BT-3102

COMPUTATIONAL METHODS FOR BUSINESS ANALYTICS

L2.2 Bayesian Networks Representation – Syntax

Aditya Karanam

Outline

- Independence
- Conditional Independence
- Bayesian Network: Syntax
  - Directed Acyclic Graph
  - Conditional Probability Tables
    - Complexity

Optional Readings

- AIMA
  - Ch 13.1: Representing Knowledge in an Uncertain Domain
  - Ch 13.2: The Semantics of Bayesian Networks

Review: Inference by Enumeration

- General case:
  - Evidence variables: $(E_1 \dots E_k) = (e_1 \dots e_k)$
  - Query variables: $Y_1 \dots Y_m$
  - Hidden variables: $H_1 \dots H_r$

$$\left.\begin{array}{l} (E_1 \dots E_k) = (e_1 \dots e_k) \\ Y_1 \dots Y_m \\ H_1 \dots H_r \end{array}\right\} \underset{\text{All variables}}{X_1, X_2, \dots X_n}$$

- We want: $P(Y_1 \dots Y_m | e_1 \dots e_k)$
- First, select the entries consistent with the evidence
- Second, sum out **H**:

$$P(Y_1 \dots Y_m, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Y_1 \dots Y_m, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- Third, normalize the remaining entries to conditionalize

Review: Complexity of Inference by Enumeration

- Space Complexity?
  - $O(d^n)$ to store the joint distribution
- Worst case time complexity: $O(d^n)$
- How to make this tractable?
  - Modeling assumptions!
    - Assumptions on the relations between random variables
  - Computational tools

S	T	R	P
summer	warm	sun	0.30
summer	warm	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	warm	sun	0.10
winter	warm	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Probabilistic Models

- Models are obtained by making reasonable assumptions about the real-world aspects that are of interest
  - May not account for every variable
  - May not account for all interactions between variables

- In this course:
  - We make reasonable assumptions about the probabilistic relations among the variables
    - "All models are not correct, but some are useful!"

Independence

- Two variables are independent *iff* for all values in the domain of X and Y:

$$P(X|Y) = P(X)$$
$$P(Y|X) = P(Y)$$

- This says that their joint distribution *factors* into a product of two simpler distributions

$$P(X,Y) = P(X)P(Y|X) = P(X)\,P(Y)$$

- Mathematical notation: $X \perp\!\!\!\perp Y$

- Why does this matter?
  - Reduces the complexity tremendously!

Independence: Example - 1

- Coin flips of N *biased independent* coins:

$$P(X_1, X_2, \dots X_n)$$

2^n

$P(X_1)$

H	0.3
T	0.7

$P(X_2)$

H	0.6
T	0.4

$\cdots$

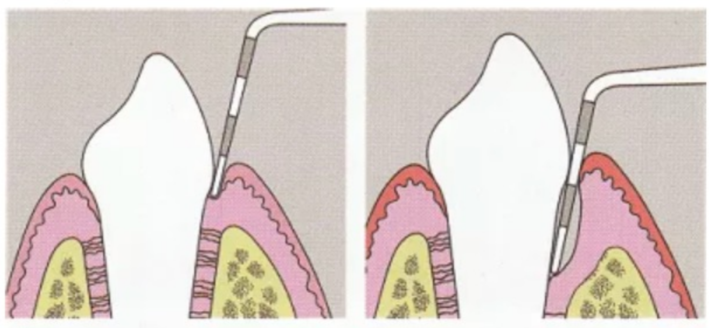
$P(X_n)$

H	0.2
T	0.8

Complexity: $O(2^n) \rightarrow O(n)$

Independence: Example – 2

- Consider that you have a toothache on a rainy day. You go to the dentist, who checks for a cavity with their probe.
- Variables: {Weather, Toothache, Catch, Cavity}
  - No. of values in the domain: Weather: 4, Other variables: 2



- How many parameters are in the joint model?
  - 4*2*2*2 – 1
- Intuitively, Weather is independent of {Toothache, Catch, Cavity}
  - How many parameters are in the independent model?
    - (4-1) + (2*2*2 – 1)

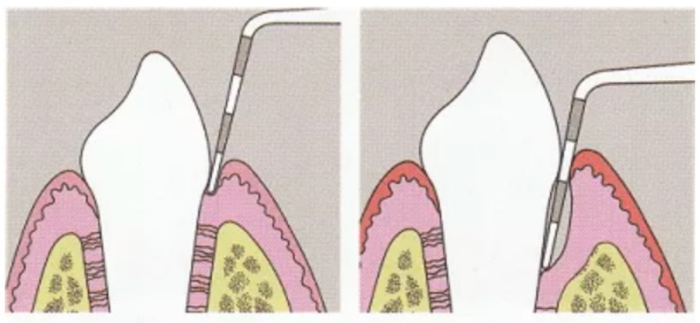
Independence: How Often We See This in Practice?

