Izmir Institute of Technology

CENG 461 – Artificial Intelligence

Probabilistic Inference

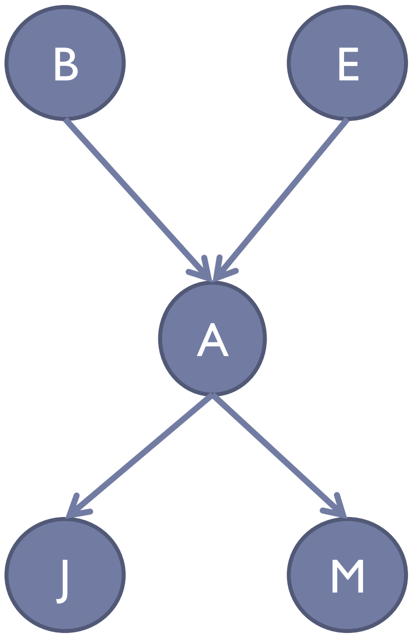
2015-16 Fall Term

Introduction

- In the previous lecture, we have learned how to represent probability distributions with Bayesian Networks.
- Once we have a representation of the joint probability distribution, we would like to ask the following kinds of questions:
  - Given the values of some random variables, what is the probability distribution of a remaining variable?
  - Given the values of some random variables, what are the most likely values of the remaining variables?
- Answers are obtained by *probabilistic inference*.

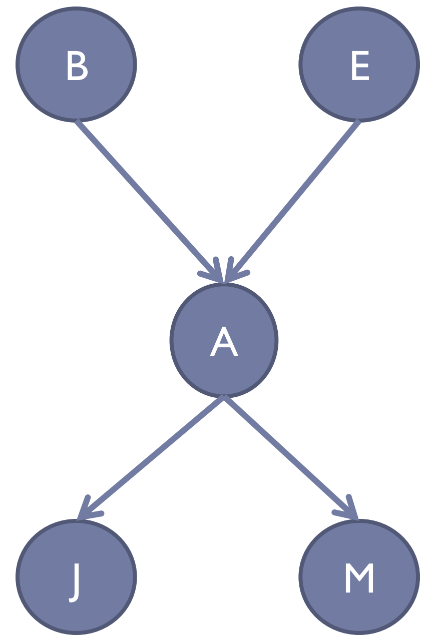
A Sample Network

- B: Burglary
- E: Earthquake
- A: Alarm
- J: John (neighbor) calls you
- M: Mary (neighbor) calls you



Query, Hidden, and Evidence Variables

- We name the variables either
  - as **evidence** variables
    (that have been observed)
  - as **query** variables
    (that we want to know about)
  - as **hidden** variables
    (that are not in query or evidence sets)
  - Note that any variable can be evidence or query.



Query, Hidden, and Evidence Variables

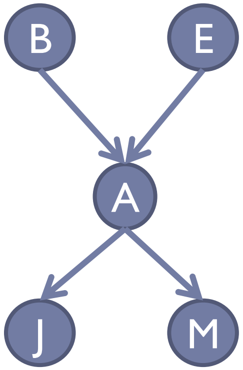
- With the above definitions, the answer to the question 'What is the probability of the query variables when evidence are given?':

$$P(Q_1, Q_2, \ldots \mid E_1 = e_1, E_2 = e_2, \ldots)$$

- The answer to the question 'Of all the possible values of query variables which combination has the highest probability?':

$$\mathrm{argmax}_q(Q_1 = q_1, Q_2 = q_2, \ldots \mid E_1 = e_1, E_2 = e_2, \ldots)$$

Complete Enumeration



- Question:

  $P(+b \mid +j, +m) = ?$

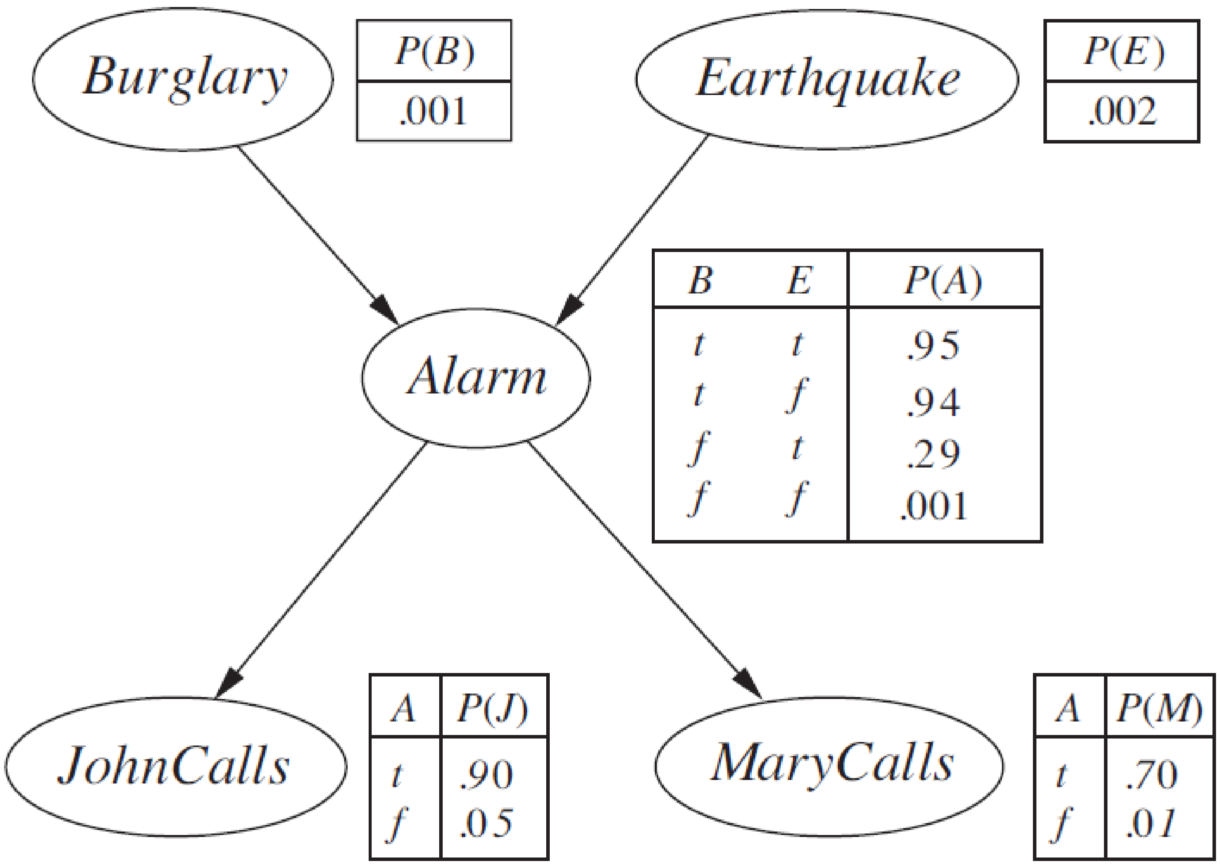
- Using the definition of conditional prob.

  which is: $P(X|Y) = P(X,Y)/P(Y)$

$$P(+b \mid +j, +m) = \frac{P(+b,+j,+m)}{P(+j,+m)} = \frac{P(+b,+j,+m)}{P(+b,+j,+m)+P(-b,+j,+m)}$$

Complete Enumeration

$$P(+b,+j,+m) = \sum_{e,a} P(+b) \cdot P(e) \cdot P(a|+b,e) \cdot P(+j|a) \cdot P(+m|a)$$



