MONASH University
Information Technology

MONASH University
Information Technology

FIT 3080: Intelligent Systems

Bayesian Networks: Representation

Chapter 14

Many slides are adapted from Stuart Russell, Andrew Moore, or Dan Klein

Assumptions about the Environment

- **Fully /partially observable**
- **Known**
- **Single/multi agent**
- **Stochastic**
- **Sequential/episodic**
- **Static**
- **Discrete/continuous**

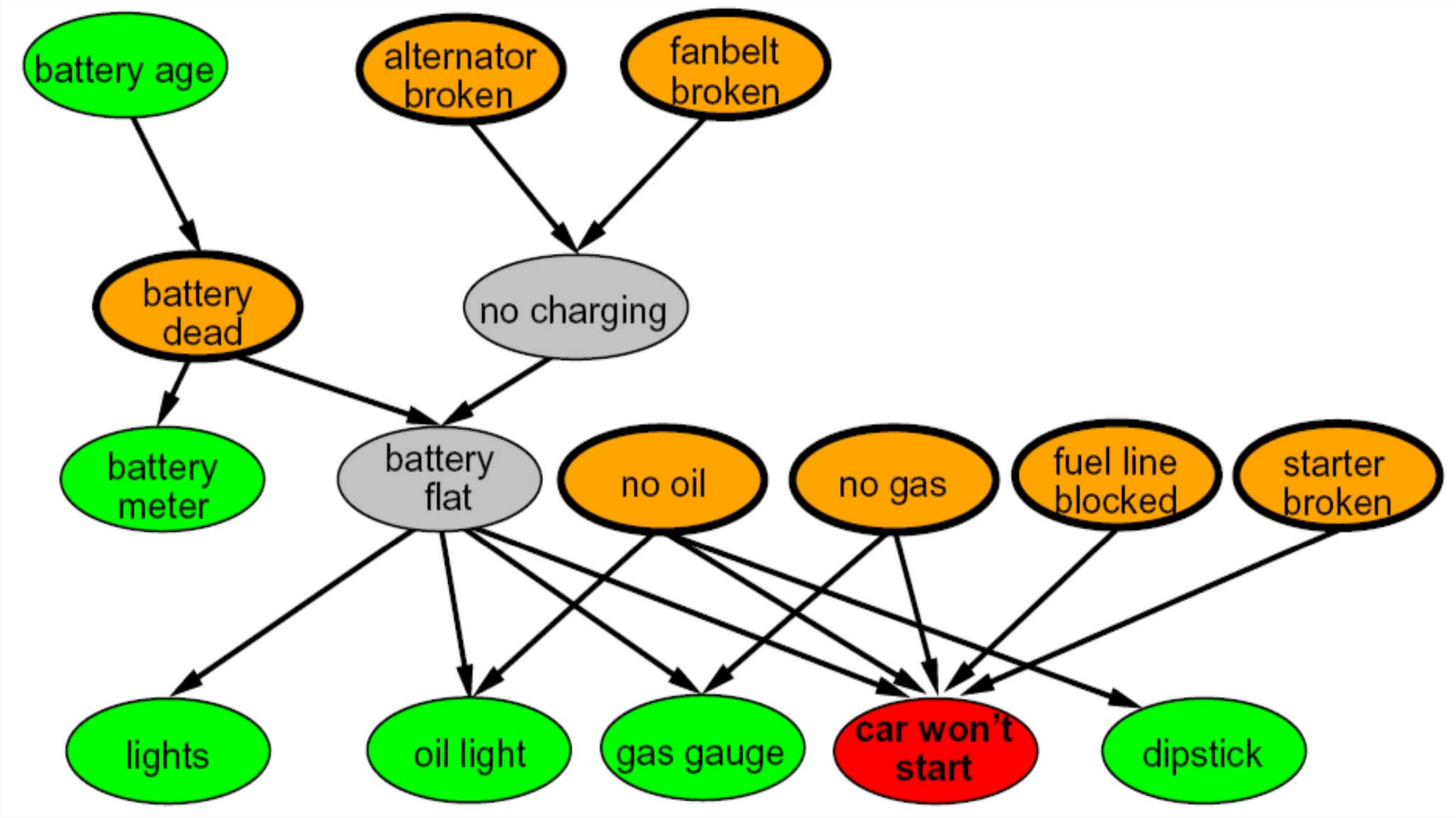
Bayesian Conception of an AI

- **An autonomous agent that**
  – has a utility structure (preferences)
  – can learn about its world and the relationship (probabilities) between its actions and future states
  – maximizes its expected utility
- **The techniques used to learn about the world are mainly statistical**
  ➔Machine learning