- Independence is a *modeling assumption*
  - In the real world, *Empirical* joint distributions (distributions obtained from data):
    - At best, are "close" to independent

- If the variables are really unrelated, why would we even model them?
  - There is no need to model weather to infer dental problems

- So, let us look at conditional independence!

Conditional Independence



- Example: {Toothache, Catch, Cavity}
  - Number of parameters in the joint model: $8 - 1 = 7$
- If I have a cavity, the probability that the probe catches doesn't depend on whether I have a toothache
  - $P(catch \mid toothache, cavity) = P(catch \mid cavity)$
- The same independence holds if I don't have a cavity:
  - $P(catch \mid toothache, \neg cavity) = P(catch \mid \neg cavity)$
- Does this mean Catch is independent of Toothache?
  - No! Catch is *conditionally independent* of Toothache given Cavity:
    - $P(Catch \mid Toothache, Cavity) = P(Catch \mid Cavity)$
- Equivalent statements:
  - $P(Toothache, Catch \mid Cavity) = P(Toothache \mid Cavity)\, P(Catch \mid Cavity)$
  - $P(Toothache \mid Catch, Cavity) = P(Toothache \mid Cavity)$

Conditional Independence

- Example: {Toothache, Catch, Cavity}

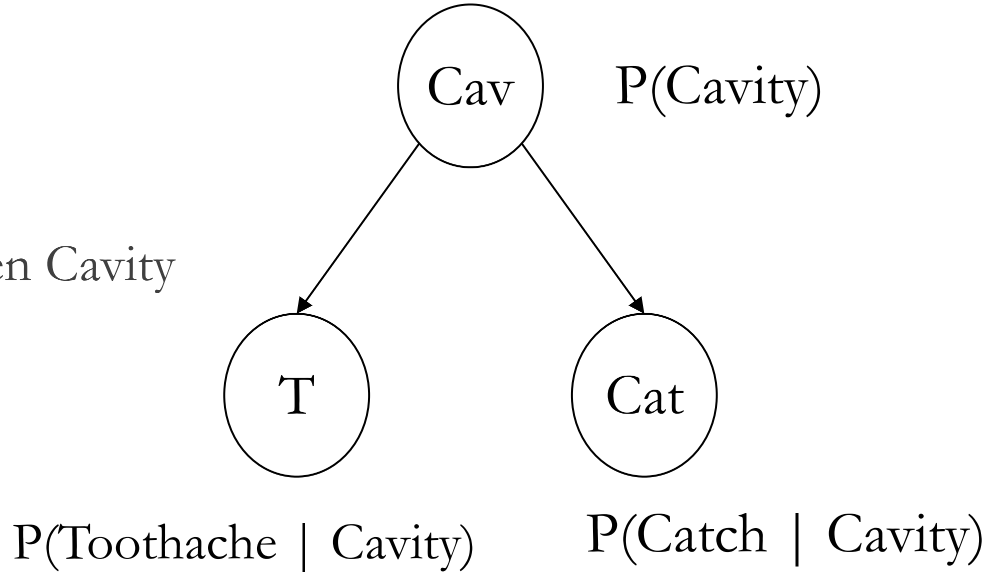
- P(Toothache, Catch, Cavity)

  = P(Cavity) P(Catch | Cavity) P(Toothache | Catch, Cavity)

  = P(Cavity) P(Catch | Cavity) P(Toothache | Cavity)

- Relationships among variables?
  - Cavity influences Toothache and Catch
  - Toothache is conditionally independent of Catch given Cavity

Pictorial Representation (more on this later):

