Laura Lützow

Problem 9.1:

a. First of all we extract the relevant information from the text. We can define three Boolean random variables: $ES = EnoughSleep$, $RE = RedEyes$ and $SC = SleepinClass$. For compactness we abbreviate the true event of the random variables with the lower case and the false one with the negated lower case (e.g. $ES = true$ becomes es and $ES = false$ becomes $\neg es$). We know that:
$P(es) = 0.7$, and thus $P(\neg es) = 0.3$;
$P(es_t|es_{t-1}) = 0.8$, and $P(es_t|\neg es_{t-1}) = 0.3$;
$P(re|es) = 0.2$, and $P(re|\neg es) = 0.7$;
$P(sc|es) = 0.1$, and $P(sc|\neg es) = 0.3$.
The hidden state of this problem is the random variable ES that has a direct influence on both the two random variables RE and SC (common cause). We can model this relation between the random variables with a Bayesian network for a generic time step t^* as in Fig.1. With such a model we can say that RE_{t^*} and SC_{t^*} are conditionally independent given ES_{t^*}.

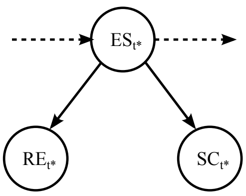


Figure 1: Bayesian network model for a generic time step t^*.

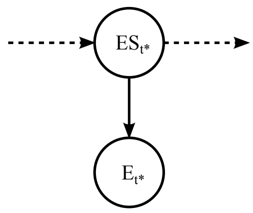
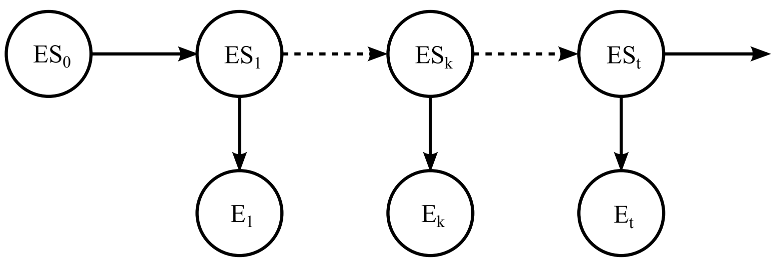


Figure 2: Bayesian network model for a generic time step t^* with only one observation variable.

We are asked to formulate an Hidden Markov Model, with only one observation variable. We thus have to manipulate the model in Fig.1 to obtain one as in Fig.2. In this case we have to introduce a new discrete random variable E with domain $\langle re \wedge sc,\ \neg re \wedge sc,\ re \wedge \neg sc,\ \neg re \wedge \neg sc \rangle$ (it covers all the combinations of events that RE and SC generate).
We can finally formulate the required Hidden Markov Model with the probability tables as in Fig.3.

$\mathbf{P}(ES_0 = true)$
0.7

ES_{k-1}	$\mathbf{P}(ES_k = true \mid ES_{k-1})$
t	0.8
f	0.3



ES_k	$\mathbf{P}(E_k = re \wedge sc \mid ES_k)$	$\mathbf{P}(E_k = \neg re \wedge sc \mid ES_k)$	$\mathbf{P}(E_k = re \wedge \neg sc \mid ES_k)$	$\mathbf{P}(E_k = \neg re \wedge \neg sc \mid ES_k)$
t	$P(re\|es)P(sc\|es) = 0.02$	$P(\neg re\|es)P(sc\|es) = 0.08$	$P(re\|es)P(\neg sc\|es) = 0.18$	$P(\neg re\|es)P(\neg sc\|es) = 0.72$
f	$P(re\|\neg es)P(sc\|\neg es) = 0.21$	$P(\neg re\|\neg es)P(sc\|\neg es) = 0.09$	$P(re\|\neg es)P(\neg sc\|\neg es) = 0.49$	$P(\neg re\|\neg es)P(\neg sc\|\neg es) = 0.21$

Figure 3: the Hidden Markov Model with its probability tables.

