Constructing Bayesian Networks

Uffe Kjærulff (uk@cs.aau.dk)

Group of Machine Intelligence
Department of Computer Science, Aalborg University

Reykjavik University, 27 April, 2005

Outline

1. Bayesian Networks
2. Constructing Bayesian networks

Bayesian Networks

- A Bayesian network is a compact model representation for reasoning under uncertainty.
  - Entities of the problem domain are represented as random variables.
  - A graphical structure describes the dependence relations between entities.
  - Conditional probability distributions specify our belief about the strengths of relations.

- Can compute posterior probabilities of unobserved variables (e.g., hypotheses) given the evidence provided by observed variables.

Variables

- The variables of a Bayesian network are represented as nodes in a causal network.
- A variable X represents a unique event or hypothesis.
- Positive examples: "Cancer", "Smoking", "Temperature".
- Negative examples: {"High temperature", "Low temperature"}, {"Error occured", "No error"}.
- Each variable has a finite set of mutually exclusive *states* (events, propositions, levels) $X = \{x_1, \ldots, x_n\}$.
- Variables have *parents* and *children*.
- A variable should pass the clarity test: There must be a state for each possible value of the variable.

Definition of Bayesian Networks

A Bayesian network $N = (G, \mathcal{P})$ consists of:

- A set of variables and a set of directed edges between variables.
- The variables together with the directed edges form a directed acyclic graph (DAG).
- Each variable has a finite set of states.
- Attached to each variable X with parents $Y_1, \ldots, Y_n$ there is a conditional probability table $P(X \mid Y_1, \ldots, Y_n)$.

A Bayesian network is a representation of knowledge for reasoning under uncertainty.

Model Specification

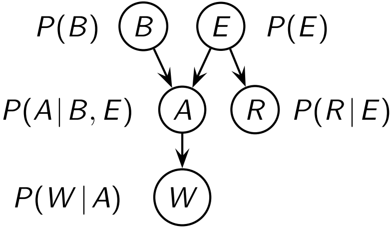
- A Bayesian network $N = (G, \mathcal{P})$ consists of a
  - *qualitative part*, the graph structure (DAG $G = (V, E)$) and a
  - *quantitative part*, the conditional probability distributions, $\mathcal{P} = \{P(child \,|\, parents)\}$.

Model Specification

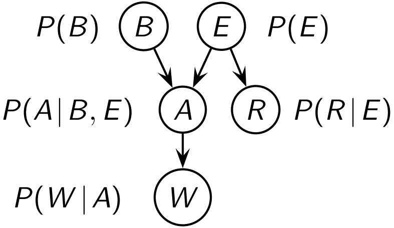
- A Bayesian network $N = (G, \mathcal{P})$ consists of a
  - *qualitative part*, the graph structure (DAG $G = (V, E)$) and a
  - *quantitative part*, the conditional probability distributions, $\mathcal{P} = \{P(\mathit{child} \mid \mathit{parents})\}$.

- Model specification consists of two parts:
  - First, specify the structure of the network and validate it (possibly leading to iterative revision).
  - Second, specify $P(X \mid \mathrm{pa}(X))$ for each variable X.

