VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

School of Engineering and Computer Science

COMP 307 — Lecture 15

Uncertainty and Probability 3

Introduction to Bayesian Networks

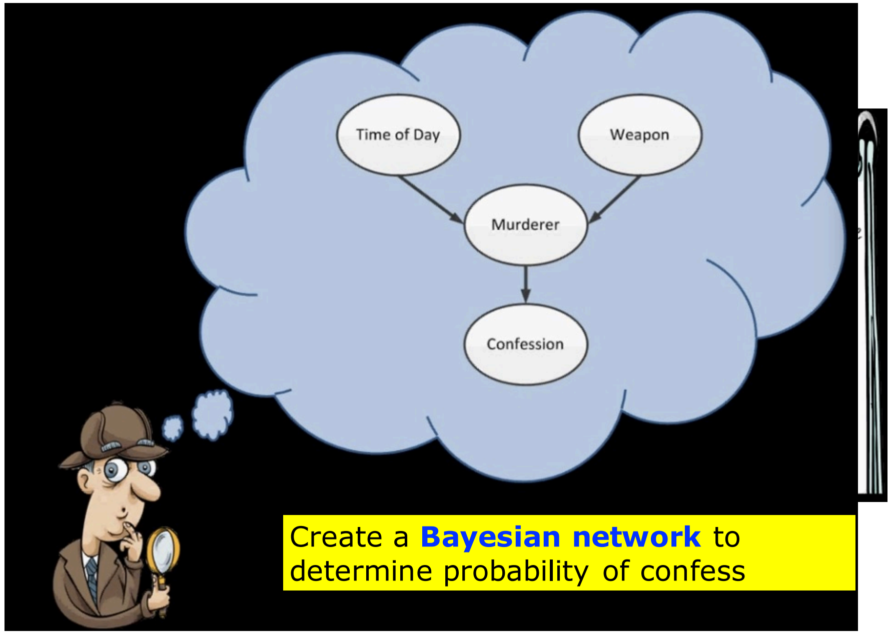
Dr Bing Xue

bing.xue@ecs.vuw.ac.nz

Outline

- Rules from previous lectures
- What is Bayesian Networks
- Why Bayesian Networks
- Cause — Effect
- Multiple causes
- Semantics of Bayesian Networks
- Summary

Thomas Bayes (/berz/; c. 1701 – 7 April 1761)



Previous Lectures

- Product Rule:
  P(X,Y)=P(X)*P(Y|X)
- Sum Rule:
  $P(X) = \sum_y P(X, Y)$
- Normalisation:
  $\sum_x P(X) = 1$
  $\sum_x P(X/Y) = 1$
- Independence
  - ↔ P(X|Y) = P(X)
  - ↔ P(X, Y) = P(X) * P(Y)

- Bayes Rules:

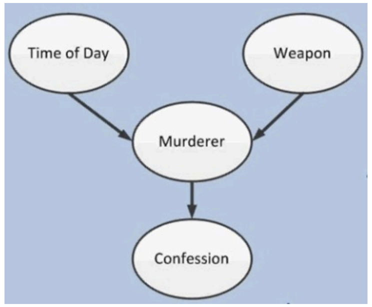
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

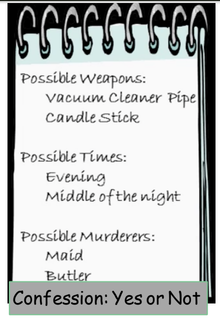
- Naive Baye
- Conditionally Independent
  - X and Y are conditionally independent given Z
  - ↔ P(Y,X|Z)=P(Y|Z)*P(X|Z)
  - ↔ P(X|Y,Z) = P(X|Z)

Bayesian Networks

- Bayesian networks(BNs) : a graphical representation of a probabilistic dependency model
  - also known as Belief networks (or Bayes nets for short)
  - Belong to the family of probabilistic graphical models (GMs).
- These graphical structures are used to represent knowledge about an uncertain domain. In particular,
  - each node in the graph represents a random variable,
  - the edges between the nodes represent probabilistic dependencies among the corresponding random variables.
  - The conditional dependencies in the graph are often estimated by using known statistical and computational methods.
- BNs combine principles from graph theory, probability theory, computer science, and statistics.