b. We have to perform state estimation (filtering), i.e. compute $\mathbf{P}(ES_t|\mathbf{e}_{1:t})$:

$$\mathbf{P}(ES_t|\mathbf{e}_{1:t}) = \mathbf{f}_{1:t} = \alpha \underbrace{\mathbf{P}(e_t|ES_t)}_{\mathbf{O}_t} \sum_{es_{k-1}} \underbrace{\mathbf{P}(ES_k|ES_{k-1})}_{\mathbf{T}} \underbrace{\mathbf{P}(ES_{t-1}|\mathbf{e}_{1:t-1})}_{\mathbf{f}_{1:t-1}} \text{ for each } t = 1,2,3.$$

In order to solve this problem we can use the filtering procedure in matrix notation. To do this we first have to define the matrices for state transition ($\mathbf{T}$) and observations ($\mathbf{O}_k$):

$$\mathbf{T} = \begin{bmatrix} P(es_k|es_{k-1}) & P(es_k|\neg es_{k-1}) \\ P(\neg es_k|es_{k-1}) & P(\neg es_k|\neg es_{k-1}) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix},$$

and for $E_1 = \neg re \wedge \neg sc$, $E_2 = re \wedge \neg sc$ and $E_3 = re \wedge sc$ we get respectively:

$$\mathbf{O}_1 = \begin{bmatrix} P(E_1 = \neg re \wedge \neg sc|es_1) & 0 \\ 0 & P(E_1 = \neg re \wedge \neg sc|\neg es_1) \end{bmatrix} = \begin{bmatrix} 0.72 & 0 \\ 0 & 0.21 \end{bmatrix},$$

$$\mathbf{O}_2 = \begin{bmatrix} P(E_2 = re \wedge \neg sc|es_2) & 0 \\ 0 & P(E_2 = re \wedge \neg sc|\neg es_2) \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0 & 0.49 \end{bmatrix},$$

$$\mathbf{O}_3 = \begin{bmatrix} P(E_3 = re \wedge sc|es_3) & 0 \\ 0 & P(E_3 = re \wedge sc|\neg es_3) \end{bmatrix} = \begin{bmatrix} 0.02 & 0 \\ 0 & 0.21 \end{bmatrix}.$$

The vector for the prior probability is:

$$\mathbf{f}_0 = \begin{bmatrix} P(es_0) \\ P(\neg es_0) \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix}.$$

Defining the vector $\mathbf{f}_{1:t} = \mathbf{P}(ES_t|\mathbf{e}_{1:t}) = \begin{bmatrix} P(es_t|\mathbf{e}_{1:t}) \\ P(\neg es_t|\mathbf{e}_{1:t}) \end{bmatrix}$, we can answer the question:

$$\mathbf{f}_{1:1} = \alpha \mathbf{O}_1 \mathbf{T} \mathbf{f}_0 = \alpha \begin{bmatrix} 0.4680 \\ 0.0735 \end{bmatrix} = \tfrac{1}{0.4680+0.0735} \begin{bmatrix} 0.4680 \\ 0.0735 \end{bmatrix} = \begin{bmatrix} 0.8643 \\ 0.1357 \end{bmatrix},$$

$$\mathbf{f}_{1:2} = \alpha \mathbf{O}_2 \mathbf{T} \mathbf{f}_{1:1} = \alpha \begin{bmatrix} 0.1318 \\ 0.1313 \end{bmatrix} = \tfrac{1}{0.1318+0.1313} \begin{bmatrix} 0.1318 \\ 0.1313 \end{bmatrix} = \begin{bmatrix} 0.5010 \\ 0.4990 \end{bmatrix},$$

$$\mathbf{f}_{1:3} = \alpha \mathbf{O}_3 \mathbf{T} \mathbf{f}_{1:2} = \alpha \begin{bmatrix} 0.0110 \\ 0.0944 \end{bmatrix} = \tfrac{1}{0.0110+0.0944} \begin{bmatrix} 0.0110 \\ 0.0944 \end{bmatrix} = \begin{bmatrix} 0.1045 \\ 0.8955 \end{bmatrix}.$$

c. We have to perform smoothing to calculate $\mathbf{P}(ES_k|\mathbf{e}_{1:3})$ for each $k = 1,2$ (we have already $\mathbf{P}(ES_3|\mathbf{e}_{1:3})$ in b. as $\mathbf{f}_{1:3}$).

