COMP90051 **Statistical Machine Learning**

Semester 2, 2015

Probability Inference on Bayesian Network

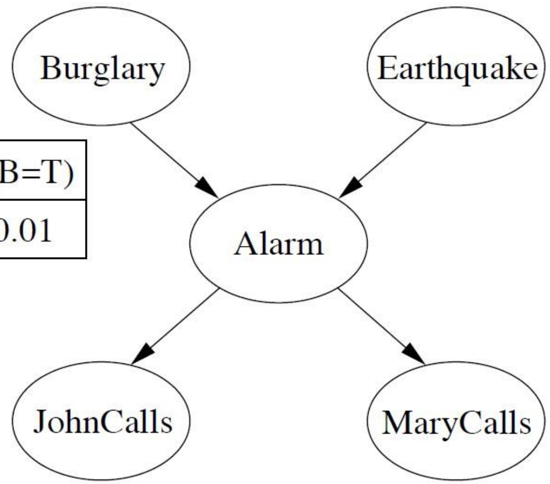
THE UNIVERSITY OF MELBOURNE

POSTERA CRESCAM LAUDE

PGM: Probability Inference

- Three different categories of random variables given the query $P(E = T \mid J = T, M = T) = ?$
  - Query *variables*: Earthquake
  - *Evidence* (*observed*) variables and their values: JohnCalls, MaryCalls
  - *Unobserved* (hidden/latent) variables: Burglary, Alarm
- Inference problem: answer questions about the query variables given the evidence variables
- Approaches
  - Enumeration
  - Elimination Algorithm
  - MCMC (an approximation method)

Example: Probability Inference



P(B=T)
0.01

P(E=T)
0.02

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

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

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

Compute the probability that there is an earthquake given both John and Mary call.

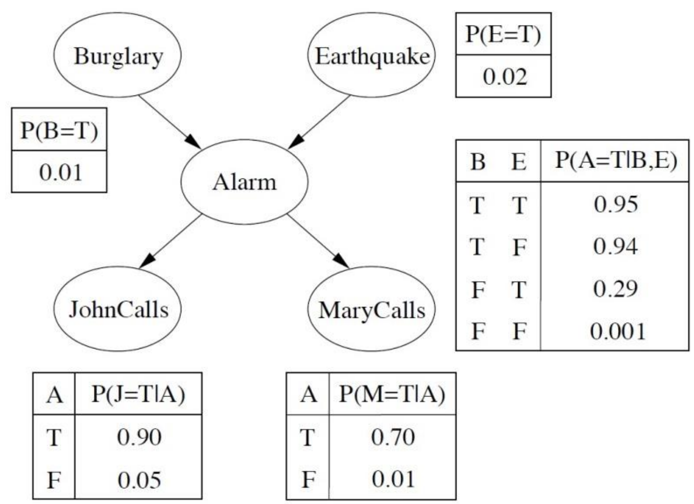
$$P(E = T \mid J = T, M = T) = ?$$

Preliminary

- $P(E|j,m) \ = \ \alpha P(E,j,m)$
  - Lower case j,m denote $J = Ture$ and $m = Ture$
  - We use $P(E|j,m)$ instead of $P(e|j,m)$ here because we calculate both *E = true* and *E = false* together
  - α denotes normalization constant
- $P(E,j,m) \ = \ \sum_a \sum_b P(E,j,m,b,a)$
  - Lower case a,b denote the marginalisation of all possible values of *A* and *B*

Joint Likelihood

- We obtain the joint likelihood of the random variables from our PGM structure:



- $P(E,j,m) = \sum_a \sum_b P(E,j,m,b,a) =$
  $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a)$

Naïve Enumeration Approach

- $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a) =$
  $P(b)P(E)P(a|b,E)P(j|a)P(m|a) \; +$

Naïve Enumeration Approach

- $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a) =$
$$P(b)P(E)P(a|b,E)P(j|a)P(m|a)$$
$$+\ P(\neg b)P(E)P(a|\neg b,E)P(j|a)P(m|a)\ +$$

