Why probability. Reason about uncertainty.

\[ p(\text{wall} = 2m \mid \text{sonar} = 1.2m) = ? \]
\[ p(\text{cat} = \text{true} \mid \text{whisker} = \text{true}, \text{fur} = \text{true}, \text{legs} = 4) = ? \]

- Conditional probability has a weakness.

  We require the "probability measure."

Suppose \( X_0, X_1, X_2, \ldots, X_n \) boolean. \( 2^n \) worlds

It's impractical to store complete probability measure

Idea: given a variable \( X \) there is usually few variables that affect \( X \) let's call them \( Zs \)

Any other variable is irrelevant to knowing \( X \) if we are given \( Zs \)

Formally: \( p(X \mid Zs) = p(X \mid Y, Zs) \) Independence

\( X \) is conditionally independent of \( Y \) given \( Zs \)

\( \text{e.i. } x \in \text{domain}(X), y, y' \in \text{domain}(Y), z \in \text{domain}(Z) \)

\[ p(X = x \mid Z = z) = p(X = x \mid Y = y \land Z = z) \]
\[ = p(X = x \mid Y = y' \land Z = z) \]

Side Note: \( X \) and \( Y \) are unconditionally independent when:

\[ p(X, Y) = p(X) \cdot p(Y) \]

\[ \prod^n p(X_i \mid X_1, \ldots, X_{i-1}) = \prod^n p(X_i \mid X_1, \ldots, X_{i-1}) \]
Why:

\[
P(X_0, X_1, X_2, X_3 \ldots X_n) = \prod_{i=0}^{n} P(X_i | X_1, \ldots, X_{i-1})
\]
by the chain rule.

Joint probability distribution.

• Suppose for variable \( X \), \( X \) depends on \( \text{parents}(X) \).

• Let's order the variables so that for every \( X \), \( \text{parents}(X) \) are predecessors of \( X \).

\[
X_j \quad \text{parents}(X_j) \subseteq (X_0, X_1, \ldots, X_{j-1})
\]

\[
P(X_0, X_1, X_2, \ldots, X_n) = \prod_{i=0}^{n} P(X_i | \text{parents}(X_i))
\]

Example: Domain of student Bob:

Bob attends class, Bob reviews his notes, Bob answers to the test, Bob's Grades.

Variables: \( A, R, T, G \)

\[
P(A, R, T, G)
\]

- \( A \) is independent \( \quad \) \( \text{parents}(A) = \{ \} \)
- \( R \) is independent \( \quad \) \( \text{parents}(R) = \{ \} \)
- \( T \) is dependent on \( A \) and \( R \) \( \quad \) \( \text{parents}(T) = \{ A, R \} \)
- \( G \) is dependent on \( T \) \( \quad \) \( \text{parents}(G) = \{ T \} \)

\[
P(A, R, T, G) = P(A) \cdot P(R | A) \cdot P(T | A, R) \cdot P(G | A, R, T)
\]

\[
= P(A) \cdot P(R) \cdot P(T | A, R) \cdot P(G | T) \quad \text{easier to store and to obtain.}
\]

\[
P(A \mid \text{Grade} = A)
\]

\[
P(R \mid \text{Grade} = A \land A = \text{Low})
\]

Bayesian Network / Belief Network:

- A Directed Acyclic Graph: represent our assumptions of variable dependency
• A Directed Acyclic Graph: represent our assumptions of variable dependency
  - Nodes are labeled by variables
  - There is an arc from every member of parents (X) to X
  - A domain for each variable.
  - A set of conditional probability tables,
    \( p(X \mid \text{parents}(X)) \)

Example: Alexa, Google Home, "flu-app"

- Listen to "Achoo" sound: depends on sneezing.
- Sneezing could be cause by allergies and flu.
- The flu causes fever.
- Allergies are dependent on pollen.
- Both pollen and flu are dependent on seasons.

Bayesian Network:

\[
P(S, F, V, L, A, Z, U) = \prod \begin{cases} p(S) & p(V \mid F) \\ p(L \mid S) & p(A \mid L) \\ p(F \mid S) & p(Z \mid A, F) \\ p(U \mid Z) \end{cases}
\]

Computing new probabilities
\[ \equiv \text{Probabilistic Inference} \equiv \]

\[ P(\text{flu} = \text{true} \mid \text{sneeze} = \text{true}) = ? \]
\[ P(\text{season} = \text{winter} \mid \text{pollen} = \text{true}) = ? \]
\[ P(\text{flu} = \text{true} \mid \text{fever} = \text{false} \land \text{allergies} = \text{false}) = ? \]
Example #2: Smart house:
- Fire alarm: external camera;
  - Fire alarm can be tampered, ring when there is a fire
  - Fire produces smoke
  - When fire alarm rings, people leave the building.
  - When people leave the building, you see people in the doorbell camera.

Belief/Bayesian Network:

\[
\begin{align*}
\text{P(Tampering)} &= 0.02 \\
\text{P(Fire)} &= 0.01 \\
\text{P(Alarm|Tampering, Fire)} &= \begin{bmatrix} 0.5 & 0.85 \\ 0.99 & 0.0001 \end{bmatrix} \\
\text{P(Smoke|Fire)} &= 0.9 \\
\text{P(Leaving|Alarm)} &= 0.88 \\
\text{P(Crowd|Leaving)} &= 0.78
\end{align*}
\]

The camera detects a crowd at the door: Crowd = true

\[
\begin{align*}
\text{P(fire|Crowd)} &= 0.23 \\
\text{P(Tampering|Crowd)} &= 0.39 \\
\text{P(Smoke|Crowd)} &= 0.21
\end{align*}
\]

- Suppose Smoke = true
  \[
  \begin{align*}
  \text{P(fire|Smoke)} &= 0.476 \\
  \text{P(Tampering|Smoke)} &= 0.02
  \end{align*}
  \]

- Suppose Crowd = true but Smoke = false
  \[
  \begin{align*}
  \text{P(fire|crowd=true \land smoke=false)} &= 0.029 \\
  \text{P(Tampering|crowd=true \land smoke=false)} &= 0.50
  \end{align*}
  \]

Effects of observation on a belief network.
Effects of observation on a belief network:

- Observe a variable \( Y \); which probabilities change:
  - The descendants of \( Y \) change
  - The ancestors of \( Y \) change.