Burglary or Earthquake

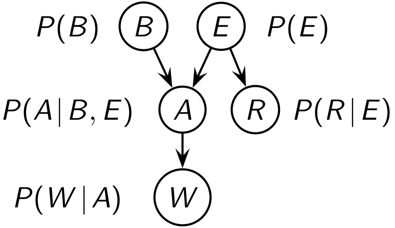


Burglary or Earthquake



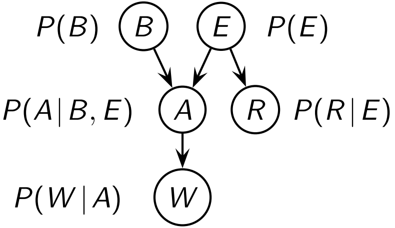
$$P(E) = (0.99, 0.01),$$

Burglary or Earthquake



$$P(E) = (0.99, 0.01), \qquad P(B) = (0.9, 0.1)$$

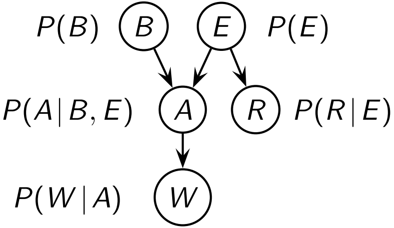
Burglary or Earthquake



$$P(E) = (0.99, 0.01), \qquad P(B) = (0.9, 0.1)$$

	E	
R	n	y
n	0.999	0.01
y	0.001	0.99

Burglary or Earthquake

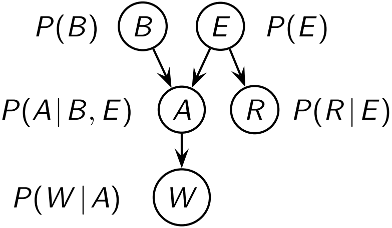


$$P(E) = (0.99, 0.01), \qquad P(B) = (0.9, 0.1)$$

	E	
R	n	y
n	0.999	0.01
y	0.001	0.99

	A	
W	n	y
n	0.9	0.1
y	0.1	0.9

Burglary or Earthquake

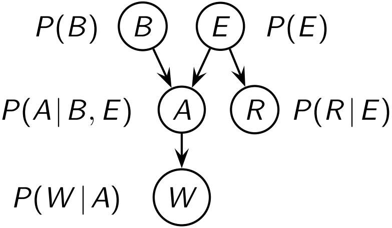


$P(E) = (0.99, 0.01), \quad P(B) = (0.9, 0.1)$

R	E = n	E = y
n	0.999	0.01
y	0.001	0.99

W	A = n	A = y
n	0.9	0.1
y	0.1	0.9

	E = n		E = y	
A	B = n	B = y	B = n	B = y
n	0.99	0.1	0.1	0.01
y	0.01	0.9	0.9	0.99



$$P(E) = (0.99, 0.01), \qquad P(B) = (0.9, 0.1)$$

	E	
R	n	y
n	0.999	0.01
y	0.001	0.99

	A	
W	n	y
n	0.9	0.1
y	0.1	0.9

	$E = n$		$E = y$	
A	$B = n$	$B = y$	$B = n$	$B = y$
n	0.99	0.1	0.1	0.01
y	0.01	0.9	0.9	0.99

Now, we have a fully specified Bayesian network.

Bayesian Network Construction

- What do I want a model of and how do I want to use the model?

Bayesian Network Construction

- What do I want a model of and how do I want to use the model?

- What are the background conditions, context, assumptions and restrictions, what is in the model, what is NOT in the model?

Bayesian Network Construction

- What do I want a model of and how do I want to use the model?
- What are the background conditions, context, assumptions and restrictions, what is in the model, what is NOT in the model?
- Work in a topdown fashion by identifying parts (components, groups)

Bayesian Network Construction

- What do I want a model of and how do I want to use the model?

- What are the background conditions, context, assumptions and restrictions, what is in the model, what is NOT in the model?

- Work in a topdown fashion by identifying parts (components, groups)

- Iterate, iterate, iterate

Bayesian Network Construction

Repeat

- Identify possible variables (information, hypothesis, intermediate).
- Identify possible states of variables.
- Identify structural relations between variables.
- Consider modeling tricks.
- Determine conditional probability distributions.

Until satisfied (e.g., model verification).

The Chain Rule

- Let $U = \{X_1, \ldots, X_n\}$ be a universe of variables.
  - From the joint $P(U)$, we can compute, e.g. $P(X_i)$, $P(X_i \mid \varepsilon)$, etc.
- Let $N = (G, \mathcal{P})$ be a Bayesian network over U.

The Chain Rule

- Let $U = \{X_1, \ldots, X_n\}$ be a universe of variables.
  - From the joint $P(U)$, we can compute, e.g. $P(X_i)$, $P(X_i \,|\, \varepsilon)$, etc.
- Let $N = (G, \mathcal{P})$ be a Bayesian network over U.
- For any ordering $(X_1, X_2, \ldots, X_n)$ (using the fundamental rule repeatedly),

$$P(X_1, .., N_n) = P(X_1)P(X_2 \,|\, X_1)P(X_3 \,|\, X_1, X_2) \cdot\cdot P(X_n \,|\, X_1, .., X_{n-1}).$$

The Chain Rule

- Let $U = \{X_1, \ldots, X_n\}$ be a universe of variables.
  - From the joint $P(U)$, we can compute, e.g. $P(X_i)$, $P(X_i \mid \varepsilon)$, etc.
- Let $N = (G, \mathcal{P})$ be a Bayesian network over U.
- For any ordering $(X_1, X_2, \ldots, X_n)$ (using the fundamental rule repeatedly),

$$P(X_1, .., N_n) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2) \cdot\cdot P(X_n \mid X_1, .., X_{n-1}).$$

- Let $(X_1, X_2, \ldots, X_n)$ be a topological ordering wrt. N, i.e., $\mathrm{pa}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}\ \forall i$. d-separation yields $X_i \perp\!\!\!\perp \mathrm{nd}(X_i) \mid \mathrm{pa}(X_i)$, where $\mathrm{pa}(X)$ are the parents of X and $\mathrm{nd}(X)$ the non-descendants of X.

