ROB311 Quiz 2

Hanhee Lee

February 9, 2025

Contents

1 Bayesian Networks 2
- 1.1 Junction 2
  - 1.1.1 Causal Chain 2
  - 1.1.2 Common Cause 2
  - 1.1.3 Common Effect 3
- 1.2 Dependence Separation 4
  - 1.2.1 Blocked 4
  - 1.2.2 Blocked Undirected Path 4
  - 1.2.3 Independence 4
  - 1.2.4 Consequence of Dependence Separation 4

2 Probabilistic Inference 5
- 2.1 Problem Setup 5
- 2.2 Method 1: Bayesian Network Inference 5
  - 2.2.1 Choosing an Elimination Ordering 5
  - 2.2.2 Heuristics for Elimination Ordering 5
- 2.3 Method 2: Inference via Sampling 6
  - 2.3.1 Inference via Sampling with Likelihood Weighting 6
- 2.4 Canonical Problems: 6
  - 2.4.1 Inference via Sampling 8

Probabilistic Inference Problems

1 Bayesian Networks

Definition: Vertices represent random variables and edges represent dependencies between variables.

1.1 Junction

Definition: A **junction** $\mathcal{J}$ consists of three vertices, X_1, X_2, and X_3, connected by two edges, e_1 and e_2:

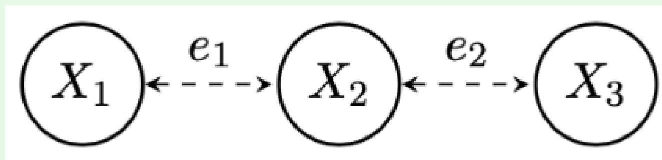

Figure 1

- X_1 and X_2 are not independent, X_2 and X_3 are not independent, but when is X_1 and X_3 independent?

1.1.1 Causal Chain

Definition: A causal chain is a junction $\mathcal{J}$ s.t.

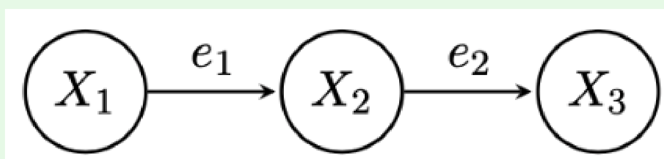

Figure 2

- X_1 and X_3 are not independent (unconditionally), but are independent given X_2.

Notes:
- **Analogy:** Given X_2, X_1 and X_3 are independent. Why? X_2's door closes when you know X_2, so X_1 and X_3 are independent.
- **Distinction b/w Causal and Dependence:** X_1 and X_2 are dependent. However, from a causal perspective, X_1 is influencing X_2 (i.e. $X_1 \to X_2$).

Warning: X_1 is influeincing X_2 and X_2 is influencing X_3.

1.1.2 Common Cause

Definition: A common cause is a junction $\mathcal{J}$ s.t.

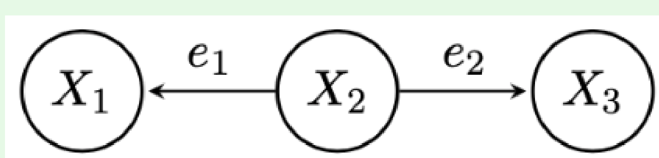

Figure 3

- X_1 and X_3 are not independent (unconditionally), but are independent given X_2.

Notes:
- **Analogy:** Given X_2, X_1 and X_3 are independent. Why? Consider the following example:
  - Let X_2 represent whether a person smokes or not, X_1 represent whether they have yellow teeth, X_3 represent whether they have lung cancer.
- Without knowing X_2, observing X_1 provides information about X_3 because yellow teeth are associated with smoking, which in turn increases the likelihood of lung cancer.
- If X_2 is known, then knowing whether a person has yellow teeth provides no additional information about whether they have lung cancer beyond what is already known from smoking status.

1.1.3 Common Effect

Definition: A common effect is a junction $\mathcal{J}$ s.t.

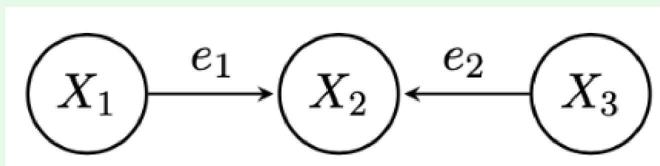

Figure 4

- X_1 and X_3 are independent (unconditionally), but are not independent given X_2 or any of X_2's descendents.

