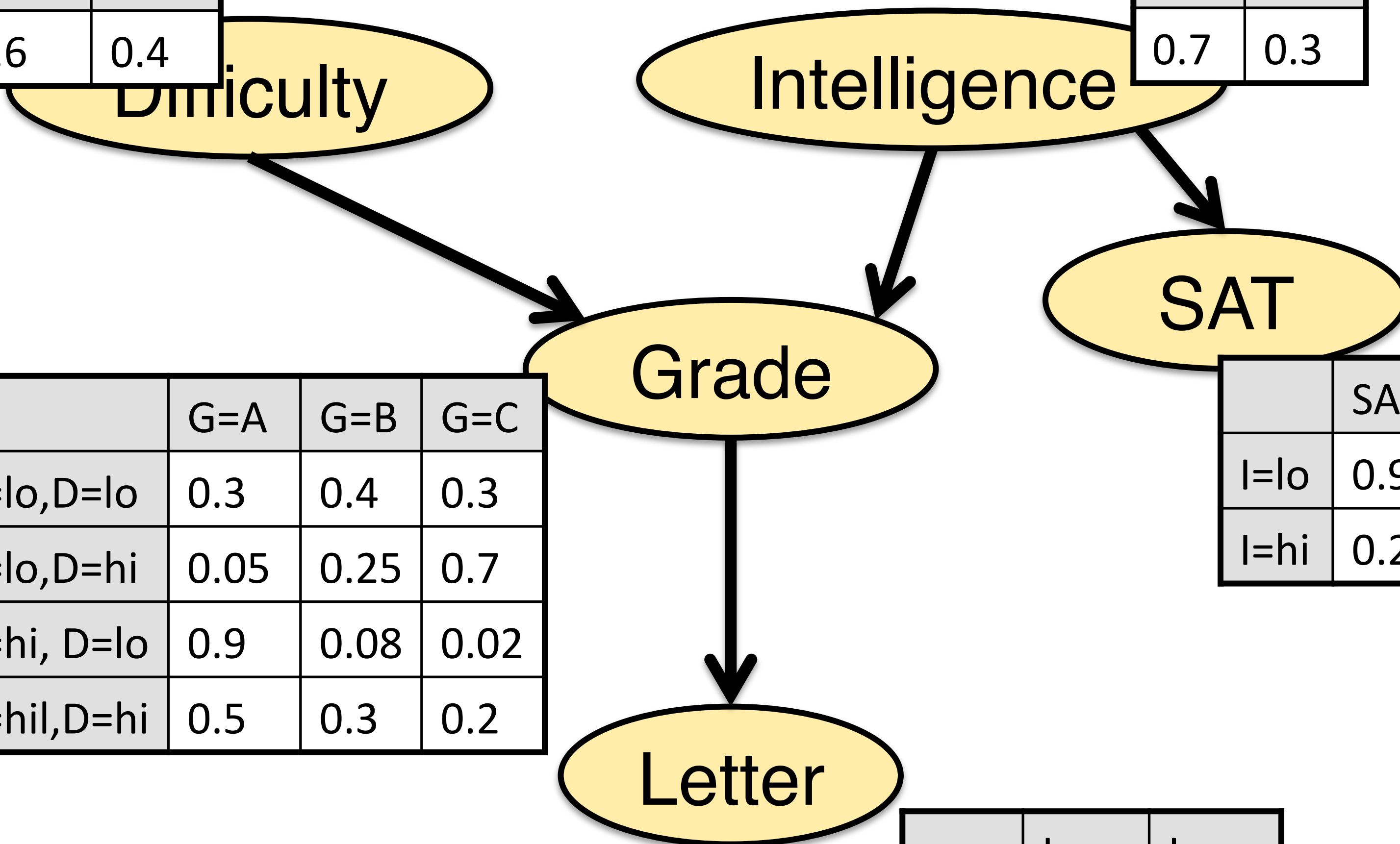
CS440/ECE448: Intro to Artificial Intelligence

