Homework 3

Group 30

a) States

$X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$

Actions

$A = \{p, s\}$; $p =$ "play", $s =$ "stop"

Observations

$Z = \{g, w, t, e\}$

Transition probabilities:

$$P_p = \begin{array}{c} \begin{array}{cccccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{array} \\ \begin{bmatrix} 0.2 & 0.4 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0.4 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0.4 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0.4 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0.4 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0.4 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0.4 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0.4 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0 & 0.8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$P_s = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Observation probabilities

$$O_p = O_s = \begin{array}{c} \begin{array}{cccc} g & w & t & e \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Immediate cost function

$$C = \begin{array}{c} \begin{array}{cc} p & s \end{array} \\ \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

.b)

.a)

$$\begin{array}{ccccccccccccc} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ \alpha_0 = \mu_0 = & \Big[\frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \Big] \end{array}$$

$$\hat{\alpha}_1 = \alpha_0 \, P_p = \left[\frac{1}{12} \ \frac{1}{30} \ \frac{1}{15} \ \frac{1}{15} \ \frac{1}{15} \ \frac{1}{15} \ \frac{11}{60} \ \frac{1}{12} \ \frac{1}{30} \ \frac{1}{15} \ \frac{1}{15} \ \frac{1}{15} \ \frac{1}{15} \ \frac{11}{60} \right]$$

$$\alpha_1^T = \hat{\alpha}_1 \, \mathrm{diag}\left(O_p (W|\cdot) \right)$$

$$= \hat{\alpha}_1 \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$= \left[0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \frac{1}{15} \ 0 \ 0 \ 0 \right]$$

normalized $\Rightarrow = \left[0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \right]$

.b)

$$\begin{array}{ccccccccccccccc} & & & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ .b) & \alpha_0 = \mu_0 = & [& \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \;] \end{array}$$

$i \to w$

$p \to e$

$p \to e$

(b) After feeling the web with its legs and playing two turns, assuming that the spider made no observation after each step (i.e., it makes two empty observations).

$$\hat{\alpha}_1 = \alpha_0 P_p = \left[\tfrac{1}{12}\ \tfrac{1}{30}\ \tfrac{1}{15}\ \tfrac{1}{15}\ \tfrac{1}{15}\ \tfrac{1}{15}\ \tfrac{11}{60}\ \tfrac{1}{12}\ \tfrac{1}{30}\ \tfrac{1}{15}\ \tfrac{1}{15}\ \tfrac{1}{15}\ \tfrac{1}{15}\ \tfrac{11}{60}\right]$$

$$\alpha_1^T = \hat{\alpha}_1 \, diag(O_p(w|\cdot))$$

$$= \hat{\alpha}_1 \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$= \left[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{15}\ 0\ 0\ 0\right]$$

$$\hat{\alpha}_2 = \alpha_1 P_p$$

$$\alpha_2^T = \hat{\alpha}_2 \, diag(O_p(e|\cdot))$$

$$\hat{\alpha}_3 = \alpha_2 P_p$$

$$\alpha_3^T = \hat{\alpha}_3 \, diag(O_p(e|\cdot))$$

$$\hat{\alpha}_2 = \alpha_1 P_p$$

$$= \left[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{15}\ 0\ 0\ 0\right] \begin{bmatrix} .2&.4&.4&0&0&0&0&0&0&0&0&0 \\ .2&0&.4&.4&0&0&0&0&0&0&0&0 \\ .2&0&0&.4&.4&0&0&0&0&0&0&0 \\ .2&0&0&0&.4&.4&0&0&0&0&0&0 \\ .2&0&0&0&0&.8&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&.2&.4&.4&0&0&0 \\ 0&0&0&0&0&0&.2&0&.4&.4&0&0 \\ 0&0&0&0&0&0&0&0&.2&.4&.4&0 \\ 0&0&0&0&0&0&0&0&.2&0&.4&.4 \\ 0&0&0&0&0&0&0&0&.2&0&0&.8 \\ 0&0&0&0&0&0&0&0&0&0&0&1 \end{bmatrix}$$

$$= \left[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{75}\ \tfrac{2}{75}\ \tfrac{2}{75}\ 0\right]$$

$$\alpha_2^T = \hat{\alpha}_2 \, diag(O_p(e|\cdot))$$

$$= \left[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{1}{75}\ \tfrac{2}{75}\ \tfrac{2}{75}\ 0\right] \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$= \left[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \tfrac{2}{75}\ \tfrac{2}{75}\ 0\right]$$

$$\hat{\alpha}_3 = \alpha_2 P_p$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{75} & \frac{2}{75} & 0 \end{bmatrix} \begin{bmatrix} .2 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{4}{375} & 0 & \frac{4}{345} & \frac{4}{125} \end{bmatrix} \overset{\text{normalized}}{\Rightarrow} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{3}{5} \end{bmatrix}$$

$$\alpha_3^T = \hat{\alpha}_3 \, \mathrm{diag}(Op(e|.))$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5} & 0 \end{bmatrix}$$

