Exact Inference

3007/7059 Artificial Intelligence

Slides by Lingqiao Liu, Wei Zhang

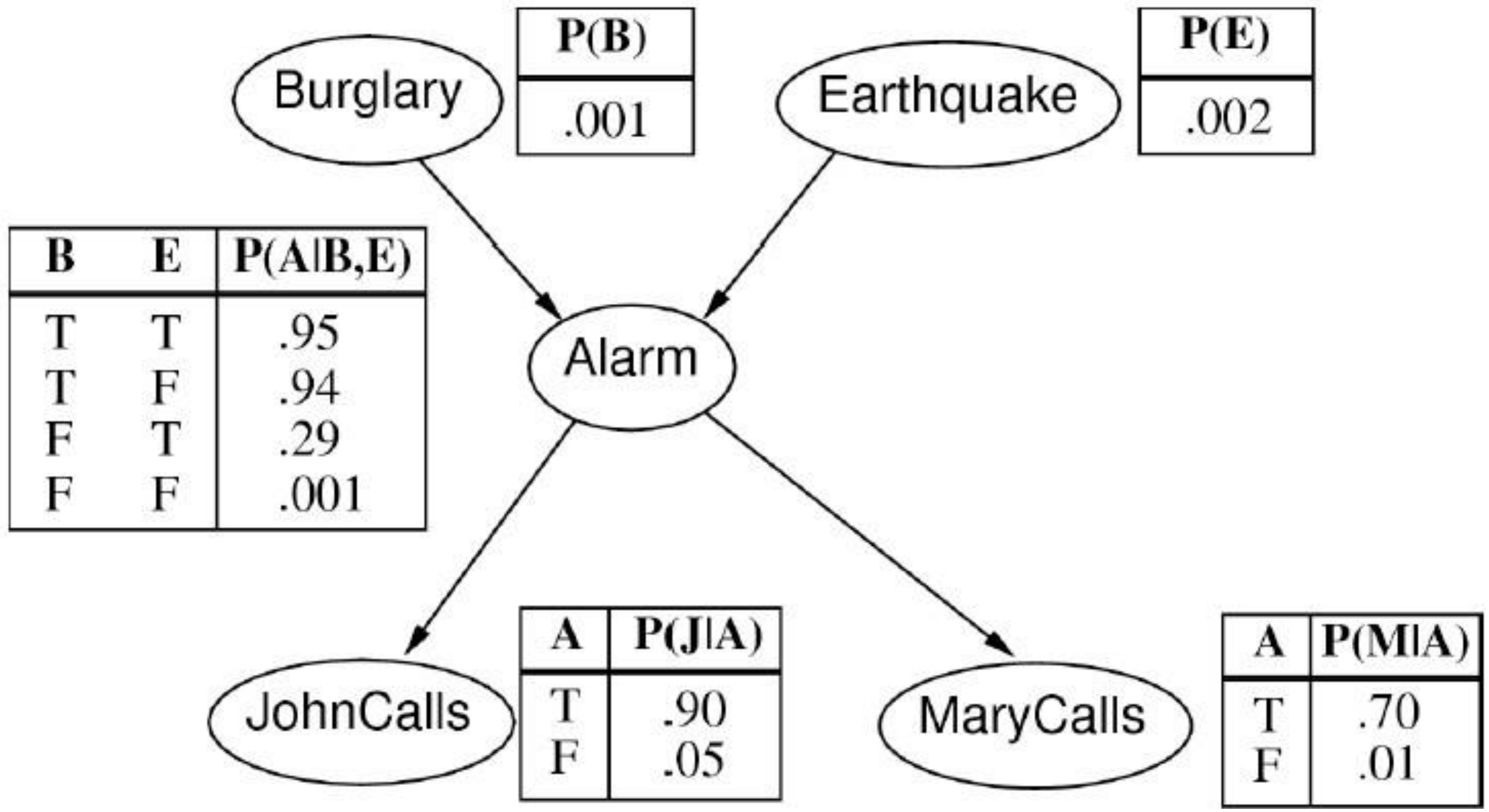
School of Computer Science
University of Adelaide

Outline

- Recap of Bayesian Networks
- Inference by enumeration
- Inference by variable elimination

Recap:Inference problem

I'm at work, neighbour John called to say my alarm is ringing, but neighbour Mary didn't call. Sometimes it's set off by minor earthquakes. Is there a burglar? $P(b|j, \neg m)$



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Recap:Global semantic of a Bayesian network

A Bayesian Network encodes our knowledge on how the variables interact. This is specified in terms of conditional independence assertions.

The global semantics of a network define a joint distribution of all variables as the product of local conditional distributions.

The joint distribution defined by a Bayesian Network with variables $X_1, \ldots, X_n$ is:

$$\begin{aligned} P(X_1, \ldots, X_n) &= P(X_1|Parents(X_1)) \times P(X_2|Parents(X_2)) \\ &\quad \times \cdots \times P(X_n|Parents(X_n)) \\ &= \prod_{i=1}^{n} P(X_i|Parents(X_i)) \end{aligned}$$

where $Parents(X_i)$ are parents of X_i as specified by the particular Bayesian Network.