The Chain Rule

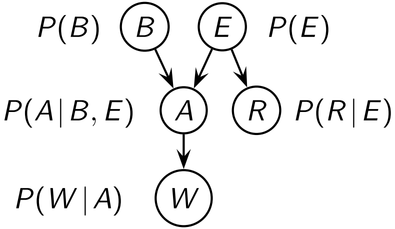
- Let $U = \{X_1, \ldots, X_n\}$ be a universe of variables.
  - From the joint $P(U)$, we can compute, e.g. $P(X_i)$, $P(X_i \,|\, \varepsilon)$, etc.
- Let $N = (G, \mathcal{P})$ be a Bayesian network over U.
- For any ordering $(X_1, X_2, \ldots, X_n)$ (using the fundamental rule repeatedly),

$$P(X_1, .., N_n) = P(X_1)P(X_2 \,|\, X_1)P(X_3 \,|\, X_1, X_2) \cdot\cdot P(X_n \,|\, X_1, .., X_{n-1}).$$

- Let $(X_1, X_2, \ldots, X_n)$ be a topological ordering wrt. N, i.e., $\mathrm{pa}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\} \; \forall i$. d-separation yields $X_i \perp\!\!\!\perp \mathrm{nd}(X_i) \,|\, \mathrm{pa}(X_i)$, where $\mathrm{pa}(X)$ are the parents of X and $\mathrm{nd}(X)$ the non-descendants of X.
- The joint probability distribution of U is then

$$P(U) = \prod_{i=1}^{n} P(X_i \,|\, X_1, \ldots, X_{i-1}) = \prod_{i=1}^{n} P(X_i \,|\, \mathrm{pa}(X_i)).$$

Burglary or Earthquake



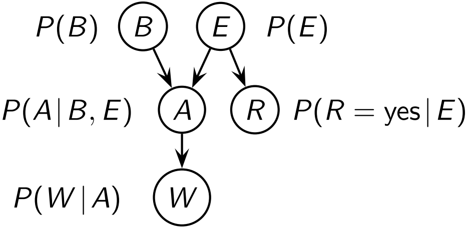
From the Chain Rule:

$$P(B, A, E, R, W) = P(W \mid A)P(E)P(B)P(A \mid B, E)P(R \mid E)$$

Evidence

An *instantiation* of a variable X is an observation of a the exact state of X.

Example: Burglary or Earthquake



$P(R = y \mid E) =$

R \ E	n	y
n	0	0
y	0.001	0.99

Now we have a representation of $P(B, A, E, R = y, W)$.

Probabilistic Inference

- What do we want to do with Bayesian networks?

Probabilistic Inference

- What do we want to do with Bayesian networks?
  - We want to update our beliefs in the light of observations $P(X \mid \varepsilon)$.

Probabilistic Inference

- What do we want to do with Bayesian networks?
  - We want to update our beliefs in the light of observations $P(X \mid \varepsilon)$.
  - The probability of the evidence $P(\varepsilon)$.

Probabilistic Inference

- What do we want to do with Bayesian networks?
  - We want to update our beliefs in the light of observations $P(X \mid \varepsilon)$.
  - The probability of the evidence $P(\varepsilon)$.
  - The joint probability $P(X_i, \ldots, X_j)$.

Probabilistic Inference

- What do we want to do with Bayesian networks?
  - We want to update our beliefs in the light of observations $P(X \mid \varepsilon)$.
  - The probability of the evidence $P(\varepsilon)$.
  - The joint probability $P(X_i, \ldots, X_j)$.
  - What is the configuration of the variables with the highest probability (max-propagation)?

Probabilistic Inference

- What do we want to do with Bayesian networks?
  - We want to update our beliefs in the light of observations $P(X \mid \varepsilon)$.
  - The probability of the evidence $P(\varepsilon)$.
  - The joint probability $P(X_i, \ldots, X_j)$.
  - What is the configuration of the variables with the highest probability (max-propagation)?
  - What is $P(X \mid \varepsilon \setminus \varepsilon_X)$ for $X \in \varepsilon$?

Probabilistic Inference

- What do we want to do with Bayesian networks?
  - We want to update our beliefs in the light of observations $P(X \,|\, \varepsilon)$.
  - The probability of the evidence $P(\varepsilon)$.
  - The joint probability $P(X_i, \ldots, X_j)$.
  - What is the configuration of the variables with the highest probability (max-propagation)?
  - What is $P(X \,|\, \varepsilon \setminus \varepsilon_X)$ for $X \in \varepsilon$?
  - $\ldots$

Probabilistic Inference

- What do we want to do with Bayesian networks?
  - We want to update our beliefs in the light of observations $P(X \mid \varepsilon)$.
  - The probability of the evidence $P(\varepsilon)$.
  - The joint probability $P(X_i, \ldots, X_j)$.
  - What is the configuration of the variables with the highest probability (max-propagation)?
  - What is $P(X \mid \varepsilon \setminus \varepsilon_X)$ for $X \in \varepsilon$?
  - ...
- Computations performed in a secondary structure.

