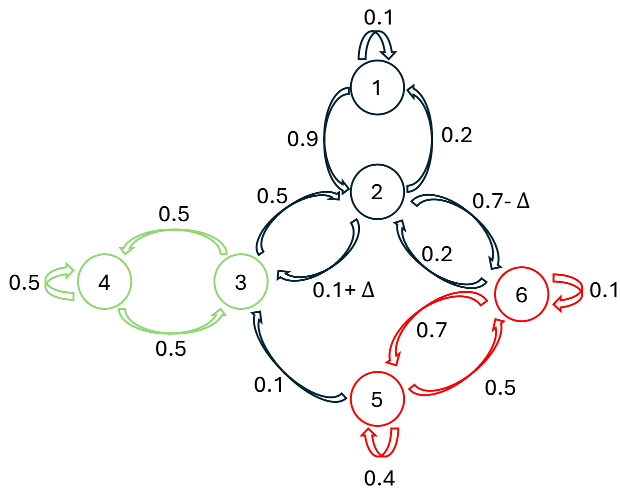
Homework Assignment for AIF

Consider the hidden Markov model defined by states $\{s\}$, observation alphabets $\{o\}$, initial probability matrix D, observation matrix A, and transition matrix B and policy set $\{\pi\}$. The graph below shows the model structure where red states can be high risk and green states can be safe.



Suppose $D \triangleq \begin{bmatrix} p_0(s=1) \\ \cdot \\ \cdot \\ p_0(s=6) \end{bmatrix} = \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$. Let $O = \{\alpha, \beta, \gamma, \mu\}$ be the observation set, then the observation or the likliehood matrix at t=0 is given by:

$$A_0^T = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccc} \alpha & \beta & \gamma & \mu \end{array} \\ \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.35 & 0.35 & 0.15 & 0.15 \\ 0.40 & 0.40 & 0.10 & 0.10 \\ 0.40 & 0.40 & 0.10 & 0.10 \\ 0.05 & 0.05 & 0.20 & 0.70 \\ 0.10 & 0.05 & 0.35 & 0.50 \end{bmatrix} \end{array}$$

Note that A_0^T is the transpose of A_0. State transition matrix B at time τ is given by B_τ^π where t is for time and π is a policy, and $B_\tau^\pi(i,j)$ is the transition probability where columns are states at time τ and rows are states at $\tau + 1$. $B_0^\pi(i,j)$ can be obtained from the above Markov graph.

Our agent is able to set policies, where policy set $\{\pi\}$ is defined by Δ which reduces transition probability (i.e., tightening a flow gate) from $2 \rightarrow 6$ by Δ and increase (i.e., more opening of flow gate) from $2 \rightarrow 3$ by Δ. Suppose that time ticks discretely, and only one observation can be made between any two time ticks. For now we will think of one policy only which is described below.

<u>You will be able to use the Python code provided to you earlier to generate observations or outcomes for the above Hidden Markov Model.</u>

Suppose that the agent runs policy π_1 $\{\Delta = 0.10\}$ for the next $T = 100$ time epochs and observes the observation set $\boldsymbol{O} = \{\boldsymbol{o_0}, \boldsymbol{o_1}, \dots \boldsymbol{o_T}\}$. Note that you can generate this observation set using the Python code provided to you.

A. Write your code to calculate the state beliefs $\{q(s_t|\pi) = \boldsymbol{S}_{\pi t}\} \triangleq \begin{bmatrix} q\,(s=1) \\ \cdot \\ \cdot \\ q\,(s=6) \end{bmatrix}$ following each observation in the above episode. The relationship for belief calculation is:

$$\varepsilon_{\pi,t} \leftarrow \frac{1}{2}\left(\ln\left(\boldsymbol{B}_{\pi,t-1}s_{\pi,t-1}\right) + \ln\left(\boldsymbol{B}^{\dagger}_{\pi,t}s_{\pi,t+1}\right)\right) + \ln\ \boldsymbol{A}^{\mathrm{T}}o_t - \ln\ s_{\pi,t}$$

$$v_{\pi,t} \leftarrow v_{\pi,t} + \varepsilon_{\pi,t}\ ;\quad s_{\pi,t} \leftarrow \sigma\left(v_{\pi,t}\right)$$

You can run it for several episodes and create a box graph (or show the probability ranges on a graph).

B – Suppose the agent switches to policy π_2 {Δ= 0.4} which is expected to lead to less risky states. Generate an eposide of $\boldsymbol{O} = \{\boldsymbol{o_0}, \boldsymbol{o_1}, \ldots \boldsymbol{o_T}\}$ for T=100. Repeate II.a using this episode and compare the results to II.a. Provide your own interpretation of results; for instance, if and how beliefs change and settle with respect to risky operation. You may try several different episodes of II.a and II.b to ease up your interpretation.

C – Suppose the agent prefers outcomes according to $\boldsymbol{C}_t = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, that is our agent prefers to receive α, β and none of the other two. Based on A_1^T, states {1,2,3,4} are more likely to produce our desirable outcomes. The agent would like to calculate its risks at time $t = T = 100$ under the two policies of π_1 and π_2 where the risk function is a component of Expected Free Energy and is given by

$$\mathbf{A}s_{\pi=1,\tau} \cdot \left(\ ln\,\mathbf{A}\,\mathbf{s}_{\pi=1,\tau} -\ ln\,\mathbf{C}_\tau\right).$$

D – Under the likelihood matrix A_0^T, some states have more precise distributions with respect to the outcomes, for instance, states 5 or 6 provides more precise information compared to states 1 or 2. As such, we expect outcome prediction errors to drive selection of the policy that will lead the agent toward state 5 or 6 insteand of state 1 or 2. The term

$$\mathrm{diag}\left(\mathbf{A}^{\mathrm{T}}\mathrm{l\,n}\,\mathbf{A}\right) \cdot s_{\pi=1,\tau}$$

of Expected Free Energy computes the ambiguity of a policy. Using this formula compute the ambiguity of the two policies at time $t = T$=100.

E – Suppose the current time is t =100 and the process has been running according to part (**A**). Assume that there are four possible actions to take, namely,
$\{u_1, u_2, u_3, u_4\}$, where $u_1 \triangleq (\Delta = 0), u_2 \triangleq (\Delta = 0.1), u_3 \triangleq (\Delta = -0.1), u_4 \triangleq (\Delta = 0.2)$.
An action is applied per a time epoch.

Also, suppose we have three possible action policies, namely, $\pi_1 = \{u_4, u_1, u_2, u_3\}, \pi_2 = \{u_1, u_4, u_3, u_3\}$ and $\pi_3 = \{u_1,\ u_1,\ u_2,\ u_3\}$ that can apply for any given four consecutive time epochs. Any combination of these policies can repeat until the end of a given time horizon.

For $\boldsymbol{C}_{t=110} = \begin{bmatrix} .25 \\ .25 \\ .25 \\ .25 \end{bmatrix}$ find the optimal set of actions, so basically, you are looking for "optimal" actions that take you from $t = 100$ to $t = 110$ ending at states that produce equally likely outcomes.

For $\boldsymbol{C}_{t=110} = \begin{bmatrix} .45 \\ .45 \\ .05 \\ .05 \end{bmatrix}$ find the optimal set of actions, so basically, you are looking for "optimal" actions that take you from $t = 100$ to $t = 110$ where the first two outcomes are more preferred to the other two.

Important Note: For every change of action or policy, set the parameter WarmUp=50. That means the model will run for 50 time epochs for warm up and only at the last epoch an observation will be collected.