CSIT5900
Reasoning Under Uncertainty

Fangzhen Lin

Department of Computer Science and Engineering
Hong Kong University of Science and Technology

Uncertainty

- Ordinary commonsense knowledge quickly moves away from categorical absolute statements like "lawyers are always rich:"

$$\forall x \; Lawyer(x) \Rightarrow Rich(x).$$

  A lawyer might not be rich because of
  - lack of customers
  - bankrupted business
  - addiction to gambling
  - frequent divorces
  - ...
- There are many ways in which we can come to less than categorical information:
  - things *usually (occasionally, seldomly)* are in a certain way.
  - fuzzy judgments e.g. barely rich, a poor example of chair, not very tall, etc.
  - imprecision of sensors.
  - reliability of sources of information, e.g. "most of time he's right".

Probabilities

- There are many possible ways to address the difficulties, a reasonable first resort would be to use *probabilities*: quantifying things using probability
  - The probability that John is rich is 0.1.
  - The probability that John, who is a lawyer, is rich is 0.8.

Language

- use an extension of propositional logic to represent facts.
- If A is a proposition standing for "John is rich", then $P(A) = 0.3$ means that there is a 0.3 chance that John is rich, and $P(\neg A) = 0.7$ means that there is a 0.7 chance that John is not rich.
- If B is a proposition standing for "John is a miser", then $P(A \wedge B) = 0.2$ means that there is a 0.2 chance that John is both rich and a miser.
- A random variable is a variable with a set of possible values.
- If X is a random variable, and v is one of its possible values, then $P(X = v)$ is the probability that X is equals to v:

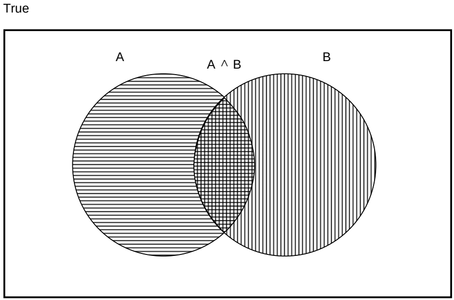
$$P(Weather = Sunny) = 0.7,$$
$$P(Weather = Rain) = 0.2,$$
$$P(Weather = Snow) = 0.1$$

Axioms for Probability

- All probabilities are between 0 and 1: $0 \leq P(A) \leq 1$
- Necessarily true propositions have probability 1: $P(True) = 1$
- Necessarily false propositions have probability 0: $P(False) = 0$
- The probability of a disjunction:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

Set-theoretic view:



Prior and Posterior Probabilities

- *Before* some evidence is obtained, the probability assessment of a proposition gives the *prior probability* (or unconditional probability).
- *After* the evidence is obtained, the probability assessment gives the *posterior probability* (or conditional probability).
- Prior probability:
  - $P(A)$ denotes the probability that proposition A is true.
  - Example:
    $P(\mathit{Tomorrow} = \mathit{Rainy}) = 0.1$
- Posterior probability:
  - $P(A \mid B)$ denotes the probability that proposition A is true given that proposition B is true.
  - Example:
    $P(\mathit{Tomorrow} = \mathit{Rainy} \mid \mathit{Today} = \mathit{Rainy}) = 0.7$

Product Rule

- Product rule (chain rule):

$$P(A \wedge B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

  Thus if $P(A), P(B) > 0$, then

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)} \qquad P(B \mid A) = \frac{P(B \wedge A)}{P(A)}$$

- More general form for multivalued random variables:

$$\begin{aligned} P(X = x_i, Y = y_j) &= P(X = x_i \mid Y = y_j)P(Y = y_j) \\ &= P(Y = y_j \mid X = x_i)P(X = x_i) \end{aligned}$$

  Or simply $P(X, Y) = P(X \mid Y)P(Y) = P(Y \mid X)P(X)$.

- The above equations express conditional probabilities in terms of joint probabilities, which are hard to come by.
- Modern probabilistic reasoning systems try to sidestep joint probabilities and work directly with conditional probabilities.

Bayes' Rule

- Bayes' rule:

$$P(A \mid E) \quad = \quad \frac{P(E \mid A)P(A)}{P(E)}$$

- More general form for multivalued random variables:

$$P(Y = y_j \mid X = x_i) \quad = \quad \frac{P(X = x_i \mid Y = y_j)P(Y = y_j)}{P(X = x_i)}$$

  Or simply,

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

- Conditioning on some background evidence E:

$$P(Y \mid X, E) = \frac{P(X \mid Y, E)P(Y \mid E)}{P(X \mid E)}$$

Medical Diagnosis Example

- Given:

$$P(Cough \mid Pneumonia) = 0.8$$
$$P(Pneumonia) = 0.005$$
$$P(Cough) = 0.05$$

  Find: $P(Pneumonia \mid Cough)$

- Applying Bayes' rule:

$$\begin{aligned} P(Pneumonia \mid Cough) &= \frac{P(Cough \mid Pneumonia)P(Pneumonia)}{P(Cough)} \\ &= \frac{0.8 \times 0.005}{0.05} = 0.08. \end{aligned}$$

- $P(Pneumonia \mid Cough)$, as diagnostic knowledge, is usually not directly available in domain knowledge. Instead, $P(Cough|Pneumonia)$, as causal knowledge, is the more commonly available form.