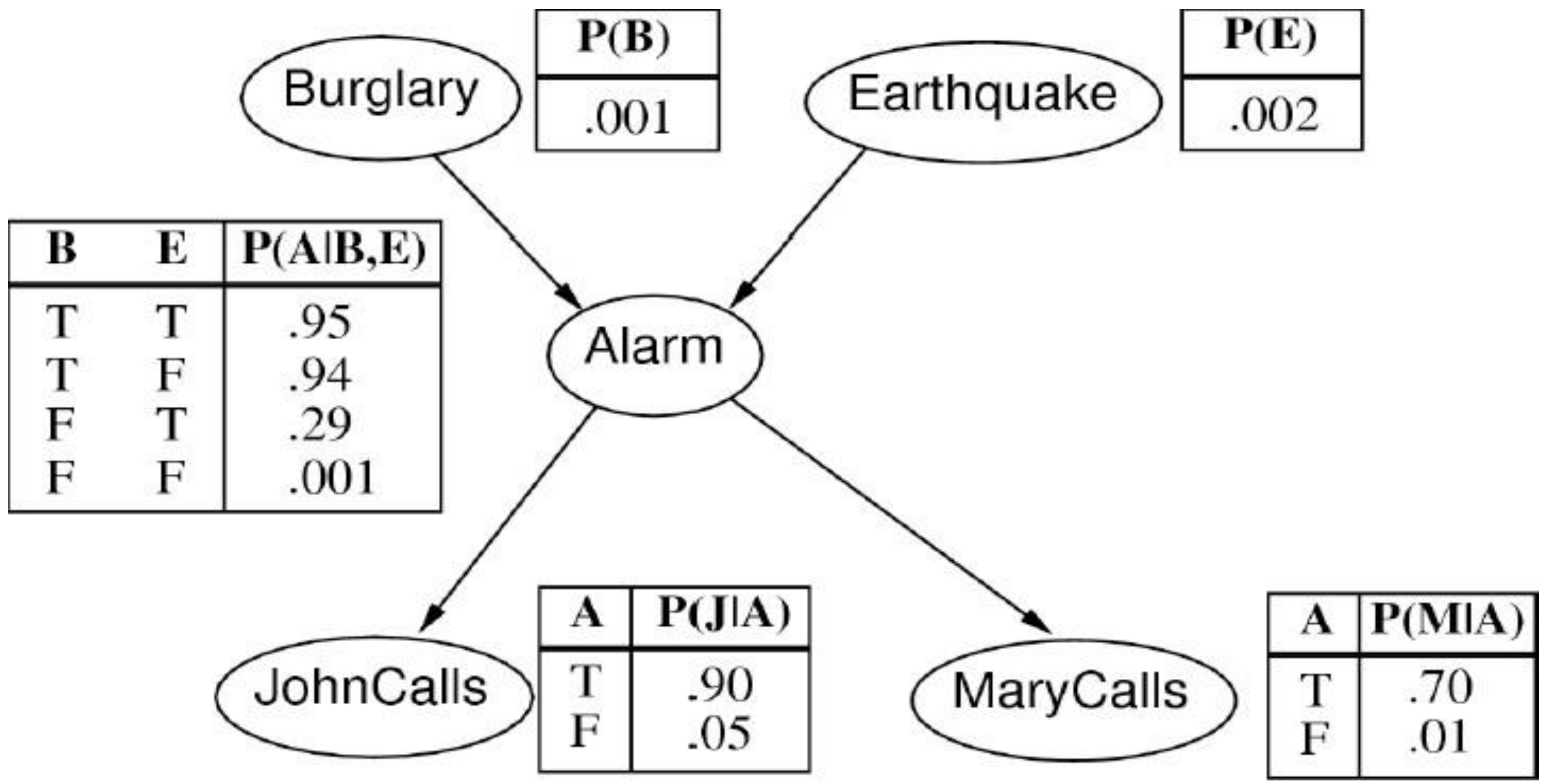
Recap: Inference

Recall the general rule of statistical inference:

$$P(X|e) = \alpha \sum_{\forall Y} P(X, e, Y)$$

where X is the query variable, e the observed values for the evidence variables, and Y the unobserved variables. As usual α is a normalisation constant that we solve for at the end.

Performing inference



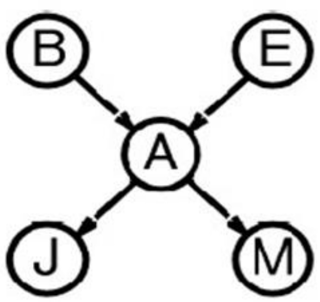
$$P(B, E, A, J, M)? \qquad P(b|j, \neg m)?$$

$$P(J, M, A, B, E) = P(J|A)P(M|A)P(A|B, E)P(B)P(E)$$

Performing inference on Bayesian networks

A direct application of the general rule yields

$$P(b|j,\neg m) = \alpha \sum_{E} \sum_{A} P(b,j,\neg m,E,A)$$



Observe that the summands are joint probabilities of all the variables. Hence, we introduce the global semantics of the network:

$$P(b|j,\neg m) \quad = \quad \alpha \sum_{E} \sum_{A} P(b)P(E)P(A|b,E)P(j|A)P(\neg m|A)$$