Also in this case we use the procedure in matrix notation. Just to recall, in the lecture we have seen that: $\mathbf{P}(X_k|\mathbf{e}_{1:t}) = \alpha \underbrace{\mathbf{P}(X_k|\mathbf{e}_{1:k})}_{\mathbf{f}_{1:k}} \times \underbrace{\mathbf{P}(e_{k+1:t}|X_k)}_{\mathbf{b}_{k+1:t}}$ where X_k is the state at k with $0 \leq k < t$ and $\mathbf{e}_{1:t}$ is the evidence from 1 to t.

This procedure is composed of a backward recursion (to compute $\mathbf{b}_{k+1:t}$) and a forward one (to compute $\mathbf{f}_{1:k}$). We obtain the backward recursion with $\mathbf{b}_{k+1:t} = \mathbf{T}^T \mathbf{O}_{k+1} \mathbf{b}_{k+2:t}$, and we have already the result of the forward recursion (see b.).
In order to perform the backward recursion we always initialize with a 1-vector, e.g. $\mathbf{b}_{4:3} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. For additional explanation on this see **russel**.

<u>$k = 2:$</u>

$$\mathbf{b}_{3:3} = \mathbf{T}^T \mathbf{O}_3 \mathbf{b}_{4:3} = \begin{bmatrix} 0.0580 \\ 0.1530 \end{bmatrix},$$

<u>$k = 1:$</u>

$$\mathbf{b}_{2:3} = \mathbf{T}^T \mathbf{O}_2 \mathbf{b}_{3:3} = \begin{bmatrix} 0.0233 \\ 0.0556 \end{bmatrix}.$$

We can now calculate the required probabilities with $\mathbf{P}(ES_k|\mathbf{e}_{1:3}) = \alpha \mathbf{f}_{1:k} \times \mathbf{b}_{k+1:3}$ for $k = 2,1$.

<u>$k = 2:$</u>

$$\mathbf{P}(ES_2|\mathbf{e}_{1:3}) = \alpha\mathbf{f}_{1:2} \times \mathbf{b}_{3:3} = \alpha \begin{bmatrix} 0.5010 \\ 0.4990 \end{bmatrix} \times \begin{bmatrix} 0.0580 \\ 0.1530 \end{bmatrix} = \begin{bmatrix} 0.2757 \\ 0.7243 \end{bmatrix},$$

$\underline{k = 1:}$

$$\mathbf{P}(ES_1|\mathbf{e}_{1:3}) = \alpha\mathbf{f}_{1:1} \times \mathbf{b}_{2:3} = \alpha \begin{bmatrix} 0.8643 \\ 0.1357 \end{bmatrix} \times \begin{bmatrix} 0.0233 \\ 0.0556 \end{bmatrix} = \begin{bmatrix} 0.7277 \\ 0.2723 \end{bmatrix}.$$

d. In order to find the most likely sequence of states we use the Viterbi's algorithm.
Just to recall, we want to find a sequence of states $x_1 \dots x_t$ that maximizes $\mathbf{P}(x_1, \dots x_t, X_{t+1}|\mathbf{e}_{1:t+1})$. In the lecture we have seen that:
$\underset{x_1 \dots x_t}{max}\mathbf{P}(x_1, \dots x_t, X_{t+1}|\mathbf{e}_{1:t+1}) = \alpha\mathbf{P}(e_{t+1}|X_{t+1})\underset{x_t}{max}(\mathbf{P}(X_{t+1}|x_t)\underset{x_1 \dots x_{t-1}}{max} P(x_1, \dots x_t|\mathbf{e}_{1:t}))$.