Naïve Enumeration Approach

- $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a) =$
$$P(b)P(E)P(a|b,E)P(j|a)P(m|a)$$
$$+\ P(\neg b)P(E)P(a|\neg b,E)P(j|a)P(m|a)$$
$$+\ P(b)P(E)P(\neg a|b,E)P(j|\neg a)P(m|\neg a)\ +$$

Naïve Enumeration Approach

- $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a) =$

$$
\begin{aligned}
& P(b)P(E)P(a|b,E)P(j|a)P(m|a) \\
& + \; P(\neg b)P(E)P(a|\neg b,E)P(j|a)P(m|a) \\
& + \; P(b)P(E)P(\neg a|b,E)P(j|\neg a)P(m|\neg a) \\
& + \; P(\neg b)P(E)P(\neg a|b,E)P(j|\neg a)P(m|\neg a)
\end{aligned}
$$

Naïve Enumeration Approach

- $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a) =$

$$P(b)P(E)P(a|b,E)P(j|a)P(m|a)$$
$$+\ P(\neg b)P(E)P(a|\neg b,E)P(j|a)P(m|a)$$
$$+\ P(b)P(E)P(\neg a|b,E)P(j|\neg a)P(m|\neg a)$$
$$+\ P(\neg b)P(E)P(\neg a|b,E)P(j|\neg a)P(m|\neg a)$$

Done! Plug in the numbers from the CPTs

Naïve Enumeration Approach

- $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a) =$

$$P(b)P(E)P(a|b,E)P(j|a)P(m|a)$$
$$+ \; P(\neg b)P(E)P(a|\neg b,E)P(j|a)P(m|a)$$
$$+ \; P(b)P(E)P(\neg a|b,E)P(j|\neg a)P(m|\neg a)$$
$$+ \; P(\neg b)P(E)P(\neg a|b,E)P(j|\neg a)P(m|\neg a)$$

Done! Plug in the numbers from the CPTs

Slow, the top-down computation has duplicate subcomputation, e.g.

$$P(j|a) \times P(m|a)$$

Improve it a Little Bit

- Well, as you can see $P(E)$ is just a constant for $\sum_a \sum_b$, we can save a little bit time by reordering the multiplications and pushing the $\sum$s inward:

- $\sum_a \sum_b P(b)P(E)P(a|b,E)P(j|a)P(m|a) = P(E)\sum_b P(b)\sum_a P(a|b,E)P(j|a)P(m|a)$

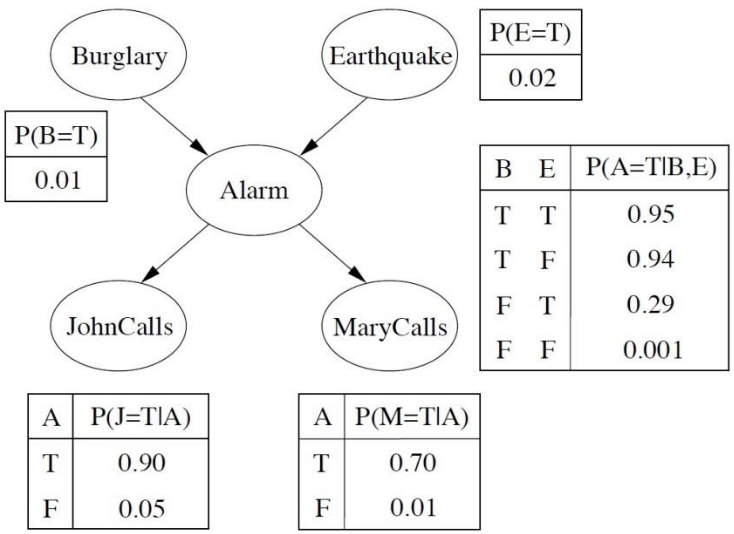
$$= P(E)(\\ P(b)(P(a|b,E)P(j|a)P(m|a) + \\ P(\neg a|b,E)P(j|\neg a)P(m|\neg a)) + \\ P(\neg b)(P(a|\neg b,E)P(j|a)P(m|a) + \\ P(\neg a|\neg b,E)P(j|\neg a)P(m|\neg a)) \\ )$$