Bayesian Networks and Examples





- Each node or variable may take one of a number of possible states or values.
- The belief in, or certainty of, each of these values is determined from the belief in each possible value of every node directly connected to it and its relationship with each of these nodes.
- The belief in each state of a node is updated whenever the belief in each state of any directly connected node changes.

Why Bayesian Networks ?

- Several advantages for data analysis:
  - the model encodes dependencies among all variables, it readily handles situations where some data entries are missing.
  - a Bayesian network can be used to learn causal relationships, and hence can be used to gain understanding about a problem domain and to predict the consequences of intervention.
  - the model has both a causal and probabilistic semantics, it is an ideal representation for combining prior knowledge (which often comes in causal form) and data.
  - Bayesian statistical methods in conjunction with bayesian networks offer an efficient and principled approach for avoiding the overfitting of data.

Heckerman, David. A tutorial on learning with Bayesian networks. Springer Netherlands, 1998.

Cause — Effect

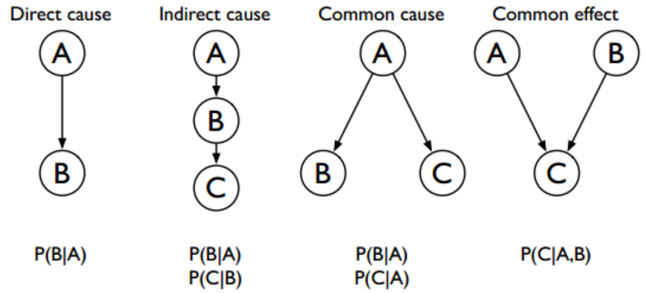
- A Cause is why something happens. An Effect is what actually happens.
  - A ball is dropped so it hits the ground.
    - Hitting the ground is an effect.
    - Dropping the ball is a cause of it hitting the ground
  - Cause: flue, Effect: High Temperature
- Causal Reasoning: solving a problem where only cause is known
  - P(Effect |Cause)
- Diagnostic Reasoning: reasoning about Cause when Effect is known
  - P(Cause | Effect)
- Inter-causal Reasoning: reasoning about the interactions between multiple causes influences

Cause — Effect

- P(Cause | Effect) = P(Cause) * P(Effect | Cause) / P(Effect)
  - P(Cause) often called prior , which is before the evidence came along.
  - P(Cause | Effect) is known as the posterior, meaning the belief after the evidence.
  - P(Effect | Cause) known as the likelihood.
- Bayes Rules:

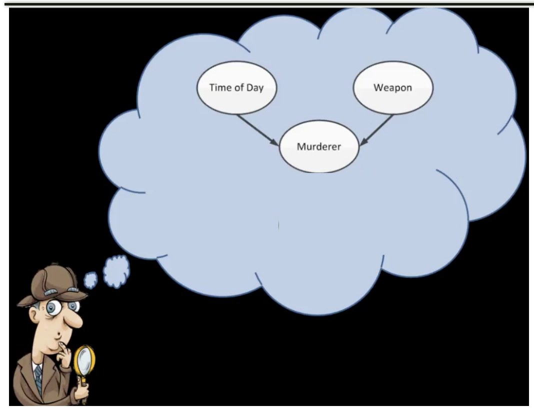
$$\underbrace{P(A|B)}_{\text{Posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}}\overbrace{P(A)}^{\text{Prior}}}{\underbrace{P(B)}_{\text{Marginal likelihood or Just nomarlisation}}}$$

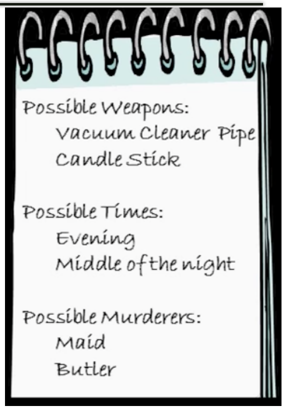
Cause — Effect



- Common effect (multiple causes, or "explaining away" — Suppose that there are exactly two possible causes of a particular effect, represented by a v-structure)