Bayesian Networks: The Big Picture

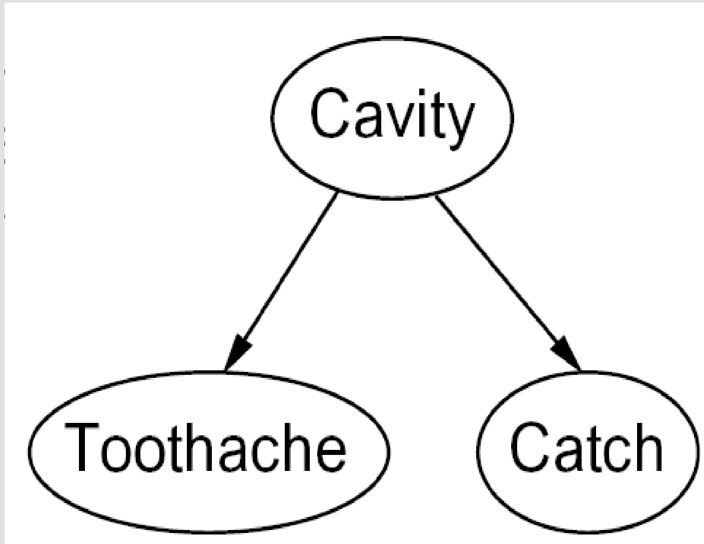
- **Two problems with using full joint distribution tables as our probabilistic models:**
  - Unless there are only a few variables, the joint is too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time
- **Bayes nets (aka graphical models): a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)**
  - Describe how variables interact locally
    - Local interactions chain together to give global, indirect interactions

Example Bayesian Network: Car



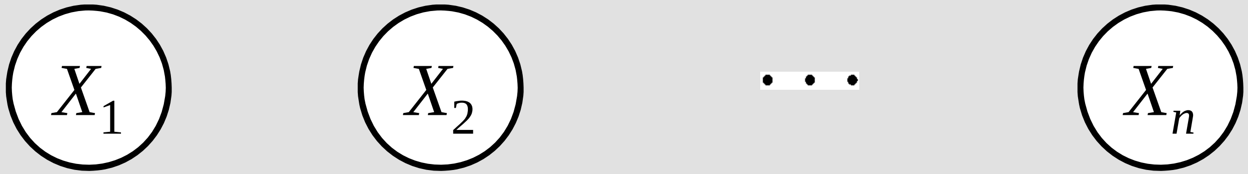
Graphical Model – Notation

- **Nodes: variables (with domains)**
  - Can be assigned (observed) or unassigned (unobserved)
- **Arcs: interactions**
  - Similar to CSP constraints
  - Indicate "direct influence" between variables
  - Formally: encode conditional independence
- **For now, imagine that arrows mean direct causation**



Example: Coin Flips (I)

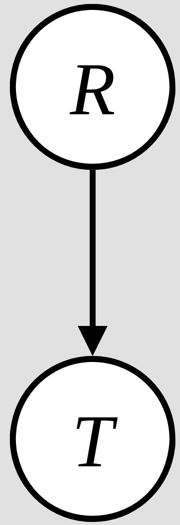
- **N independent coin flips**



- **No interactions between variables: absolute independence**

Example: Traffic (I)

- **Variables:**
  - R: It rains
  - T: There is traffic



- **Model 1: independence**

- **Model 2: rain causes traffic**

- **Why is model 2 better?**

Bayesian Networks – Definition (I)