Variable Elimination Algorithm

- A form of dynamic programming approach
  - Using *factor tables* to store the immediate results
- Two key operations:
  - Multiplication
  - Marginalisation
- $P(E)\sum_b P(b)\sum_a P(a|b,E)P(j|a)P(m|a) = f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E)f_J(A)f_M(A)$

Factor Tables

- The initial factor tables are the reformatted CPTs:



P(B=T)
0.01

P(E=T)
0.02

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

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

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

B	$f_B(B)$
T	0.01
F	0.99

E	$f_E(E)$
T	0.02
F	0.98

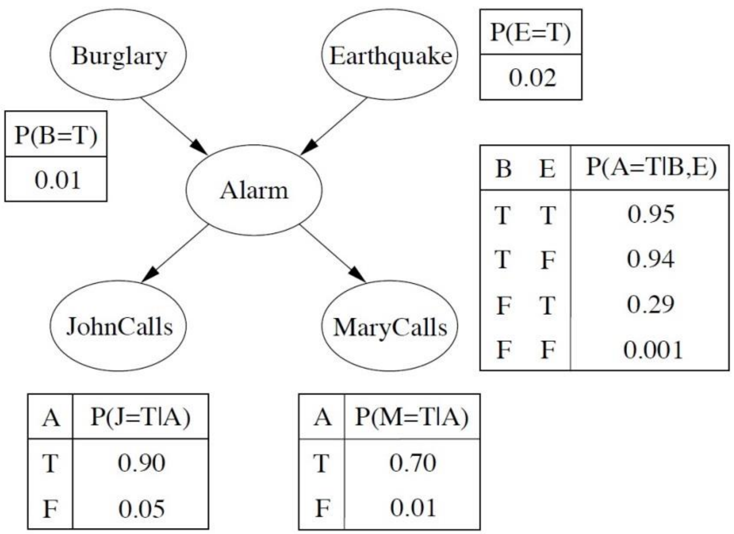
A	B	E	$f_A(A,B,E)$
T	T	T	0 .95
T	T	F	0 .94
T	F	T	0 .29
T	F	F	0 .001
F	T	T	0 .05
F	T	F	0 .06
F	F	T	0 .71
F	F	F	0 .999

A	$f_J(A)$
T	0.9
F	0.05

A	$f_M(A)$
T	0.7
F	0.01

Factor Tables

- The initial factor tables are the reformatted CPTs:



B	$f_B(B)$
T	0.01
F	0.99

E	$f_E(E)$
T	0.02
F	0.98

A	B	E	$f_A(A,B,E)$
T	T	T	0 .95
T	T	F	0 .94
T	F	T	0 .29
T	F	F	0 .001
F	T	T	0 .05
F	T	F	0 .06
F	F	T	0 .71
F	F	F	0 .999

A	$f_J(A)$
T	0.9
F	0.05

A	$f_M(A)$
T	0.7
F	0.01

We only have one random variable in $f_J(A)$ and $f_M(A)$ because *j* and *m* are known values (evidence)

Bottom-up Computation:

$$f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E)f_J(A)f_M(A)$$
$$= f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E)f_{JM}(A)$$

The Computation

$$
\begin{aligned}
& f_E(E)\textstyle\sum_b f_B(B)\sum_a f_A(A,B,E)f_J(A)f_M(A) \\
=\ & f_E(E)\textstyle\sum_b f_B(B)\sum_a {\color{red} f_A(A,B,E)f_{JM}(A)} \\
=\ & f_E(E)\textstyle\sum_b f_B(B)\sum_a {\color{red} f_{AJM}(A,B,E)}
\end{aligned}
$$

The Computation

$$f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E)f_J(A)f_M(A)$$
$$= f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E)f_{JM}(A)$$
$$= f_E(E)\sum_b f_B(B)\sum_a f_{AJM}(A,B,E)$$
$$= f_E(E)\sum_b f_B(B)f_{\bar{A}JM}(A,B,E)$$