Inference by enumeration

Expanding by enumerating the summands we obtain

$$
\begin{aligned}
P(b|j,\neg m) &= \alpha \sum_E \sum_A P(b)P(E)P(A|b,E)P(j|A)P(\neg m|A) \\
&= \alpha\,[P(b)P(e)P(a|b,e)P(j|a)P(\neg m|a) && \text{e, a} \\
&\quad +P(b)P(e)P(\neg a|b,e)P(j|\neg a)P(\neg m|\neg a) && \text{e, -a} \\
&\quad +P(b)P(\neg e)P(a|b,\neg e)P(j|a)P(\neg m|a) && \text{-e, a} \\
&\quad +P(b)P(\neg e)P(\neg a|b,\neg e)P(j|\neg a)P(\neg m|\neg a)] && \text{-e, -a}
\end{aligned}
$$

For each of the components on the RHS, we read its value from the appropriate CPT, yielding $P(b|j,\neg m) = \alpha 0.00025677$.

Note that the result does not yet amount to a probability value as we haven't solved for α.

Inference by enumeration

Calculate the scaling factor

To compute $\alpha = \frac{1}{P(j,\neg m)}$ we obtain the marginal probability

$$P(j, \neg m) = \sum_B \sum_E \sum_A P(B, E, A, j, \neg m)$$

Inference by enumeration

An alternative is to realise that $\langle P(b|j,\neg m), P(\neg b|j,\neg m)\rangle$ is a probability distribution and that α is a normalisation constant that ensures that the probability distribution sums to 1.

Hence, compute $P(\neg b|j,\neg m) = \alpha 0.0498$, using the global semantics and enumerating the summands as before, yielding

$$\langle P(b|j,\neg m), P(\neg b|j,\neg m)\rangle = \alpha\langle 0.00025677, 0.0498\rangle = \langle 0.0051, 0.9949\rangle$$

where α is solved as $\frac{1}{0.00025677+0.0498}$.

Complexity of inference by enumeration

In the burglar-alarm example, we evaluate the expression

$$\begin{aligned} P(b|j,\neg m) &= \alpha \sum_E \sum_A P(b)P(E)P(A|b,E)P(j|A)P(\neg m|A) \\ &= \alpha\, [P(b)P(e)P(a|b,e)P(j|a)P(\neg m|a) \\ &\quad +P(b)P(e)P(\neg a|b,e)P(j|\neg a)P(\neg m|\neg a) \\ &\quad +P(b)P(\neg e)P(a|b,\neg e)P(j|a)P(\neg m|a) \\ &\quad +P(b)P(\neg e)P(\neg a|b,\neg e)P(j|\neg a)P(\neg m|\neg a)] \end{aligned}$$

by adding up 4 terms, each obtained by multiplying 5 numbers—
In total we need 16 multiplications and 3 additions (excludes the contribution due to term α).

Complexity of inference by enumeration

In the burglar-alarm example, we evaluate the expression

$$P(b|j, \neg m) \;=\; \alpha \sum_{E} \sum_{A} P(b)P(E)P(A|b, E)P(j|A)P(\neg m|A)$$

In the worst case, where we have to sum out almost all of the n variables (where we assume they are all Boolean), the complexity of inference by enumeration is $\mathcal{O}(n2^n)$.

This means we will not be able to perform inference by enumeration except for the smallest networks!

Depth-first Evaluation

An improvement can be achieved by observing that $P(b)$ is a constant that can be moved outside the summations over E and A, while $P(e)$ can be moved outside the summation over A:

$$P(b|j,\neg m) = \alpha P(b) \sum_E P(E) \sum_A P(A|b,E) P(j|A) P(\neg m|A)$$

Thus we evaluate by looping through the variables in order, multiplying CPT entries as we go.