.b)

.c) (c) After starting and playing 3 times, assuming that the spider made no observation after each step (i.e., it makes three empty observations).

$$\hat{\alpha}_1 = \alpha_0 P_p = [1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0] \begin{bmatrix} .2 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \left[\tfrac{1}{5}\ \tfrac{2}{5}\ \tfrac{2}{5}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right]$$

$$\alpha_1^T = \hat{\alpha}_1 \ \mathrm{diag}(O_p \mid e(.))$$

$$= \left[\tfrac{1}{5}\ \tfrac{2}{5}\ \tfrac{2}{5}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right] \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \left[0\ \tfrac{2}{5}\ \tfrac{2}{5}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right]$$

$$\hat{\alpha}_2 = \alpha_1 P_p$$

$$= \left[0\ \tfrac{2}{5}\ \tfrac{2}{5}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right] \begin{bmatrix} .2 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \left[\tfrac{4}{25}\ 0\ \tfrac{4}{25}\ \tfrac{8}{25}\ \tfrac{4}{25}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right]$$

normalized $\Rightarrow$ $= \left[\tfrac{1}{5}\ 0\ \tfrac{1}{5}\ \tfrac{2}{5}\ \tfrac{1}{5}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\right]$

$$\alpha_2^T = \hat{\alpha}_2 \, diag(O_p \mid e(.))$$

$$= \begin{bmatrix} \frac{1}{5} & 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{\alpha}_3 = \alpha_2 P_p$$

$$= \begin{bmatrix} 0 & 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} .2 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{4}{25} & 0 & 0 & \frac{2}{25} & \frac{6}{25} & \frac{8}{25} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

normalized

$$\Rightarrow = \begin{bmatrix} \frac{1}{5} & 0 & 0 & \frac{1}{10} & \frac{3}{10} & \frac{4}{10} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\alpha_3^T = \hat{\alpha}_3 \, diag(O_p \mid e(.))$$

$$= \begin{bmatrix} \frac{1}{5} & 0 & 0 & \frac{1}{10} & \frac{3}{10} & \frac{4}{10} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & \frac{1}{10} & \frac{3}{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \overset{\text{normalize}}{\Rightarrow} \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

c)

Belief: [0.2 0.08 0.24 0.32 0.16 0 0 0 0 0 0 0]

<u>MLS</u>

- For MLS we have to identify the most likely state according to the belief
- The highest probability is $b(3) = 0.32$ so state 3 is the most likely state
- To know the best action to take we have to calculate which action has the lowest cost between playing and stopping, since with this heuristic we will never do the "identify" action.

→ The cost-to-go for this policy would be:

$$J^{\pi}(3) = c_{\pi}(3) + \gamma \sum_{y \in \mathcal{X}} P_{\pi}(y \mid \underset{\substack{\downarrow \\ \text{state we currently are}}}{3}) \times J^{\pi}(y)$$

$$J^{\pi}(3) = 1 + 0.9\,(0.2\,J^{\pi}(0) + 0.4\,J^{\pi}(4) + 0.4\,J^{\pi}(5))$$

$J^{\pi}(5) = 0$, because according to the homework any action after reaching level 5 will have a cost of 0

$$J^{\pi}(4) = 1 + 0.9\,(0.2\,J^{\pi}(0) + 0.8\,\overbrace{J^{\pi}(5)}^{0}) =$$

$$= 1 + 0.9 \times 0.2\,J^{\pi}(0) = 1 + 0.18\,J^{\pi}(0)$$

$$J^{\pi}(3) = 1 + 0.9(0.2\,J^{\pi}(0) + 0.4\,J^{\pi}(4) + 0.4\,J^{\pi}(5)) =$$

$$= 1 + 0.9(0.2\,J^{\pi}(0) + 0.4(1 + 0.18\,J^{\pi}(0))) = 1.36 + 0.2448\,J^{\pi}(0)$$

$$J^{\pi}(2) = 1 + 0.9\,(0.2\,J^{\pi}(0) + 0.4\,(1.36 + 0.2448\,J^{\pi}(0)) + 0.4\,(1 + 0.18\,J^{\pi}(0))) =$$

