Artificial Intelligent Systems

(BMA-EL-IZB-LJ-RE 1. year 2024/2025)

Probabilistic Reasoning

Simon Dobrišek

Lecture Topics

- Probabilistic reasoning
- Bayesian reasoning revisited
- Bayesian networks

Probabilistic Reasoning

- Probabilistic reasoning gives probabilistic results that summarizes uncertainty from various sources.
- Sources of uncertainty are
  - Uncertain inputs (missing/noisy data)
  - Uncertain knowledge (incomplete knowledge of the causalities in the domain)
  - Uncertain outputs (abductive and inductive reasoning are inherently uncertain)

The Main Forms of Inference Revisited

Deduction: Reasons from causes to effects		
Major premise:	*All coins in the box are silver*	$P \to Q$
Minor premise:	*These coins are from the box*	P
Conclusion:	*These coins are silver*	Q

Abduction: Reasons from effects to causes		
Rule:	*All coins in the box are silver*	$P \to Q$
Observation:	*These coins are silver*	Q
Explanation:	*These coins are from the box*	Possibly P

Induction: Reasons from specific cases to general rules		
Case:	*These coins are from the box*	Whenever P
Observation:	*These coins are silver*	then Q
Hypothesized rule:	*All coins in the box are silver*	Possibly $P \to Q$

Abductive and Inductive Reasoning

- In abductive and inductive reasoning, "conclusions" are hypotheses, not theorems (may be false even if facts and/or rules are true).
- In abductive reasoning, there may be multiple plausible hypotheses, e.g.,

  Given the rules $P \to Q$ and $R \to Q$, as well as the fact Q, both P and R are plausible hypotheses.

- Hypotheses can be ranked by their plausibility (if it can be determined)
- Abduction and induction are inherently uncertain.

Hypothesize-and-Test Cycle

- Hypothesize: Postulate possible hypotheses, any of which would explain the given facts (or at least most of the important facts).
- Test: Test the plausibility of all or some of these hypotheses.
- One way to test a hypothesis H is to ask whether something that is currently unknown – but can be predicted from H – is actually true.
- Support for different hypotheses is increased or decreased if the predicted fact is actually true.

Non-Monotonicity of Reasoning

- The plausibility of hypotheses in abductive and inductive reasoning can thus increase/decrease as new facts are collected.
- We say that such a reasoning is non-monotonic.
- In contrast, deductive inference is monotonic: it never changes a sentence's truth value, once known.
- In abductive and inductive reasoning, some hypotheses may be discarded, and new ones formed, when new observations are made.

Decision Making With Uncertainty

- Decision making that is based on rational behavior follow the following rules.
  - For each possible action, identify the possible outcomes.
  - Compute the probability of each outcome
  - Compute the utility of each outcome
  - Compute the probability-weighted (expected) utility over possible outcomes for each action
  - Select the action with the highest expected utility (principle of Maximum Expected Utility).

Bayesian Reasoning Revisited

- Bayesian reasoning is an application of probability theory to inductive and abductive reasoning.
- It relies on an interpretation of probabilities as expressions of an agent's uncertainty about the world.
- Other reasoning schemes that deal with uncertainties are Default reasoning (Nonmonotonic logic), Evidential reasoning (Dempster-Shafer theory), as well as Fuzzy reasoning.
- Bayesian inference uses probability theory and information about independence, as well as reasoning from evidence to conclusions or from causes to effects.

Generalizing Modus Ponens

- The common rule of inference

$$\begin{array}{l} P \rightarrow Q \\ \underline{P \qquad\quad} \\ Q \end{array}$$

can be generalized to Bayes' rule

$$\begin{array}{l} \text{If } P \text{ then sometimes } Q \\ \underline{P \qquad\qquad\qquad\qquad\qquad} \\ \text{Maybe } Q \end{array}$$

- Bayes' rule lets us calculate the probability of Q, taking P into account.