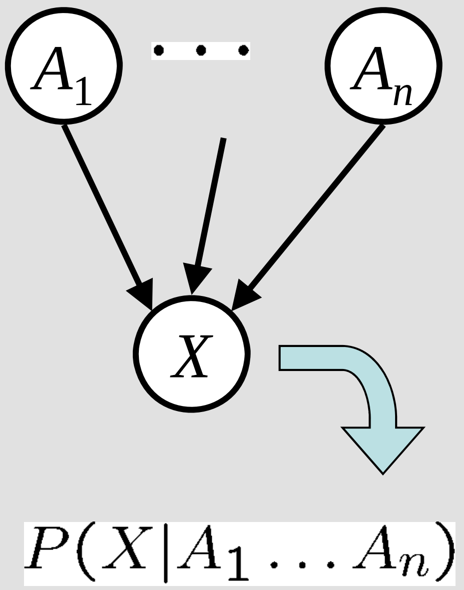
- **A data structure that represents the dependence between random variables**
- **A Bayesian Network is a directed acyclic graph (DAG) in which the following holds:**
  1. A set of random variables makes up the nodes in the network
  2. A set of directed links connects pairs of nodes
  3. Each node has a probability distribution that quantifies the effects of its ***parent nodes***
- **Gives a concise specification of the *joint probability distribution* of the variables**

Bayesian Networks – Definition (II)

- **The probability distribution for each node X is a collection of distributions over X, one for each combination of its parents' values**

$$\mathbf{Pr}(\boldsymbol{X}|\boldsymbol{a_1}, \dots, \boldsymbol{a_n})$$

  - described by a ***Conditional Probability Table*** **(*CPT*)**
  - describes a "noisy" causal process

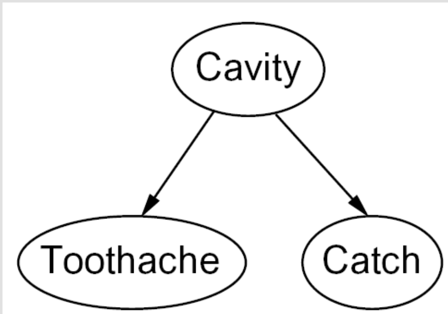


Bayesian network = Topology (graph) + Local Conditional Probabilities

Probabilities in BNs

- **Bayes nets implicitly encode joint distributions**
  - As a product of local conditional distributions
- **To see what probability a BN gives to a full assignment, multiply all the relevant conditionals**

$$\Pr(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \Pr(x_i \mid \text{parents}(X_i))$$



  - Example: Pr(+cavity, +catch, ¬toothache)
- **This lets us reconstruct any entry of the full joint distribution**

Example: Coin Flips (II)

X_1 X_2 $\cdots$ X_n

$\Pr(X_1)$

h	0.5
t	0.5

$\Pr(X_2)$

h	0.5
t	0.5

$\cdots$

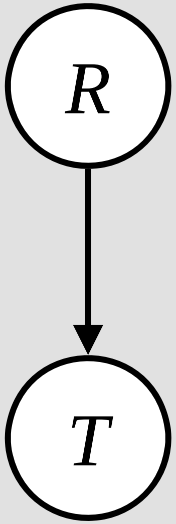
$\Pr(X_n)$

h	0.5
t	0.5

$$\Pr(h, t, t, h) =$$

Only distributions whose variables are independent can be represented by a Bayes net with no arcs

Example: Traffic (II)



$\Pr(R)$

+r	1/4
¬r	3/4

$\Pr(T|R)$

+r	+t	3/4
+r	¬t	1/4
¬r	+t	1/2
¬r	¬t	1/2

$$\Pr(+r, \neg t) =$$

Example – Lung Cancer Diagnosis

A patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer.

The doctor knows that other diseases, such as tuberculosis and bronchitis are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive Xray would indicate either TB or lung cancer.

Nodes and Values

Q: What do the nodes represent and what values can they take?

