CS6462

Probabilistic and Explainable AI

Lesson 17

Bayesian Networks

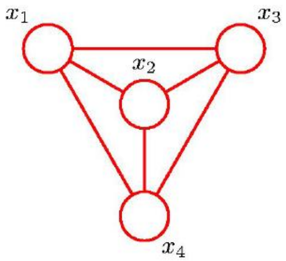
*

Theoretical Foundations of Bayesian Networks

by Emil Vassev

March 10, 2025

Theoretical Background



Probability and uncertainty:

- *uncertainty*: lack of certainty or sureness of an event
- *probability*: a number that reflects the chance or likelihood that an event will happen
- *real-life example*: know the probability of a specific disease if we observe symptoms
- *reversed conditional probability*: probability of the evidence given the cause

Bayesian approach:

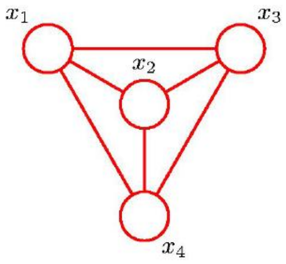
- Bayesian networks: useful tools for dealing with uncertainty, complexity and causality

Definition:

- A Bayesian network represents the causal probabilistic relationship among a set of *random variables*, their *conditional dependences*, and it provides a compact representation of a *joint probability distribution*.*

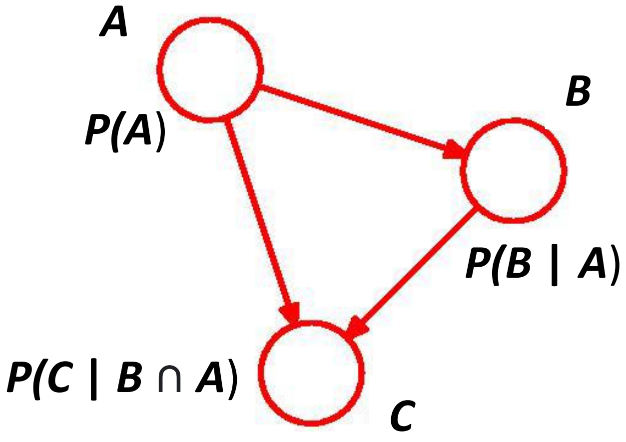
* Murphy K. (1998): A Brief Introduction to Graphical Models and Bayesian Networks.

Structure

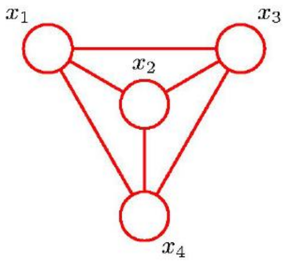


Bayesian network:

- a directed acyclic graph: **DAG:** $\mathcal{G}(V, E)$ - nodes V and edges E
  - a set of random variables represented by nodes $V \equiv X = \{A, B, C\}$
  - directed edges – connect two random variables with causal probabilistic dependency
  - directed edge from a node A to a node B :
    random variable A causes random variable B
  - cycles not allowed in the graph



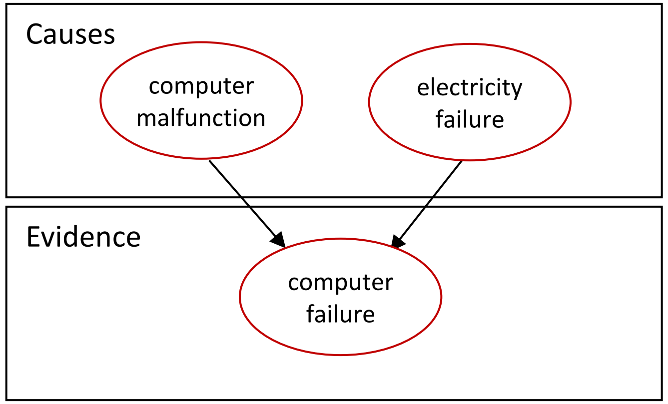
- a set of conditional probability distributions
  - conditional probability distribution is defined for each node in the graph
  - the conditional probability distribution of a node (random variable) is defined for every possible outcome of the preceding causal node(s)



Example

*Computer failure**:*

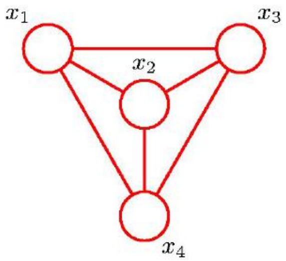
- computer does not start (observation/evidence)
- possible causes of failure – 1) electricity failure; and 2) computer malfunction
- directed acyclic graph representing two independent possible causes of a computer failure



$$P(Cause|Evidence) = \frac{P(Evidence|Cause) * P(Cause)}{P(Evidence)}$$

- electricity failure and computer malfunction are ancestors/parents of computer failure
- objective (inference): calculate the posterior conditional probability distribution of each of the possible unobserved causes given the observed evidence

