1. Consider the computer failure example described in the book by Cowell et al. (Chap. 2.9, p. 17-19), with two DAG models given in Fig. 2.5 and 2.6 respectively, and with the following specifications of variables and probabilities:

C : Computer failure?, E : Electricity failure?, M : Malfunction?, L : Light failure?.

$$\begin{aligned} P(E = \text{yes}) &= 0.1, & P(M = \text{yes}) &= 0.2, \\ P(C = \text{yes} \mid E = \text{no}, M = \text{no}) &= 0, & P(C = \text{yes} \mid E = \text{no}, M = \text{yes}) &= 0.5, \\ P(C = \text{yes} \mid E = \text{yes}, M = \text{no}) &= 1, & P(C = \text{yes} \mid E = \text{yes}, M = \text{yes}) &= 1, \\ P(L = \text{yes} \mid E = \text{yes}) &= 1, & P(L = \text{yes} \mid E = \text{no}) &= 0.2. \end{aligned}$$

Use belief propagation to find answers to the following questions:

(a) What is $P(C = \text{yes})$ under the model in Fig. 2.5?

(b) Suppose we find the computer fails (the event $C = \text{yes}$ occurs). What are

$$P(E = \text{yes} \,|\, C = \text{yes}), \quad P(M = \text{yes} \,|\, C = \text{yes})$$

under the model in Fig. 2.5?

(c) Suppose we find also that the light does not work ($L = \text{yes}$). What are

$$P(E = \text{yes} \,|\, C = \text{yes}, L = \text{yes}), \quad P(M = \text{yes} \,|\, C = \text{yes}, L = \text{yes})$$

under the model in Fig. 2.6?

(d) What is the most probable configuration of (E, M) under the condition of (c)?

(Please include all the intermediate steps in your solution.)

Solution: Let $\lambda_{uv}(\text{no}), \lambda_{uv}(\text{yes})$ denote the two elements of a λ-message from a node u to a node v, and let λ_{uv} denote the message in vector form $\big[\, \lambda_{uv}(\text{no}) \,,\, \lambda_{uv}(\text{yes}) \,\big]$. We use similar notation for π-messages.

(a) Node C receives the messages from E and M,

$\pi_{EC} = \big[\, P(E = \text{no}) \,,\, P(E = \text{yes}) \,\big] = \big[\, 0.9 \,,\, 0.1 \,\big], \quad \pi_{MC} = \big[\, P(M = \text{no}) \,,\, P(M = \text{yes}) \,\big] = \big[\, 0.8 \,,\, 0.2 \,\big],$

from which it can calculate its belief $P(C = \text{yes})$ as

$$P(C = \text{yes}) = \sum_{a \in \{\text{yes,no}\}} \sum_{b \in \{\text{yes,no}\}} P(C = \text{yes} \mid E = a, M = b) \cdot \pi_{EC}(a) \cdot \pi_{MC}(b) = 0.19. \tag{1}$$

(b) Let $\mathbf{e}$ denote the evidence/event $\{C = \text{yes}\}$. Node C sends the messages to E and M,

$$\begin{aligned} \lambda_{CE} &= \big[\, P(C = \text{yes} \mid E = \text{no}) \,,\, P(C = \text{yes} \mid E = \text{yes}) \,\big] = \big[\, 0.1 \,,\, 1 \,\big], \\ \lambda_{CM} &= \big[\, P(C = \text{yes} \mid M = \text{no}) \,,\, P(C = \text{yes} \mid M = \text{yes}) \,\big] = \big[\, 0.1 \,,\, 0.55 \,\big], \end{aligned}$$

which are composed by using the following formula: for $a \in \{\text{yes}, \text{no}\}$,

$$\lambda_{CE}(a) = \sum_{b \in \{\text{yes,no}\}} P(C = \text{yes} \mid E = a, M = b) \cdot \pi_{MC}(b), \tag{2}$$

$$\lambda_{CM}(a) = \sum_{b \in \{\text{yes,no}\}} P(C = \text{yes} \mid E = b, M = a) \cdot \pi_{EC}(b). \tag{3}$$