Probability Theory

- The concepts of conditional probability, product rule and marginalization (summing out) are used in Bayesian inference.

$$P(A \land B) = P(A)\,P(B|A) = P(B)\,P(A|B)$$

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

$$P(A) = P(A \land B) + P(A \land \neg B)$$

$$P(A) = \sum_i P(B_i)P(A|B_i)$$

Examples of Bayesian Inferences

- P(alarm) = 0.19
  P(burglary | alarm) = 0.47
  P(alarm | burglary) = 0.9

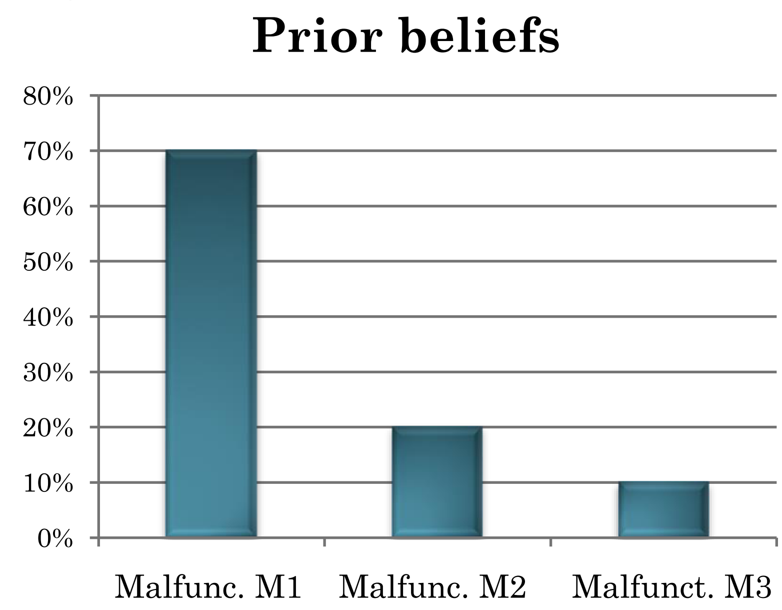
- P(burglary ∧ alarm) =
  P(alarm) P(burglary | alarm) = 0.19 * 0.47 = 0.09

- P(burglary | alarm) =
  P(burglary ∧ alarm) / P(alarm)
  = 0.09 / 0.19 = 0.47

- P(alarm) =
  P(alarm ∧ burglary) +
  P(alarm ∧ ¬burglary) = 0.09 + 0.1 = 0.19

An example of Bayesian inference Revisited

- A technician deals with a car engine that has some unusual symptoms that reflect in unusual sounds.
- Technician's prior beliefs are, that 70% of car engines with such symptoms has a malfunction M_1, 20 % a malfunction M_2 and 10 % a malfunction M_3.
- Technician's beliefs about the malfunction can be illustrated by the diagram on the right.



An example of Bayesian inference Revisited

- The technician captures sound samples of the engine and analyses them using a special diagnostics application.

- In the specifications of the application, it is stated that 60% of the engines with the malfunction M_1, 90% of the engines with the malfunction M_2 and 100% of the engines with the malfunction M_3 cause a positive test result.