In order to increase readability, we define the following vector:
$\mu_{t+1}(X_{t+1}) = \underset{x_1 \dots x_t}{max}\mathbf{P}(x_1, \dots x_t, X_{t+1}|\mathbf{e}_{1:t+1})$.

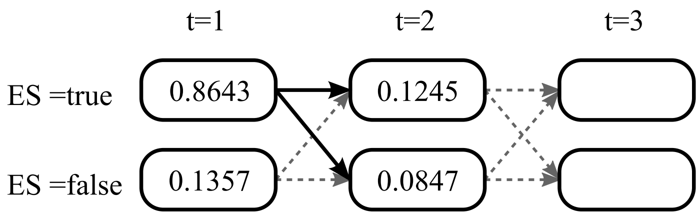
Let us implement the Viterbi's algorithm. We first initialize the procedure with:
$\mu_1(ES_1) = \mathbf{f}_{1:1} = \begin{bmatrix} 0.8643 \\ 0.1357 \end{bmatrix}$.
In the following we omit the optional normalization.

$\underline{Step 1:}$ we calculate $\mu_2(ES_2) = \mathbf{P}(E_2|ES_2)\underset{ES_1}{max}\mathbf{P}(ES_2|ES_1)\mu_1(ES_1)$.

– $ES_1 = true$: $\mathbf{P}(ES_2|es_1)\mu_1(es_1) = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} \cdot 0.8643 = \begin{bmatrix} \underline{0.6914} \\ \underline{0.1729} \end{bmatrix}$

– $ES_1 = false$: $\mathbf{P}(ES_2|\neg es_1)\mu_1(\neg es_1) = \begin{bmatrix} 0.3 \\ 0.7 \end{bmatrix} \cdot 0.1357 = \begin{bmatrix} 0.0407 \\ 0.0950 \end{bmatrix}$

$$\mu_2(ES_2) = \begin{bmatrix} 0.18 \\ 0.49 \end{bmatrix} \times \begin{bmatrix} \max(0.6914, 0.0407) \\ \max(0.1729, 0.0950) \end{bmatrix} = \begin{bmatrix} 0.1245 \\ 0.0847 \end{bmatrix}.$$

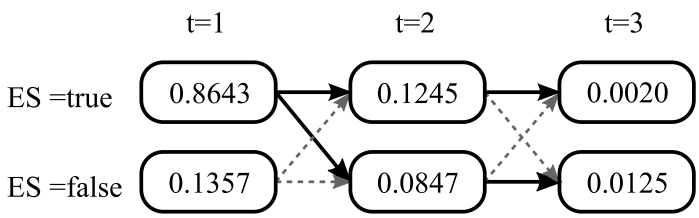


$\underline{Step 2:}$ we calculate $\mu_3(ES_3) = \mathbf{P}(E_3|ES_3)\underset{ES_2}{max}\mathbf{P}(ES_3|ES_2)\mu_2(ES_2)$.

– $ES_2 = true$: $\mathbf{P}(ES_3|es_2)\mu_2(es_2) = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} \cdot 0.1245 = \begin{bmatrix} \underline{0.0996} \\ 0.0249 \end{bmatrix}$

– $ES_2 = false$: $\mathbf{P}(ES_3|\neg es_2)\mu_2(\neg es_2) = \begin{bmatrix} 0.3 \\ 0.7 \end{bmatrix} \cdot 0.0847 = \begin{bmatrix} 0.0254 \\ \underline{0.0593} \end{bmatrix}$

$$\mu_3(ES_3) = \begin{bmatrix} 0.02 \\ 0.21 \end{bmatrix} \times \begin{bmatrix} \max(0.0996, 0.0254) \\ \max(0.0249, 0.0593) \end{bmatrix} = \begin{bmatrix} 0.0020 \\ 0.01245 \end{bmatrix}.$$



Following the bold arrows from the most likely state in μ_3, we get the most likely sequence: *true-false-false*.