Receiving the message $\lambda_{CE} = \big[\, 0.1 \,,\, 1 \,\big]$ from C, node E then calculates

$$P(E = \text{no}, \mathbf{e}) = P(E = \text{no}) \cdot \lambda_{CE}(\text{no}) = 0.09, \quad P(E = \text{yes}, \mathbf{e}) = P(E = \text{yes}) \cdot \lambda_{CE}(\text{yes}) = 0.10,$$

and obtains by normalization that

$$P(E = \text{yes} \mid \mathbf{e}) = \frac{P(E = \text{yes}, \mathbf{e})}{P(E = \text{no}, \mathbf{e}) + P(E = \text{yes}, \mathbf{e})} = 0.5263. \tag{4}$$

Similarly, receiving the message $\lambda_{CM} = \left[\, 0.1\,,\, 0.55 \,\right]$ from C, node M calculates

$$P(M = \text{no}, \mathbf{e}) = P(M = \text{no}) \cdot \lambda_{CM}(\text{no}) = 0.08, \quad P(M = \text{yes}, \mathbf{e}) = P(M = \text{yes}) \cdot \lambda_{CM}(\text{yes}) = 0.11,$$

and obtains by normalization that

$$P(M = \text{yes} \mid \mathbf{e}) = \frac{P(M = \text{yes}, \mathbf{e})}{P(M = \text{no}, \mathbf{e}) + P(M = \text{yes}, \mathbf{e})} = 0.5789. \tag{5}$$

The normalization constant in both cases is the same; it is the probability of $\mathbf{e}$:

$$P(\mathbf{e}) = P(C = \text{yes}) = \sum_{a \in \{\text{yes,no}\}} P(E = a, \mathbf{e}) = \sum_{a \in \{\text{yes,no}\}} P(M = a, \mathbf{e}) = 0.19,$$

which is consistent with the result of part (a), as it should be.

(c) Let $\mathbf{e}'$ denote the new total evidence $\{C = \text{yes}, L = \text{yes}\}$ and $\mathbf{e}'_L$ the partial evidence $\{L = \text{yes}\}$. Node L sends a message to E:

$$\lambda_{LE} = \left[\, P(L = \text{yes} \mid E = \text{no})\,,\, P(L = \text{yes} \mid E = \text{yes}) \,\right] = \left[\, 0.2\,,\, 1 \,\right].$$

Combining this with the message $\lambda_{CE} = \left[\, 0.1\,,\, 1 \,\right]$ (from part (b)), node E can then calculate

$$\begin{aligned} P(E = \text{no}, \mathbf{e}') &= P(E = \text{no}) \cdot \lambda_{CE}(\text{no}) \cdot \lambda_{LE}(\text{no}) = 0.018, \\ P(E = \text{yes}, \mathbf{e}) &= P(E = \text{yes}) \cdot \lambda_{CE}(\text{yes}) \cdot \lambda_{LE}(\text{yes}) = 0.100, \end{aligned}$$

and by normalization,

$$P(E = \text{yes} \mid \mathbf{e}') = \frac{P(E = \text{yes}, \mathbf{e}')}{P(E = \text{no}, \mathbf{e}') + P(E = \text{yes}, \mathbf{e}')} = 0.8475. \tag{6}$$

At the same time, node E sends a new message to C:

$$\pi_{EC} = \left[\, P(E = \text{no}, \mathbf{e}'_L)\,,\, P(E = \text{yes}, \mathbf{e}'_L) \,\right] = \left[\, 0.18\,,\, 0.10 \,\right],$$

which is composed by the formula

$$\pi_{EC}(a) = P(E = a) \cdot \lambda_{LE}(a), \quad a = \text{yes, no}.$$

Subsequently, C sends a new message to M, composed by using the formula in Eq. (3),

$$\lambda_{CM} = \left[\, P(\mathbf{e}' \mid M = \text{no})\,,\, P(\mathbf{e}' \mid M = \text{yes}) \,\right] = \left[\, 0.10\,,\, 0.19 \,\right].$$

Node M now updates its belief by calculating

$$P(M = \text{no}, \mathbf{e}') = P(M = \text{no}) \cdot \lambda_{CM}(\text{no}) = 0.08, \quad P(M = \text{yes}, \mathbf{e}') = P(M = \text{yes}) \cdot \lambda_{CM}(\text{yes}) = 0.038,$$

and obtains by normalization that

$$P(M = \text{yes} \mid \mathbf{e}') = \frac{P(M = \text{yes}, \mathbf{e}')}{P(M = \text{no}, \mathbf{e}') + P(M = \text{yes}, \mathbf{e}')} = 0.3220. \tag{7}$$

For both E and M, the normalization constant is

$$P(\mathbf{e}') = 0.118.$$

(d) The π^*-messages that E sends to C coincide with the π-messages it sends in part (c) (because no variables need to be eliminated in the sub-polytree containing E). This is also the case with the π^*-message from M to C and with the λ^*-message from L to E.