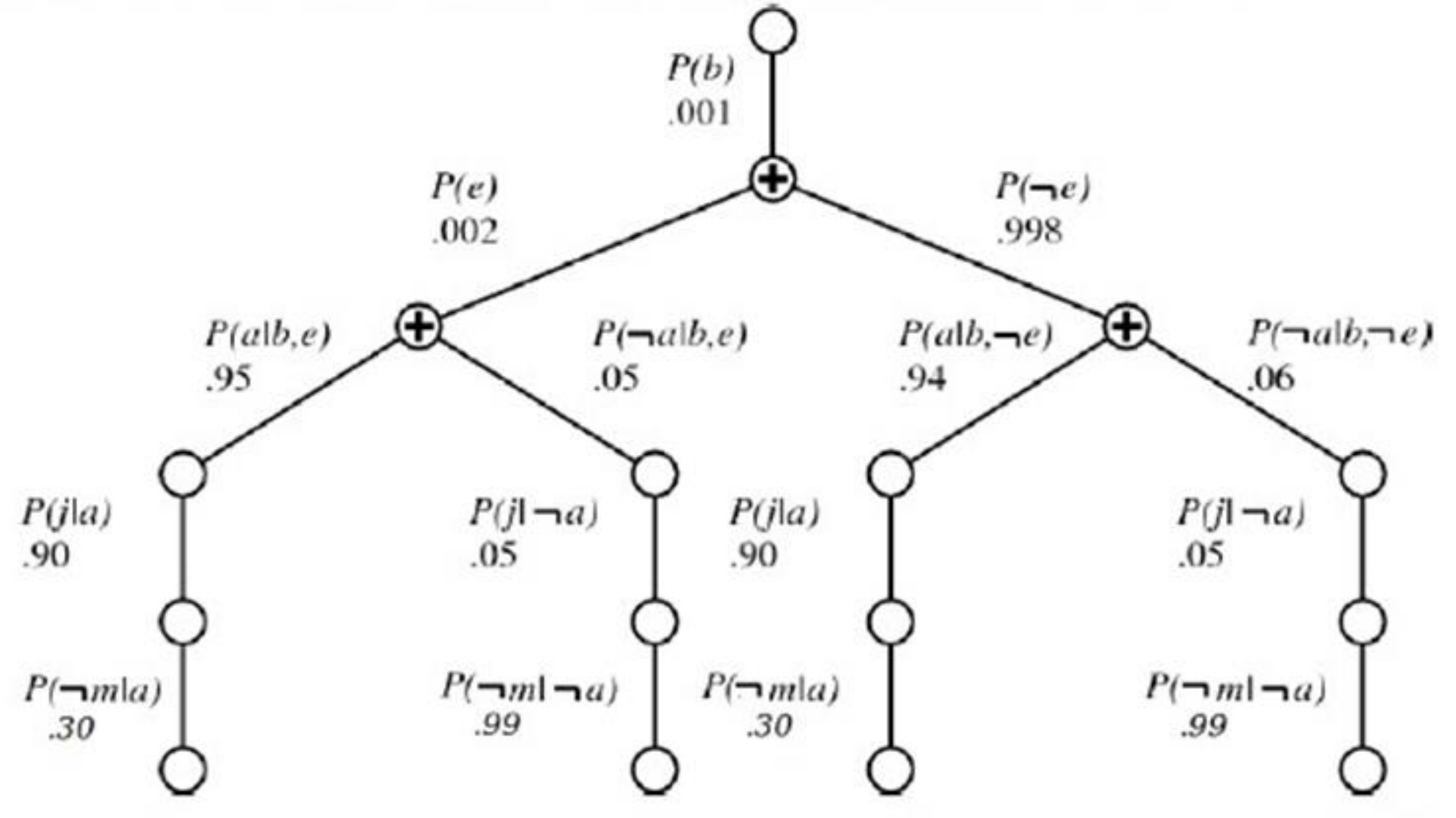
Now evaluating the expression involves 11 multiplications and 3 additions (again excluding the contribution due to term α).

$$= \alpha P(b) \sum_E P(E)[P(a|b,E)P(j|a)P(\neg m|a) + P(\neg a|b,E)P(j|\neg a)P(\neg m|\neg a)]$$

Depth-first Evaluation

$$P(b|j,\neg m) = \alpha P(b) \sum_E P(E) \sum_A P(A|b,E) P(j|A) P(\neg m|A)$$

The process can be illustrated as an evaluation tree.



The evaluation proceeds top-down, multiplying values along each path and summing at the “+” nodes

Complexity of Depth-first Evaluation

$$\begin{aligned}
P(b|j,\neg m) &= \alpha \sum_E \sum_A P(b)P(E)P(A|b,E)P(j|A)P(\neg m|A) \\
&= \alpha\,[P(b)P(e)P(a|b,e)P(j|a)P(\neg m|a) \\
&\quad +P(b)P(e)P(\neg a|b,e)P(j|\neg a)P(\neg m|\neg a) \\
&\quad +P(b)P(\neg e)P(a|b,\neg e)P(j|a)P(\neg m|a) \\
&\quad +P(b)P(\neg e)P(\neg a|b,\neg e)P(j|\neg a)P(\neg m|\neg a)]
\end{aligned}$$