$P(B)$
.001

$P(E)$
.002

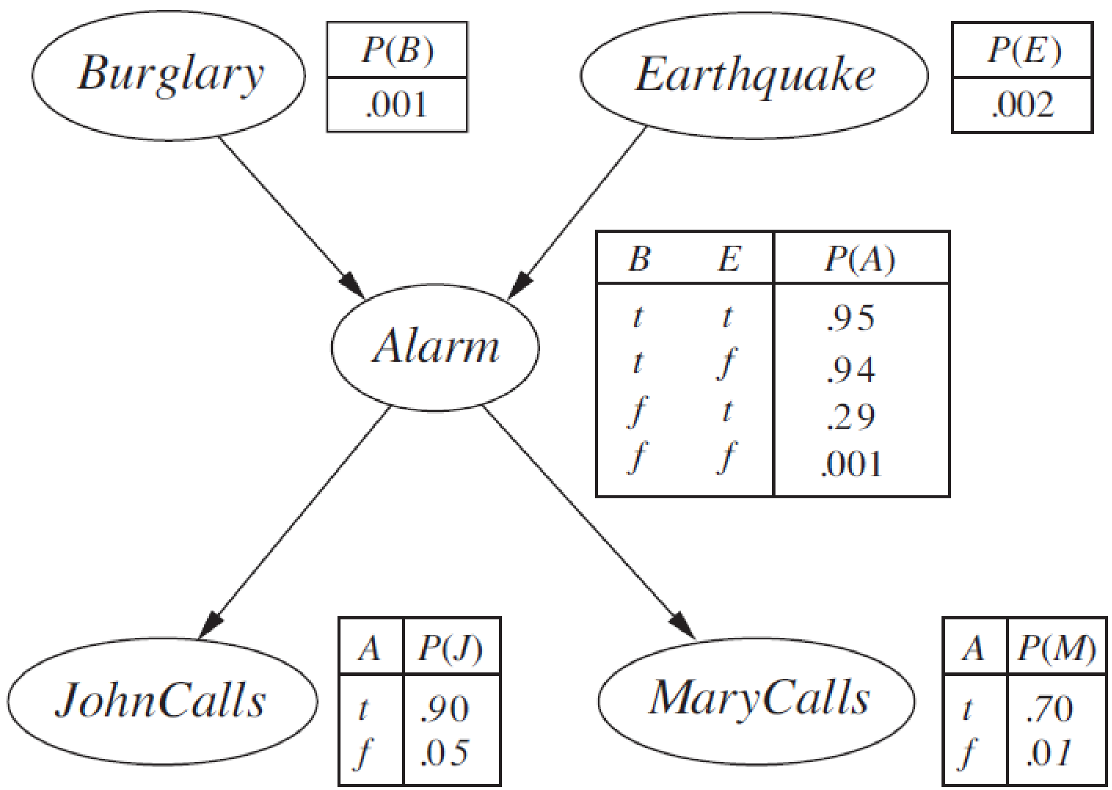
B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01

Complete Enumeration

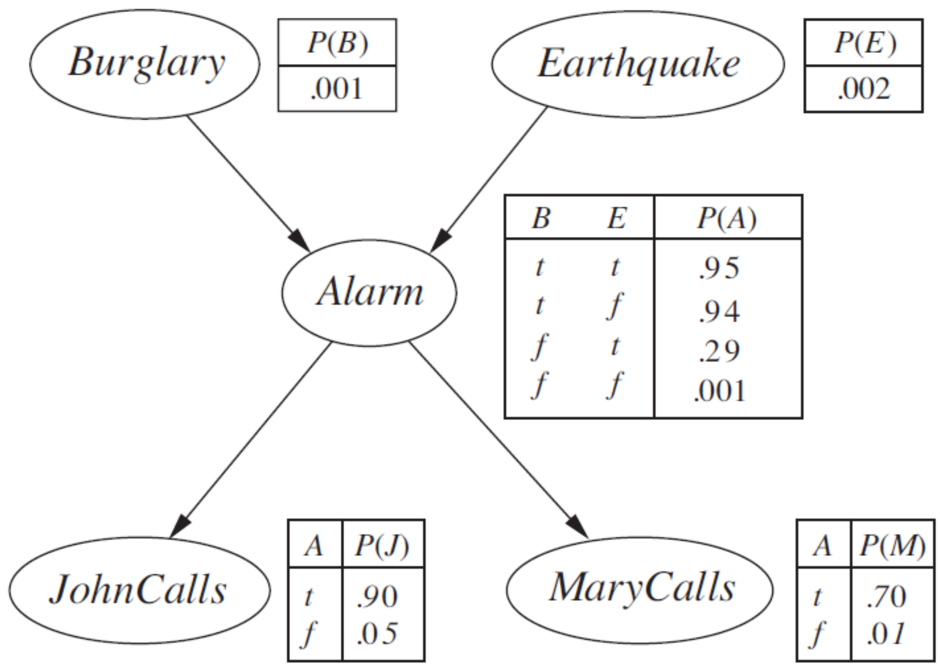
$$
\begin{aligned}
P(+b,+j,+m) = \; & P(+b)P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a) \\
& +P(+b)P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a) \\
& +P(+b)P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a) \\
& +P(+b)P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a) = 0.0005922
\end{aligned}
$$



