Bayesian networks

Syntax, Semantics, and Modeling

Representing and Reasoning under Uncertainty

- Probability Theory provides a framework for representing and reasoning under uncertainty
- Joint probability distributions allow one to model uncertain beliefs and to answer any questions about the domain.

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Joint probability distributions

- A joint probability distribution has an exponential size in the number of variables of interest
  - Computational viewpoint: computing marginal and conditional probabilities poses a complexity challenge
  - **Modelling viewpoint**: requires a large number of probabilities that can be impossible to obtain directly in certain situations.

Human reasoning

- Joint probability distributions are inadequate for representing human reasoning: human good at low order mariginal and conditional probabilities, much difficult to judge joint probability
- Pure numerical representation of probabilistic information is lack of psychological meaningfulness
- This suggests that the elementary building block of human knowledge are not entries of a joint-distribution table
- People can easily and confidently detect dependencies, even though they may not be able to provide precise numerical estimates of probabilities

Capturing Dependencies

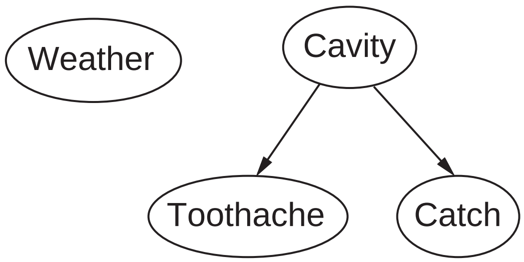
- The notions of relevance, **dependence**, and **causality** are far more basic to human reasoning than the numerical values attached to probability judgments
- Commonsense judgments are issued qualitatively, without reference to numerical probabilities
- How to make sure that the distribution captures the dependence beliefs of a domain expert?
- A reasoning system for representing probabilistic information should allow assertions about dependency relationships to be expressed qualitatively, directly, and explicitly
- Bayesian networks is a graphical modelling tool for specifying probability distributions which effectively address these problems

Bayesian networks

- **Bayesian networks (BNs)** is a graphical modelling tool for specifying probability distributions
  - Encode conditional independence assertions and causal relationships explicitly
  - Provides a compact representation of joint distribution
  - Support efficient algorithms for answering probabilistic queries
- BNs have emerged as the method of choice for uncertainty reasoning.

Bayesian networks

- Bayesian network is a directed acyclic graph (DAG)
  - Nodes: random variables of interest
  - Edges: **direct** (causal) influences
  - Each node is annotated with a conditional distribution $\mathbf{P}(X_i|Parents(X_i))$
  - Each variable is asserted to be conditionally independent of its non-descendants given its parents.

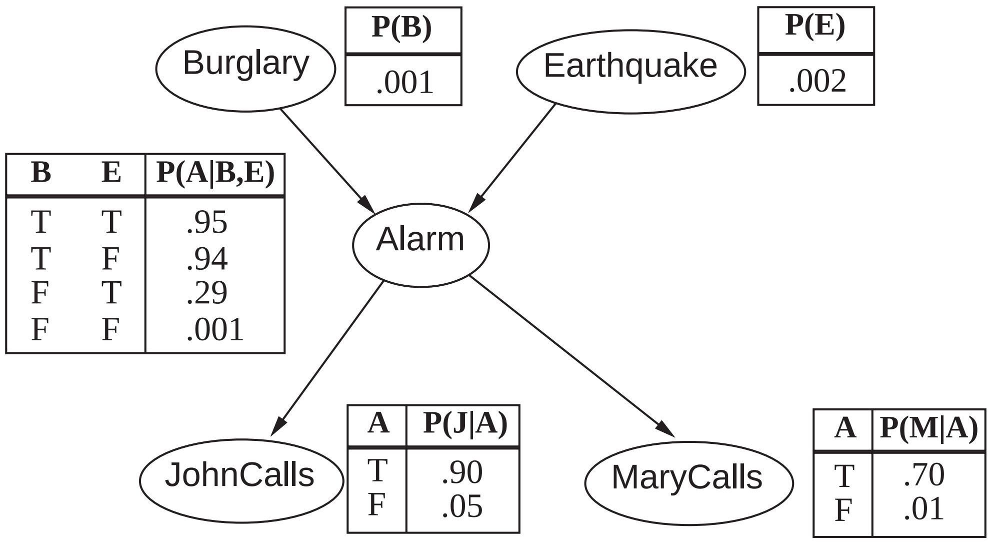


Example

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
Network topology reflects "causal" knowledge:
- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

Example contd.