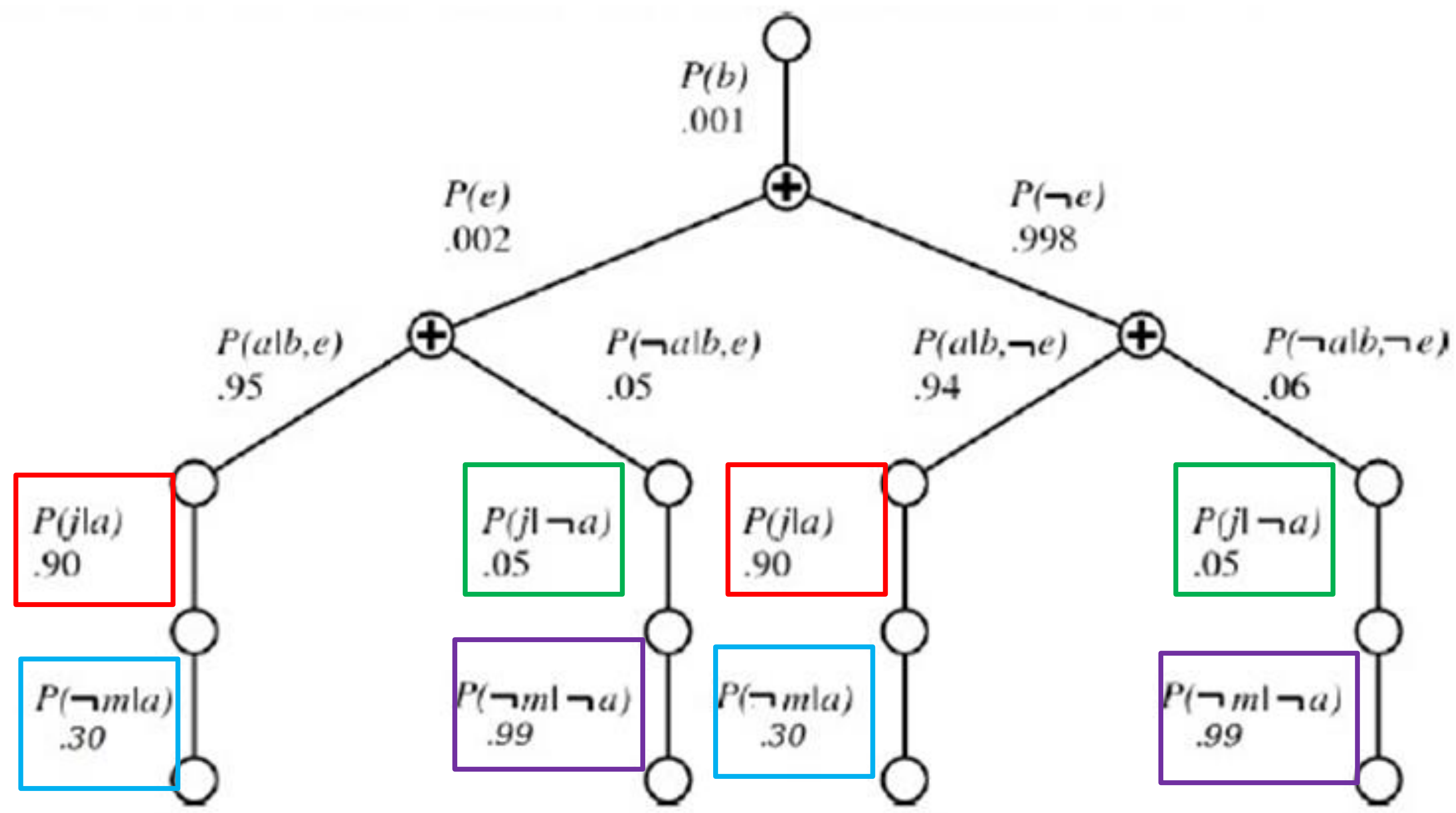
$O(n\,2^n)$

$$P(b|j,\neg m) = \alpha P(b)\sum_E P(E)\sum_A P(A|b,E)P(j|A)P(\neg m|A)$$

$O(2^n)$

Problem of DF Evaluation

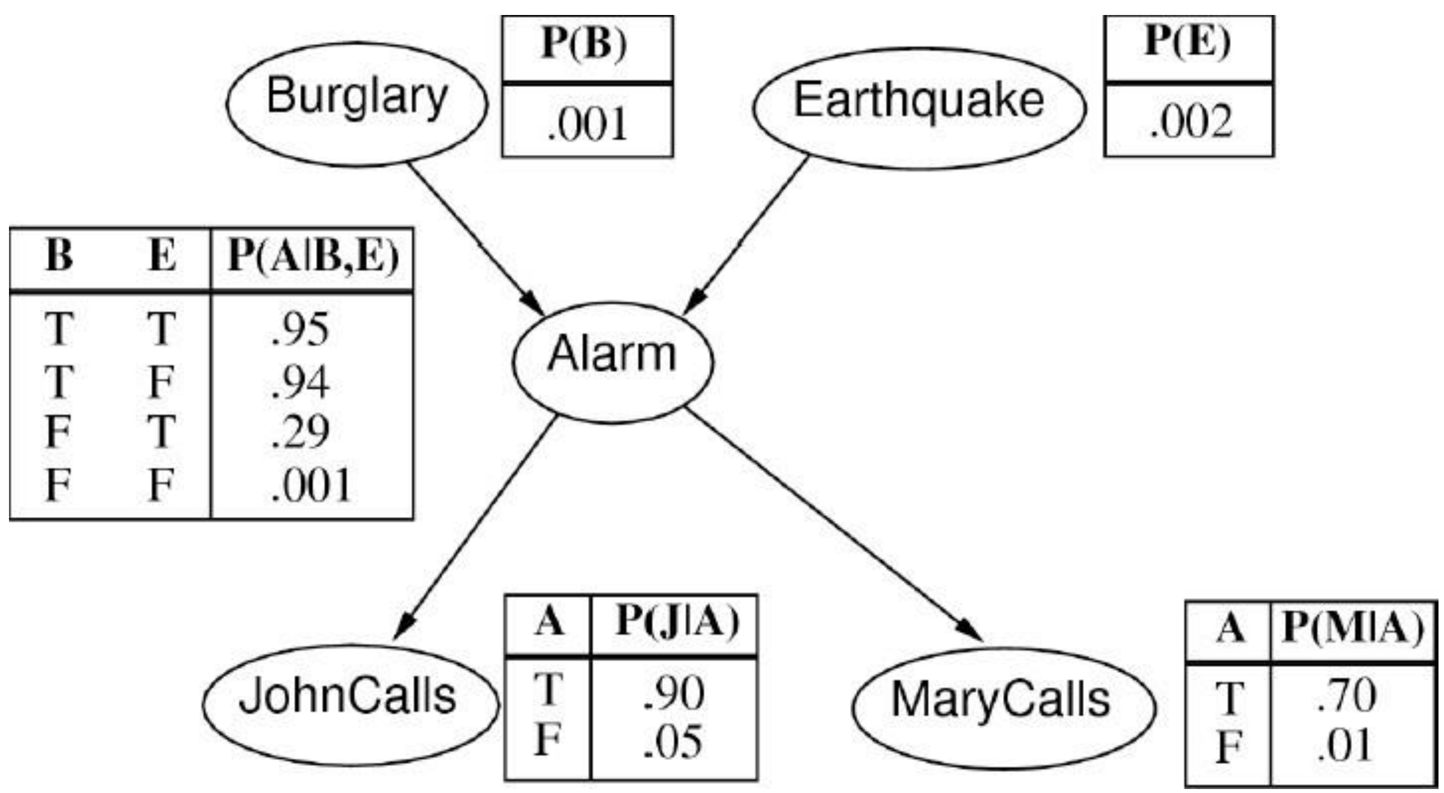
$$P(b|j,\neg m) = \alpha P(b)\sum_E P(E)\sum_A P(A|b,E)P(j|A)P(\neg m|A)$$