e. This fixed value represents the steady state of the probability distribution of ES as the number of days increase and the professor always observes that the student sleeps in the class and has red eyes. At the steady state, the probability that a student got enough sleep given this evidence is not zero because we have a model of the observation such that $P(E = re \wedge sc|es) \neq 0$. Additionally, even if we start with a prior probability such that we are sure that the student didn't get enough sleep, according to the transition model, there is the possibility that a student sleeps enough the next day.

f. The probability is $\mathbf{f}_{1:k} = \mathbf{P}(ES_k|\mathbf{e}_{1:k})$ for $k \to \infty$, considering that the observation remains $E = re \wedge sc$ for all days.
Let's consider the state estimation procedure in matrix notation:
$\mathbf{f}_{1:k} = \alpha \mathbf{O}_3 \mathbf{T} \mathbf{f}_{1:k-1}$,
in steady state we can write $\mathbf{f}_{1:k} = \mathbf{f}_{1:k-1} = \mathbf{f}_{ss}$ and thus
$\mathbf{f}_{ss} = \alpha \mathbf{O}_3 \mathbf{T} \mathbf{f}_{ss}$.
To find $\mathbf{f}_{ss}$, we write:
$\underbrace{\frac{1}{\alpha}}_{\lambda} \mathbf{f}_{ss} = \underbrace{O_3 T}_{A} \mathbf{f}_{ss}$,

and we calculate the eigenvalues and eigenvectors of $\mathbf{A} = \begin{bmatrix} 0.0160 & 0.0060 \\ 0.0420 & 0.1470 \end{bmatrix}$.

$det(\lambda \mathbf{I} - \mathbf{A}) = det\left(\begin{bmatrix} \lambda - 0.0160 & -0.0060 \\ -0.0420 & \lambda - 0.1470 \end{bmatrix}\right) = \lambda^2 - 0.1630\lambda + 0.0021 = 0,$

$\lambda_1 = \frac{0.1630+\sqrt{0.02657-0.0084}}{2} = 0.1489,$

$\lambda_2 = \frac{0.1630-\sqrt{0.02657-0.0084}}{2} = 0.0141.$

We can now compute the corresponding eigenvectors $\mathbf{v}_1 = \left[v_{1x},\, v_{1_y}\right]^T$ and $\mathbf{v}_2 = \left[v_{2x},\, v_{2_y}\right]^T$.

$(\lambda_1 \mathbf{I} - \mathbf{A})\mathbf{v}_1 = \begin{bmatrix} 0.1489 - 0.0160 & -0.0060 \\ -0.0420 & 0.1489 - 0.1470 \end{bmatrix} \mathbf{v}_1 = \mathbf{0},$

$0.1329\, v_{1x} - 0.0060\, v_{1y} = 0$
We set e.g. $v_{1y} = 1$ and we get $v_{1x} = \frac{0.0060}{0.1329} = 0.0451$.

$(\lambda_2 \mathbf{I} - \mathbf{A})\mathbf{v}_2 = \begin{bmatrix} 0.0141 - 0.0160 & -0.0060 \\ -0.0420 & 0.0141 - 0.1470 \end{bmatrix} \mathbf{v}_2 = \mathbf{0},$

$-0.0420\, v_{2x} - 0.1329\, v_{2y} = 0$
We set e.g. $v_{2y} = 1$ and we get $v_{2x} = -\frac{0.1329}{0.0420} = -3.1643$.