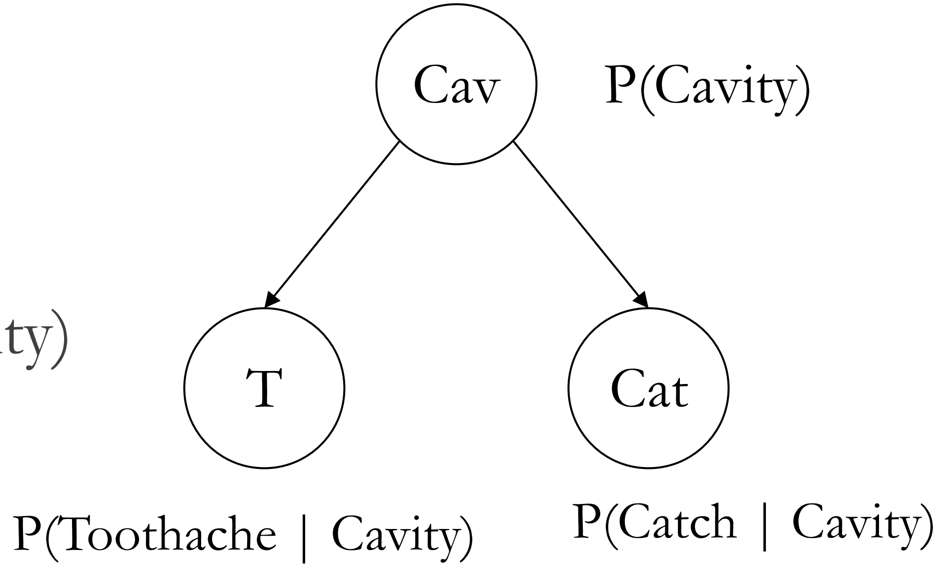


Conditional Independence: Usefulness

- Example: {Toothache, Catch, Cavity}

- P(Toothache, Catch, Cavity)

  = P(Toothache | Cavity) P(Catch | Cavity) P(Cavity)



- How many parameters with the conditional independence assumptions?

  $\Rightarrow$ **1+2+2 = 5**

Conditional Independence: Example – 2

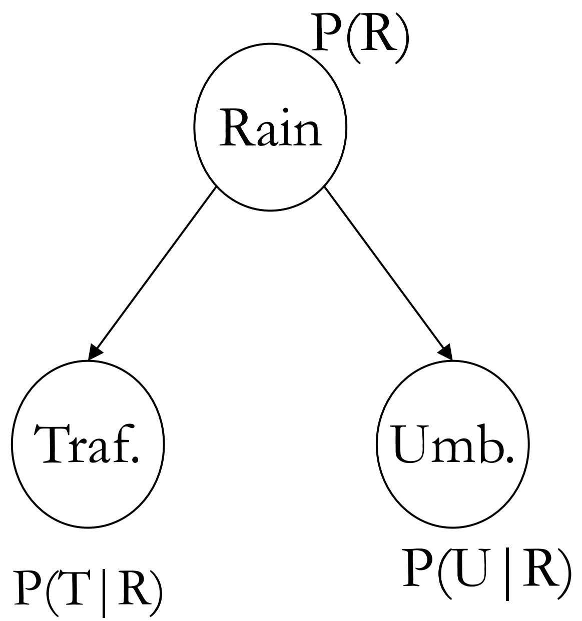
- Three Variables: {Traffic, Rain, Umbrella}
- Trivial decomposition:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- With conditional independence:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

Pictorial Representation:



Conditional Independence

- X is *conditionally independent* of Y given Z iff for all values in the domain of X, Y and Z:

$$P(X|Y,Z) = P(X|Z)$$

$$P(X,Y|Z) = P(X|Z)P(Y|Z)$$

- Notation: $X \perp\!\!\!\perp Y|Z$

- Conditional independence is extremely powerful and reduces the storage complexity from exponential to linear!!

- Conditional independence is the most basic and robust form of knowledge about uncertain environments

Bayesian Networks: Motivation

- Two problems with using full joint distributions for probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to estimate anything empirically about more than a few variables at a time

- Independence and conditional independence helps us to resolve these issues

- How to keep track of all the independence assumptions across variables, and how to represent their probabilities?