$$= 1.8496 + 0.3329\,J^{\pi}(0)$$

$$J^{\pi}(1) = 1 + 0.9\,(0.2\,J^{\pi}(0) + 0.4\,J^{\pi}(2) + 0.4\,J^{\pi}(3)) = 2.1554 + 0.3880\,J^{\pi}(0)$$

$$J^{\pi}(0) = 1 + 0.9\,(0.2\,J^{\pi}(0) + 0.4\,J^{\pi}(1) + 0.4\,J^{\pi}(2)) = 2.4418 + 0.4396\,J^{\pi}(0)$$

$$J^{\pi}(0) = 2.4418 + 0.4396\,J^{\pi}(0) \Leftrightarrow J^{\pi}(0) = 4.3572$$

$$J^{\pi}(3) = 1.36 + 0.2448\,J^{\pi}(0) = 2.4266 \rightarrow \text{cost for policy play}$$

Now for the action "stop"

$$J^{\pi}_{stop}(3) = 0 + 0.9\,(1 \times J^{\pi}(0))$$

We will assume that in all other states the policy is optimal, and will do the "play" action, for state 0

$$J^{\pi}(3) = 0.9 \times 4.3572 = 3.921$$

→ According to ALS heuristic and with the given belief the best action to make is "play" since $\underbrace{J^{play}(3)}_{2.4266} < \underbrace{J^{stop}(3)}_{3.921}$

QMDP:

We have to calculate the action that minimizes $\sum_{x \in \mathcal{X}} b(x)\, Q^*_{MDP}(x,a)$

calculated the same way as above

$$Q^*(0, \text{Play}) = 1 + 0.9(0.2\,J^*(0) + 0.4\,J^*(1) + 0.4\,J^*(2)) =$$

$$= 1 + 0.9\,(0.2 \times 4.3572 + 0.4 \times 3.8460 + 0.4 \times 2.4953 = 4.0672$$

$$Q^*(0, \text{stop}) = 1 + 0.9 \times (1 \times J^*(0)) = 4.92148$$

for state 2 the optimal policy is choosing the stop action

$$Q^*(1, \text{Play}) = 1 + 0.9(0.2\,J^*(0) + 0.4\,J^*(2) + 0.4\,J^*(3)) =$$

$$= 1 + 0.9\,(0.2 \times 4.3572 + 0.4 \times 2.4953 + 0.4 \times 2.4266 = 3.5562$$

$$Q^*(1, \text{stop}) = 0 + 0.9 \times (1 \times J^*(0)) = 3.92148$$

$$Q^*(2, \text{Play}) = 1 + 0.9(0.2\,J^*(0) + 0.4\,J^*(3) + 0.4\,J^*(4)) =$$

$$= 1 + 0.9\,(0.2 \times 4.3572 + 0.4 \times 2.4266 + 0.4 \times 1.7843 = 3.3002$$

$$Q^*(2, \text{stop}) = 0 + 0.9 \times (1 \times J^*(8)) = 2.4953$$

$$Q^*(3, \text{Play}) = 1 + 0.9(0.2\,J^*(0) + 0.4\,J^*(4) + 0.4\,J^*(5)) =$$

$$= 1 + 0.9\,(0.2 \times 4.3572 + 0.4 \times 1.7843 + 0.4 \times 0 = 2.4266$$

$$Q^*(3, \text{stop}) = 0 + 0.9 \times (1 \times J^*(0)) = 3.92148$$

$$Q^*(4, \text{Play}) = 1 + 0.9(0.2\,J^*(0) + 0.8\,J^*(5))$$

$$= 1 + 0.9\,(0.2 \times 4.3572 + 0) = 1.7843$$

$$Q^*(4, \text{Stop}) = 0 + 0.9 \times (1 \times J^*(0)) = 3.92148$$

$$\sum_{x \in \mathcal{X}} b(x)\, Q^*_{MDP}(x, \text{Play}) = 0.2 \times 4.0672 + 0.08 \times 3.5562 + 0.24 \times 3.3002 + 0.32 \times 2.4266 + 0.16 \times 1.7843 = 2.9520$$

$$\sum_{x \in \mathcal{X}} b(x)\, Q^*_{MDP}(x, \text{Stop}) = 0.2 \times 4.9215 + 0.08 \times 3.9215 + 0.24 \times 2.4953 + 0.32 \times 3.9215 + 0.16 \times 3.9215 = 3.7792$$

- Since our objective is to minimize that function, the QMDP heuristic will choose the action "play" for the given belief