Notes:
- **Analogy:** Consider the following example:
  - Let X_2 represent whether the grass is wet, X_1 represent whether it rained, X_3 represent whether the sprinkler was on.
- Without knowing whether the grass is wet (X_2), the occurrence of rain (X_1) and the sprinkler being on (X_3) are independent events. The rain may occur regardless of the sprinkler, and vice versa.
- However, once we observe that the grass is wet (X_2), the two events become dependent:
  - If we learn that the sprinkler was not on, then the wet grass must have been caused by rain.
  - If we learn that it did not rain, then the wet grass must have been caused by the sprinkler.

1.2 Dependence Separation

1.2.1 Blocked

Definition: $\mathcal{J} = (\{X_1, X_2, X_3\}, \{e_1, e_2\})$ is **blocked** given $\mathcal{K} \subseteq \mathcal{V}$ if X_1 and X_3 are independent given $\mathcal{K}$.

1.2.2 Blocked Undirected Path

Definition: An undirected path,

$$p = \langle (X_1, e_1, X_2), \ldots, (X_{|p|-1}, e_{|p|-1,|p|}, X_{|p|}) \rangle,$$

is **blocked** given $\mathcal{K} \subseteq \mathcal{V}$ if any of its junctions,

$$\mathcal{J}^{(n)} = \{(X_{n-1}, X_n, X_{n+1}), (e_{n-1}, e_n)\},$$

is blocked given $\mathcal{K}$.

1.2.3 Independence

Theorem: Any two variables, X_1 and X_2, in a Bayesian network, $\mathcal{B} = (\mathcal{V}, \mathcal{E})$, are independent given $\mathcal{K} \subseteq \mathcal{V}$ if every undirected path is blocked.

1.2.4 Consequence of Dependence Separation

Theorem: For any variable, $X \in \mathcal{V}$, it can be shown that X is independent of X's non-descendants, $\mathcal{V} \setminus \mathrm{des}(X)$, given X's parents, $\mathrm{pts}(X)$.

Notes:

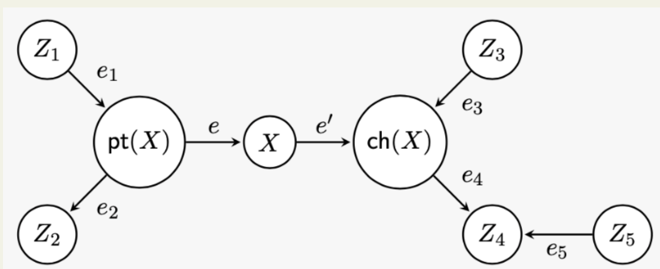

Figure 5

2 Probabilistic Inference

2.1 Problem Setup

Definition: Given a Bayesian network, $\mathcal{B} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{X_1, \ldots, X_{|\mathcal{V}|}\}$, we want to find the value of:

$$\mathrm{pr}(\mathbf{Q} \mid \mathbf{E}) := \mathrm{pr}(Q_1, \ldots, Q_{|\mathbf{Q}|} \mid E_1, \ldots, E_{|\mathbf{E}|}) = \frac{\sum_{\mathcal{V} \setminus (\mathbf{Q} \cup \mathbf{E})} p(X_1, \ldots, X_{|\mathcal{V}|})}{\sum_{\mathcal{V} \setminus \mathbf{E}} p(X_1, \ldots, X_{|\mathcal{V}|})}$$

- $\mathbf{Q} = \{Q_1, \ldots, Q_{|\mathbf{Q}|}\}$
- $\mathbf{E} = \{E_1, \ldots, E_{|\mathbf{E}|}\} \subseteq \mathcal{V}$
- $\mathbf{Q} \cap \mathbf{E} = \emptyset$.

