Uncertainty
Modelling

Introduction to Artificial Intelligence

Chandra Gummaluru

Department of Electrical and Computer Engineering
University of Toronto

Our goal is to quantify the uncertainty in an uncertain situation. One way to do this as follows:

Definition (Relative Frequency)

Suppose we have an experiment whose possible outcomes are given by the set, Ω. sample space

The **probability** of any one outcome ω is given by

number of times ω occurs

$$\Pr\{\omega\} := \lim_{N\to\infty} \frac{N(\omega)}{N}.$$

number of trials

probability of ω

Definition (Event)

An **event**, $A \subseteq \Omega$ is a subset of outcomes. A occurs if any one of its outcomes, $\omega \in A$, occurs.

The set of all events is the power-set of the sample space, i.e., 2^Ω.

Definition (Marginal Probabilities)

The **marginal probability** of the event $A \subseteq \Omega$ is the sum of the probabilities of each outcome $\omega \in A$, i.e.,

$$\Pr\{A\} = \sum_{\omega \in A} \Pr\{\omega\}.$$

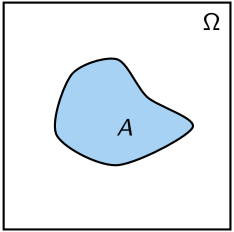


Figure 1: The probability of an event, $A \subset \Omega$, is the area taken up by A, relative to that of Ω.

Definition (Joint Probabilities)

The **joint probability** of the events A and B is the probability of the event $A \cap B$, i.e.,

$$\Pr\{A \text{ and } B\} = \sum_{\omega \in A \cap B} \Pr\{\omega\}$$

In general, we cannot determine $\Pr\{A \cap B\}$ from $\Pr\{A\}$ and $\Pr\{B\}$.

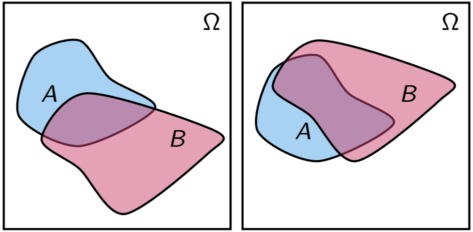


Figure 2: The area of A and B are the same in both cases, but that of $A \cap B$ is different.

Definition (Partition)

A **partition** of Ω, is a disjoint set of events, $B_1, \ldots, B_n$, such that $\bigcup_{i=1}^{n} B_i = \Omega$.

Theorem (Marginalization)

For any event A and partition, $B_1, \ldots, B_n$,

$$\Pr\{A\} = \sum_{i=1}^{n} \Pr\{A \text{ and } B_i\},$$

This is called **marginalization**.

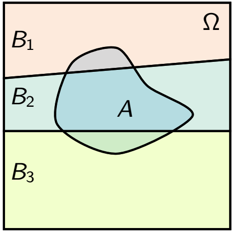


Figure 3: $A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$

Definition (Conditional Probabilities)

The **conditional probability** of the event A given that the event B has occurred is the probability of A relative to B, i.e.,

$$\Pr\{A \text{ given } B\} = \frac{\Pr\{A \text{ and } B\}}{\Pr\{B\}}$$

Of course, repeating an experiment infinitely many times to compute probabilities is not actually possible, and so, we must assign the values ourselves.

The exact values are irrelevant, but they must satisfy a few axioms.

Definition (Probability Measure)

A **probability measure** is a function, $P : 2^{\Omega} \to [0, 1]$, such that for each event, A,

$$\Pr\{A\} := P(A),$$

and that is:

1. **non-negative**: $P(A) \geq 0$, for any event, A
2. **additive**: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \varnothing$
3. **normalized**: $P(\Omega) = 1, P(\varnothing) = 0$
4. **complimentary**: $P(\neg A) = 1 - P(A)$

We call $(\Omega, 2^{\Omega}, P)$ a **probability space**.

We would like to know exactly how the knowledge of an event influences the probability of another.

Theorem (Bayes' Theorem)

For any two events, A and B,

$$\underbrace{P(A|B)}_{\text{posterior of } A \text{ given } B} = \underbrace{P(A)}_{\text{prior of } A}\overbrace{\frac{P(B|A)}{P(B)}}^{\text{factor by which } B \text{ reinforces } A},$$

provided $P(B) \neq 0$.