Combining Evidence

- Find: $P(A \mid E_1, E_2)$
- One approach: Absorbing all evidence at once.

$$P(A \mid E_1, E_2) = \frac{P(E_1, E_2 \mid A)P(A)}{P(E_1, E_2)}$$

- Another approach: Absorbing evidence one piece at a time.
  - When only E_1 is available:

$$P(A \mid E_1) = P(A)\,\frac{P(E_1 \mid A)}{P(E_1)}$$

  - When E_2 also becomes available:

$$\begin{aligned} P(A \mid E_1, E_2) &= P(A \mid E_1)\,\frac{P(E_2 \mid E_1, A)}{P(E_2 \mid E_1)} \\ &= P(A)\,\frac{P(E_1 \mid A)}{P(E_1)}\,\frac{P(E_2 \mid E_1, A)}{P(E_2 \mid E_1)} \qquad (1) \end{aligned}$$

  - This process is order independent.

Conditional Independence

- Estimating $P(E_1, E_2 \mid A)$ or $P(E_2 \mid E_1, A)$ may be computationally unattractive, especially when many pieces of evidence have to be combined.
- Conditional independence can help.
- Independence:
  - Two random variables X and Y are *independent* if
    - ⋆ $P(X|Y) = P(X)$, or equivalently
    - ⋆ $P(Y|X) = P(Y)$.
  - Knowledge about one variable contains no information about the other.
- Conditional independence:
  - Two random variables X and Y are *conditionally independent* given a third variable Z if
    - ⋆ $P(X|Y, Z) = P(X|Z)$, or equivalently
    - ⋆ $P(Y|X, Z) = P(Y|Z)$.
  - Given background knowledge Z, knowledge about one variable contains no information about the other.

Making Use of Conditional Independence

- If E_1 and E_2 are conditionally independent given A, then

$$P(A \mid E_1, E_2) = \frac{P(E_1, E_2 \mid A)P(A)}{P(E_1, E_2)} = \frac{P(E_1 \mid A)P(E_2 \mid A)P(A)}{P(E_1, E_2)}$$

or

$$\begin{aligned} P(A \mid E_1, E_2) &= P(A)\,\frac{P(E_1 \mid A)}{P(E_1)}\,\frac{P(E_2 \mid E_1, A)}{P(E_2 \mid E_1)} \\ &= P(A)\,\frac{P(E_1 \mid A)P(E_2 \mid A)}{P(E_1, E_2)}. \end{aligned}$$

- $\frac{1}{P(E_1,E_2)}$ is the *normalization constant* and

$$\begin{aligned} P(E_1, E_2) &= P(E_1, E_2 \mid A)P(A) + \\ &\quad P(E_1, E_2 \mid \neg A)P(\neg A) \\ &= P(E_1 \mid A)P(E_2 \mid A)P(A) + \\ &\quad P(E_1 \mid \neg A)P(E_2 \mid \neg A)P(\neg A) \end{aligned}$$

Reasoning with Joint Probability

Example: Alarm

- Story: In LA, burglary and earthquake are not uncommon. They both can cause alarm. In case of alarm, two neighbors John and Mary may call.
- Variables: Burglary (B), Earthquake (E), Alarm (A), JohnCalls (J), MaryCalls (M).
- Problem: Estimate the probability of a burglary based on who has or has not called.
- Problem can be solved if *joint probability* is available.

$$P(B, E, A, J, M)$$

Reasoning with Joint Probability

B	E	A	J	M	Prob	B	E	A	J	M	Prob
y	y	y	y	y	.00001	n	y	y	y	y	.0002
y	y	y	y	n	.000025	n	y	y	y	n	.0004
y	y	y	n	y	.000025	n	y	y	n	y	.0004
y	y	y	n	n	.00000	n	y	y	n	n	.0002
y	y	n	y	y	.00001	n	y	n	y	y	.0002
y	y	n	y	n	.000015	n	y	n	y	n	.0002
y	y	n	n	y	.000015	n	y	n	n	y	.0002
y	y	n	n	n	.0000	n	y	n	n	n	.0002
y	n	y	y	y	.00001	n	n	y	y	y	.0001
y	n	y	y	n	.000025	n	n	y	y	n	.0002
y	n	y	n	y	.000025	n	n	y	n	y	.0002
y	n	y	n	n	.0000	n	n	y	n	n	.0001
y	n	n	y	y	.00001	n	n	n	y	y	.0001
y	n	n	y	n	.00001	n	n	n	y	n	.0001
y	n	n	n	y	.00001	n	n	n	n	y	.0001
y	n	n	n	n	.00000	n	n	n	n	n	.996