Bayesian networks - Qualitative Part

- We formally interpret each DAG G as a compact representation of the following independence statements:

  $\{I(V, Parents(V), Non{-}Descendants(V))$: for all variables V in $G\}$

  – Every variable is conditionally independent of its non-descendants given its parents.
- This set of independence statements are often referred to as the **local Markovian assumptions** of DAG G

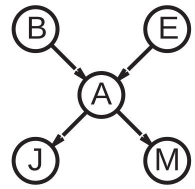
BN as a Knowledge Base

- Since the joint distribution must satisfy the independence assumptions, the chain rule

$$Pr(x_1, \ldots, x_n) = \prod_i Pr(x_i|pa_i)$$

e.g., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$\begin{aligned} &= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &\approx 0.00063 \end{aligned}$$

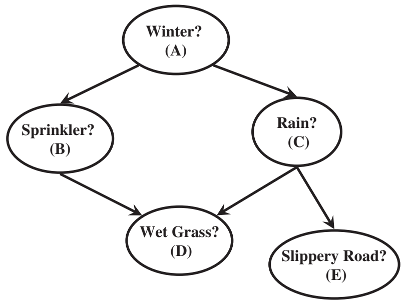


- The joint distribution can be constructed by specifying the local conditional distributions $Pr(x_i|pa_i)$'s

Parameterizing BNs

- The joint distribution can be constructed by specifying the local CPDs $P(X_i|Pa_i)$'s
- A **parameterization** Θ of the DAG G:
  - Θ consists of a set of parameters for each local CPD
  - For discrete random variables: conditional probability tables (CPTs), giving the distribution over X for each combination of parent values
  - Denote $Pr(x_i|pa_i)$ by $\theta_{x_i|pa_i}$, a **parameter**

A Bayesian network



A	Θ_A
true	.6
false	.4

A	B	$\Theta_{B\|A}$
true	true	.2
true	false	.8
false	true	.75
false	false	.25

A	C	$\Theta_{C\|A}$
true	true	.8
true	false	.2
false	true	.1
false	false	.9

B	C	D	$\Theta_{D\|B,C}$
true	true	true	.95
true	true	false	.05
true	false	true	.9
true	false	false	.1
false	true	true	.8
false	true	false	.2
false	false	true	0
false	false	false	1

C	E	$\Theta_{E\|C}$
true	true	.7
true	false	.3
false	true	0
false	false	1

Bayesian network

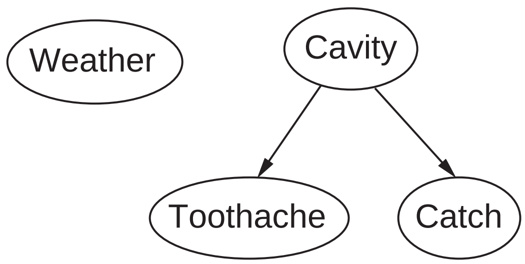
- A **Bayesian network** over a set of variables $X_1, \ldots, X_n$ is a pair (G, Θ) such that

$$Pr(x_1, \ldots, x_n) = \prod_i \theta_{x_i|pa_i}$$

- A BN provides a compact representation of joint distribution: $O(n * d^{k+1})$ vs. $O(d^n)$ (every variable takes up to d values and has at most k parents)
  For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)
- BNs support efficient algorithms for answering probabilistic queries

BN as modeling tool

- Human good at low order mariginal and conditional probabilities, much difficult to judge joint probability
- The parents of X are those variables judged to be **direct causes** of X or have **direct influence** on X
- The parameters requested from model builders are conditional probabilities that quantify conceptual relationships in one's mind, e.g., cause-effect relations, which are psychologically meaningful, and may be obtained by direct measurement



Constructing Bayesian networks

Given a distribution Pr, can we construct a BN?