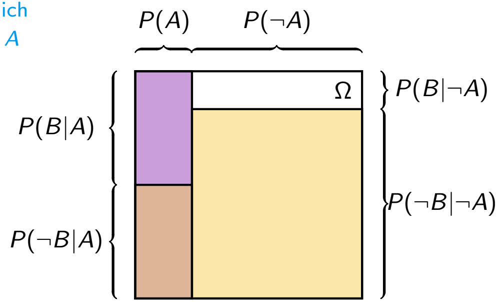


Figure 4: A visual representation of the various conditional probabilities of B given A.

Proof :
By definition,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

But also, by definition, $P(A \cap B) = P(A)P(B|A)$. Therefore,

$$P(A|B) = P(A)\frac{P(B|A)}{P(B)}.$$

■

We need a systematic way to represent dependence relationships.

Definition (Bayesian Network)

A **Bayesian network** is a directed acyclic graph, $(\mathcal{V}, \mathcal{E})$, in which the vertices represent the related random variables, and the edges represent the dependence relationships between them:

- If $(V_1, V_2) \in \mathcal{E}$, where $V_1, V_2 \in \mathcal{V}$, then $V_1 \not\perp V_2$ (but the converse need not be true).
- The directions of the arcs is technically irrelevant since dependence is commutative, but they typically indicate the direction of causality.

Figure 5: Both networks above indicate that X_1 and X_2 are dependent, but the left suggests that X_1 causes X_2, while the right suggests that X_2 causes X_1.

Given a Bayesian network, $\mathcal{B} = (\mathcal{V}, \mathcal{E})$, our task is to determine if $V_1 \perp V_2 | \mathcal{K}$, where $\mathcal{K} \subset \mathcal{V} \setminus \{V_1, V_2\}$.

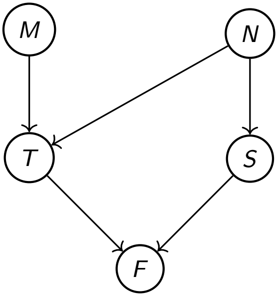
Example (Modelling Uncertain Situations)

The variables of interest are:

- whether we catch the flight or not, F
- when we get to the airport, T
- how long it takes to get through security, S
- how we get to the airport, M
- how many bags we have, N

The causal relationships are that:

- F is directly influenced by T and S
- T is directly influenced by M and N
- S is directly influenced by N



Our goal now is to determine whether any two variables in a Bayesian network are dependent or not.

We first consider the simplest non-trivial case.

Definition (Junction Network)

A **junction**, $\mathcal{J} = (\{X_1, X_2, X_3\}, \{(e_1, e_2)\})$ is a Bayesian network that consists of three random variables, X_1, X_2, X_3, connected by two arcs, e_1 and e_2.

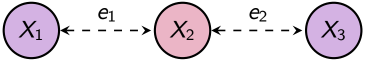


Figure 6: There are three distinct types of junctions.

We call X_1 and X_3, the **outer variables**, and X_2 the **central variable**.

In a junction network, the central variable and each outer variable are always dependent.

We want to know when the outer variables are dependent.

Definition (Causal Chain Junction)

In a **causal chain** junction, X_1 directly influences X_2, which in turn, directly influences X_3.

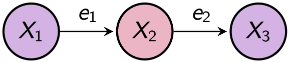


Figure 7: A causal chain junction.

In this case, $X_1 \not\perp X_3$ and $X_1 \perp X_3 | X_2$.

We say that $\mathcal{J}$ is closed given X_2, and open given no evidence.

Example (Causal Chain Relationship)

Rain causes the grass to get wet, which then causes worms to come out.

Definition (Common Cause Junction)

In a **common cause** junction, X_2 directly influences X_1 and X_3.

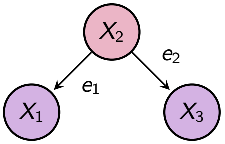


Figure 8: A common cause junction.

In this case, $X_1 \not\perp X_3$ and $X_1 \perp X_3 | X_2$.

We say $\mathcal{J}$ is closed given X_2, and open given no evidence.

Example (Common Cause Relationship)

Rain causes the grass to get wet, and the road to be slippery.

Definition (Common Effect Junction)

In a **common effect** junction, X_1 and X_3 both directly influence X_2.