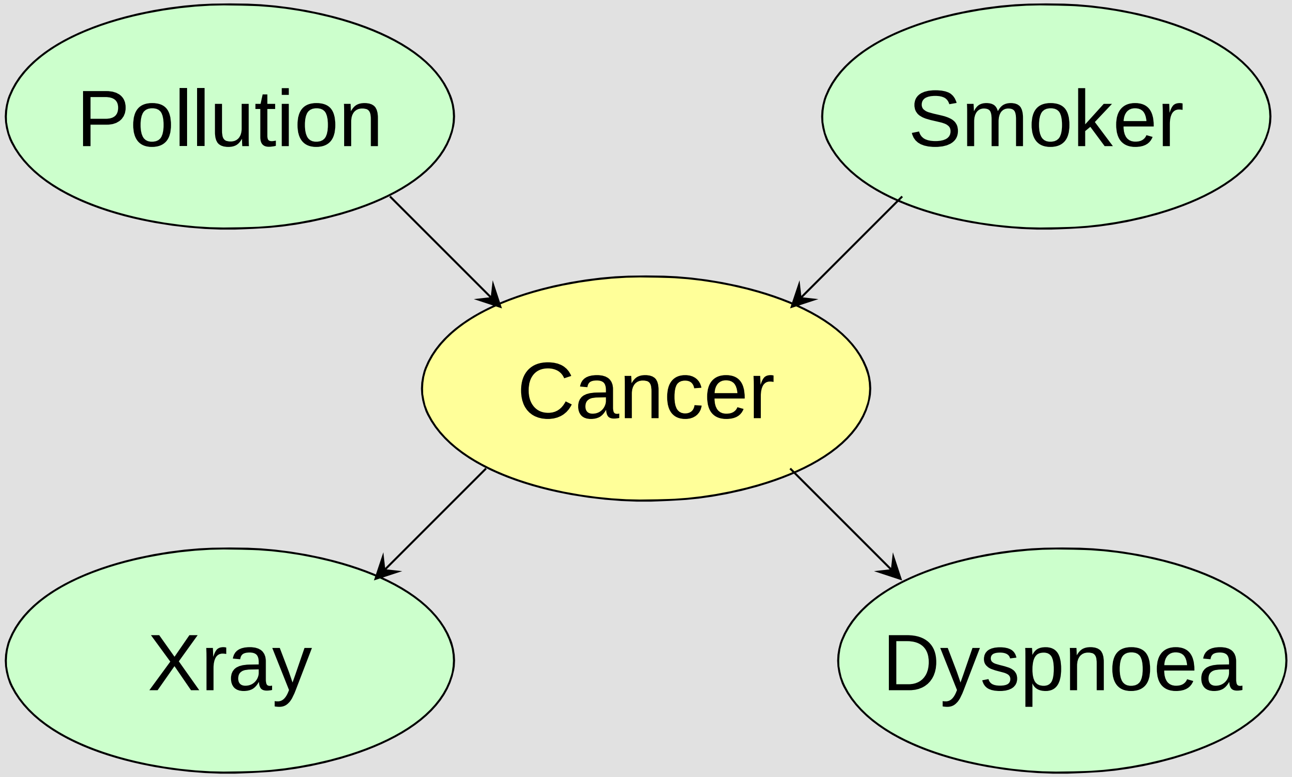
A: Nodes can be discrete or continuous

- **Binary values**
  – Boolean nodes (special case)
  Example: *Cancer* node represents proposition *"the patient has cancer"*
- **Ordered values**
  – Example: *Pollution* node with values *low, medium, high*
- **Integral values**
  – Example: *Age* with possible values 1-120

Lung Cancer Example: Nodes and Values

Node name	Type	Values
Pollution	Binary	{*low,high*}
Smoker	Boolean	{T,F}
Cancer	Boolean	{T,F}
Dyspnoea	Boolean	{T,F}
Xray	Binary	{*pos,neg*}

Lung Cancer Example: Network Structure

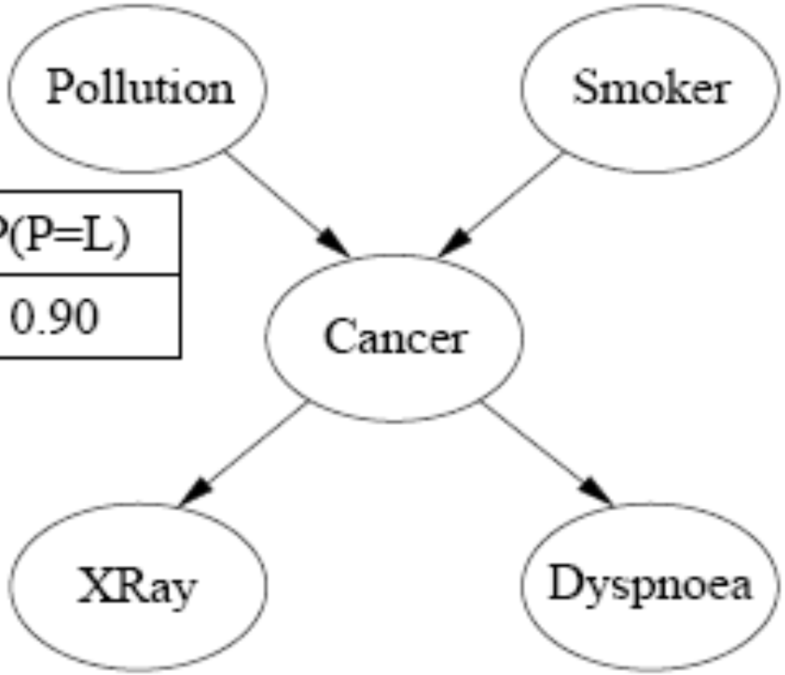


Conditional Probability Tables (CPTs)