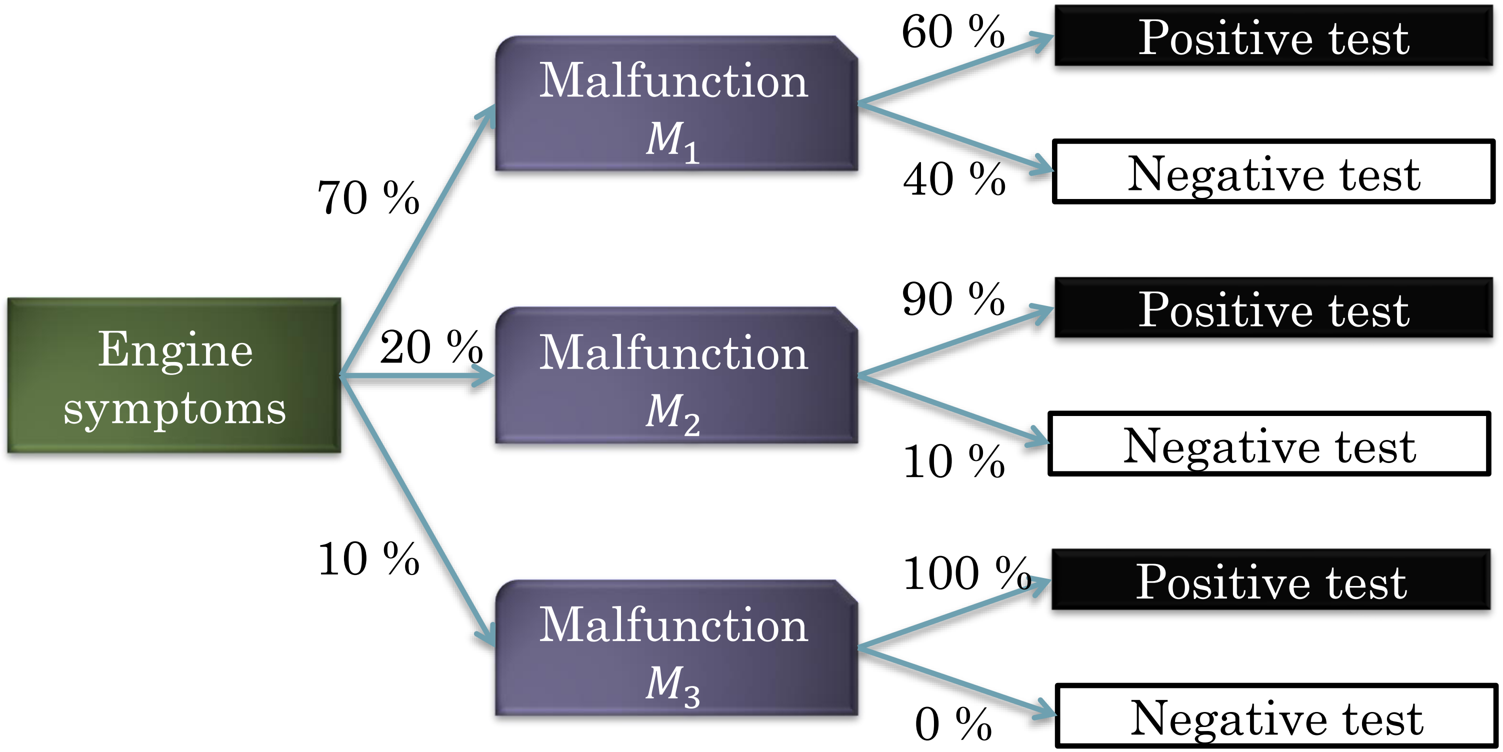
$$P(E|M_1) = 0{,}6 \qquad P(E|M_2) = 0{,}9 \qquad P(E|M_3) = 1{,}0$$

- Let as assume that the application gave the positive test result (evidence E).

- What are now technician's beliefs in what malfunction is causing the unusual engine symptoms.

An example of Bayesian inference Revisited

- A tree with possible causes and facts with their probabilities



An example of Bayesian inference Revisited

- A joint probability that we have a technician dealing with an engine that has one of the malfunctions (M_1, M_2, M_3) and the test is positive (E), is thus,

$$P(\mathrm{E},\mathrm{M}_1) = P(M_1)\cdot P(E|M_1) = 0{,}7\cdot 0{,}6 = 0{,}42$$

$$P(\mathrm{E},\mathrm{M}_2) = P(M_2)\cdot P(E|M_2) = 0{,}2\cdot 0{,}9 = 0{,}18$$