Bayesian Networks: Big Picture

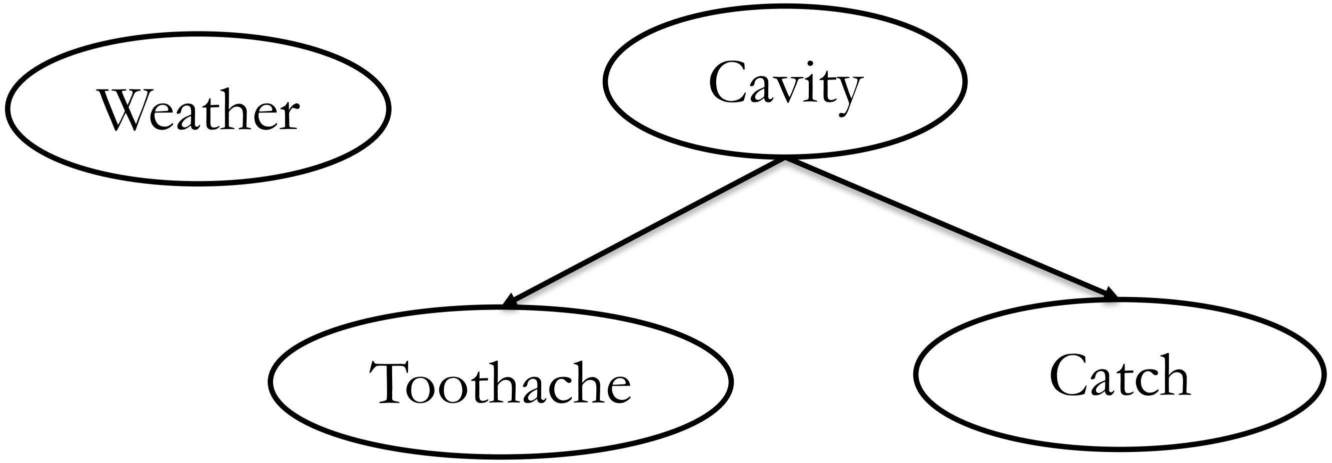
- Bayesian Networks (a.k.a. Bayes nets)
  - A type of probabilistic graphical models
- Graphical component: represents the direct influence of one variable on other
- Probabilistic component: captures the local interactions
  - Local distributions can be used to derive the complex joint distribution over all the variables in the network
- CS community recognized the person for developing these models with the Turing Award!

J. Pearl

Bayes Nets: Graphical Model Notation

- Nodes: Variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Directed edge/link/arc: Indicate "direct influence" between variables
  - Intuitively arrows represent a "noisy causal" process
- Acyclic graph

- Example: {Toothache, Catch, Cavity, Weather}

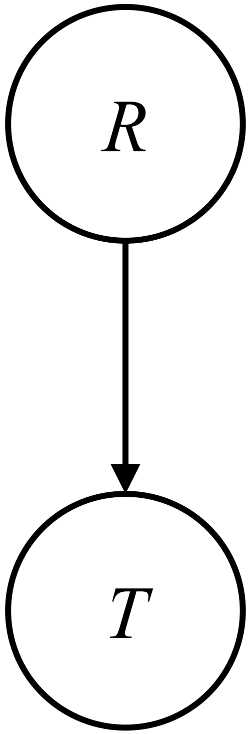


Graphical Model: Coin Flips Example

- N independent coin flips: $\{X_1, X_2, \dots, X_n\}$
- No interactions between variables: absolute independence

Graphical Model: Simple Traffic

- Variables:
  - R: It rains
  - T: There is traffic
- Model 1: Independence
- Model 2: Rain causes Traffic



Graphical Model: Traffic – 2

- Let' s build the graph of Bayes' Net

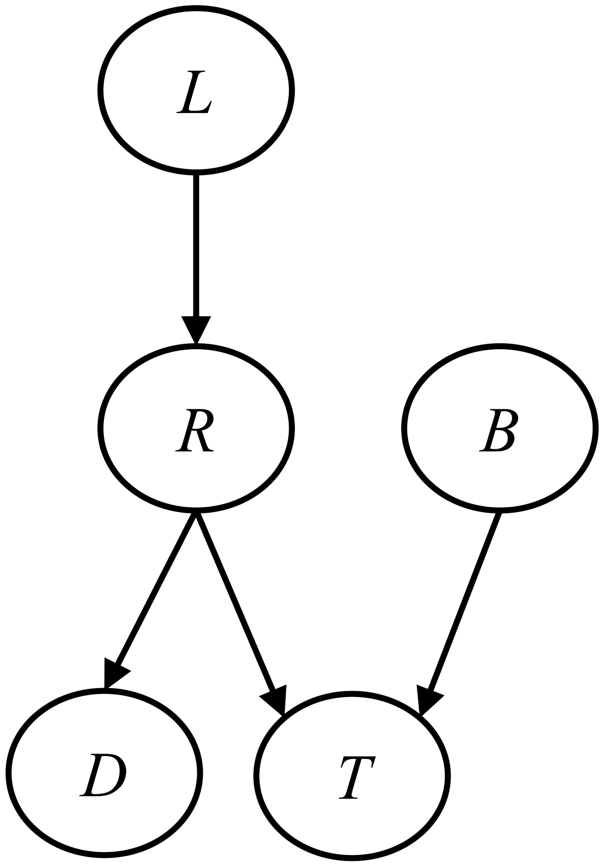
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame

- Different orders can give you different graph
  - Easy if you think of arrows as "causal"

Graphical Model: Traffic – 2

- Let’s build the graph of bayes’ net

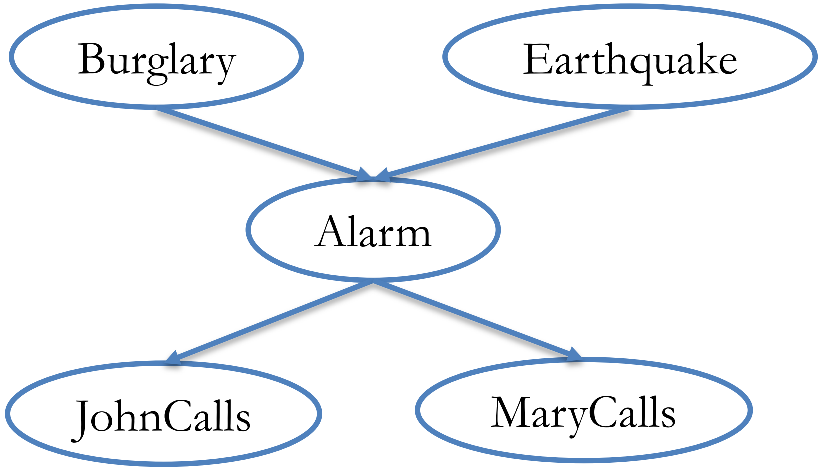
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame



- Different order can give you different graph
  - Easy if you think arrows as “causal”

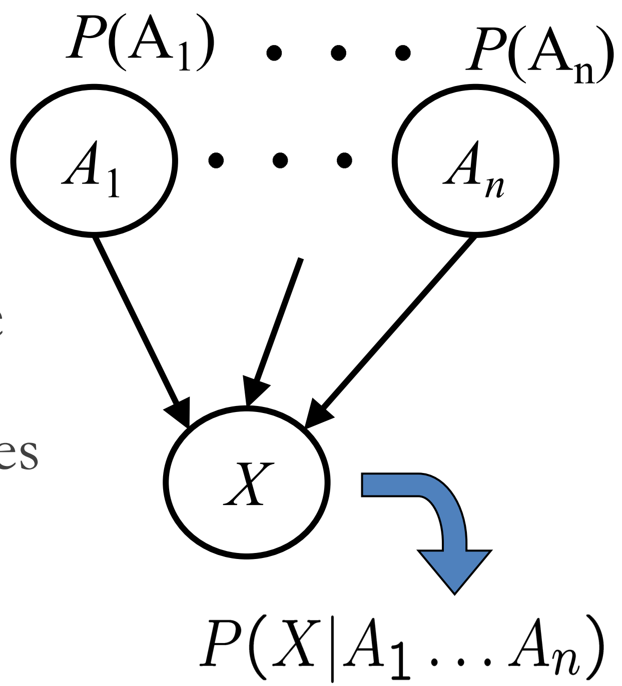
Graphical Model: Burglary and Earthquake

- J. Pearl lives in Los Angeles (high-crime and earthquake prone region).
  - Pearl installed an alarm that sets off when there is a burglary
  - He asked his neighbors (John and Mary) to call when they hear the alarm. So that he could come home immediately.
  - Alarm also goes off when there are minor earth-quakes.
- Graphical Notation:
  - Variables: Burglary (B), Alarm (A), JohnCalls (J), MaryCalls (M), and Earthquake (E)



Bayes' Networks Syntax

- **Graph Component:** A directed, **acyclic** graph
  - Each node represents a variable X
- **Prob. Component:** Conditional distribution for each node
  - E.g.: Distribution of X, for each combination of its Parents' values
  - CPT: conditional probability table



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

This reduces the complexity immensely!