After specifying topology, must specify the CPT for each discrete node

- **Each row contains the conditional probability of each node value for each possible combination of values in its parent nodes**
- **Each row must sum to 1**
- **A CPT for a Boolean variable with n Boolean parents contains 2^{n+1} probabilities**
- **A node with no parents has one row (its prior probabilities)**

Lung Cancer Example: CPTs



P(S=T)
0.30

P(P=L)
0.90

P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

C	P(X=pos|C)
T	0.90
F	0.20

C	P(D=T|C)
T	0.65
F	0.30

Understanding Bayesian Networks

- **Understand how to construct a network**
  - A (more compact) representation of the joint probability distribution, which encodes a collection of conditional independence statements
- **Understand how to design inference procedures**
  - Encode a collection of conditional independence statements
  - Apply the ***Markov property***
    - > *There are no direct dependencies in the system being modeled which are not already explicitly shown via arcs*
    - > Example: smoking can influence dyspnoea only through causing cancer

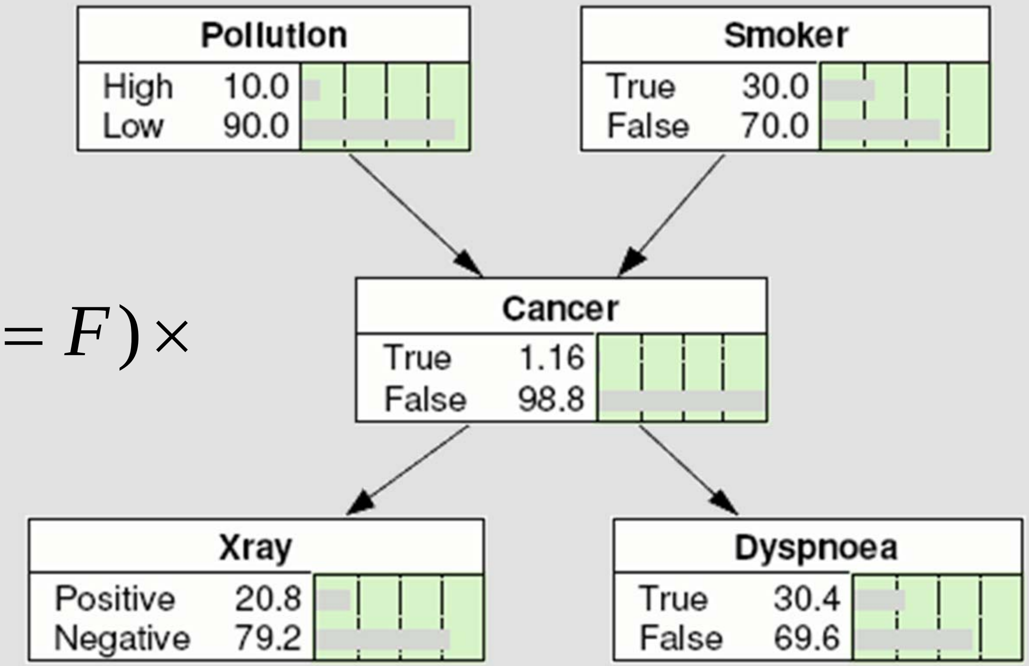
Representing Joint Probability Distribution: Example