After $\mathbf{e}'$ occurs, node C has the following π^*-messages from E and M respectively:

$$\pi^*_{EC} = \big[\, 0.18\,,\, 0.10 \,\big], \quad \text{(from part (c))}; \qquad \pi^*_{MC} = \big[\, 0.8\,,\, 0.2 \,\big], \quad \text{(from part (a))}.$$

Node C sends a λ^*-message to its parent E:

$$\lambda^*_{CE} = \big[\, P^*(C = \text{yes} \mid E = \text{no})\,,\, P^*(C = \text{yes} \mid E = \text{yes}) \,\big] = \big[\, 0.1\,,\, 0.8 \,\big],$$

which is composed by using the formula

$$\lambda^*_{CE}(a) = \max_{b \in \{\text{yes,no}\}} \Big\{ P(C = \text{yes} \mid E = a, M = b) \cdot \pi^*_{MC}(b) \Big\}, \quad a = \text{yes, no}.$$

And it sends a λ^*-message to its parent M:

$$\lambda^*_{CM} = \big[\, P^*(\mathbf{e}' \mid M = \text{no})\,,\, P^*(\mathbf{e}' \mid M = \text{yes}) \,\big] = \big[\, 0.10\,,\, 0.10 \,\big],$$

which is composed by using the formula

$$\lambda^*_{CM}(a) = \max_{b \in \{\text{yes,no}\}} \Big\{ P(C = \text{yes} \mid E = b, M = a) \cdot \pi^*_{EC}(b) \Big\}, \quad a = \text{yes, no}.$$

For node E, using C's message and L's message $\lambda^*_{LE} = \big[\, 0.2\,,\, 1 \,\big]$ (from part (c)), E can calculate

$$P^*(E = \text{no}, \mathbf{e}') = P(E = \text{no}) \cdot \lambda^*_{CE}(\text{no}) \cdot \lambda^*_{LE}(\text{no}) = 0.018,$$
$$P^*(E = \text{yes}, \mathbf{e}') = P(E = \text{yes}) \cdot \lambda^*_{CE}(\text{yes}) \cdot \lambda^*_{LE}(\text{yes}) = 0.08,$$

and from which it can conclude that $E = \text{yes}$ in any most probable configuration given $\mathbf{e}'$.

Similarly for node M. Using C's message, M can calculate

$$P^*(M = \text{no}, \mathbf{e}') = P(M = \text{no}) \cdot \lambda^*_{CM}(\text{no}) = 0.08,$$
$$P^*(M = \text{yes}, \mathbf{e}') = P(M = \text{yes}) \cdot \lambda^*_{CM}(\text{yes}) = 0.02,$$

and from which it can conclude that $M = \text{no}$ in any most probable configuration given $\mathbf{e}'$.

Since both functions $P^*(M = \cdot, \mathbf{e}')$, $P^*(E = \cdot, \mathbf{e}')$ have a unique maximum, the most probable configuration of (E, M) is unique and given by $(E = \text{yes}, M = \text{no})$. Furthermore,

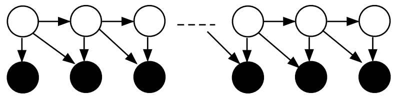
$$P(E = \text{yes}, M = \text{no}, \mathbf{e}') = \max_{a,b \,\in\, \{\text{yes,no}\}} P(E = a, M = b, \mathbf{e}') = P^*(E = \text{yes}, \mathbf{e}') = P^*(M = \text{no}, \mathbf{e}') = 0.08,$$

and the probability of the most probable configuration can be calculated as

$$P(E = \text{yes}, M = \text{no} \mid \mathbf{e}') = \frac{P^*(M = \text{no}, \mathbf{e}')}{P(\mathbf{e}')} = \frac{P^*(E = \text{yes}, \mathbf{e}')}{P(\mathbf{e}')} = \frac{0.08}{0.118} = 0.678,$$

where $P(\mathbf{e}')$ is obtained at the end of part (c). □

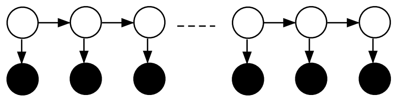
2. Consider the HMM below, in which the observation variables (black) depend on both the current and previous states (white). Is the DAG singly connected? Explain how you would apply the belief propagation algorithms for singly connected networks to inference under this model.