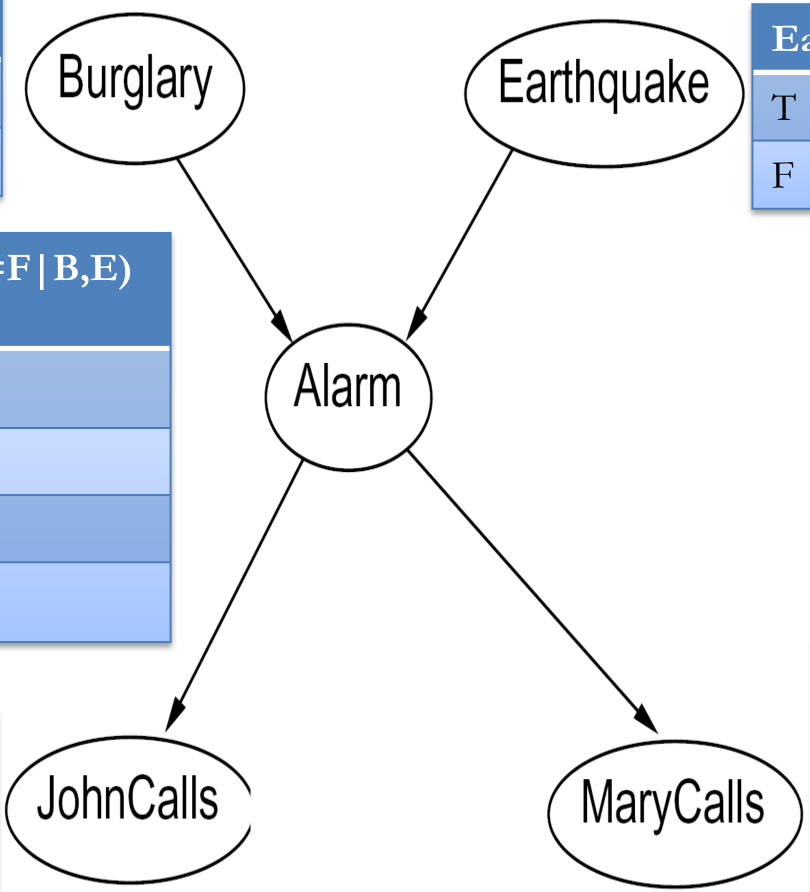
Prob. Component of BN: Burglars and Earthquakes

CPT Representation: Probability distribution for each node given its parents' values

Burglary	P(B)
T	0.001
F	0.999

Earthquake	P(E)
T	0.002
F	0.998

B	E	P(A=T | B,E)	P(A=F | B,E)
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999



A	P(J=T | A)	P(J=F | A)
T	0.90	0.10
F	0.05	0.95

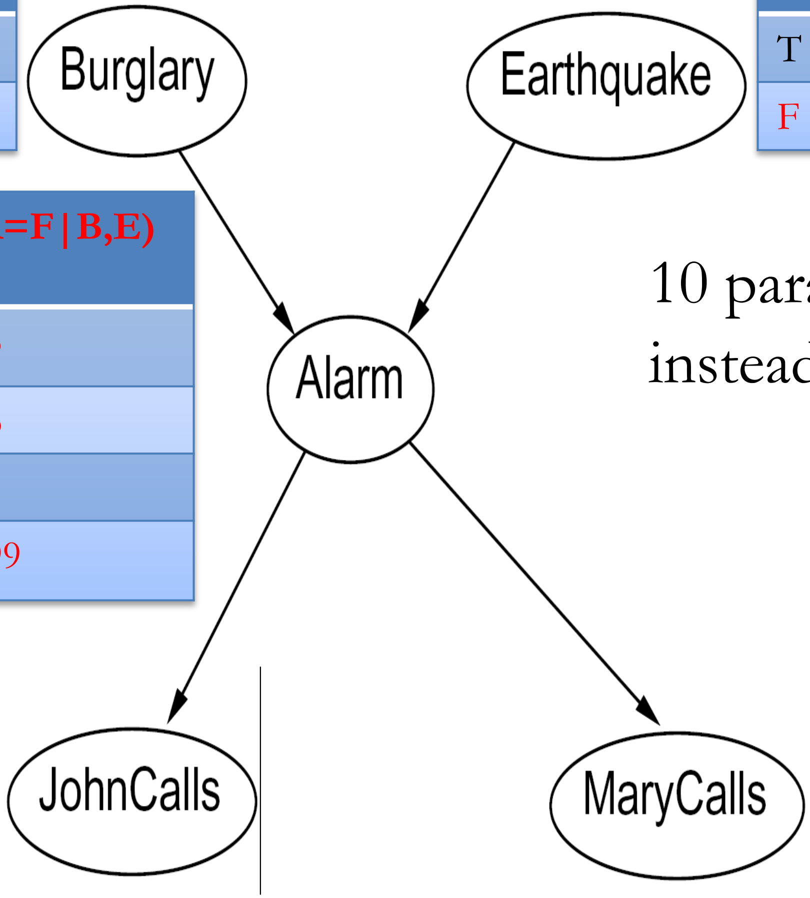
A	P(M=T | A)	P(M=F | A)
T	0.70	0.30
F	0.01	0.99

Prob. Component of BN: Burglars and Earthquakes

- Number of parameters for full Joint vs Bayes' net?