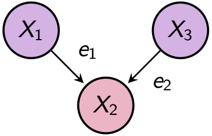


Figure 9: A common effect junction.

In this case, $X_1 \perp X_3$ and $X_1 \perp X_3 | K$, where $K \subseteq \{X_2\} \cup \mathrm{des}(X_2)$.

We say $\mathcal{J}$ is open given X_2 and/or any of its descendants, and closed otherwise.

Example (Common Effect Relationship)

Both rain and watering the garden can cause the grass to be wet.

Theorem (Dependence Separation)

Let $\mathcal{B} = (\mathcal{V}, \mathcal{E})$ be some Bayesian network and $\mathcal{K} \subseteq \mathcal{V}$.

Let a path, p, in $\mathcal{B}$ be an ordered list of junctions,

$$\mathcal{J}^{(i)} = \left(\left\{ X_1^{(i)}, X_2^{(i)}, X_3^{(i)} \right\}, \left\{ (e_1^{(i)}, e_2^{(i)}) \right\} \right)$$

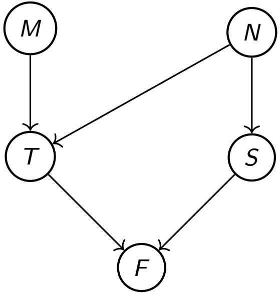
such that $X_1^{(i+1)} = X_2^{(i)}$ and $X_2^{(i+1)} = X_3^{(i)}$, or equivalently, $e_1^{(i+1)} = e_2^{(i)}$.

The path, p, is **blocked** under K if $\mathcal{J}^{(i)}$ is blocked under $\mathcal{K}$ for some i.

For any $V_1, V_2 \in \mathcal{V}$ and $\mathcal{K} \subseteq \mathcal{V} \setminus \{V_1, V_2\}$, we have $V_1 V_2 | \mathcal{K}$ if and only if every path between V_1 and V_2 is blocked under $\mathcal{K}$.

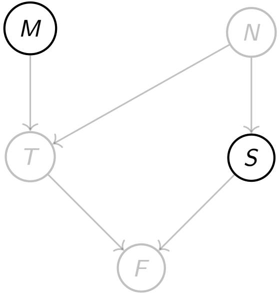
Example (Dependence Separation)

Determine whether S and M are independent or not, given no evidence.



Example (Dependence Separation)

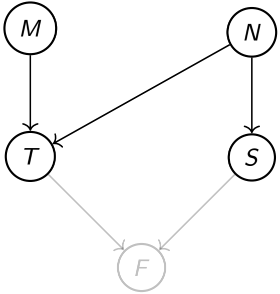
Suppose we want to know whether S and M are independent or not given no evidence.



There are two paths from S to M; $S \perp M$ if and only if both are closed.

Example (Dependence Separation)

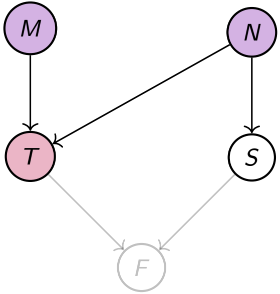
Suppose we want to know whether S and M are independent or not given no evidence.



One path from M to S is $p_1 = (\{M, T, N\}, \{T, N, S\})$.

Example (Dependence Separation)

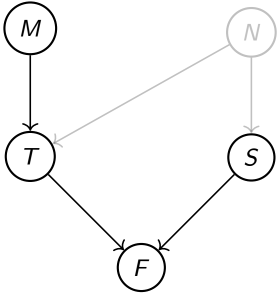
Suppose we want to know whether S and M are independent or not given no evidence.



$\{M, T, N\}$ is a common effect junction and is **closed** given no evidence; thus p_1 is **closed**.

Example (Dependence Separation)

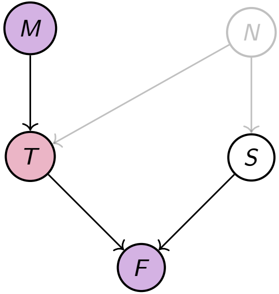
Suppose we want to know whether S and M are independent or not given no evidence.



The other path from M to S is $p_2 = (\{M, T, F\}, \{T, F, S\})$.

Example (Dependence Separation)

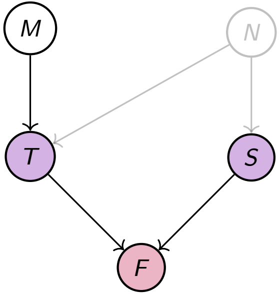
Suppose we want to know whether S and M are independent or not given no evidence.