$$\Pr(P = low \wedge S = F \wedge C = T \wedge X = pos \wedge D = T) =$$

$$\Pr(P = low) \times$$
$$\Pr(S = F) \times$$

$$\Pr(C = T \mid P = low, S = F) \times$$

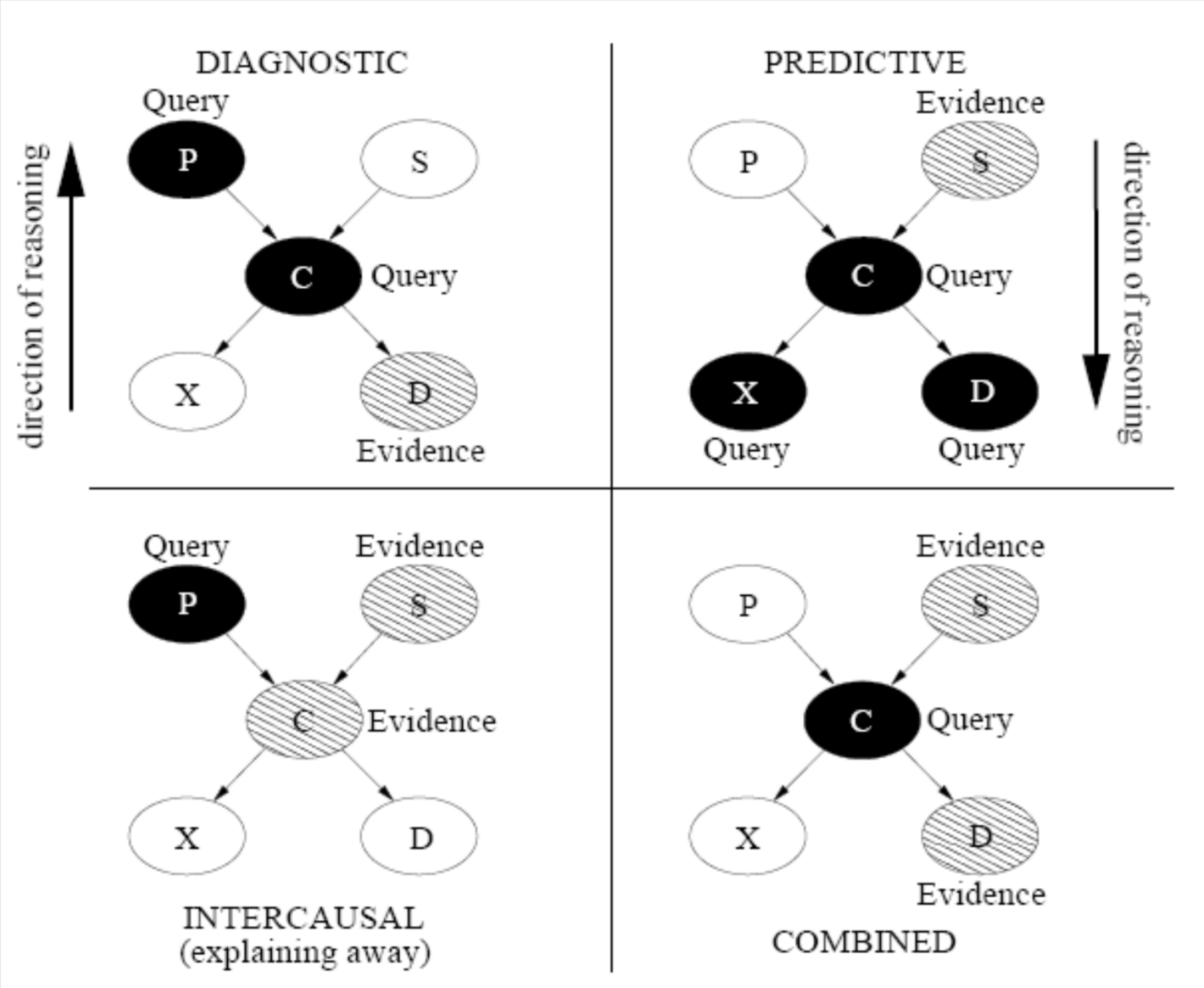
$$\Pr(X = pos \mid C = T) \times$$
$$\Pr(D = T \mid C = T)$$



Reasoning with Bayesian Networks

- **Basic task for any probabilistic inference system:**
  Compute the posterior probability distribution for a set of *query variables*, given new information about some *evidence variables*
- **Also called *conditioning* or *belief updating* or *inference***