A Lazy Detective





- Three variables:
  - Time: E-evening, N-night
  - Weapon: V-vacuum, S-candle stick
  - Murder: M-maid, B-bulter

A Lazy Detective

- ***Time****: E-evening, N-night;* ***Weapon****: V-vacuum, S-candle stick*
- ***Murder****: M-maid, B-bulter*
- Report from the Lab:
  - P(T=E)=0.05, P(T=N)=0.95
  - P(W=V)=0.8, P(W=S) =0.2
- From Detective

Time	Weapon	P(M=M|T,W)
E	V	0.9
E	S	0.55
N	V	0.35
N	S	0.05

- Time:
  - The ***maid*** not likely committed a murder in the middle of the night. The butler is quite likely
- Weapon:
  - a vacuum cleaner - maid
  - a candle stick - butler
  - although it is possible for one employee to use the others's tool to commit the murder, it is not likely

A Lazy Detective

- Probability of Maid is the murder

P(M=M) = P(M,E,V)+P(M,E,S)+P(M,N,V)+P(M,N,S)

= P(M|E,V)*P(E,V) + P(M|E,S)*P(E,S)

+P(M|N,V)*P(N,V) + P(M|N,S)*P(N,S)

= P(M|E,V)*P(E)*P(V) + P(M|E,S)*P(E)*P(S)

+ P(M|N,V)*P(N)*P(V) + P(M|N,V)*P(N)*P(S)

= P(M|E,V)*P(E)*P(V) + P(M|E,S)*P(E)*P(S)

+ P(M|N,V)*P(N)*P(V) + P(M|N,V)*P(N)*P(S)

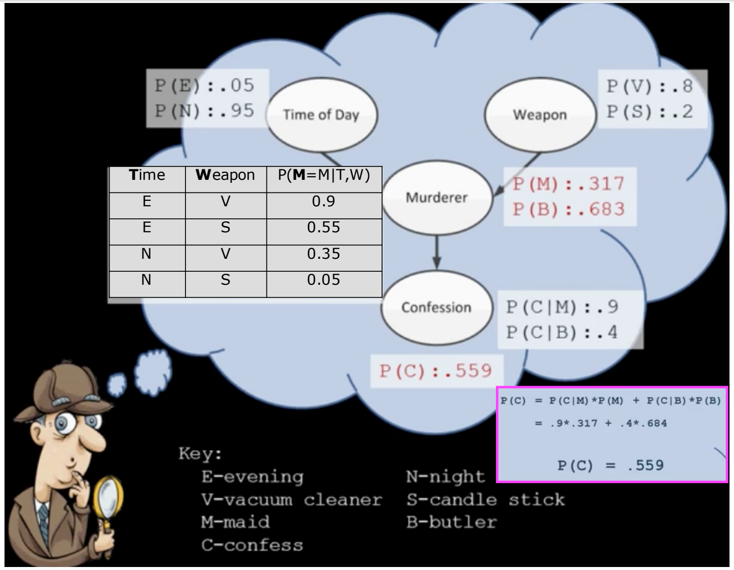
= 0.9*0.5*0.8 + 0.55*0.5*0.2

+0.35*0.95*0.8 + 0.05*0.9580.2

= 0.317

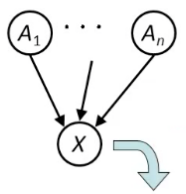
- Which rules are used here ?
- What is P(M=B) ?

A Lazy Detective



Semantics of Bayesian Networks

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution table for each node
  - a collection of distributions over X, one for each combination of parents values
    $P(X|a_1 \ldots a_n)$
  - CPT: conditional probability table
  - (usually) description of a "causal" process



$P(X|A_1 \ldots A_n)$

A Bayes Net = Topology (graph) + Local Conditional Probabilities

Bayesian Networks

- Weapon and Time are independent if Murder is unknown.
- Are Weapon and Time still independent if Murder is know ?
- Next Lecture: probablitis in BN and How to Build a BN