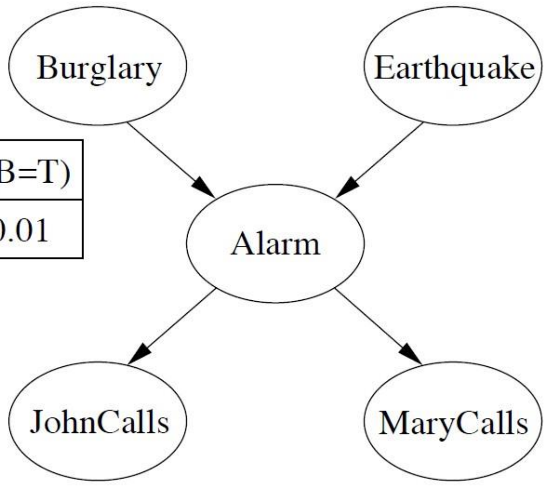
COMP90051 Statistical Machine Learning
Semester 2, 2015

Probability Inference on PGM

Earthquake Example

Example statement: *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.*



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

Given evidence about who has and hasn't called, you'd like to estimate the probability of an earthquake (from Pearl (1988)):

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

(1) The Earthquake Example:

To compute the probability that there is an earthquake given both John and Mary call:

$$P(E|j,m) = \alpha P(E,j,m)$$

Please note: We use $P(E|j,m)$ instead of $P(e|j,m)$ here because we calculate both $E = true$ and $E = false$ together (where e means $E = true$), so that we do not need to worry about the denominator. Symbol α will disappear after normalization.

We use marginalisation to sum out unwanted random variables (or called hidden variables):

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

The lower case of a and b denote the marginalisation of all possible values of A and B.

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

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

Then, a basic method would be using the **na ïve enumeration approach**:

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

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

All you need to do now is to plug in the numbers from the conditional probability tables (CPTs).

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 as much as possible:

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

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

We can still see there are some repeated calculations in the above equation: $P(j|a)P(m|a)$ *and* $P(j|\neg a)P(m|\neg a)$. Actually, it is very hard to avoid the wasted computations by using a top-down approach. Here, we introduce a smarter approach to address this issue, namely, **variable elimination algorithm**.

The variable elimination algorithm uses a bottom-up approach. It stores immediate results using *factor tables*, and use them for the future calculations. It is a form of dynamic programming. A variable elimination algorithm has two key operations: multiplication and marginalisation.

In our example, we use following *factor tables* to store our immediate results:

$P(E)\sum_b P(b)\sum_a P(a|b,E)P(j|a)P(m|a) = f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E)f_J(A)f_M(A)$

The initial factor tables are just the CPTs by reformatting them a little bit. Just put all the random variables on the left-hand side of a factor table.

E	$f_E(E)$
T	0.02
F	0.98

B	$f_B(B)$
T	0.01
F	0.99

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

Note: we only have one random variable in $f_J(A)$ and $f_M(A)$ because *j* and *m* are known values.

The computation is performed from right to left, i.e. bottom-up approach.

$f_E(E)\sum_b f_B(B)\sum_a f_A(A,B,E)f_J(A)f_M(A)$

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

$= f_E(E)\sum_b f_B(B)\sum_a f_{AJM}(A,B,E)$

$= f_E(E)\sum_b f_B(B)f_{\underline{A}JM}(A,B,E)$

$= f_E(E)\sum_b f_{B\underline{A}JM}(B,E)$

$= f_E(E)f_{\underline{BA}JM}(E)$

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

Step by step:

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	$.95 \times .63$
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: $\sum_a f_{AJM}(A,B,E) = f_{\underline{A}JM}(A,B,E)$ **(add up the numbers with same B,E but different A)**

B	E	$f_{\underline{A}JM}(A,B,E)$
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	$.01 \times .5985$
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: $f_{\underline{BA}JM}(E)$