Solution: This DAG is not singly connected. But we can consider an equivalent model with a singly connected DAG, to which we can apply the belief propagation algorithms. Let $(X_1, \dots, X_n)$ and $(Y_1, \dots Y_n)$ be the state and observation variables, respectively, in the given HMM. Define

$$\widehat{X}_i = (X_{i-1}, X_i), \quad i = 1, \dots, n$$

with X_0 being a dummy variable. Then $(\widehat{X}_1, \dots, \widehat{X}_n)$ is a Markov chain and the joint distribution of $(\widehat{X}_1, \dots, \widehat{X}_n)$ and $(Y_1, \dots Y_n)$ factorizes recursively according to the following singly connected DAG:



(The transition and observation probability distributions are given by $P(\widehat{X}_{i+1} \mid \widehat{X}_i) = P(X_{i+1} \mid X_i)$ and $P(Y_i \mid \widehat{X}_i) = P(Y_i \mid X_{i-1}, X_i)$ respectively.) □

Problem 3 (on the next page) is from Exercises 4.1 and 5.1 of Pearl's 1988 book. It may help to program the message passing algorithms to solve this problem.

3. A language L has a four-character vocabulary $V = \{\epsilon, A, B, C\}$ where ϵ is the empty symbol. The probability that character v_i will be followed by v_j is given by the following matrix:

$$P(v_j \,|\, v_i) = \begin{array}{c|cccc} v_i \backslash v_j & \epsilon & A & B & C \\ \hline \epsilon & 1/4 & 1/4 & 1/4 & 1/4 \\ A & 1/2 & 0 & 1/4 & 1/4 \\ B & 1/8 & 1/2 & 1/8 & 1/4 \\ C & 1/4 & 1/8 & 1/2 & 1/8 \end{array}$$

In transmitting messages from L, some characters may be corrupted by noise and be confused with others. The probability that the transmitted character v_j will be interpreted as v_k is given by the following confusion matrix:

$$P_c(v_k \,|\, v_j) = \begin{array}{c|cccc} v_j \backslash v_k & \epsilon & A & B & C \\ \hline \epsilon & .9 & .1 & 0 & 0 \\ A & .1 & .8 & .1 & 0 \\ B & 0 & .1 & .8 & .1 \\ C & 0 & .1 & .1 & .8 \end{array}$$

The string, $\epsilon, \epsilon, B, C, A, \epsilon, \epsilon$ is received, and it is known that the transmitted string begins and ends with ϵ.

(a) Find the probability that the ith transmitted symbol is C, for $i = 1, 2, \ldots, 7$.

(b) Find the probability that the string transmitted is the one received.

(c) Find the probability that no message (a string of ϵ's) was transmitted.

(d) Find the message most likely to have been transmitted.

(e) Find the most likely seven-symbol string in L that starts and ends with ϵ.

Solution: (a) Let $\mathbf{e}$ denote the evidence that the string, $\epsilon, \epsilon, B, C, A, \epsilon, \epsilon$ is received and the transmitted string begins and ends with ϵ. For $i = 1, \ldots, 7$, $P(V_i = C \,|\, \mathbf{e})$ are given by

i:	1	2	3	4	5	6	7
	0	0	0.1033	0.7881	0.0444	0	0

(b) The probability that the string transmitted is the one received is 0.4139.

(c) The probability that a string of ϵ's was transmitted is 0.

(d) The message most likely to have been transmitted is the received string, $\epsilon\epsilon BCA\epsilon\epsilon$.

(e) There seem to be six most likely strings of length 7 and starting and ending with ϵ:

$$\epsilon CBA\epsilon A\epsilon, \quad \epsilon A\epsilon CBA\epsilon, \quad \epsilon BA\epsilon BA\epsilon, \quad \epsilon BACBA\epsilon, \quad \epsilon CBABA\epsilon, \quad \epsilon CBCBA\epsilon.$$

(The third string differs from the fourth in the fourth symbol, and so does the fifth string from the sixth string.) Acknowledgement: this non-uniqueness of the most probable strings was pointed out to me by a participant of the class. □