Answering a Query

- What is the probability of burglary given a phone call from Dr. Watson and a radio report on earthquake?

Answering a Query

- What is the probability of burglary given a phone call from Dr. Watson and a radio report on earthquake?

$$P(B \mid \varepsilon) = \sum_{A,E,R,W} P(A, B, E, R, W)$$

Answering a Query

- What is the probability of burglary given a phone call from Dr. Watson and a radio report on earthquake?

$$
\begin{aligned}
P(B \mid \varepsilon) &= \sum_{A,E,R,W} P(A, B, E, R, W) \\
&= \sum_{A,E,R,W} P(B)P(E)P(A \mid B, E)P(R = \mathrm{y} \mid E)P(W = \mathrm{y} \mid A)
\end{aligned}
$$

Answering a Query

- What is the probability of burglary given a phone call from Dr. Watson and a radio report on earthquake?

$$
\begin{aligned}
P(B \mid \varepsilon) &= \sum_{A,E,R,W} P(A, B, E, R, W) \\
&= \sum_{A,E,R,W} P(B)P(E)P(A \mid B, E)P(R = \mathrm{y} \mid E)P(W = \mathrm{y} \mid A) \\
&= P(B) \sum_{E} P(E) \sum_{A} P(A \mid B, E) \sum_{R} P(R = \mathrm{y} \mid E) \\
&\quad \sum_{W} P(W = \mathrm{y} \mid A).
\end{aligned}
$$

Answering a Query

- What is the probability of burglary given a phone call from Dr. Watson and a radio report on earthquake?

$$
\begin{aligned}
P(B \mid \varepsilon) &= \sum_{A,E,R,W} P(A, B, E, R, W) \\
&= \sum_{A,E,R,W} P(B)P(E)P(A \mid B, E)P(R = \mathrm{y} \mid E)P(W = \mathrm{y} \mid A) \\
&= P(B) \sum_{E} P(E) \sum_{A} P(A \mid B, E) \sum_{R} P(R = \mathrm{y} \mid E) \\
&\quad \sum_{W} P(W = \mathrm{y} \mid A).
\end{aligned}
$$

- In general, probabilistic inference is NP-hard.
- HUGIN uses a secondary structure for performing inference.

Complexity of Inference

Computing $P(B \mid \varepsilon)$ like

$$P(B \mid \varepsilon) = \sum_{A,E,R,W} P(B)P(E)P(A \mid B, E)P(R = \mathrm{y} \mid E)P(W = \mathrm{y} \mid A)$$

involves 160 multiplications, 32 additions, and a table with 32 numbers.

Complexity of Inference

Computing $P(B \mid \varepsilon)$ like

$$P(B \mid \varepsilon) = \sum_{A,E,R,W} P(B)P(E)P(A \mid B, E)P(R = \mathrm{y} \mid E)P(W = \mathrm{y} \mid A)$$

involves 160 multiplications, 32 additions, and a table with 32 numbers.

$$P(B \mid \varepsilon) = P(B) \sum_{E} P(E) \sum_{A} P(A \mid B, E) \sum_{R} P(R = \mathrm{y} \mid E) \sum_{W} P(W = \mathrm{y} \mid A).$$

involves 32 multiplications, 12 additions, and two tables with 4 numbers and one with 8 numbers.

Complexity of Inference

Computing $P(B \mid \varepsilon)$ like

$$P(B \mid \varepsilon) = \sum_{A,E,R,W} P(B)P(E)P(A \mid B, E)P(R = \mathrm{y} \mid E)P(W = \mathrm{y} \mid A)$$

involves 160 multiplications, 32 additions, and a table with 32 numbers.

$$P(B \mid \varepsilon) = P(B) \sum_{E} P(E) \sum_{A} P(A \mid B, E) \sum_{R} P(R = \mathrm{y} \mid E) \sum_{W} P(W = \mathrm{y} \mid A).$$

involves 32 multiplications, 12 additions, and two tables with 4 numbers and one with 8 numbers.

The key to efficient inference lies in finding a good summation order (also called an *elimination order* or *elimination sequence*).

Outline

1. Bayesian Networks
2. Constructing Bayesian networks
   1. Types of variables
   2. Undirected relations
   3. Measurement error
   4. Simple bayes
   5. Independence of causal influence
   6. Parent divorcing
   7. Experts disagreement
   8. Structural uncertainty
   9. Functional uncertainty
   10. Model verification

Types of Variables

- *Hypothesis variables:* Unobservable variables for which posterior probabilities are wanted.