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

Step 6: $f_{E\underline{B}AJM}(E)$

E	$f_{EBAJM}(E)$
T	.02 ×.1872= .0037
F	.98 ×.0070= .0069

=

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

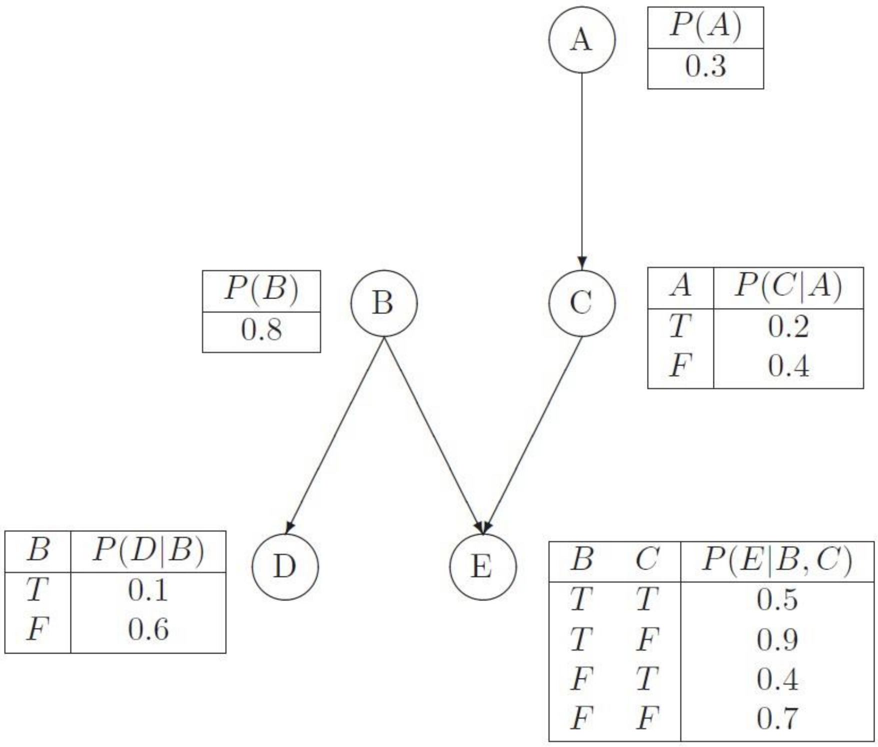
E	$f_{B\underline{A}JM}(B,E)$
T	.1872
F	.0070

Well, one more step: normalisation:

$P(e|j,m)$ = . 0037/ (.0037+ .0069) = 0.3491

(2) **Exercises**

Given the following Bayesian network of Boolean variables:



compute the following probabilities and show your work.

(1) ***P(e|d)***

(2) ***P(d|¬a)***

(3) ***P(¬d & e|c)***

(4) ***P(e|¬a)***

(5) ***P(c|b & ¬d)***

(2) Exercises

1) $P(e|d)$

$P(E|d) = \alpha P(E, d)$
$= \alpha \sum_a \sum_b \sum_c P(a,b,c,E,d)$
$= \alpha \sum_a P(a) \sum_c P(c|a) \sum_b P(b)P(d|b)P(E|b,c)$

$P(e|d)$ = 0.6644

2) $P(d|\neg a)$ = 0.2

3) $P(\neg d \;\&\; e|c)$ = 0.392

4) $P(e|\neg a)$ = 0.708

5) $P(c|b \;\&\; \neg d)$ = 0.34

Discussions:

(1) There are many ways to explain the variable elimination algorithm. An alternative way is the matrix multiplication in our lecture note.
(2) Regarding variable ordering of the variable elimination process: every choice of ordering yields a valid algorithm, but different orderings cause different factor tables. Finding the optimal ordering is a NP-hard problem. One useful hint about choosing the order: "eliminate whichever variable minimises the size of the next factor to be constructed". (please refer to [1] for more details)

References:

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

[2] Example from: Kevin B. Korb and Ann E. Nicholson. Bayesian Artificial Intelligence

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