The probability vector $\mathbf{f}_{ss}$ can only be proportional to the first eigenvector $\mathbf{v}_1$ because the second one has elements with different sign. In conclusion, to find $\mathbf{f}_{ss}$ we normalize $\mathbf{v}_1$:

$\mathbf{f}_{ss} = \alpha\, \mathbf{v}_1 = \frac{1}{0.0451+1} \begin{bmatrix} 0.0451 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.0432 \\ 0.9568 \end{bmatrix}.$

Problem 9.2:

a. Let's first define the state at time step k as X_k that is as a discrete random variable with domain $\langle S_1,\ S_2,\ S_3,\ S_4,\ S_5,\ S_6\rangle$. We formulate now the state transition model as in Fig.4

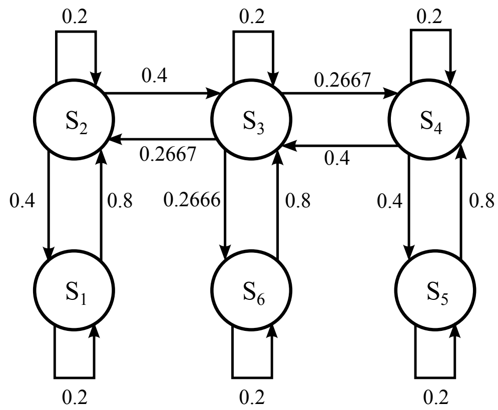


Figure 4: State transition model of Ex. 9.1.

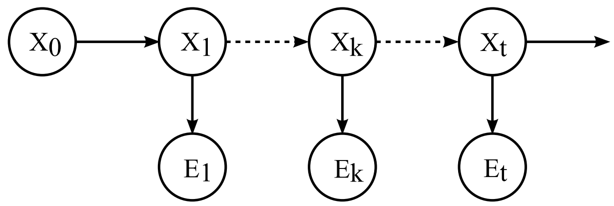


Figure 5: the Hidden Markov Model of Ex. 9.1.

In the following we provide the prior probability of X_0, the transition matrix and the observation matrices for the required cases.

$\mathbf{f}_0 = \mathbf{P}(X_0) = \left[\frac{1}{6},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6}\right]^T.$

Recall: $T_{ij} = P(X_k = S_i | X_{k-1} = S_j)$ and $\begin{cases} (O_{ij})_k = P(e_k | X_k = S_i),\ \ if\ \ j = i \\ \quad 0,\ otherwise. \end{cases}$

$$\mathbf{T} = \begin{bmatrix} 0.2 & 0.4 & 0 & 0 & 0 & 0 \\ 0.8 & 0.2 & 0.2667 & 0 & 0 & 0 \\ 0 & 0.4 & 0.2 & 0.4 & 0 & 0.8 \\ 0 & 0 & 0.2667 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 0.4 & 0.2 & 0 \\ 0 & 0 & 0.2666 & 0 & 0 & 0.2 \end{bmatrix}$$

For the observation matrices we first consider the case for $E_k = SWE$:
$P(E_k = SWE | X_k = S_1) = 1 - 0.1 = 0.9,$
$P(E_k = SWE | X_k = S_2) = 0,$
$P(E_k = SWE | X_k = S_3) = 0,$
$P(E_k = SWE | X_k = S_4) = 0,$
$P(E_k = SWE | X_k = S_5) = 1 - 0.1 = 0.9,$
$P(E_k = SWE | X_k = S_6) = 1 - 0.1 = 0.9.$

$$\mathbf{O}_{k,SWE} = \begin{bmatrix} 0.9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9 \end{bmatrix}$$

We now consider the case for $E_k = NE$:
(Recall: $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$)

$P(E_k = NE | X_k = S_1) = 0,$
$P(E_k = NE | X_k = S_2) = 0,$
$P(E_k = NE | X_k = S_3) = P(fpEast | X_k = S_3)\,(1 - P(fpSouth \vee fpWest | X_k = S_3)) = 0.1\left(1 - 0.1 - 0.1 + 0.1^2\right) = 0.081,$
$P(E_k = NE | X_k = S_4) = (1 - P(fpSouth \vee fpWest | X_k = S_4)) = \left(1 - 0.1 - 0.1 + 0.1^2\right) = 0.81,$

$P(E_k = NE|X_k = S_5) = 0$,
$P(E_k = NE|X_k = S_6) = 0$.
(note: we abbreviated a false positive in a direction with $fpDir$).