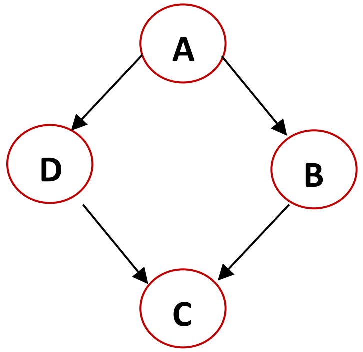
* Cowel R. G., Dawid A. P., Lauritzen S. L., Spiegelhalter D. J. (1999): Probabilistic Networks and Expert Systems.



Local Markov Property

Bayesian networks: satisfy the property known as the Local Markov Property:

- a node is conditionally independent of its non-descendants, given its parents
- *example*: $P(D|A \cap B)$ is equal to $P(D|A)$ because D is independent of its non-descendent B

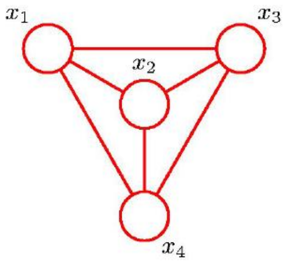


$\mathrm{D} \perp \mathrm{B}$ – notation for independent variables

joint distribution: $\boldsymbol{P}$ over $V \equiv X = \{X_1, X_2, \dots, X_n\}$

- simplifies the Joint Distribution
- leads to the concept of a Markov Random Field which is a random field around a variable that is said to follow Markov properties

Joint Probability Distribution (Revision)

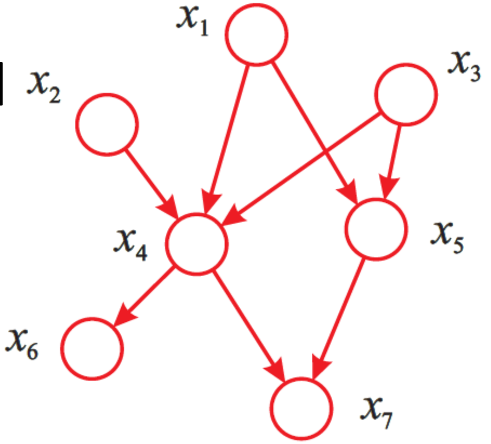


Features:

- *node*: conditionally independent of its non-descendants given that node's parents
- joint probability distribution of all random variables in the graph:
  - factorizes into a series of conditional probability distributions of random variables given their parents
  - joint distribution over ***k*** random variables factorizes:

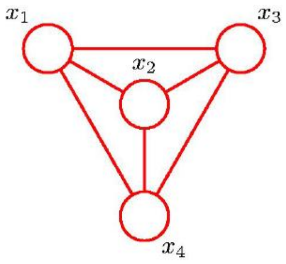
$$P(X_1 \cap X_2 \cap \ldots \cap X_k) = P(X_1) * P(X_1|X_2) * P(X_3|X_1 \cap X_2) * \ldots * P(X_k|X_1 \cap X_2 \cap \ldots \cap X_{k-1})$$

- *absence of links (Local Markov Property)*: graph is not fully connected



$$P(X_1 \cap X_2 \cap \ldots \cap X_7) = P(X_1) * P(X_2) * P(X_3) * P(X_4|X_1 \cap X_2 \cap X_3) *$$
$$P(X_5|X_1 \cap X_3) * P(X_6|X_4) * P(X_7|X_4 \cap X_5)$$

Joint Probability Distribution (Revision-cont.)



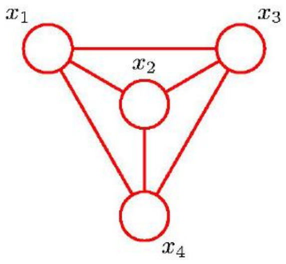
Factorization property:

- joint distribution defined by a graph is given by the product of conditional probability distributions for each node conditioned on its parents:

$$P(X_1 \cap X_2 \cap \ldots \cap X_k) = P(X_1) * P(X_1|X_2) * P(X_3|X_1 \cap X_2) * \ldots. * P(X_k|X_1 \cap X_2 \cap \ldots \cap X_{k-1})$$

$$P(X_1 \cap X_2 \cap \cdots \cap X_k) = \prod_{i=1}^{k} P(X_i|Parents(X_i))$$

- $Parents(X_i)$ is a function that returns the joint probability distribution of the parents of node X_i
- this equation expresses a key factorization property of the joint distribution for Bayesian Networks



Marginal Distribution (Revision)