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n
   add X_i to the network
   identify a minimal subset $Parents(X_i)$ from $X_1, \ldots, X_{i-1}$ such that
   $\mathbf{P}(X_i|Parents(X_i)) = \mathbf{P}(X_i|X_1, \ldots, X_{i-1})$

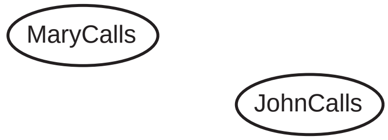
Need a series of locally testable assertions of conditional independence

This choice of parents guarantees the global semantics:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i|Parents(X_i)) \quad \text{(by construction)}\end{aligned}$$

Example

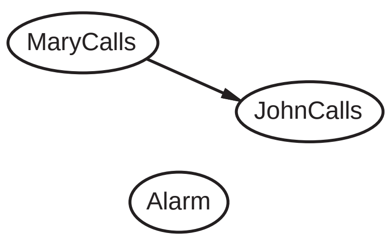
Suppose we choose the ordering M, J, A, B, E



$P(J|M) = P(J)$?

Example

Suppose we choose the ordering M, J, A, B, E

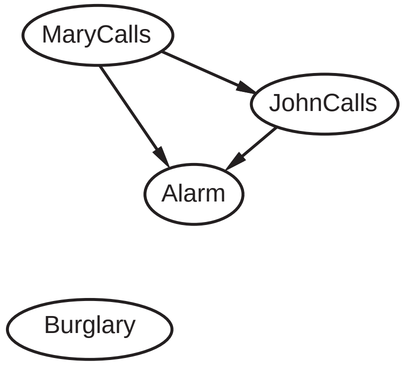


$P(J|M) = P(J)$? No

$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$?

Example

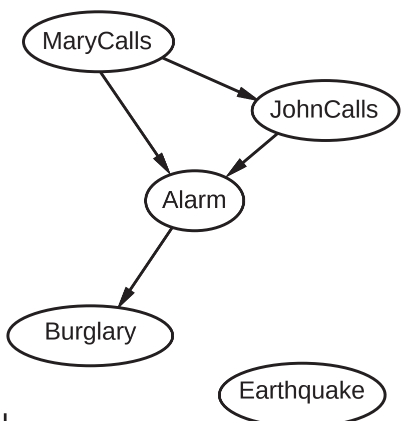
Suppose we choose the ordering M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No
$P(B|A, J, M) = P(B|A)$?
$P(B|A, J, M) = P(B)$?

Example

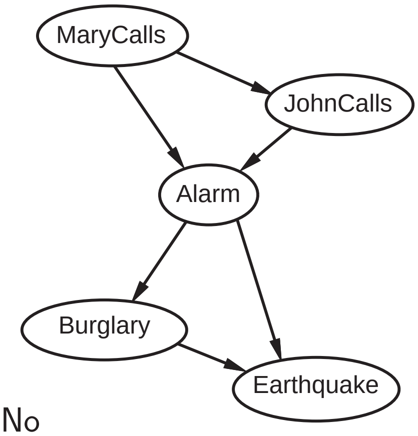
Suppose we choose the ordering M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No
$P(B|A, J, M) = P(B|A)$? Yes
$P(B|A, J, M) = P(B)$? No
$P(E|B, A, J, M) = P(E|A)$?
$P(E|B, A, J, M) = P(E|A, B)$?

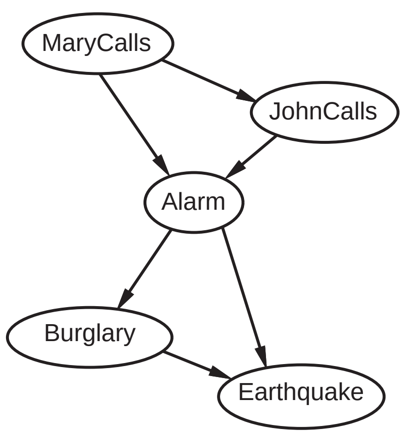
Example

Suppose we choose the ordering M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No
$P(B|A,J,M) = P(B|A)$? Yes
$P(B|A,J,M) = P(B)$? No
$P(E|B,A,J,M) = P(E|A)$? No
$P(E|B,A,J,M) = P(E|A,B)$? Yes

Example contd.



Deciding conditional independence is hard in noncausal directions

(Causal models and conditional independence seem hardwired for humans!)