Burglary	P(B)
T	0.001
F	0.999

Earthquake	P(E)
T	0.002
F	0.998



B	E	P(A=T|B,E)	P(A=F|B,E)
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

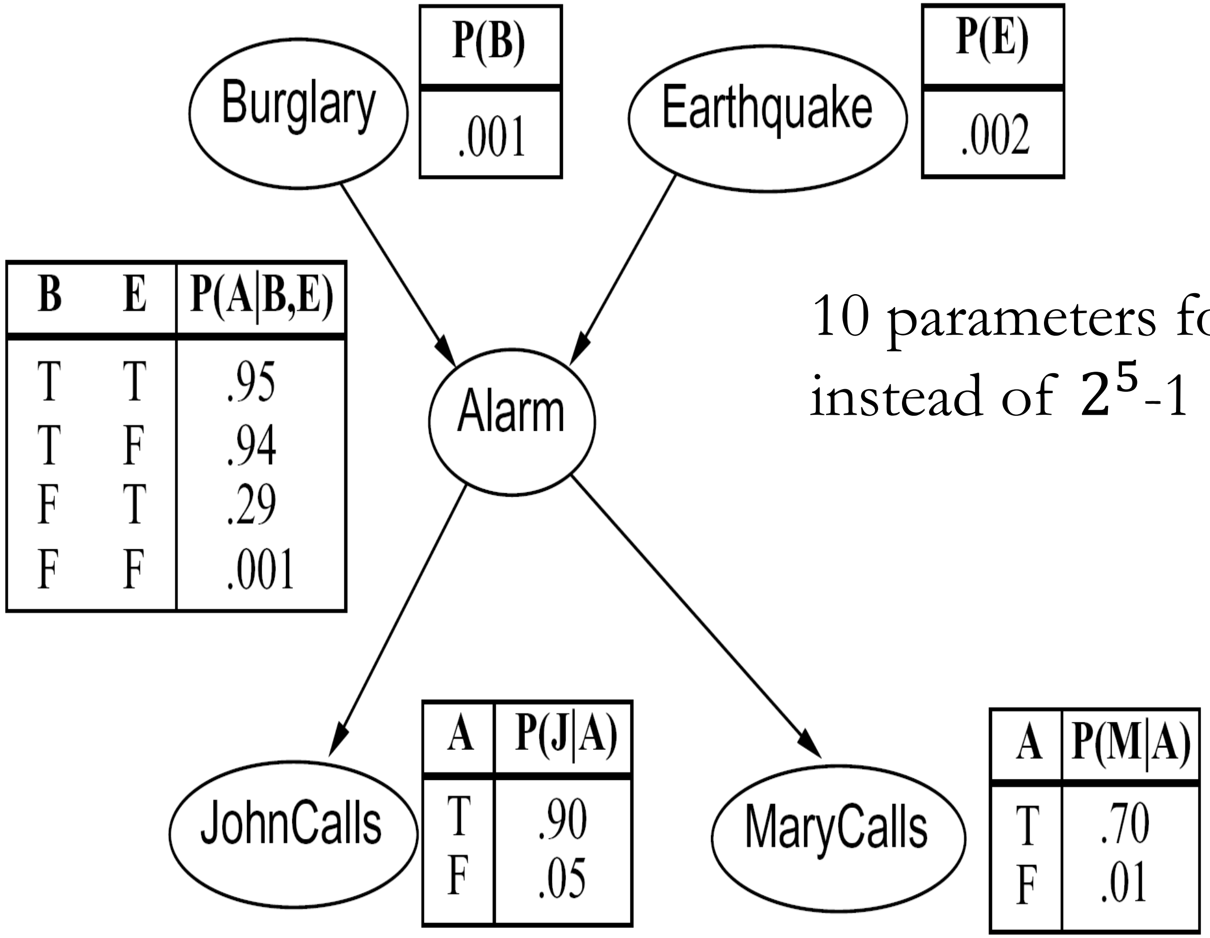
10 parameters for Bayes Net instead of 2^5-1 for the full joint