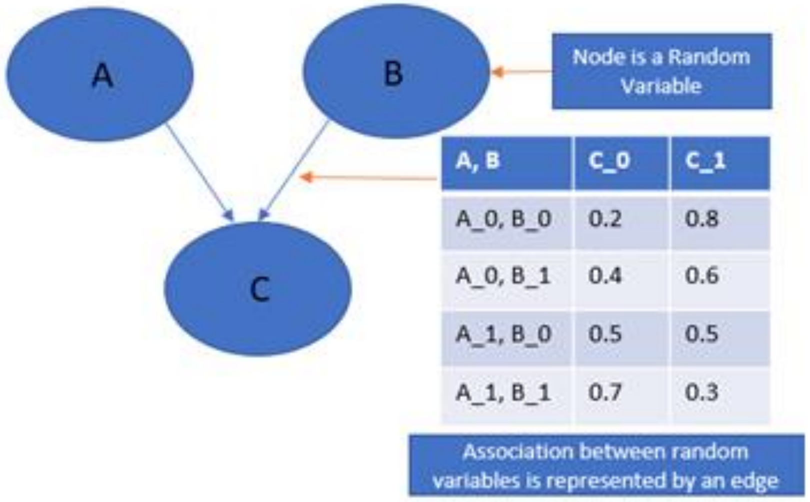
Definition & calculation:

- *definition*: distribution over a single random variable (or a subset of variables) from a larger set of variables, without any reference to an observed set of variables
- *calculation*: the marginal distribution for variable $\boldsymbol{X_i}$ is:

$$P(X_i) = \sum_{X_j \neq X_i,} P(X_1 \cap X_2 \cap \cdots \cap X_k), \text{ for } \forall\, X_j = \{x_1, \ldots.\ x_n\}$$

where the sum is over the states of all variables $\boldsymbol{X_j \neq X_i}$ and can be computed by the sum of the products

Example: calculate the marginal distribution for ***C***

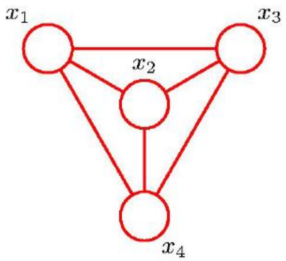


A, B	C_0	C_1
A_0, B_0	0.2	0.8
A_0, B_1	0.4	0.6
A_1, B_0	0.5	0.5
A_1, B_1	0.7	0.3

$$P(C) = 0.3 + 0.24 + 0.12 + 0.34 = 1$$

A	B	C	Probability
A_0	B_0	C_0	0.15
A_0	B_0	C_1	0.15
A_0	B_1	C_0	0.08
A_0	B_1	C_1	0.16
A_1	B_0	C_0	0.05
A_1	B_0	C_1	0.07
A_1	B_1	C_0	0.02
A_1	B_1	C_1	0.32

A	B	Probability
A_0	B_0	0.3
A_0	B_1	0.24
A_1	B_0	0.12
A_1	B_1	0.34



Bayesian Network Inference

Given a Bayesian network, what queries might we want to ask?

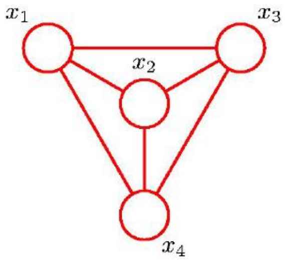
- ***Parents(X) = evidence***
- *general query*: What's the marginal probability of each node?
- *conditional probability query*: ***P(X|Parents(X))*** - What is the probability for $\boldsymbol{X}$ for having a particular value $\boldsymbol{x}$ of ∈ $\boldsymbol{X}$?
- *max posterior probability*: How is maximizing the posterior probability ***P(X|evidence)*** equivalent to maximizing the likelihood ***P(evidence|X)*** ? What value of ***X*** maximizes the ***P(evidence|X)***?

Example:

- Given the joint distribution over three variables, we can answer any question about the probability of a single value held by another variable by summing (or marginalizing) over the first three variables.

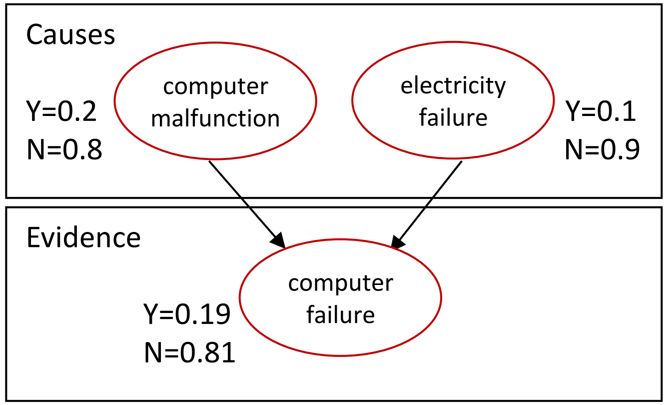
$$\boldsymbol{P(d_2) = \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} P(A = a \cap B = b \cap C = c \cap D = d_2)}$$

Inference (cont.)