Assessing conditional probabilities is hard in noncausal directions

Network is less compact: $1+2+4+2+4=13$ numbers needed (vs. 10)

Role of Causality

- The interpretation of directed acyclic graphs as carriers of independence assumptions does not necessarily imply causation
- The ubiquity of DAG models in statistical and AI applications stems (often unwittingly) primarily from their causal interpretation
- In practice, DAG models are rarely used in any variable ordering other than those which respect the direction of time and causation

Role of Causality

The advantages of building DAG models around causal rather than associational information

- The judgments required in the construction of the model are more meaningful, more accessible, and hence more reliable.
- Conditional independence judgments are accessible (hence reliable) only when they are anchored onto more fundamental building blocks of our knowledge, such as causal relationships.
- If conditional independence judgments are byproducts of stored causal relationships, then representing those relationships directly would be a more natural way of expressing what we know or believe about the world -the philosophy behind **causal Bayesian networks**.

BNs as a Logic of Dependences

- A BN can be used as an inference instrument for deducing new independence relationships from those used in constructing the network.
- Assertions about dependence can be inferred qualitatively/logically without reference to numerical quantities
- Input independence statements, the local Markovian assumptions,

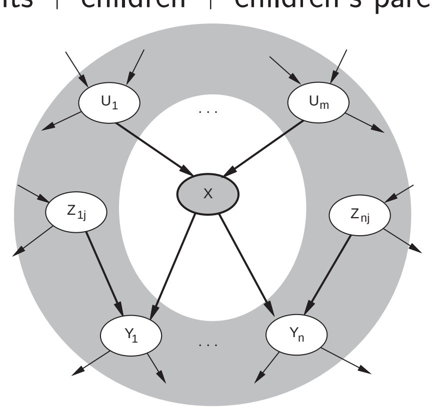
$$\{I(X_i, PA_i, \{X_1, X_2, \ldots, X_{i-1}\} - PA_i)\}$$

- Additional independence statements can be deduced by logical inference rules, captured using a graphical test known as **d-separation** – Two variables are independent if all paths between them are blocked by evidence.

Markov blanket

A **Markov Blanket** for X is a set of variables B which, when known, will render every other variable irrelevant to X, i.e., $I(X, B, R)$, where R is the set of all variables other than X and $B \rightarrow$ feature selection

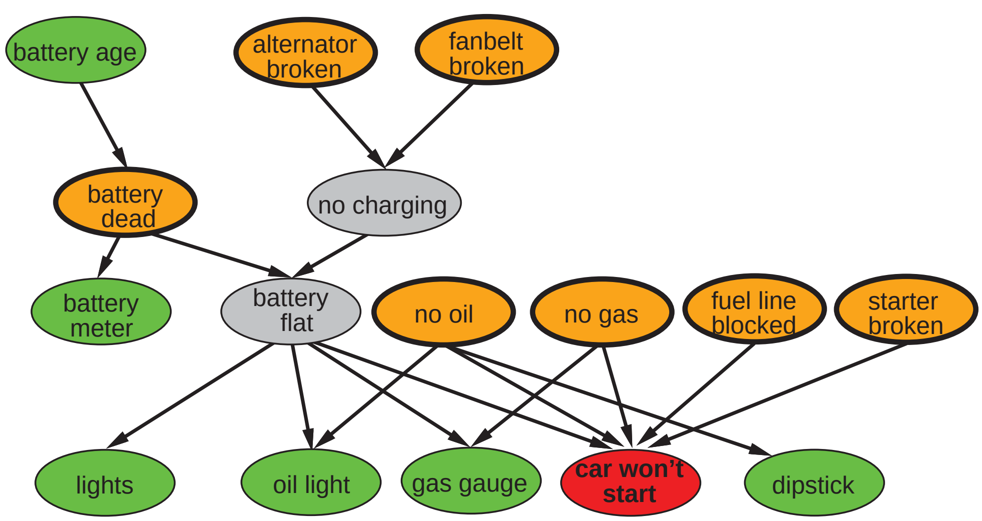
Each node is conditionally independent of all others given its
Markov blanket: parents + children + children's parents