  Examples: Classification and diagnosis variables

Types of Variables

- *Hypothesis variables:* Unobservable variables for which posterior probabilities are wanted.

  Examples: Classification and diagnosis variables

- *Information variables:* Variables for which observations are available, and which can provide information relevant for hypothesis variables

  Examples: Sensor readings, background information, test results, etc.

Types of Variables

- *Hypothesis variables:* Unobservable variables for which posterior probabilities are wanted.

  Examples: Classification and diagnosis variables

- *Information variables:* Variables for which observations are available, and which can provide information relevant for hypothesis variables

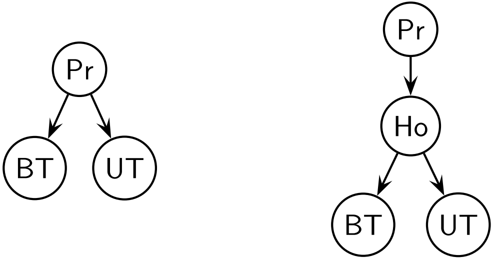
  Examples: Sensor readings, background information, test results, etc.

- *Mediating variables:* Unobservable variables for which posterior probabilities are not wanted, but which play important roles for achieving
  - correct conditional independence and dependence properties and
  - efficient inference.

Mediating Variables

Mediating variables important for achieving correct conditional independence and dependence properties.

Example: Hormonal state (Ho): Blood test (BT) and urine test (UT) are not independent given pregnancy state (Pr).



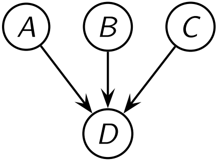
The model without the variable Ho is wrong, since Ho does not depend deterministically on Pr, and BT and UT are independent given Ho.

Undirected Relations

Let $R(A, B, C)$ be a relation in numbers such that $R(a, b, c) \in \{0, 1\}$. How can R be represented in a BN?

Undirected Relations

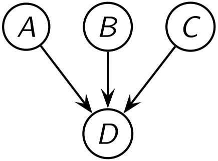
Let $R(A, B, C)$ be a relation in numbers such that $R(a, b, c) \in \{0, 1\}$. How can R be represented in a BN?



- Introduce variable D with states "on" and "off".

Undirected Relations

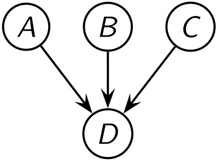
Let $R(A, B, C)$ be a relation in numbers such that $R(a, b, c) \in \{0, 1\}$. How can R be represented in a BN?



- Introduce variable D with states "on" and "off".
- Let $P(D = \text{on} \mid A, B, C) = R(A, B, C)$ and $P(D = \text{off} \mid A, B, C) = 1 - R(A, B, C)$.

Undirected Relations

Let $R(A, B, C)$ be a relation in numbers such that $R(a, b, c) \in \{0, 1\}$. How can R be represented in a BN?



- Introduce variable D with states "on" and "off".
- Let $P(D = \text{on} \,|\, A, B, C) = R(A, B, C)$ and $P(D = \text{off} \,|\, A, B, C) = 1 - R(A, B, C)$.
- Clamp the state of D to on.

An Example: Constraints

- Assume we have four items, $S_1, \ldots, S_4$, where $\mathrm{dom}(S_i) = \{t_1, t_2\}$, such that two items must be of type t_1 and two must be of type t_2.

- This constraint can be encoded in the CPT of a common child variable, C, with states on and off:

$$P(C = \text{on} \,|\, s_1, s_2, s_3, s_4) = \begin{cases} 1 & \text{if } |\{s_i = t_1\}| = 2 \\ 0 & \text{otherwise.} \end{cases}$$

- By clamping C to "on", the constraint gets enforced.

A Constraints Example

- Assume we have 2 pairs of socks, $S_1, \ldots, S_4$, where $\mathrm{dom}(S_i) = \{t_1, t_2\}$.

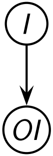
- The "2 pairs" constraint can be encoded through the CPT

S_1	t_1								t_2							
S_2	t_1				t_2				t_1				t_2			
S_3	t_1		t_2		t_1		t_2		t_1		t_2		t_1		t_2	
S_4	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2
on	0	0	0	1	0	1	1	0	0	1	1	0	1	0	0	0
off	1	1	1	0	1	0	0	1	1	0	0	1	0	1	1	1

- By clamping to state "on", the "2 pairs" constraint gets enforced: The CPT ensures that only configurations involving exactly 2 socks of type t_1 (and 2 of type t_2) have non-zero probability.

Measurement Error

- Often the true value of an information variable, I, cannot be obtained due to measurement error.
- Measurement error modeled by introducing a new variable, say OI, representing the observed value of I.
- Size of measurement error represented in $P(OI \mid I)$.
- I turned into mediating variable (i.e., unobservable).