Reasoning with Joint Probability

- $P(B{=}y|M{=}y)$?
- Compute *marginal probability*:

$$P(B,M) = \sum_{E,A,J} P(B,E,A,J,M)$$

B	M	Prob
y	y	.000115
y	n	.000075
n	y	.00015
n	n	.99971

-

$$\begin{aligned} P(B{=}y|M{=}y) &= \frac{P(B{=}y, M{=}y)}{P(M{=}y)} \\ &= \frac{.000115}{.000115 + .00015} = 0.43 \end{aligned}$$

Reasoning with Joint Probability

Advantages:

- Clear semantics
- In theory, can perform arbitrary inference among the variables.

Difficulties: Complexity

- In the Alarm example:
  - 31 numbers needed,
  - quite unnatural to assess: e.g.

$$P(B = y, E = y, A = y, J = y, M = y)$$

  - Many additions in inference.
- In general, $P(X_1, X_2, \ldots, X_n)$ needs at least $2^n - 1$ numbers to specify the joint probabilities.
  - Knowledge acquisition (complex, unnatural)
  - Storage
  - Inference

Reasoning with Joint Probability

Solution:

- Use product rule:

$$
\begin{aligned}
P(B,E,A,J,M) &= P(B,E,A,J)P(M|B,E,A,J) \\
&= \dots \\
&= P(B)P(E|B)P(A|B,E) \\
&\quad P(J|B,E,A)P(M|B,E,A,J).
\end{aligned}
$$

- Conditional independence:
  - $P(E|B) = P(E)$.
  - $P(J|B,E,A) = P(J|A)$.
  - $P(M|B,E,A,J) = P(M|A)$.
- Thus

$$P(B,E,A,J,M) = P(B)P(E)P(A|B,E)P(J|A)P(M|A).$$

- Probabilities:

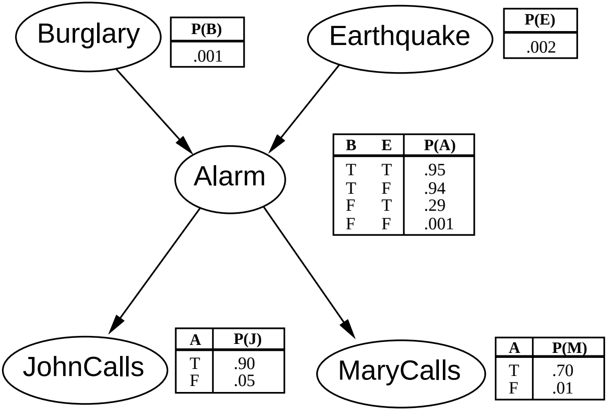
B	$P(B)$	E	$P(E)$	A	J	$P(J\|A)$	A	M	$P(M\|A)$
y	.001	y	.002	y	y	.9	y	y	.7
				n	y	.05	n	y	.01

B	E	A	$P(A\|B,E)$	B	E	A	$P(A\|B,E)$
y	y	y	.95	n	y	y	.29
y	n	y	.94	n	n	y	.001

- Only 10 numbers are needed. These numbers are natural to assess.

Belief Networks

- Draw an arc to each variable from each of its conditioning variables.



- Attach to each variable its conditional probability table (CPT).
- Result: A *belief network*.
- Also know as Bayesian networks, Bayesian belief networks, probabilistic influence diagrams.

Belief Networks

- A belief network is a directed acyclic graph (DAG):
  - Nodes: variables
  - Links: direct probabilistic dependencies between variables
  - Node attributes: conditional probability tables (CPT)
- It is a compact graphical representation of the joint probability distribution of the random variables:
  - The *global* joint probability is expressed as the product of *local* conditional probabilities:

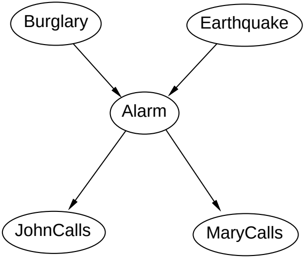
$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

    where $Parent(X_i)$ is the set of parent nodes of X_i in the network.
  - Example:

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) =$$
$$P(J|A)P(M|A)P(A|\neg B \wedge \neg E)P(\neg B)P(\neg E) =$$
$$0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062$$

Belief Networks - Qualitative Level

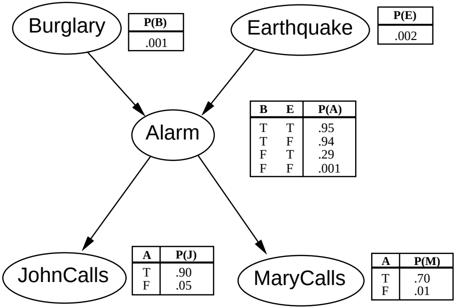
Qualitative level: Network structure



- A depends on B and E.
- J and M are conditionally independent given the event A.