Complete Enumeration

$$
\begin{aligned}
P(-b,+j,+m) = \; & P(-b)P(+e)P(+a|-b,+e)P(+j|+a)P(+m|+a) \\
& + P(-b)P(+e)P(-a|-b,+e)P(+j|-a)P(+m|-a) \\
& + P(-b)P(-e)P(+a|-b,-e)P(+j|+a)P(+m|+a) \\
& + P(-b)P(-e)P(-a|-b,-e)P(+j|-a)P(+m|-a) = 0.0015
\end{aligned}
$$

$$P(+b|+j,+m) = \frac{P(+b,+j,+m)}{P(+b,+j,+m)+P(-b,+j,+m)} = \frac{0.000592}{0.000592+0.0015} = 0.283$$



Speeding Up Enumeration

- We can re-order the computations to save some time

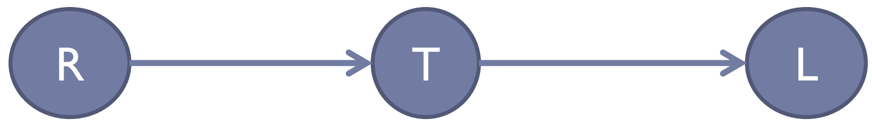
$$P(+b,+j,+m) = \sum_{e,a} P(+b)P(e)P(a|+b,e)P(+j|a)P(+m|a)$$

$$P(+b,+j,+m) = P(+b)\sum_{e} P(e)\sum_{a} P(a|+b,e)P(+j|a)P(+m|a)$$

But we still have the same number of variables in the tables, and computation load is still high.

Variable Elimination (by an example)

- Variables R: Rain, T: Traffic, L: Late
  and corresponding probability tables



P(R)	
+R	0.1
-R	0.9

P(T|R)		
+R	+T	0.8
+R	-T	0.2
-R	+T	0.1
-R	-T	0.9

P(L|T)		
+T	+L	0.3
+T	-L	0.7
-T	+L	0.1
-T	-L	0.9

$P(+L) = \sum_R \sum_T P(+L|T)P(T|R)P(R)$ .... that is enumeration

For large problems, variable elimination is more effective than enumeration

Variable Elimination (by an example)

Construct the table of joint probability P(R,T)

P(R,T)		
+R	+T	0.08
+R	-T	0.02
-R	+T	0.09
-R	-T	0.81

P(L|T)		
+T	+L	0.3
+T	-L	0.7
-T	+L	0.1
-T	-L	0.9

Marginalize over R to get P(T)

P(T)	
+T	0.17
-T	0.83

Variable Elimination (by an example)

Simplified network

P(T)	
+T	0.17
-T	0.83

P(L|T)		
+T	+L	0.3
+T	-L	0.7
-T	+L	0.1
-T	-L	0.9

Table of joint probability P(T,L)

P(T,L)		
+T	+L	0.051
+T	-L	0.119
-T	+L	0.083
-T	-L	0.747

Marginalize over T

P(L)	
+L	0.134
-L	0.866

Approximate Inference by Sampling

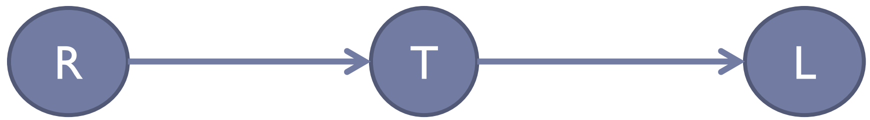
- Estimate the distributions in question by sampling from the distribution.
- By looking at many samples, compute joint probability distributions.
- E.g.: flipping two coins repeatedly counting the occurrence of each joint event.