Repeat computation!

Problem with DF evaluation

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



Problem with DF evaluation

We need to do this sum both for $E = true$ and $E = false$. For the first case:

$$P(e) \left(\begin{array}{c} P(b)\left(P(a|b,e)P(j|a)P(m|a) + P(\neg a|b,e)P(j|\neg a)P(m|\neg a)\right) + \\ P(\neg b)\left(P(a|\neg b,e)P(j|a)P(m|a) + P(\neg a|\neg b,e)P(j\neg a)P(m|\neg a)\right) \end{array} \right)$$

Problem with DF evaluation

We need to do this sum both for $E = true$ and $E = false$. For the first case:

$$P(e)\left(\begin{array}{c} P(b)\,(P(a|b,e)\boxed{P(j|a)P(m|a)} + P(\neg a|b,e)\boxed{P(j|\neg a)P(m|\neg a))} + \\ P(\neg b)\,(P(a|\neg b,e)\boxed{P(j|a)P(m|a)} + P(\neg a|\neg b,e)\boxed{P(j\neg a)P(m|\neg a))} \end{array}\right)$$

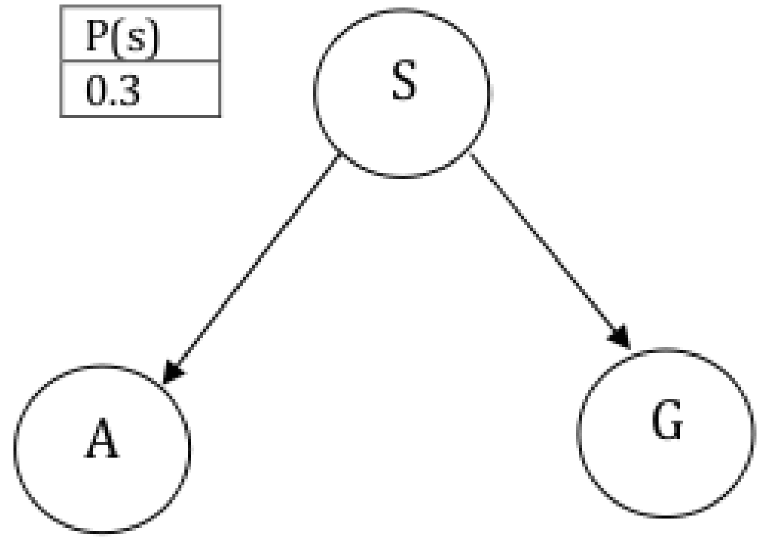
Repeat computation!

Variable Elimination

- Recursively eliminate variables and merge terms into factors
- Store factors to avoid repeating computations