$\{M, T, F\}$ is a causal chain junction and is **open** given no evidence.

Example (Dependence Separation)

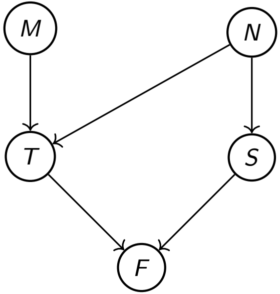
Suppose we want to know whether S and M are independent or not given no evidence.



$\{T, F, S\}$ is a common effect junction and is **closed** given no evidence; thus p_2 is closed.

Example (Dependence Separation)

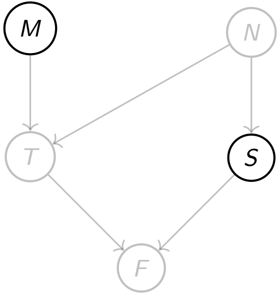
Suppose we want to know whether S and M are independent or not given no evidence.



Since p_1 and p_2 are both closed, it follows that $S \perp M$.

Example (Dependence Separation)

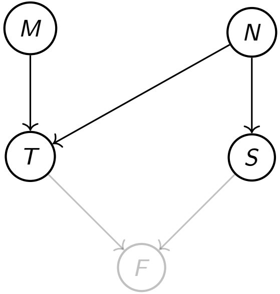
Suppose we want to know whether S and M are independent or not given F.



There are two paths from S to M; $S \perp M|F$ if and only if both are closed given F.

Example (Dependence Separation)

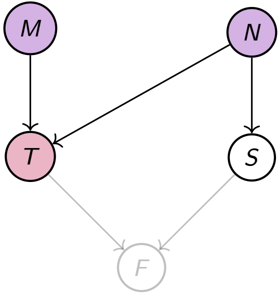
Suppose we want to know whether S and M are independent or not given F.



One path from M to S is $p_1 = (\{M, T, N\}, \{T, N, S\})$.

Example (Dependence Separation)

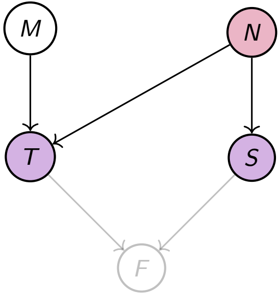
Suppose we want to know whether S and M are independent or not given F.



$\{M, T, N\}$ is a common effect junction and is **open** given F.

Example (Dependence Separation)

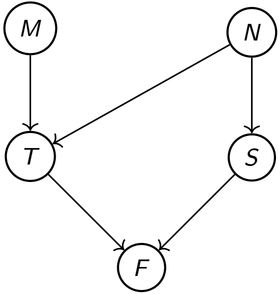
Suppose we want to know whether S and M are independent or not given F.



$\{T, N, S\}$ is a common cause junction and is **open** given F; thus p_1 is open.

Example (Dependence Separation)

Suppose we want to know whether S and M are independent or not given no evidence.



Since p_1 is open, it follows that $S \not\perp M | F$.

Dependence separation allows us to prove an even more important result.

Theorem (Independence Relationships)

If X is a variable in a Bayesian network, $(\mathcal{V}, \mathcal{E})$, then

$$P(X|\mathcal{S} \cup \text{pts}(X)) = P(X|\text{pts}(X))$$

X is independent of its non-descendants given its parents.

where $\mathcal{S} \subseteq \mathcal{V} \setminus \text{des}(X)$.

Proof :
Observe that every path from X to any of its *non-descendants* must contain either:

- a causal chain junction through one of X's parents
- a common cause junction through one of X's parents
- a common effect junction through one of X's descendants

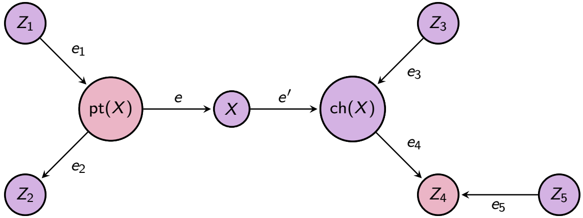


Figure 10: The various paths between X and its non-descendants.

Thus, all paths to X from any of its non-descendants are blocked given pts(X). ■