$$P(\mathrm{E},\mathrm{M}_3) = P(M_3)\cdot P(E|M_3) = 0{,}1\cdot 1{,}0 = 0{,}1$$

- Technicians' prior beliefs after the positive test thus changes to the posterior beliefs

$$P(E) = \sum_i P(M_i)P(E|M_i) = \\ = 0{,}42 + 0{,}18 + 0{,}1 = 0{,}7$$

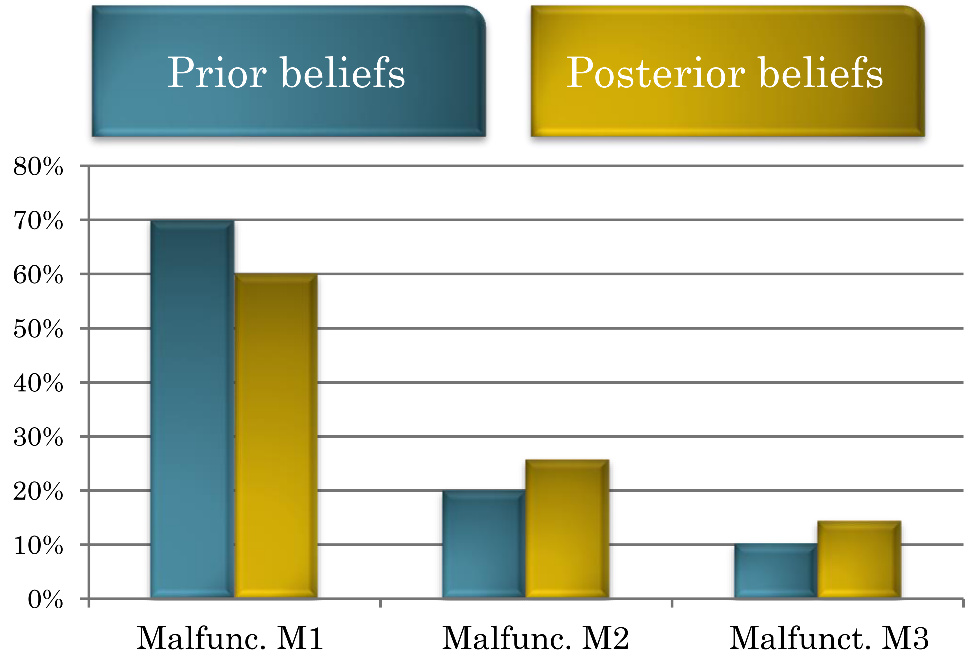
$$P(M_1|E) = \frac{0{,}42}{0{,}42+0{,}18+0{,}1} = 0{,}6$$

$$P(M_2|E) = \frac{0{,}18}{0{,}42+0{,}18+0{,}1} = 0{,}257$$

$$P(M_3|E) = \frac{0{,}1}{0{,}42+0{,}18+0{,}1} = 0{,}143$$

An example of Bayesian inference Revisited

- We can say that technician inferred about malfunction from the input evidence.



Inference From the Joint Distribution

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	0.01	0.08	0.001	0.009
¬burglary	0.01	0.09	0.01	0.79

$P(\text{burglary} \mid \text{alarm}) = ?$ $\qquad P(\neg\text{burglary} \mid \text{alarm}) = ?$

$$P(\text{burglary} \mid \text{alarm}) = \frac{P(\text{alarm} \wedge \text{burglary})}{P(\text{alarm})}$$

$P(\text{alarm}) = P(\text{alarm} \wedge \text{burglary} \wedge \text{earthquake}) + P(\text{alarm} \wedge \text{burglary} \wedge \neg\text{earthquake}) +$
$P(\text{alarm} \wedge \neg\text{burglary} \wedge \text{earthquake}) + P(\text{alarm} \wedge \neg\text{burglary} \wedge \neg\text{earthquake}) =$
$0.01 + 0.08 + 0.01 + 0.09 = 0.19$