Variable Elimination

P(s)
0.3



S	P(a|S)
t	0.6
f	0.3

S	P(g|S)
t	0.7
f	0.2

$$P(A) = \sum_S \sum_G P(A, S, G)$$

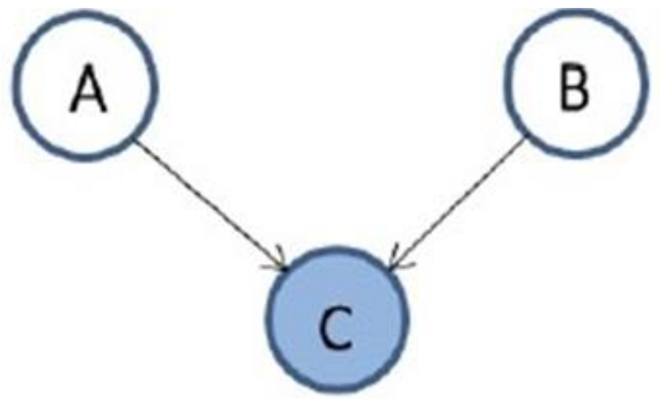
$$= \sum_S \sum_G P(S)P(A|S)P(G|S)$$

$$= \sum_S \left(P(S)P(A|S) \sum_G P(G|S) \right)$$

Eliminate G

$$= \sum_S P(S)P(A|S)$$

Variable Elimination



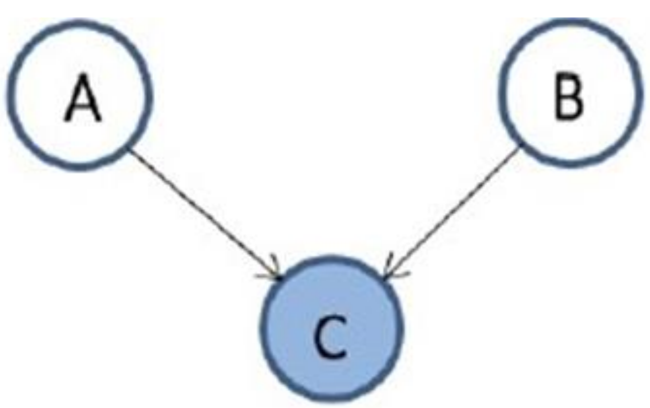
$$
\begin{aligned}
P(C) &= \sum_{A} \sum_{B} P(A, B, C) \\
&= \sum_{A} \sum_{B} P(A) P(B) P(C|A, B) \\
&= \sum_{B} P(B) \sum_{A} P(A) P(C|A, B) \\
&= \sum_{B} P(B) f_1(B, C) \\
&= f_2(C)
\end{aligned}
$$

Eliminate A

Eliminate B

Variable Elimination-Factor

- Factor associate a real value for each setting of its arguments.
- Factor in BN is corresponding to conditional probability distributions.
- The joint distribution is a product of factors.



$$P(A, B, C) = P(A)P(B)P(C|A, B)$$

Variable Elimination -Factor Operation

- Let X, Y and Z are three random variables, and $\phi_1(X,Y)$ and $\phi_2(Y,Z)$ are two factors, their product is a new factor:

$$\psi(X,Y,Z) = \phi_1(X,Y)\phi_2(Y,Z)$$

- An Example:

ϕ_1 has 3*2 = 6 entries

ϕ_2 has 2*2 = 4 entries

yields:

ψ has 3*2*2 = 12 entries

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

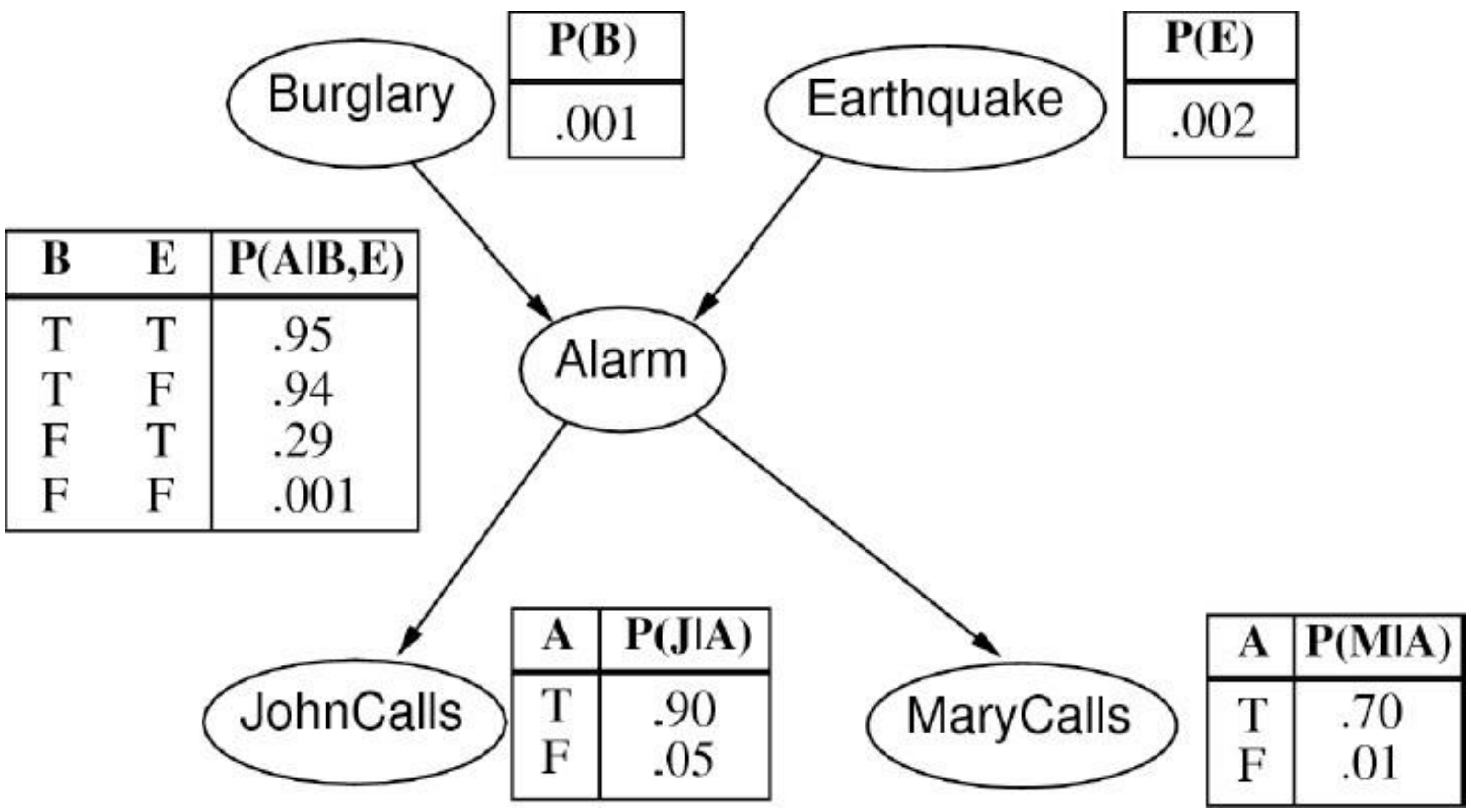
a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