Types of Reasoning



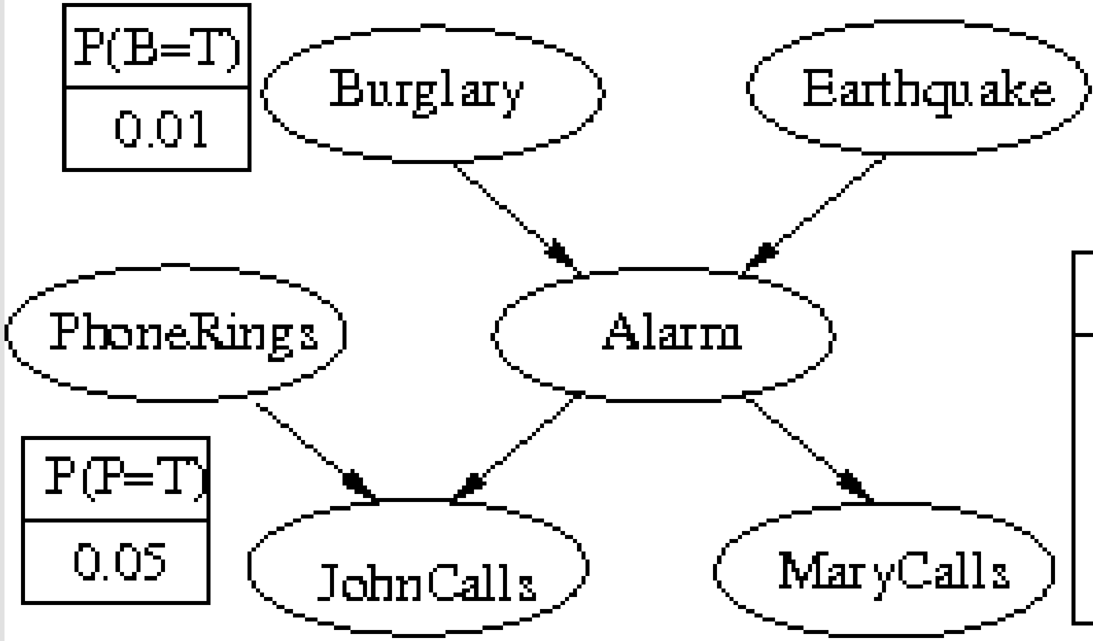
Example – Earthquake (Pearl 1988)

You have a new burglar alarm installed. It reliably detects burglary, but also responds to minor earthquakes. Two neighbours, John and Mary, promise to call the police when they hear the alarm. John always calls when he hears the alarm, but sometimes confuses the alarm with the phone ringing and calls then also. On the other hand, Mary likes loud music and sometimes doesn't hear the alarm. Given evidence about who has and hasn't called, you'd like to estimate the probability of a burglary.

Earthquake Example: Nodes and Values

Node name	Type	Values
B: Burglary	Boolean	{T,F}
A: Alarm (goes off)	Boolean	{T,F}
M: Mary calls	Boolean	{T,F}
J: John calls	Boolean	{T,F}
P: Phone rings	Boolean	{T,F}
E: Earthquake	Boolean	{T,F}

BN for Earthquake Example



P(B=T)
0.01

P(E=T)
0.02

P(P=T)
0.05

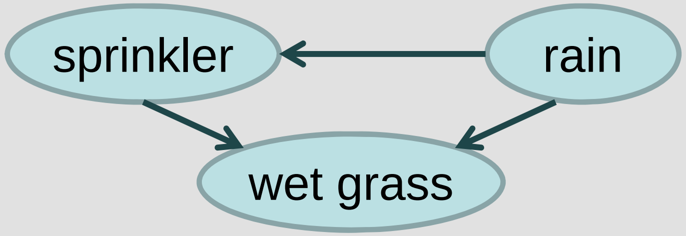
B	E	P(A=T|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

P	A	P(J=T|P,A)
T	T	0.95
T	F	0.50
F	T	0.90
F	F	0.01

A	P(M=T|A)
T	0.70
F	0.01

Causality?

- **When Bayesian networks reflect causal patterns:**
  – Often simpler (nodes have fewer parents)
  – Often easier to think about
  – Often easier to elicit from experts
- **BNs need not actually be causal, but it is good practice**



  – Arrows reflect correlation, not causation
- **What do the arrows really mean?**
  – Topology may happen to encode causal structure
  – Topology really encodes conditional independence

Reading

- **Russell, S. and Norvig, P. (2010), *Artificial Intelligence – A Modern Approach* (3nd ed), Prentice Hall**
  - Chapter 14, Sections 14.1-14.2