$P(\text{alarm} \wedge \text{burglary}) = P(\text{alarm} \wedge \text{burglary} \wedge \text{earthquake}) +$
$P(\text{alarm} \wedge \text{burglary} \wedge \neg\text{earthquake}) = 0.01 + 0.08 = 0.09$

$$P(\text{burglary} \mid \text{alarm}) = \frac{0.09}{0.19} \doteq 0.474$$

$P(\neg\text{burglary} \mid \text{alarm}) = 1 - P(\text{burglary} \mid \text{alarm}) \doteq 0.526$

Example – The Monty Hall problem

- There are three doors
- Behind one is a car, and
behind the other two are goats.
- You get to keep whatever
is behind the door you choose.

- You choose a door (say, the door number three).
- The host opens one of the other doors (say, the door number one), which reveals a goat.
- The host says, "Would you like to select the **other** door?"
- **Should you switch?**

Probabilistic Independence

- When two sets of propositions do not affect each others' probabilities, we call them independent, and can easily compute their joint and conditional probability.

$$Independent(A, B) \leftrightarrow \; P(A \wedge B) = P(A)\, P(B), P(B|A) = P(B)$$

- For example, {moon-phase, light-level} might be independent of {burglary, alarm, earthquake}
- Then again, it might not be so, as burglars might be more likely to burglarize houses when there's a new moon (and hence little light).
- Once we're burglarized, light level doesn't affect whether the alarm goes off.
- A more complex notion of independence, and methods for reasoning about these kinds of relationships is needed.

CONDITIONAL INDEPENDENCE

- A and B are absolutely independent if and only if $P(A \wedge B) = P(A)P(B)$; equivalently, $P(A) = P(A|B)$ and $P(B) = P(B|A)$.
- A and B are conditionally independent given C if and only if $P(A \wedge B|C) \ = \ P(A|C)\, P(B|C)$.
- In a similar way the joint probability of three events can be decomposed $P(A \wedge B \wedge C) = P(A|C)P(B|C)P(C)$.
- Conditional independence is weaker than absolute independence, but still useful in decomposing the full joint probability distribution.

Bayesian Networks

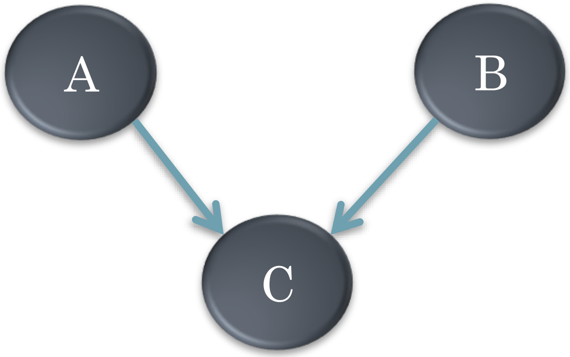
- Bayesian Networks are a practical way to manage probabilistic inference when multiple variables (perhaps many) are involved.
- Reasoning about events involving many parts or contingencies generally requires that a joint probability distribution is known.
- Such a distribution might require a lot of parameters.
- Modelling at this level of detail is typically not practical.

Bayes Networks

- Bayes Networks require making assumptions about the relevance of some conditions to others.
- Once the assumptions are made, the joint distribution can be "factored" so that there are fewer parameters that must be specified.
- A Bayesian network specifies a joint distribution in a structured form.
- It represent dependence/independence via a directed graph, where:
    - Nodes represent random variables, and
    - Edges represent direct dependencies