Variable Elimination

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

$$= \alpha \underbrace{P(E)}_{E} \sum_b \underbrace{P(b)}_{B} \sum_a \underbrace{P(a|b,E)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

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



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Variable Elimination

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

$$\alpha \underbrace{P(E)}_{E} \sum_b \underbrace{P(b)}_{B} \sum_a \underbrace{P(a|b,E)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

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

E	$f_E(E)$
T	.002
F	.998

B	$f_B(B)$
T	.001
F	.999

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

A	$f_J(A)$
T	.9
F	.05

A	$f_M(A)$
T	.7
F	.01

Variable Elimination

$$
\begin{aligned}
P(E|j,m) &= \alpha P(E,j,m) \\
&= \alpha P(E) \sum_b P(b) \sum_a P(a|b,E) P(j|a) P(m|a) \\
&= \alpha f_E(E) \textstyle\sum_b f_B(B) \sum_a f_A(A,B,E) f_J(A) f_M(A) \quad \text{factorize}
\end{aligned}
$$

$$
\mathbf{P}(E \mid j,m) = \alpha \mathbf{f}_E(E) \times \sum_b (\mathbf{f}_B(B) \times \sum_a (\mathbf{f}_A(A,B,E) \times \mathbf{f}_J(A) \times \mathbf{f}_M(A)))
$$

Variable Elimination

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

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

$$= \alpha f_E(E) \textstyle\sum_b f_B(B) \sum_a f_A(A,B,E) \boxed{f_J(A) f_M(A)}$$

factorize

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

factor product

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

=

A	$f_J(A)$
T	.9
F	.05

A	$f_M(A)$
T	.7
F	.01

$$f_{JM}(A) = f_J(A) f_M(A)$$

Variable Elimination

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

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

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

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

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

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

=

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

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

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

Variable Elimination

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

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

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

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

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

$$= \alpha f_E(E) \sum_b f_B(B) \boxed{f_{\bar{A}JM}(B,E)}$$ factor marginalization, and elimitate A

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

B	E	$f_{\bar{A}JM}(B,E)$
T	T	.95 * .63 + .05 * .0005 = .5985
T	F	.94 * .63 + .06 * .0005 = .5922
F	T	.29 * .63 + .71 * .0005 = .1830
F	F	.001 * .63 + .999 * .0005 = .001129

$$f_{\bar{A}JM}(B,E) = \sum_a f_{AJM}(A,B,E)$$

Variable Elimination

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

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

$$= \alpha f_E(E) \textstyle\sum_b f_B(B) \sum_a f_A(A,B,E) f_J(A) f_M(A)$$ factorize

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

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

$$= \alpha f_E(E) \textstyle\sum_b f_B(B) f_{\bar{A}JM}(B,E)$$ Eliminate A

$$= \alpha f_E(E) \textstyle\sum_b f_{B\bar{A}JM}(B,E)$$

$$= \alpha f_E(E) f_{\bar{B}\bar{A}JM}(E)$$ Eliminate B

$$= \alpha f_{E\bar{B}\bar{A}JM}(E)$$

The process of evaluation is a process of summing out variables (right to left) from pointwise products of factors to produce new factors, eventually yielding a factor that is the solution, i.e., the posterior distribution over the query variable.

It is bottom-up in the evaluation tree.

Variable elimination

Also useful for doing inference multiple times

e.g.

$$P(B|J,M) = \alpha P(B,J,M)$$
$$P(B|J) = \alpha P(B,J)$$
$$P(B|M) = \alpha P(B,M)$$

Variable elimination

$$
\begin{aligned}
P(B|J,M) &= \alpha P(B,J,M) \\
&= \sum_a \sum_e P(B)P(e)P(a|B,e)P(J|A)P(M|a) \\
&= \alpha P(B) \sum_e P(e) \sum_a P(a|B,e)P(J|a)P(M|a) \\
&= \alpha P(B) f(J,M)
\end{aligned}
$$

$$
\begin{aligned}
P(B|J) &= \alpha P(B,J) \\
&= \sum_m \sum_a \sum_e P(B)P(e)P(a|B,e)P(J|A)P(m|a) \\
&= \alpha P(B) \sum_m \sum_e P(e) \sum_a P(a|B,e)P(J|a)P(m|a) \\
&= \alpha P(B) \sum_m f(J,m)
\end{aligned}
$$

We only need to calculate *f* once

$$
\begin{aligned}
P(B|M) &= \alpha P(B,M) \\
&= \sum_j \sum_a \sum_e P(B)P(e)P(a|B,e)P(j|A)P(M|a) \\
&= \alpha P(B) \sum_j \sum_e P(e) \sum_a P(a|B,e)P(j|a)P(M|a) \\
&= \alpha P(B) \sum_j f(j,M)
\end{aligned}
$$