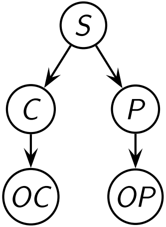


Measurement Error: Example

- Assume that socks of type t_1 are characterized by color c_1 and pattern p_2, and socks of type t_2 by color $c_2 \neq c_1$ and pattern $p_2 \neq p_1$.

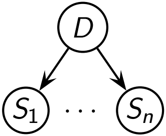
- After several washes:
  - c_1 is mistaken for c_2 in 3 out of 10 cases.
  - c_2 is mistaken for c_1 in 2 out of 10 cases.
  - p_1 is mistaken for p_2 in 2 out of 100 cases.
  - p_2 is always recognized correctly.

- Modeled as:
  - $P(OC = c_2 \mid C = c_1) = 0.3$.
  - $P(OC = c_1 \mid C = c_2) = 0.2$.
  - $P(OP = p_2 \mid P = p_1) = 0.02$.
  - $P(OP = p_1 \mid P = p_2) = 0$.



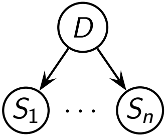
Simple (or Naïve) Bayes

- Consider a medical diagnosis situation:
  - An exhaustive set of mutually exclusive diseases $d_1, \ldots, d_n$, represented as states of a hypothesis variable D.
  - Assume that symptoms (information variables) $S_1, \ldots, S_n$ are independent when the disease is known.



Simple (or Naïve) Bayes

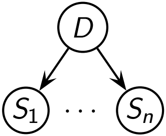
- Consider a medical diagnosis situation:
  - An exhaustive set of mutually exclusive diseases $d_1, \ldots, d_n$, represented as states of a hypothesis variable D.
  - Assume that symptoms (information variables) $S_1, \ldots, S_n$ are independent when the disease is known.



- However, the conclusion may be misleading: The assumption that information variables are independent given the hypothesis need not hold.

Simple (or Naïve) Bayes

- Consider a medical diagnosis situation:
  - An exhaustive set of mutually exclusive diseases $d_1, \ldots, d_n$, represented as states of a hypothesis variable D.
  - Assume that symptoms (information variables) $S_1, \ldots, S_n$ are independent when the disease is known.

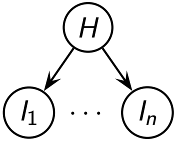


- However, the conclusion may be misleading: The assumption that information variables are independent given the hypothesis need not hold.

- Computationally and representationally a very efficient model that provides good results in many cases.

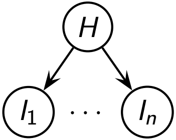
Simple Bayes Inference

- Let the possible hypotheses (e.g., diseases) be collected into one hypothesis variable H with prior $P(H)$.



Simple Bayes Inference

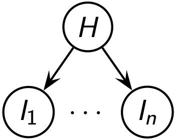
- Let the possible hypotheses (e.g., diseases) be collected into one hypothesis variable H with prior $P(H)$.



- For each information variable I, acquire $P(I \mid H) = L(H \mid I)$.

Simple Bayes Inference

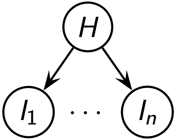
- Let the possible hypotheses (e.g., diseases) be collected into one hypothesis variable H with prior $P(H)$.



- For each information variable I, acquire $P(I \mid H) = L(H \mid I)$.
- For any set of observations $\mathcal{I} = \{i_1, \ldots, i_m\}$ calculate $L(H \mid \mathcal{I}) = \prod P(i_j \mid H)$.

Simple Bayes Inference

- Let the possible hypotheses (e.g., diseases) be collected into one hypothesis variable H with prior $P(H)$.



- For each information variable I, acquire $P(I \mid H) = L(H \mid I)$.

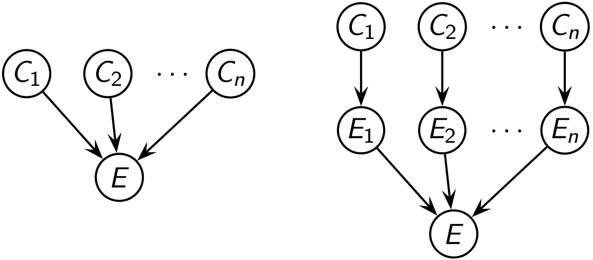
- For any set of observations $\mathcal{I} = \{i_1, \dots, i_m\}$ calculate $L(H \mid \mathcal{I}) = \prod P(i_j \mid H)$.