Out of 1000 flips		
H	H	247
H	T	251
T	H	249
T	T	253

P(H,H)=?

Approximate Inference by Sampling

- When we are given the tables, we can simulate the process.



P(R)	
+R	0.1
-R	0.9

P(T|R)		
+R	+T	0.8
+R	-T	0.2
-R	+T	0.1
-R	-T	0.9

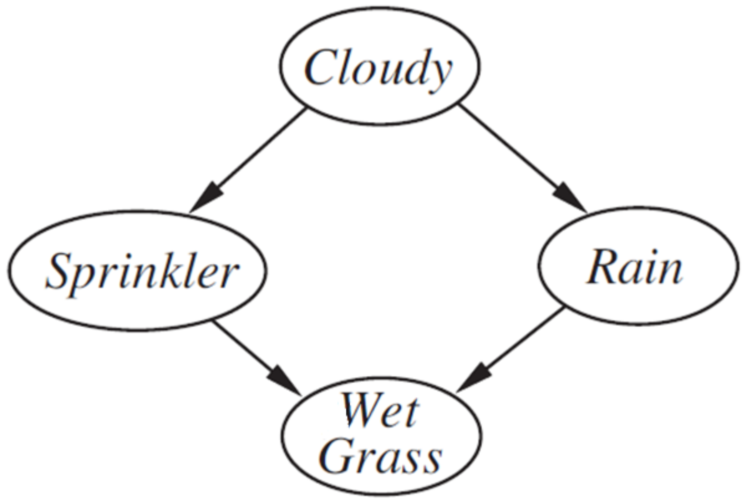
P(L|T)		
+T	+L	0.3
+T	-L	0.7
-T	+L	0.1
-T	-L	0.9

- We randomly generate a sample from the joint distribution by starting from the uppermost nodes.
- We can then sample the lower layer conditioned on the value we have obtained for the upper layers.

Example Network and Question

$P(+W \mid -C) = ?$

$P(C) = .5$



C	$P(S)$
t	.10
f	.50

C	$P(R)$
t	.80
f	.20

S	R	$P(W)$
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Example Network and Question

- P(+W | -C) = P(+W,-C) / [P(+W,-C) + P(-W,-C)]
- With joint probabilities

$$= \frac{P(+W,-C,+S,+R) + P(+W,-C,-S,+R) + P(+W,-C,-S,-R) + P(+W,-C,+S,-R)}{.... + P(-W,-C,+S,+R)+P(-W,-C,-S,+R)+P(-W,-C,-S,-R)+P(-W,-C,+S,-R)}$$

- With sampling (use a random number generator) we obtain

(-C,+S,+R,+W) ... 754 times
(-C,+S,-R,+W) ... 646 times
(-C,-S,+R,+W) ... 672 times
(-C,-S,-R,+W) ... 580 times
(-C,+S,+R,-W) ... 47 times
(-C,+S,-R,-W) ... 122 times
(-C,-S,+R,-W) ... 89 times
(-C,-S,-R,-W) ... 213 times

Then, add up and divide corresponding values to get P(+W|-C)

Approximate Inference by Sampling

- Pay attention that we only used samples with -C for oour computation, however joint distribution contains samples with +C as well.
- We simply discard those samples and use the rest, this is called ‘rejection sampling’.
- This is a simple and effective method but you will discard many samples if the observed variables have low probability (remember 0.001 probability of burglary).

Monthy Hall Problem

- A big prize behind one of the doors.
- You select Door 1.
- Door 3 opens and there is goat behind.
- Would you change your selection as Door 2?