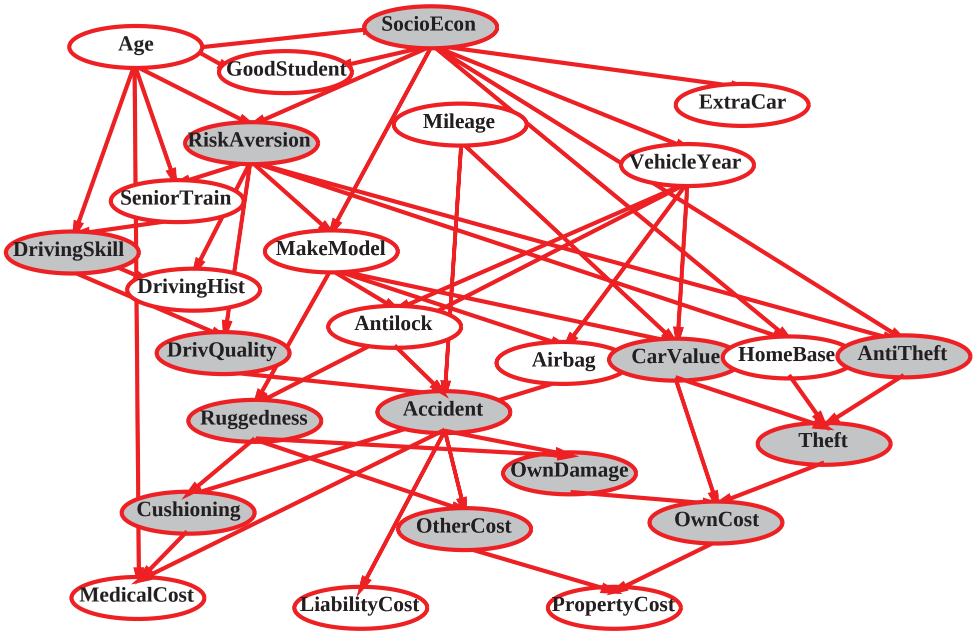
Applications

- Successful applications in a variety of domains:
  - diagnosis
  - troubleshooting
  - data mining
  - pattern recognition
  - bioinformatics/computational biology
  - . . .

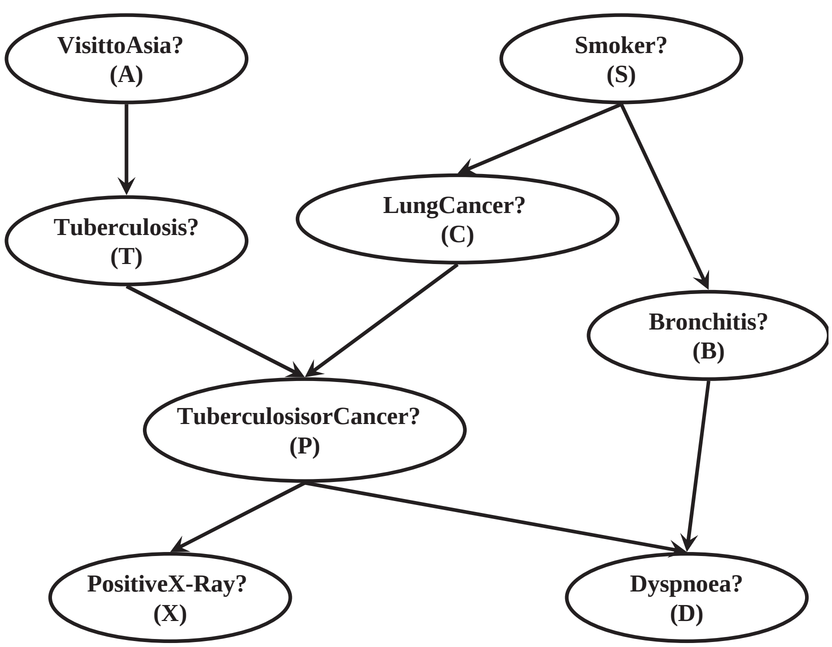
Car Diagnosis BN



Car insurance BN



A Bayesian Network



Reasoning with BNs

- What types of queries can be posed to a Bayesian network?
- Probability of Evidence query: the probability of some variable instantiation e, $Pr(e)$
$$Pr(X = yes, D = no)\ ?$$
- The variables $E = \{X, D\}$ are called **evidence variables**
- Prior and posterior marginal queries: $P(S)$, $P(S|e)$ for small $S \subset V$

Reasoning with BNs

- **Most probable explanation (MPE)**: identify an instantiation $x_1, \ldots, x_n$ for which $P(x_1, \ldots, x_n|e)$ is maximal.
- Identify the most probable instantiation of network variables given some evidence
- Choosing each value x_i so as to maximizes the probability $Pr(x_i|e)$ does not necessarily lead to a most probable explanation

Reasoning with BNs

- **Maximum a posteriori hypothesis (MAP)**: find an instantiation m of variables $M \subset V$ for which $P(m|e)$ is maximal
- Finding the most probable instantiation for a subset of network variables
- The variables in M are known as MAP variables
- A common method for approximating MAP is to compute an MPE and then project the result on the MAP variables.