The Computation

$$
\begin{aligned}
& f_E(E) \textstyle\sum_b f_B(B) \sum_a f_A(A,B,E) f_J(A) f_M(A) \\
= \; & f_E(E) \textstyle\sum_b f_B(B) \sum_a f_A(A,B,E) f_{JM}(A) \\
= \; & f_E(E) \textstyle\sum_b f_B(B) \sum_a f_{AJM}(A,B,E) \\
= \; & f_E(E) \textstyle\sum_b f_B(B) f_{\underline{A}JM}(A,B,E) \\
= \; & f_E(E) \textstyle\sum_b f_{B\underline{A}JM}(B,E)
\end{aligned}
$$

The Computation

$$
\begin{aligned}
& f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E) f_J(A) f_M(A) \\
&= f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E) f_{JM}(A) \\
&= f_E(E)\sum_b f_B(B)\sum_a f_{AJM}(A,B,E) \\
&= f_E(E)\sum_b f_B(B) f_{\underline{A}JM}(A,B,E) \\
&= f_E(E)\sum_b f_{B\underline{A}JM}(B,E) \\
&= f_E(E) f_{\underline{B}\underline{A}JM}(E)
\end{aligned}
$$

The Computation

$$
\begin{aligned}
& f_E(E)\textstyle\sum_b f_B(B)\sum_a f_A(A,B,E)f_J(A)f_M(A) \\
= \ & f_E(E)\textstyle\sum_b f_B(B)\sum_a f_A(A,B,E)f_{JM}(A) \\
= \ & f_E(E)\textstyle\sum_b f_B(B)\sum_a f_{AJM}(A,B,E) \\
= \ & f_E(E)\textstyle\sum_b f_B(B)f_{\underline{A}JM}(A,B,E) \\
= \ & f_E(E)\textstyle\sum_b f_{B\underline{A}JM}(B,E) \\
= \ & f_E(E)f_{\underline{BA}JM}(E) \\
= \ & f_{E\underline{BA}JM}(E)
\end{aligned}
$$

Step 1

- A multiplication: $f_{JM}(A) = f_J(A) f_M(A)$: **(think about the natural join on database tables)**

A	$f_{JM}(A)$
T	
F	

$=$

A	$f_J(A)$
T	0.9
F	0.05

A	$f_M(A)$
T	0.7
F	0.01

=

Step 1

- A multiplication: $f_{JM}(A) = f_J(A) f_M(A)$: **(think about the natural join on database tables)**

A	$f_{JM}(A)$
T	$.9\times.7$
F	$.05 \times .01$

=

A	$f_J(A)$
T	0.9
F	0.05

A	$f_M(A)$
T	0.7
F	0.01

=

Step 2

- $f_{AJM}(A,B,E) = f_A(A,B,E)f_{JM}(A)$

A	B	E	$f_{AJM}(A,B,E)$
T	T	T	
T	T	F	
T	F	T	
T	F	F	
F	T	T	
F	T	F	
F	F	T	
F	F	F	

=

A	B	E	$f_A(A,B,E)$
T	T	T	0 .95
T	T	F	0 .94
T	F	T	0 .29
T	F	F	0 .001
F	T	T	0 .05
F	T	F	0 .06
F	F	T	0 .71
F	F	F	0 .999

A	$f_{JM}(A)$
T	.63
F	.0005

=

Step 2

- $f_{AJM}(A, B, E) = f_A(A, B, E) f_{JM}(A)$

A	B	E	$f_{AJM}(A,B,E)$
T	T	T	$.95 \times .63$
T	T	F	$.94 \times .63$
T	F	T	$.29 \times .63$
T	F	F	$.001 \times .63$
F	T	T	$.05 \times .0005$
F	T	F	$.06 \times .0005$
F	F	T	$.71 \times .0005$
F	F	F	$.999 \times .0005$

=

=

A	B	E	$f_A(A,B,E)$
T	T	T	0 .95
T	T	F	0 .94
T	F	T	0 .29
T	F	F	0 .001
F	T	T	0 .05
F	T	F	0 .06
F	F	T	0 .71
F	F	F	0 .999

A	$f_{JM}(A)$
T	.63
F	.0005

Step 3

- A marginalisation on *A*: $\sum_a f_{AJM}(A, B, E) = f_{\bar{A}JM}(A, B, E)$ **(add up the numbers with same *B,E* but different *A*)**

