CSC242: Homework 3.5
AIMA 15-15.2

1. The basic approach to modeling uncertainty in a changing world is to view the world as a series of "snapshots" or time slices. Each time slice includes a set $\mathbf{X}$ of state variables assumed be unobservable, and a set $\mathbf{E}$ of evidence variables that can be observed.

   (a) Explain briefly what the (first-order) Markov assumption is and what it means for the state variables $\mathbf{X}$. Be formal.

   ANSWER: The first order Markov assumptions says that the current state depends on the previous state and not on any earlier states. Equivalently, a state provides enough information to make the future conditionally independent of the past. Formally:
   $$\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t-1})$$

   (b) Explain briefly what a stationary process assumption is and what it means for the state variables $\mathbf{X}$. Be formal.

   ANSWER: A stationary process assumptions assumes that the underlying process generating the states does not change, at least not over the duration with which we are concerned. In other words, a single conditional probability distribution applies to all the state variable nodes. Formally:
   $$\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t-1}) \text{ is the same for all } t$$

   (c) Explain briefly what the (first-order) sensor Markov assumption means for the evidence variables $\mathbf{E}$. Be formal.

   ANSWER: The sensor Markov assumption says that the current state is sufficient to generate the current sensor values. Formally:
   $$\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}) = \mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$$

   (d) Explain briefly what a stationary process assumption means for the evidence variables $\mathbf{E}$. Be formal.

   ANSWER: A stationary process assumption for the evidence variables says that the conditional probability of the evidence (observable) variables given the state (unobservable) variables does not change, at least not over the duration with which we are concerned. In other words, a single conditional probability distribution applies to all the evidence variable nodes. Formally:

$$\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t) \text{ is the same for all } t$$

(e) Give an example of a Bayesian network for a temporal model with a single state variable X and a single evidence variable E. Draw a picture. Include formulas describing the probabilities required by the network (no tables or numbers needed).

ANSWER: Example like Figure 15.2: Markov chain of state variables, each connected (only) to their corresponding evidence variable. A single CPT for all the state variables holds $\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t-1})$. A single CPT for all the evidence variables holds $\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$. And you need the state prior $\mathbf{P}(\mathbf{X}_0)$.

2. Consider the four inference tasks for temporal models identified in the textbook and discussed in class.

(a) Identify the four tasks and give the posterior distribution that each task needs to compute (or estimate), using the standard notation as seen in class and in the textbook.

ANSWER:

- Filtering or state estimation: $\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t})$
- Prediction: $\mathbf{P}(\mathbf{X}_{t+k} \mid \mathbf{e}_{1:t})$
- Smoothing: $\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:t})$ for some k such that $0 \leq k < t$
- Most likely explanation: $\operatorname{argmax}_{\mathbf{x}_{1:t}} P(\mathbf{x}_{1:t} \mid \mathbf{e}_{1:t})$

(b) Can the basic temporal inference tasks on Bayesian networks be computed efficiently? Why or why not (very briefly)?

ANSWER: Yes they can, thanks to the assumptions that go into the model. The resulting Bayes net structure allows the necessary posterior distribution to be updated incrementally in light of new evidence. Examples include the FORWARD update equation (AIMA Eq. 15.5) for filtering, the FORWARD-BACKWARD algorithm for smoothing (AIMA Fig 15.4), and the Viterbi algorithm for computing the most likely sequence (AIMA Eq. 15.11).

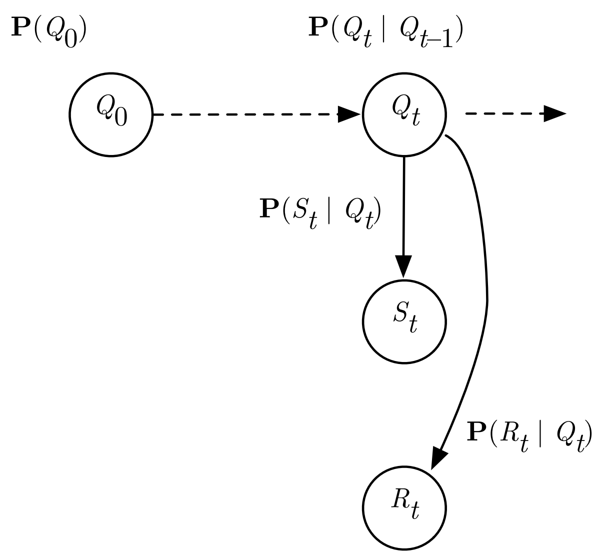
(c) Why is this important? Be very brief.