EXAMPLE OF A SIMPLE BAYESIAN NETWORK

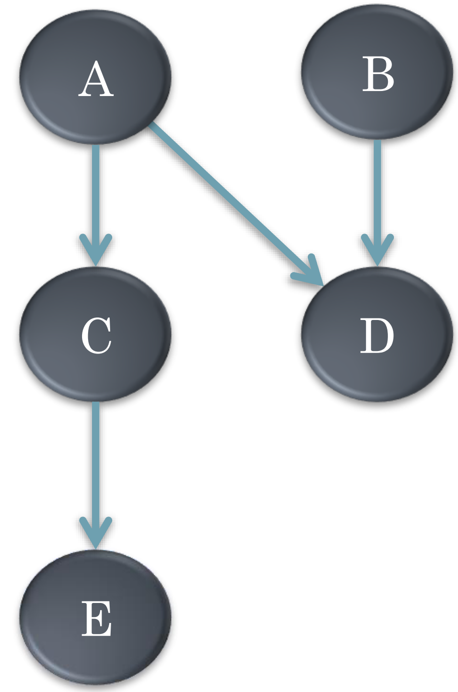
$$\mathrm{P}(A,B,C) = P(C|A,B)P(A)P(B)$$



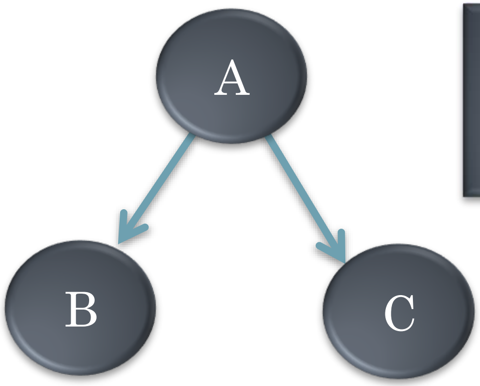
- Probability model has simple factored form
- Directed edges → direct dependence
- Absence of an edge → conditional independence
- Such networks are also known as belief networks, graphical models, causal networks.

Example of a Bayesian Network

- The independences expressed in this Bayesian net are that
  - A and B are (absolutely) independent.
  - C is independent of B given A.
  - D is independent of C given A and B.
  - E is independent of A, B, and D given C.
- The network would normally record the following probabilities:
  - $P(A), P(B)$,
  - $P(C|A), P(C|\neg A)$,
  - $P(D|A,B), P(D|A,\neg B), P(D|\neg A,B), P(D|\neg A,\neg B)$
  - $P(E|C), P(E|\neg C)$



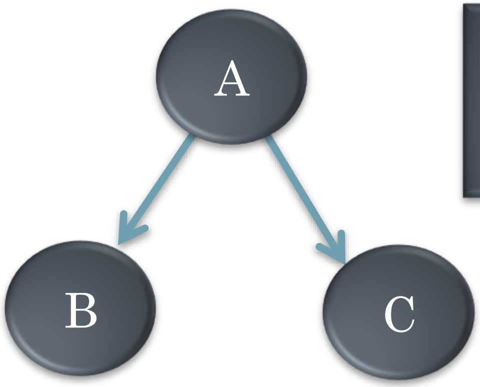
Example of a Bayesian Network



A: An accident blocked traffic on the highway.
B: Barbara is late for work.
C: Christopher is late for work.

- The given information that is recorded in the network could then be:
  - $P(A) = 0.2$
  - $P(B|A) = 0.5, \; P(B|\neg A) = 0.15$
  - $P(C|A) = 0.3, \; P(C|\neg A) = 0.1$

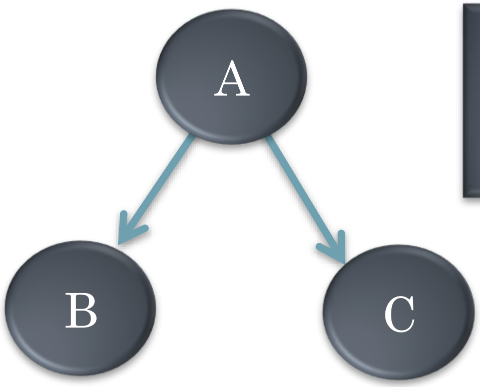
Example of a Bayesian Network



A: An accident blocked traffic on the highway.
B: Barbara is late for work.
C: Christopher is late for work.

