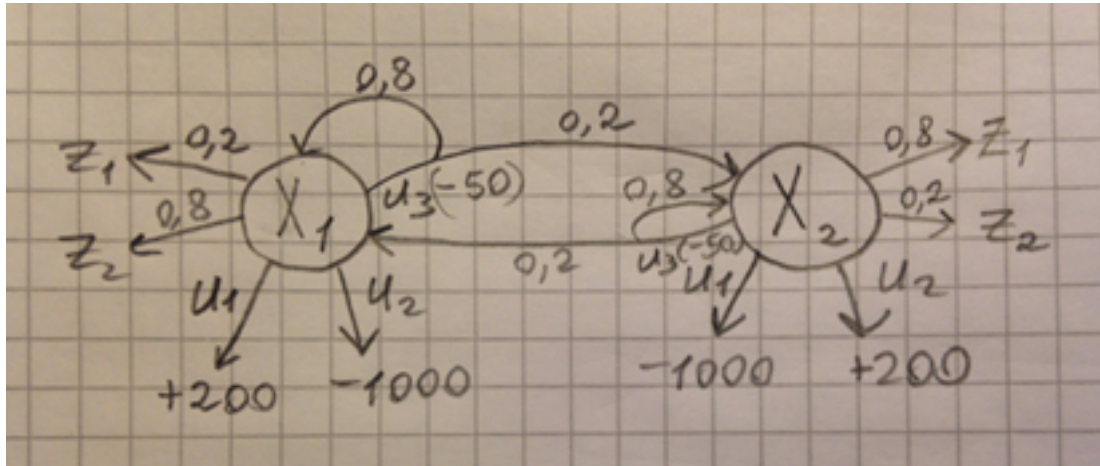


Figure 1: POMDP scheme

Task 7.3

Cumulative reward (cost) of the sequence "Listen, then open the door for which you did not hear a noise" is:

$$\begin{aligned}
 R &= r(b, u_3) + = \\
 &= -50 + 200 * 0.8 - 1000 * 0.2 = -90
 \end{aligned}$$

Person acts accordingly to the measurement after committing u_3 , thus he opens the door with best measurement with probability 0.8.

Task 7.4