A	B	E	$f_{AJM}(A,B,E)$
T	T	T	$.95 \times .63$
T	T	F	$.94 \times .63$
T	F	T	$.29 \times .63$
T	F	F	$.001 \times .63$
F	T	T	$.05 \times .0005$
F	T	F	$.06 \times .0005$
F	F	T	$.71 \times .0005$
F	F	F	$.999 \times .0005$

$\rightarrow$

B	E	$f_{\underline{A}JM}(A,B,E)$
T	T	
T	F	
F	T	
F	F	

Step 3

- A marginalisation on *A*: $\sum_a f_{AJM}(A,B,E) = f_{\bar{A}JM}(A,B,E)$ **(add up the numbers with same *B,E* but different *A*)**

A	B	E	$f_{AJM}(A,B,E)$
T	T	T	$.95 \times .63$
T	T	F	$.94 \times .63$
T	F	T	$.29 \times .63$
T	F	F	$.001 \times .63$
F	T	T	$.05 \times .0005$
F	T	F	$.06 \times .0005$
F	F	T	$.71 \times .0005$
F	F	F	$.999 \times .0005$

$\rightarrow$

B	E	$f_{\bar{A}JM}(A,B,E)$
T	T	$.95 \times .63 + .05 \times .0005$
T	F	$.94 \times .63 + .06 \times .0005$
F	T	$.29 \times .63 + .71 \times .0005$
F	F	$.001 \times .63 + .999 \times .0005$

Step 4

- $f_{B\underline{A}JM}(B,E)$

B	E	$f_{B\underline{A}JM}(B,E)$
T	T	
T	F	
F	T	
F	F	

$=$

B	$f_B(B)$
T	.01
F	.99

B	E	$f_{\underline{A}JM}(A,B,E)$
T	T	.5985
T	F	.5922
F	T	.183
F	F	.001129

Step 4

- $f_{B\underline{A}JM}(B, E)$

B	E	$f_{B\underline{A}JM}(B,E)$
T	T	$.01 \times .5985$
T	F	$.01 \times .5922$
F	T	$.99 \times .183$
F	F	$.99 \times .001129$

=

B	$f_B(B)$
T	.01
F	.99

B	E	$f_{\underline{A}JM}(A,B,E)$
T	T	.5985
T	F	.5922
F	T	.183
F	F	.001129

Step 5

- $f_{B\underline{A}JM}(E)$

B	E	$f_{B\underline{A}JM}(B,E)$
T	T	$.01 \times .5985$
T	F	$.01 \times .5922$
F	T	$.99 \times .183$
F	F	$.99 \times .001129$

$\longrightarrow$

Step 5

- $f_{\underline{B}AJM}(E)$

B	E	$f_{B\underline{A}JM}(B,E)$
T	T	$.01 \times .5985$
T	F	$.01 \times .5922$
F	T	$.99 \times .183$
F	F	$.99 \times .001129$

→

E	$f_{B\underline{A}JM}(B,E)$
T	$.01 \times .5985 + .99 \times .183 = 0.1872$
F	$.01 \times .5922 + .99 \times .001129 = 0.0070$

Step 6

- $f_{E\underline{BA}JM}(E)$

E	$f_{E\underline{BA}JM}(E)$
T	.02 × .1872= .0037
F	.98 × .0070= .0069

=

E	$f_E(E)$
T	.02
F	.98

E	$f_{B\underline{A}JM}(B,E)$
T	.1872
F	.0070

Step 6

- $f_{E\underline{BA}JM}(E)$

E	$f_{E\underline{B}AJM}(E)$
T	.02 × .1872= .0037
F	.98 × .0070= .0069

=

E	$f_E(E)$
T	.02
F	.98

E	$f_{B\underline{A}JM}(B,E)$
T	.1872
F	.0070

- One more step: normalisation:

$$P(e|j,m) = 0.0037/\,(0.0037 + 0.0069) = 0.3491$$

Reference

[1] Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach 3rd Edition.

[2] Kevin B. Korb and Ann E. Nicholson. Bayesian Artificial Intelligence 2nd Edition

[3] Some slides were derived from UIUC Artificial Intelligence (CS440/ECE448)