ANSWER: "The time and space requirements for updating must be constant if an agent with limited memory is to keep track of the current state distribution over an unbounded sequence of observations" (AIMA p. 572).

3. Suppose you are modelling an industrial process and trying to monitor (estimate) the quality of the output as acceptable or not acceptable. You can't directly measure the quality, but you have two sensors, each of which have some likelihood of detecting acceptable and unacceptable products.

   (a) Develop a temporal model to describe this process. Draw a picture. Include any distributions needed by your model, but of course you don't have the numbers.

   ANSWER: State variable is Boolean Q_t for quality. Evidence variables are Booleans S_t and R_t.



   (b) Explain how the assumptions underlying temporal models apply to this model and discuss whether they are reasonable (or not).

   ANSWER:

   - Markov assumption: whether the product is acceptable or not depends only on whether it was so at the previous time step. Not unreasonable, although the probabilities will change over a long enough time.
   - Stationary process assumption: Assumes environment doesn't change over period that we are monitoring. Risky but necessary.
   - Sensor Markov assumption: Assumes that the sensors are only affected by the quality of the product (no outside influences and do not influence each other), and only the *current* quality affects the sensors, not their history. Probably not true over a long timespan or after "extreme" readings.
   - Stationarity for sensor model: Probably false over even reasonably short timespans—sensors get worn out, *etc.*. But what else can you do? More modeling for a more accurate model...

(c) Give an explicit formula (no big sum or product symbols) for the full joint probability distribution of the model at time $t = 1$.

ANSWER: In general, a temporal model is a Bayesian network, so we have:

$$\mathbf{P}(\mathbf{X}_{0:t}, \mathbf{E}_{1:t}) = \mathbf{P}(X_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i \mid \mathbf{X}_{i-1})\mathbf{P}(\mathbf{E}_i \mid \mathbf{X}_i)$$

But that's not an "explicit" formula, so let's "unroll" the network. At time $t = 1$, there are four nodes in the unrolled network: the initial state variable for $t = 0$ and the state and evidence variables for $t = 1$. You should probably draw that picture including the specific distributions stored at the nodes (that is, with numbers instead of expressions involving t).

Thus the full joint probability distribution for our unrolled network is:

$$\mathbf{P}(\mathbf{X}_{0:1}, \mathbf{E}_1) = \mathbf{P}(\mathbf{X}_0, \mathbf{X}_1, \mathbf{E}_1) = \mathbf{P}(Q_0, Q_1, S_1, R_1)$$

Using the structure of the network, this can be factored into the product of the distributions stored at the nodes:

$$\mathbf{P}(\mathbf{X}_{0:1}, \mathbf{E}_1) = \mathbf{P}(Q_0)\,\mathbf{P}(Q_1 \mid Q_0)\,\mathbf{P}(S_1 \mid Q_1)\,\mathbf{P}(R_1 \mid Q_1)$$

(d) Give an explicit formula for the state of the process at time $t = 1$ given the observations at that time.

ANSWER: In general for a temporal model of this form:

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_1 \mid \mathbf{e}_1) &= \alpha\,\mathbf{P}(\mathbf{X}_1, \mathbf{e}_1) && \text{(Defn of cond. prob.)}\\
&= \alpha \sum_{\mathbf{x}_0} \mathbf{P}(\mathbf{x}_0, \mathbf{X}_1, \mathbf{e}_1) && \text{(Marginalize over } \mathbf{X}_0\text{)}\\
&= \alpha \sum_{\mathbf{x}_0} \mathbf{P}(\mathbf{X}_0)\,\mathbf{P}(\mathbf{X}_1 \mid \mathbf{X}_0)\,\mathbf{P}(\mathbf{e}_1 \mid \mathbf{X}_1) && \text{(Factored FJPD)}
\end{aligned}$$

For this example, we get:

$$\begin{aligned}
&= \alpha \sum_{q_0} \mathbf{P}(q_0)\,\mathbf{P}(Q_1 \mid q_0)\,\mathbf{P}(S_1 \mid W_1)\,\mathbf{P}(R_1 \mid Q_1)\\
&= \alpha[\mathbf{P}(q_0)\,\mathbf{P}(Q_1 \mid q_0)\,\mathbf{P}(S_1 \mid W_1)\,\mathbf{P}(R_1 \mid Q_1)+\\
&\qquad \mathbf{P}(\neg q_0)\,\mathbf{P}(Q_1 \mid \neg q_0)\,\mathbf{P}(S_1 \mid W_1)\,\mathbf{P}(R_1 \mid Q_1)]
\end{aligned}$$

(e) Which of the inference tasks for temporal models does this formula perform (compute)?

ANSWER: State estimation (filtering).