- The posterior $P(H \mid \mathcal{I}) = \alpha L(H \mid \mathcal{I}) P(H)$, where $\alpha = P(\mathcal{I})^{-1}$, or expressed via Bayes' rule:

$$P(H \mid \mathcal{I}) = \frac{P(\mathcal{I} \mid H) P(H)}{P(\mathcal{I})}.$$

Independence of Causal Influence

- Exploit knowledge about the internal structure of CPTs to reduce the complexity of representation and inference.



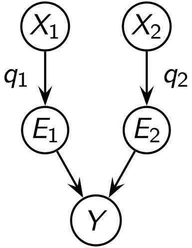
- The contribution from C_i to E is assumed to be independent of the contribution from C_j ($i \neq j$).
- C_i results in E unless it is inhibited by "something".
- The inhibitors are assumed to be independent:

$$P(\text{Inhibitor}_i, \text{Inhibitor}_j) = P(\text{Inhibitor}_i)P(\text{Inhibitor}_j) = q_i q_j.$$

The Noisy-OR Interaction Model

Let X_1, X_2 be causes of the effect variable Y and let all variable be Boolean.

		Y	
X_1	X_2	off	on
off	off	1	0
off	on	0.2	0.8
on	off	0.1	0.9
on	on	0.02	0.98



$$
\begin{aligned}
P(Y = \text{on} \mid X_1 = \text{on}, X_2 = \text{on}) &= 1 - P(Y = \text{off} \mid X_1 = \text{on}, X_2 = \text{on}) \\
&= 1 - \prod_{i=1}^{2} q_i .
\end{aligned}
$$

Example: Noisy OR

Assume that the causal influences of *cold* and *angina* on *sore throat* can be assumed to be independent. Also, assume that there is a "background" event that can cause the throat to be sore.

- The "background" event causes sore throat with probability 0.05.
- Cold causes sore throat with probability 0.4.
- Angina causes sore throat with probability 0.7.

Example: Noisy OR

Assume that the causal influences of *cold* and *angina* on *sore throat* can be assumed to be independent. Also, assume that there is a "background" event that can cause the throat to be sore.

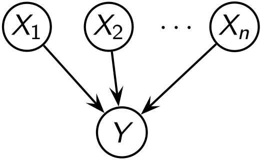
- The "background" event causes sore throat with probability 0.05.
- Cold causes sore throat with probability 0.4.
- Angina causes sore throat with probability 0.7.

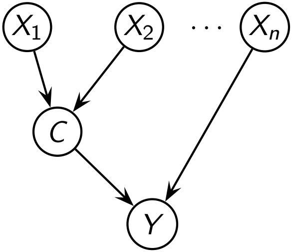
Then $P(\mathrm{sore}\,|\,\mathrm{cold}, \mathrm{angina})$ can be described as a *noisy-OR* function:

- For each cause we just need to specify a single number, namely the inhibitor probability. In the example, we have inhibitor probabilities 0.95, 0.6, and 0.3, respectively.
- The number of parameters needed grows only linearly with the number of parents.

Parent Divorcing

- Let $X_1, \ldots, X_n$ be a set of causes of Y:





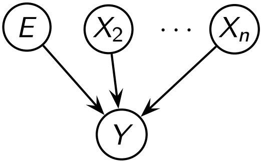
- The assumption is that the configurations of (X_1, X_2) can be partitioned into sets $c_1, \ldots, c_m$ such that $(x_1, x_2), (x'_1, x'_2) \in c_i$:

$$P(y \mid x_1, x_2, x_3, \ldots, x_n) = P(y \mid x'_1, x'_2, x_3, \ldots, x_n).$$

- Divorcing is (representationally) efficient, if $|C| < |X_1| \cdot |X_2|$.

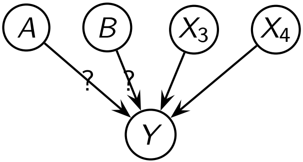
Experts Disagreement

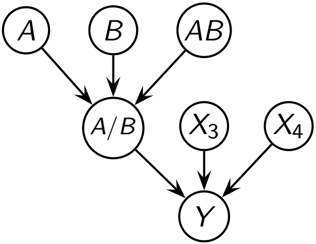
- Assume you have a number of experts $e_1, \ldots, e_m$.
- All have an opinion about $P(Y \mid X_1, \ldots, X_n)$.
  - $P(Y \mid X_1, \ldots, X_n, e_i)$ for each $i = 1, \ldots, m$.



- Your prior belief in the experts is $P(E)$ where E has one state for each expert.

- Consider a variable Y with certain parents X_3 and X_4: uncertain whether A or B is parent.



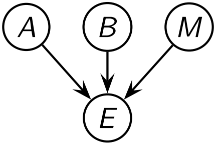


- $P(A/B \mid A, B, AB)$ is deterministic.

- $P(AB)$ encodes our belief about whether A or B is parent.

