

---

## AI Lab 2 : Agents and Environments

---

**Environment:** The environment is a village called “Binary”pur, with two categories of people: category 0 is *Kid* and category 1 is *Adult*.

**State:**

- At time  $t$ , the  $s_t \in \{0, 1\}$ , i.e., the state is  $s_t = 0$  or  $s_t = 1$ . Note that the state can assume only one of the values. Here 0 means *Kid* and 1 means *Adult*.
- State is generated with  $P(s_t = 0) = p_{\text{kid}}$ ,  $P(s_t = 1) = p_{\text{adult}} = 1 - p_{\text{kid}}$ .

**Observation:**

- $o_t = (h_t)$ , where  $h_t$  denotes height of a given person.
- The height a *Kid* is distributed between 2 to 4.5 feet as shown. The distribution is given by

Height	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	4.4
% of kids in that height	1	9	0.5	5	4.5	15	0.8	4.2	13	7	22	8	10

- Choose your own distribution for the height of an *Adult*.

**Action:** Agent observes  $o_t$  and needs to decide whether the person is *Kid* or an *Adult*. The action set  $a_t = \{0, 1\}$ , where 0 means *Kid* and 1 means *Adult*.

**Reward:** The reward  $r_t = R(s_t, a_t)$ ,  $R(0, 0) = 1$ ,  $R(1, 1) = 1$ ,  $R(0, 1) = 0$ ,  $R(1, 0) = 0$ , i.e., if the prediction is correct then reward is 1, else it is 0.

1. Produce a dataset file which contains  $t, s_t, h_t, a_t, r_t, t = 1, \dots, 1000$ . Use a new line for each  $t$ .
2. Plot histograms of the height of the *Kid* and *Adult*.
3. Measure the performance of the agent, i.e., average reward.

### 1 Dynamic Control Task: Room Cleaner Robot

Consider a robot which cleans the room which contains dirt.

**Environment:** The room is a grid with dimensions  $x_{\text{size}} \times y_{\text{size}}$ . It has walls on all sides and the robot if it tries to move out it will hit the wall and stay in the same place. 10 random locations contain dirt.

**State:** At time  $t$ , the agent is in location  $(x_t, y_t)$ .  $d_t$  is an array of size  $x_{\text{size}} \times y_{\text{size}}$ , and it contains the information on dirt.

**Observation:**  $o_t = (x_t, y_t)$ , i.e., the agent gets to observe its position. **It does not observe the dirt information.**

**Action:** Agent need to decide whether it has to move right, left, up, down or pick up the dirt. The action set  $a_t = \{\text{up,down,right,left, pick-dirt}\}$ . The agent picks one action at random.

**Reward:** The reward  $r_t = R(s_t, a_t)$ , reward is  $-1$  if the agent tries to pick-dirty in a clean grid,  $-10$  on hitting the wall and is equal to the amount of dirt when it picks the dirt.

1. Print out the activity at each time  $t = 1, \dots, 100$ , location of the agent, dirt in each location, action of the agent and the reward obtained.
2. Measure the performance of the agent, i.e., the average reward obtained.