Constructing Bayesian Networks

- How to construct BNs?
- Construct network structure based on domain knowledge
- Learn the structures of Bayesian networks from training data
  - An active research area.

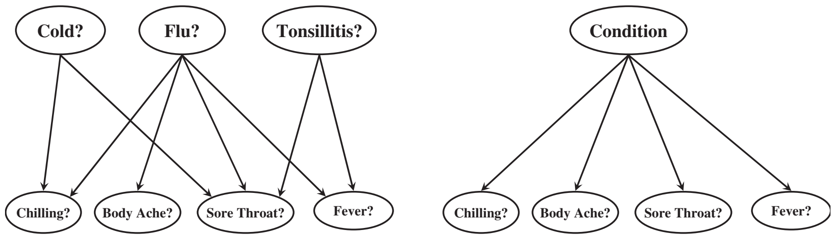
Modeling with Bayesian networks

1. Define the network variables and their values.
   - Query variables
   - Evidence variables
   - Intermediary variables
2. Define the network structure.
   - Guided by causal interpretation of network structure.
   - what is the set of variables that we regard as the direct causes of X?
3. Define the CPTs.

Modeling with Bayesian networks

Diagnosis I: medical diagnosis
The flu is an acute disease characterized by fever, body aches and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a sore throat. Tonsillitis is inflammation of the tonsils which leads to a sore throat and can be associated with fever.

BNs for medical diagnosis



- the Naive Bayes structure makes a key commitment known as the single-fault assumption: it assumes that only one condition can exist in the patient at any time
- They represent different independence claims: $I(fever, cold, sorethroat)$? $I(fever, , cold)$?

Modeling with Bayesian networks

Specification of CPTs

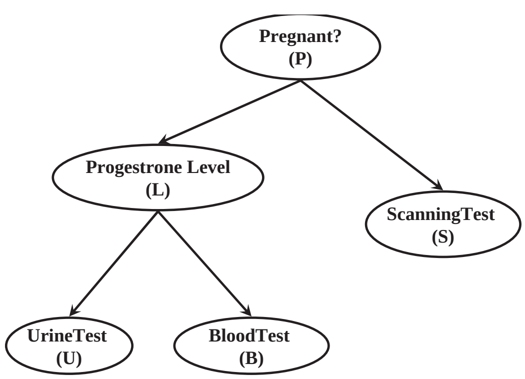
- The CPT for a condition, such as tonsillitis, must provide the belief in developing tonsillitis by a person about whom we have no knowledge of any symptoms
- The CPT for a symptom, such as chilling, must provide the belief in this symptom under the possible conditions
- The probabilities are usually obtained from a medical expert, based on known medical statistics or subjective beliefs gained through practical experience
- Another key method for specifying the CPTs is by estimating them directly from medical records of previous patients

Modeling with Bayesian networks

Diagnosis II: medicine diagnosis

A few weeks after inseminating a cow, we have three possible tests to confirm pregnancy. The first is a scanning test which has a false positive of 1% and a false negative of 10%. The second is a blood test, which detects progesterone with a false positive of 10% and a false negative of 30%. The third test is a urine test, which also detects progesterone with a false positive of 10% and a false negative of 20%. The probability of a detectable progesterone level is 90% given pregnancy, and 1% given no pregnancy. The probability that insemination will impregnate a cow is 87%.