Belief Networks - Numerical Level

Numerical level: Conditional probabilities



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

- The CPT for a Boolean variable with n Boolean parents contains 2^n rows.
- B and E have no parent nodes. The conditional probabilities thus degenerate to prior probabilities.

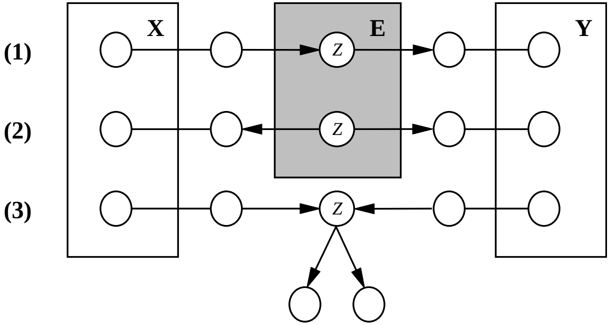
Conditional Independence in BN

- BN expresses the conditional independence of a node and its predecessors, given its parents.
- A more general question is: given any nodes X and Y, are they independent given a set of nodes E?
- If X and Y are d-separated by E, then they are conditionally independent given E.

D-Separation

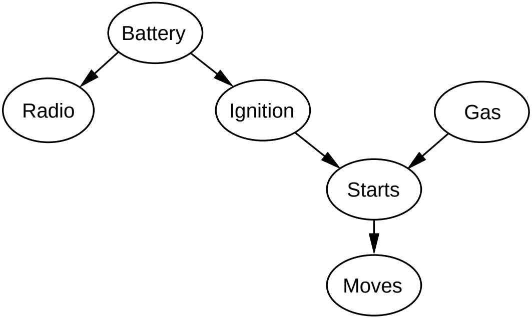
X and Y are d-separated by E if for every undirected path from X to Y, there is a node Z on the path such that one of the following conditions holds:

1. Both path arrows lead in to Z, and neither Z nor any descendants of Z are in E.
2. Z is in E, and it is not the case that both path arrows lead in to Z.



An Example of Conditional Independence

Consider the following example BN:



- *Gas* and *Radio* are conditionally independent given *Ignition*.
- *Gas* and *Radio* are conditionally independent given *Battery*.
- *Gas* and *Radio* are marginally independent (conditional independent given no evidence).
- *Gas* and *Radio* are not conditionally independent given *Starts*.

Causality and Bayesian Networks

- Arcs in a Bayesian network can usually (but not always) be interpreted as indicating cause-effect relationships. They are from causes to effects (*causal network*).
- Making use of cause-effect relationships in Bayesian network construction:
  - Choose a set of variables that describes the domain.
  - Draw an arc to a variable from each of its DIRECT causes.
  - Assess the conditional probability of each node given its parents.

Examples

1. Poole's example:

   tampering, fire, smoke, alarm, people-leaving.

2. Wet shoes

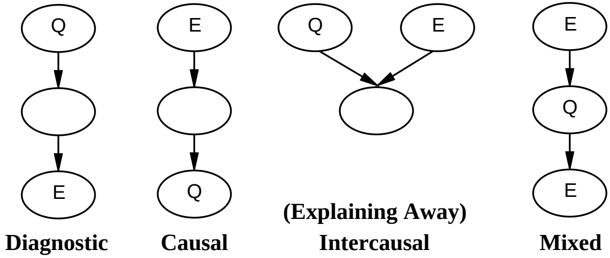
   wet-grass, rain, sprinkler, faulty-sprinkler, walking-on-grass, wet-shoe

Inference

- Any node can serve as either a *query* variable or an *evidence* variable.
- Four inference types:
  - Diagnostic inference:
    - From effects to causes.
    - Example: $P(Burglary \mid JohnCalls)$
  - Causal inference:
    - From causes to effects.
    - Example: $P(JohnCalls \mid Burglary)$
  - Intercausal inference:
    - Between causes of a common effect.
    - Example: $P(Burglary \mid Alarm, Earthquake)$
      (Even though burglaries and earthquakes are independent, the presence of one makes the other less likely.)
  - Mixed inference:
    - Combining two or more of the above.
    - Example: $P(Alarm \mid JohnCalls, \neg Earthquake)$
      (Simultaneous use of diagnostic and causal inferences)
    - Example: $P(Burglary \mid JohnCalls, \neg Earthquake)$
      (Simultaneous use of diagnostic and intercausal inferences)

Inference (cont'd)

- Examples:



- There exist efficient algorithms for all types of inference.