Example: computer failure (Bayesian inference) – **prior probability**

- notation: electricity failure = A, computer malfunction = B, computer failure = C
- A = {yes, no}; B = {yes, no}; C={yes, no}; $A \perp B$ (independent)
- P(A=yes) = 0.1; P(A=no) = 0.9; P(B=yes) = 0.2; P(B=no) = 0.8
- P (C=yes | A=no ∩ B=no) = 0
- P (C=yes | A=no ∩ B=yes) = 0.5 (possible not to work)
- P (C=yes | A=yes ∩ B=no) = 1 (100% does not work)
- P (C=yes | A=yes ∩ B=yes) = 1

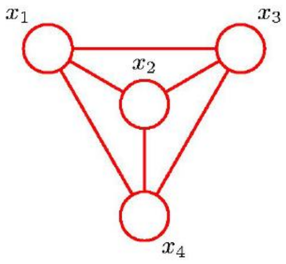


$$\boldsymbol{P(c_1) = \sum_{a \in A} \sum_{b \in B} P(A = a \cap B = b \cap C = c_1)}$$

- P (C=yes) $=\sum_{A,B}$ P(C=yes ∩ A ∩ B) = $\sum_{A,B}$ P(C = yes | A ∩ B) ∗ P(A) ∗ P(B)
- P (C=yes) = 0+0.5*0.9*0.2+1*0.1*0.8+1*0.2*0.1=0.09+0.08+0.02
- P (C=yes) = 0.19

prior probability distribution - before observing any evidence

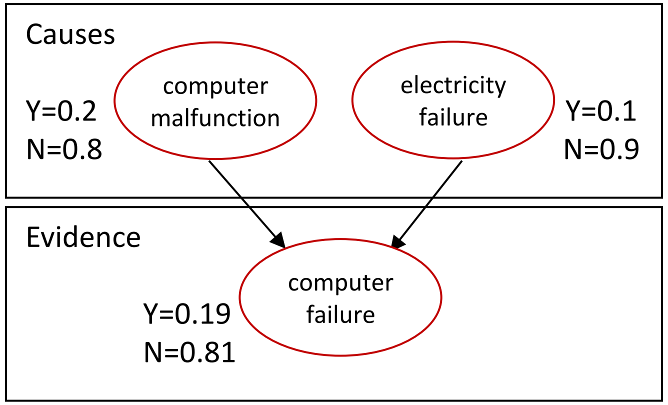
Inference (cont.)



$$P(Cause|Evidence) = \frac{P(Evidence|Cause) * P(Cause)}{P(Evidence)}$$

Example: computer failure (Bayesian inference) - **posterior probability**

- notation: electricity failure = A, computer malfunction = B, computer failure = C
- $A = \{yes, no\}$; $B = \{yes, no\}$; $C=\{yes, no\}$; $A \perp B$ (independent)
- $P(A=yes)=0.1$; $P(A=no)=0.9$; $P(B=yes)=0.2$; $P(B=no)=0.8$
- $P(C=yes)=0.19$; $P(C=no)=0.81$
- $P\,(C=yes|\ A \cap\ B\,) = 1$
- $P\,(A=yes \mid C=yes) = ?$
- $P\,(A=yes \mid C=yes) = P\,(A=yes\ \cap\ B \mid C=yes)$



$P\,(A=yes\ \cap B \mid C= yes) = \sum_{B=\{yes,no\}} P(C=yes|\ A=yes \cap B) * P(A=yes \cap B)\ /\ P(C=yes) =$

$P(C=yes|\ A=yes \cap B=yes) * P(A=yes \cap B=yes)\ /\ P(C=yes) +$

$P(C=yes|\ A=yes \cap B=no) * P(A=yes \cap B=no)\ /\ P(C=yes) =$

$$\sum_{B=\{yes,no\}} \frac{P(C=yes|\ A=yes \cap B) * P(A=yes) * P(B)}{P(C=yes)} = 1*0.1*0.2/0.19+ 1*0.1*0.8/0.19 = 0.526$$

posterior probability distribution, i.e. after observing evidence (computer failure)

Summary

Bayesian Networks – *Theoretical Foundations of Bayesian Networks*

Structure

Local Markov Property

Joint Probability Distribution

Marginal Distribution

Inference

Next Lesson:

- Bayesian Networks - Construction of Bayesian Networks

Thank You!

Questions?