A Bayesian Network



P	θ_p
yes	.87

P	S	$\theta_{s\|p}$
yes	–ve	.10
no	+ve	.01

P	L	$\theta_{l\|p}$
yes	undetectable	.10
no	detectable	.01

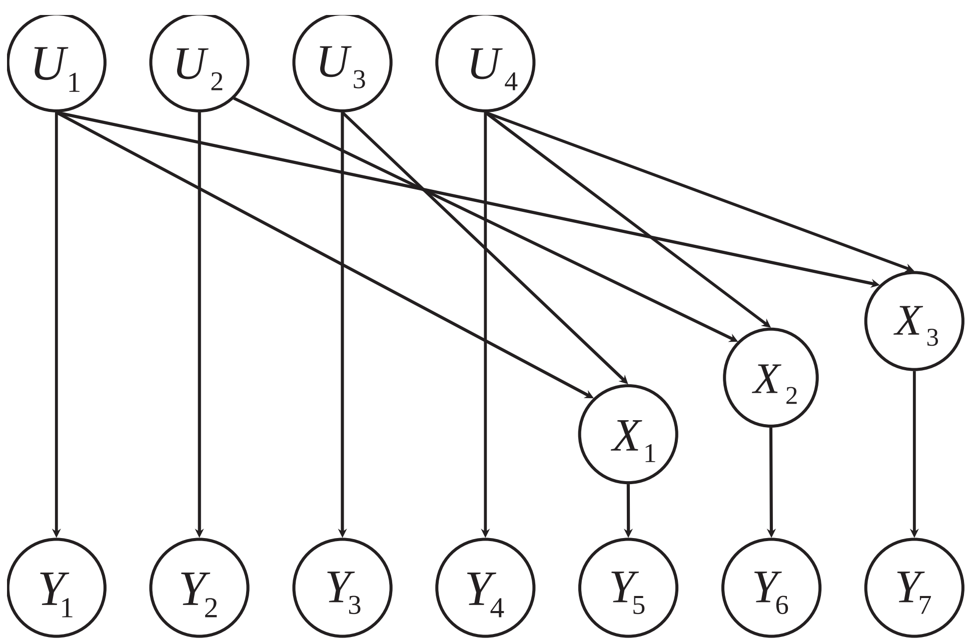
L	B	$\theta_{b\|l}$
detectable	–ve	.30
undetectable	+ve	.10

L	U	$\theta_{u\|l}$
detectable	–ve	.20
undetectable	+ve	.10

Modeling with Bayesian networks

Channel Coding
We need to send four bits U_1, U_2, U_3, and U_4 from a source S to a destination D over a noisy channel, where there is a 1% chance that a bit will be inverted before it gets to the destination. To improve the reliability of this process, we will add three redundant bits X_1, X_2, and X_3 to the message, where X_1 is the XOR of U_1 and U_3, X_2 is the XOR of U_2 and U_4, and X_3 is the XOR of U_1 and U_4. Given that we received a message containing seven bits at destination D, our goal is to restore the message generated at the source S.

BNs for Channel Coding



Channel Coding

Decoder quality measures

- Word Error Rate (WER)
- Bit Error Rate (BER)

Queries to pose

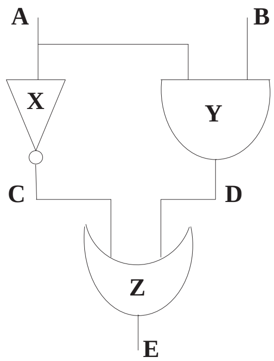
- MAP
- Posterior Marginal $Pr(u_i|y_1, \ldots, y_7)$

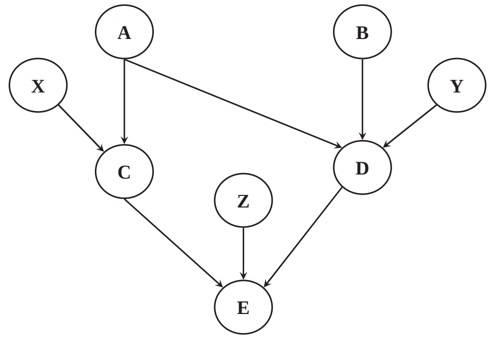
Modeling with Bayesian networks

Diagnosis III: digital circuit

Consider a digital circuit. Given some values for the circuit primary inputs and output (test vector), our goal is to decide whether the circuit is behaving normally. If not, our goal is then to decide the most likely health states of its components.

A Bayesian Network





The BN structures can be generated automatically by software

Diagnosis III: digital circuit

- The values of variables representing circuit wires (primary inputs, outputs, or internal wires): $\{low, high\}$
- The values for health variables: $\{ok, faulty\}$,

X	θ_x
ok	$.99$
$faulty$	$.01$

A	X	C	$\theta_{c\|a,x}$
$high$	ok	$high$	0
low	ok	$high$	1
$high$	$faulty$	$high$	$.5$
low	$faulty$	$high$	$.5$

- MAP queries, where MAP variables are X, Y, Z, and evidence variables are A, B, E

Summary

Bayes nets provide a natural representation for (causally induced) conditional independence, which can be read by d-separation criterion

Topology + CPTs = compact representation of joint distribution

Generally easy for (non)experts to construct