Functional Uncertainty

- Consider a situation where the parent set is known, but the functional relation is unknown.
- E is either $E = A \vee B$ or $E = A \wedge B$.



- Create node M with two states *and* and *or*.
- Let M be a model node of $P(E \mid A, B)$.
- $P(M)$ specifies our prior belief in the dependence.

Model Verification

- Very important to get the directions of the links right.
- The links and their directions determine the dependence and independence relations encoded in the DAG.
- Wrong dependence and independence relations may lead to faulty conclusions.
- The model structure can be verified by checking the dependence and independence relations.
- Use d-separation to uncover the relations.

Model Verification: A Simple Example

Assume we have the following three Boolean variables:

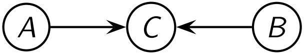
- A: Two or more PC's bought within a few days using the same credit card.
- B: Card used almost at the same time at different locations.
- C: Fraud.

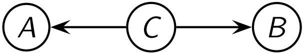
Model Verification: A Simple Example

Assume we have the following three Boolean variables:

- *A*: Two or more PC's bought within a few days using the same credit card.
- *B*: Card used almost at the same time at different locations.
- *C*: Fraud.

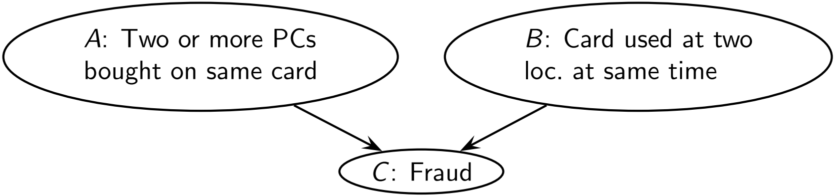
Two possible models:

- 

- 

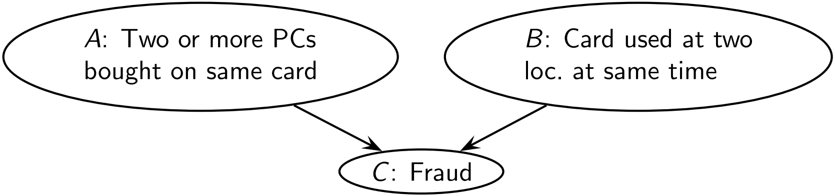
Which model is correct?

Fraud: Model 1



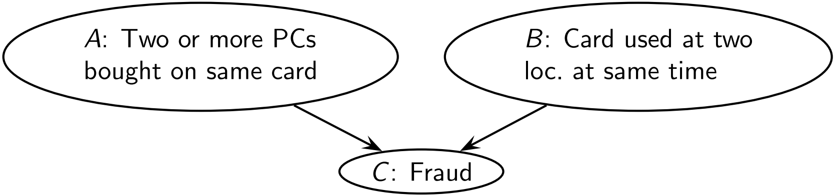
This model tells us that observing A (or B) does not provide us with any information about B (or A) when C is unknown.

Fraud: Model 1



This model tells us that observing A (or B) does not provide us with any information about B (or A) when C is unknown. *Wrong!*

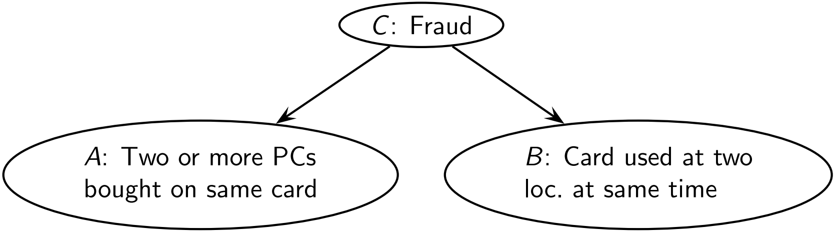
Fraud: Model 1



This model tells us that observing A (or B) does not provide us with any information about B (or A) when C is unknown. *Wrong!*

Observing A (or B) increases our belief in C, which in turn increases our belief that we might also observe B (or A). Therefore, this model gives wrong probabilities!

Fraud: Model 2



This model rightfully tells us that

- A and B are dependent when we have no hard evidence on C: Observing A (or B) will increase our belief that we will also observe B (or A), and
- A and B are independent when C is known: If we know we are considering a fraud case, then observing A (or B) will not change our belief about whether or not we are going to observe B (or A).

Summary

- Mediating variables
- Simple Bayes model
- Constructing Bayesian networks
  - Undirected relations
  - Simple bayes
  - Independence of causal influence
  - Parent divorcing
  - Experts disagreement
  - Structural uncertainty
  - Functional uncertainty
  - Model verification
- Other issues
  - Object-oriented Bayesian networks
  - Dynamic Bayesian networks
  - Continuous variables
  - ...