$$\mathbf{O}_{k,NE} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.081 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.81 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

b. We receive $E_1 = SWE$ after the first action at $k = 1$. To estimate the location of the robot we define $\mathbf{f}_{1:k} = \mathbf{P}(X_k|\mathbf{E}_{1:k})$ and we use filtering:
$\mathbf{f}_{1:1} = \alpha \mathbf{O}_{1,SWE}\mathbf{T}\mathbf{f}_0 = \frac{1}{0.09+0.09+0.07}[0.09,\ 0,\ 0,\ 0,\ 0.09,\ 0.07]^T = [0.36,\ 0,\ 0,\ 0,\ 0.36,\ 0.28]^T$.

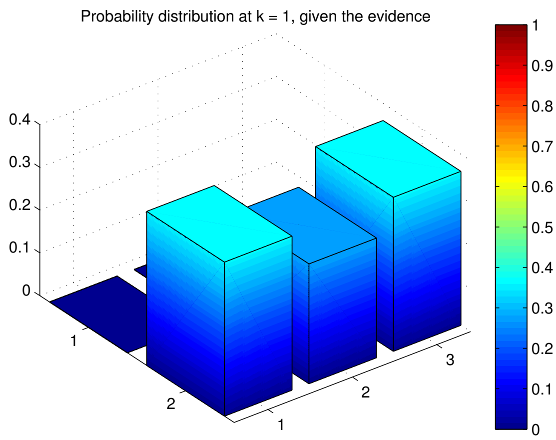


Figure 6: Probability distribution at $k = 1$ given the evidence.

c. $\mathbf{P}_{pred,k=2} = \mathbf{P}(X_2|\mathbf{E}_{1:1}) = \mathbf{T}\mathbf{f}_{1:1} = [0.072,\ 0.288,\ 0.224,\ 0.288,\ 0.072,\ 0.056]^T$

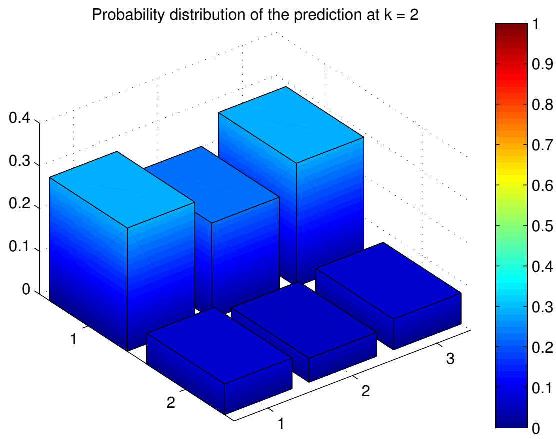


Figure 7: Probability distribution of the prediction at $k = 2$.

d. $\mathbf{f}_{1:2} = \alpha \mathbf{O}_{2,NE}\mathbf{T}\mathbf{f}_{1:1} = \frac{1}{0.0181+0.2333}[0,\ 0,\ 0.0181,\ 0.2333,\ 0,\ 0]^T = [0,\ 0,\ 0.0722,\ 0.9278,\ 0,\ 0]^T$

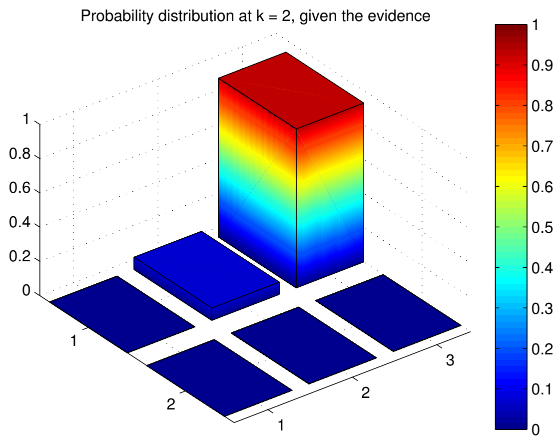


Figure 8: Probability distribution at $k = 2$ given the evidence.