- Forward propagation from causes (accident) to effects (being late for work) can then be used to calculate marginal probabilities, i.e.,
  - $P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A) = (0.5)(0.2) + (0.15)(0.8) = 0.22$
  - $P(C) = P(C|A)P(A) + P(C|\neg A)P(\neg A) = (0.3)(0.2) + (0.1)(0.8) = 0.14$
- Marginalizing means eliminating a contingency by summing the probabilities for its different cases (A and $\neg A$).

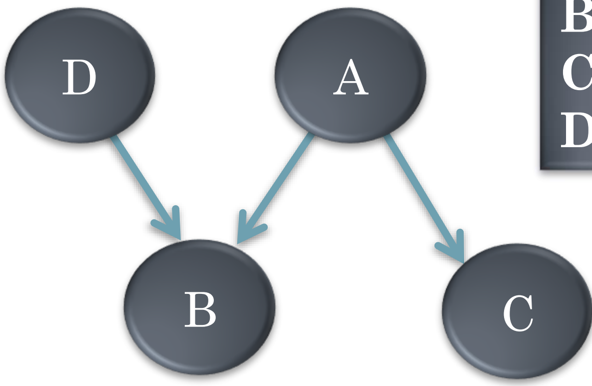
Example of a Bayesian Network



A: An accident blocked traffic on the highway.
B: Barbara is late for work.
C: Christopher is late for work.

- Backward propagation from effects (being late for work) to causes (accident) can then be used for a probabilistic inference (diagnosis).
- For example, Barbara is being late for work. What is the probability of an accident on the highway?
  - $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.5 * 0.2}{(0.5 * 0.2 + 0.15 * 0.8)} = 0.1 / 0.22 = 0.4545$

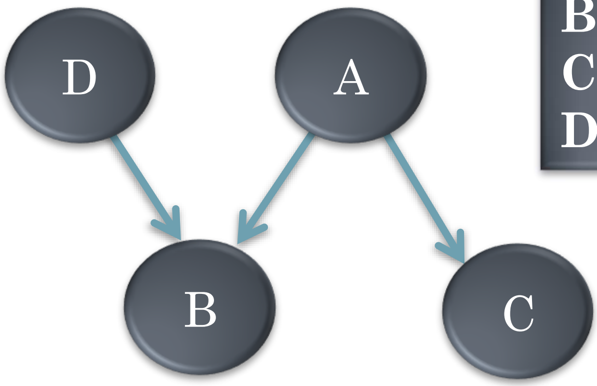
Example of a Bayesian Network



A: An accident blocked traffic on the highway.
B: Barbara is late for work.
C: Christopher is late for work.
D: Barbara has the flu (disease).

- An additional cause could be added to the network.
- The additional given information that is recorded in the network could then be:
  - $P(D) = 0.111$,
  - $P(B|A, D) = 0.9$, $P(B|A, \neg D) = 0.45$
  - $P(B|\neg A, D) = 0.75$ $P(B|\neg A, \neg D) = 0.1$

Example of a Bayesian Network



A: An accident blocked traffic on the highway.
B: Barbara is late for work.
C: Christopher is late for work.
D: Barbara has the flu (disease).

- Barbara is being late for work. What is the probability of the two possible causes?
  - $P(A|B) = 0.4545$
  - $P(B|D) = P(B|A,D)P(A) + P(B|\neg A,D)P(\neg A) = 0.78$
  - $P(D|B) = P(B|D)P(D)/P(B) = 0.3939$
- Propagating probabilities happens along the paths in the network. With a full joint probability distribution, many more computations may be needed.

QUESTIONS

- Explain why abductive and inductive reasoning are inherently uncertain.
- Explain how the Modus Ponens rule of inference is generalized to Bayes' rule.
- What is the difference between absolute and conditional probabilistic independence?
- Why the probabilistic inference from the joint probability distribution is usually computationally more demanding than using Bayesian networks?
- Give an example of a Bayesian Network