2.2 Method 1: Bayesian Network Inference

2.2.1 Choosing an Elimination Ordering

2.2.2 Heuristics for Elimination Ordering

Definition: Choose the elimination ordering to minimize the elimination width using the following heuristics:
1. Eliminate variable with the fewest parents.
2. Eliminate variable with the smallest domain for its parents, where

$$|\operatorname{dom}(\mathrm{pts}(X))| = \prod_{Z \in \mathrm{pnt}(X)} |\operatorname{dom}(Z)|.$$

3. Eliminate variable with the smallest Markov blanket.
4. Eliminate variable with the smallest domain for its Markov blanket, where

$$|\operatorname{dom}(\mathrm{mbk}(X))| = \prod_{Z \in \mathrm{embk}(X)} |\operatorname{dom}(Z)|.$$

2.3 Method 2: Inference via Sampling

Definition: Generate a large # of samples and then approximate as:

$$p(\mathbf{Q} \mid \mathbf{E}) \approx \frac{\text{\# of samples w/ } \mathbf{Q} \text{ and } \mathbf{E}}{\text{\# of samples w/ } \mathbf{E}}.$$

- As # of samples $\to \infty$, the approximation becomes exact.

2.3.1 Inference via Sampling with Likelihood Weighting

Motivation: Most of the samples are wasted since they are not consistent with the evidence.

Definition: Generate a large # of samples and then approximate as:

$$p(\mathbf{Q} \mid \mathbf{E}) \approx \frac{\text{weight of samples w/ } \mathbf{Q} \text{ and } \mathbf{E}}{\text{weight of samples w/ } \mathbf{E}}.$$

- Weight for each sample: Probability of forcing the evidence, i.e. probability of the evidence given the sample.

2.4 Canonical Problems:

Example:

1. **Given:** Caveman is deciding whether to go hunt for meat. He must take into account several factors:
   - Weather
   - Possibility of over-exertion
   - Possibility encountering lion

   These factors can result in Cavemen's death. His decision will ultimately depend on the **chances** of his death.
2. **Binary Variables:**
   - $W = \{\text{Sun}, \text{Rainy}\}$: Weather
   - H: Whether the Cavemen goes hunting or not.
   - L: Whether the Cavemen encounters a lion or not.
   - T: Whether the Cavement is tired or not.
   - D: Whether the Cavemen dies or not
3. **Problem:** Cavemen must decide whether to go hunting or not.
   - He must consider the conditional probabilities (i.e. dependence) of each event.

Warning: Have to be discrete.

Process:

1.

Example:

1. **Given:** Bayesian network.

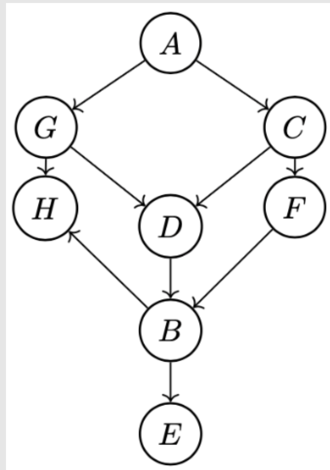

Figure 6

2. **Problem:** A and E are
    - independent if $\mathcal{K} =$
    - not necessarily independent for $\mathcal{K} =$

2.4.1 Inference via Sampling

Definition:

Process:

1.

Example: Bayesian Inference

1. **Given:**

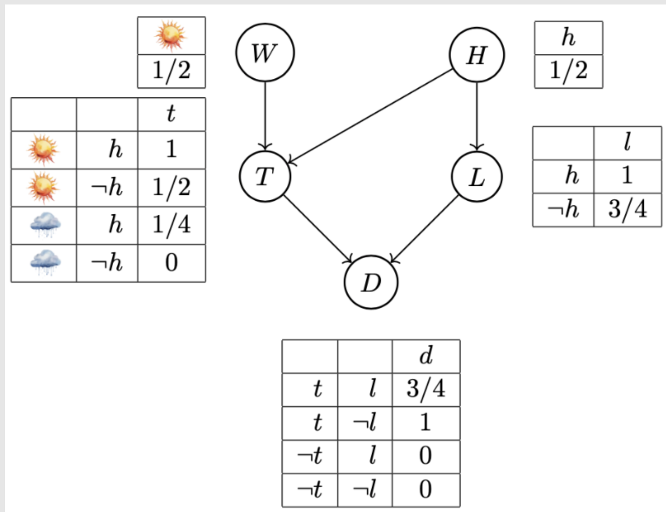

Figure 7

2. **Problem:**

Process:
1.

Example: Inference via Sampling
1. **Given:**
2. **Problem:**