A	P(J=T|A)	P(J=F|A)
T	0.90	0.10
F	0.05	0.95

A	P(M=T|A)	P(M=F|A)
T	0.70	0.30
F	0.01	0.99

Prob. Component of BN: Burglars and Earthquakes

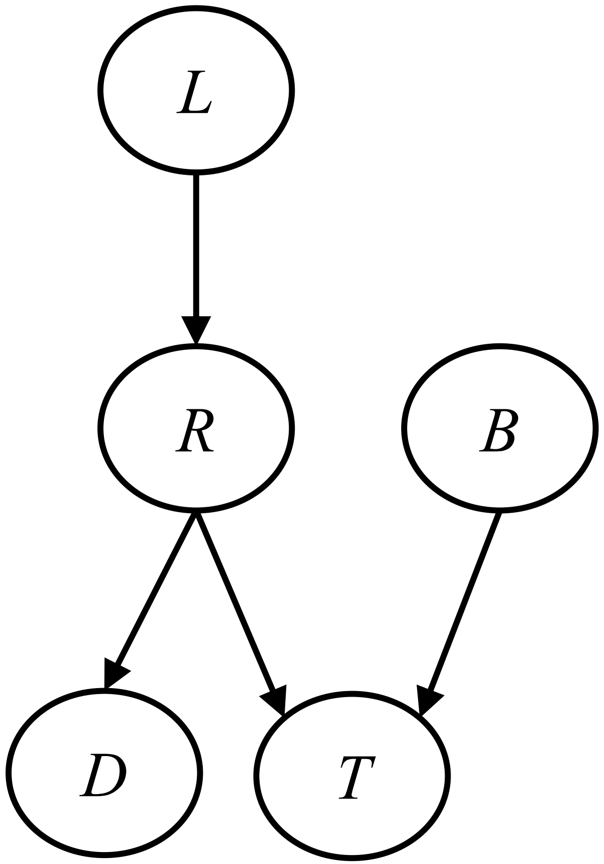
- Number of parameters for full Joint vs Bayes' net?



10 parameters for Bayes Net instead of 2^5-1 for the full joint

Prob. Component of BN: Traffic

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame



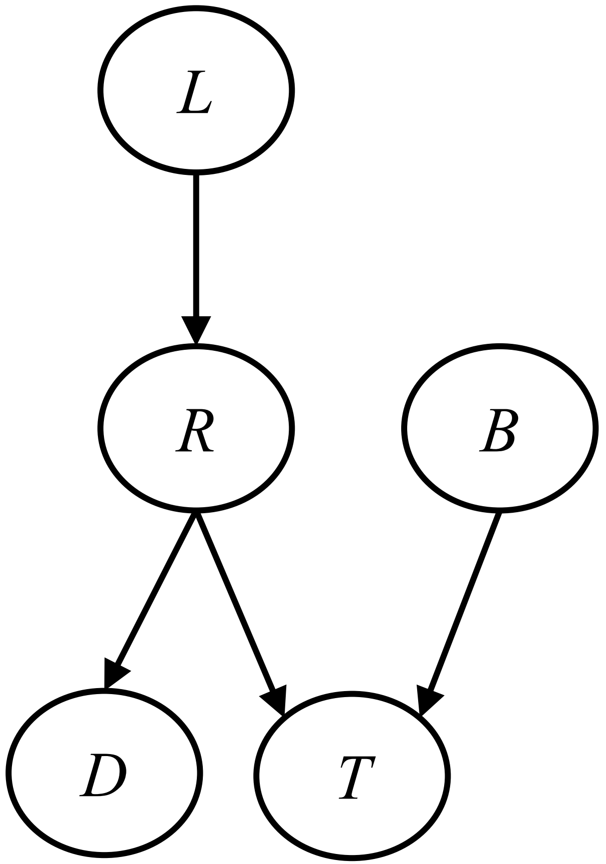
- No. of parameters for full joint distribution?
- No. of parameters for Bayes' net?

Prob. Component of BN: Traffic

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame

- No. of parameters for full joint distribution?
  - 2*2*2*2*2 – 1
- No. of parameters for Bayes' net?
  - 1+1+2+2+2*2

- What is causing this reduction?
  - Conditional independence assumptions



Next Class

- BN models have intuitive representation and reduce the space complexity!

- We do have some unanswered questions about the representation:
  - Are the links Causal or just Correlations?
  - Why can’t there be cycles?

- How can we obtain joint distribution from Bayes Nets?

- How do we infer conditional independence relationships over long-range relationships in the Bayesian Network?

Thank You