Lecture 17: More on exact inference in Bayes Nets

Prof. Julia Hockenmaier
juliahmr@illinois.edu

http://cs.illinois.edu/fa11/cs440



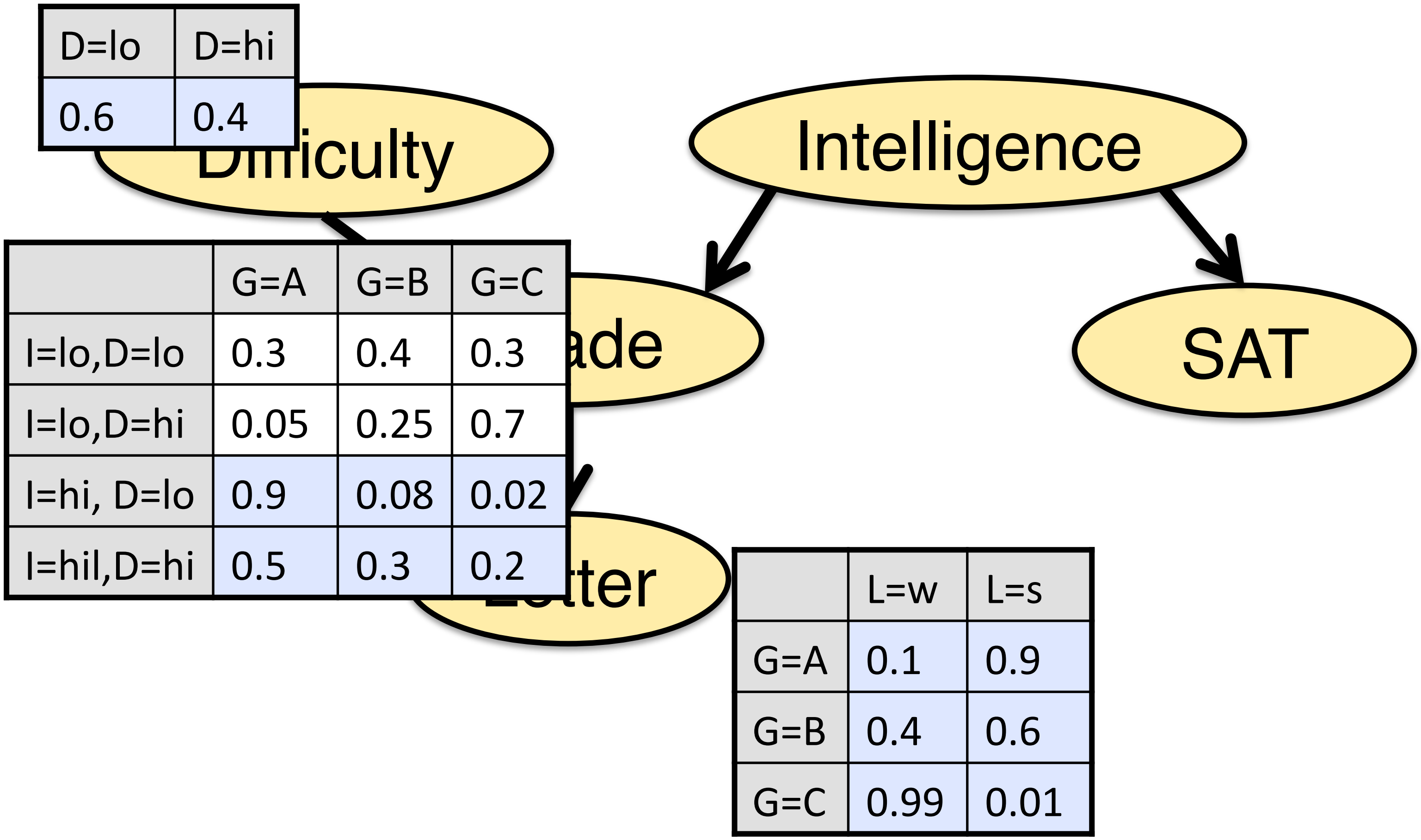
D=lo	D=hi
0.6	0.4

I=lo	I=hi
0.7	0.3

	G=A	G=B	G=C
I=lo,D=lo	0.3	0.4	0.3
I=lo,D=hi	0.05	0.25	0.7
I=hi, D=lo	0.9	0.08	0.02
I=hil,D=hi	0.5	0.3	0.2

	SAT=lo	SAT=hi
I=lo	0.95	0.05
I=hi	0.2	0.8

	L=w	L=s
G=A	0.1	0.9
G=B	0.4	0.6
G=C	0.99	0.01



D=lo	D=hi
0.6	0.4

	G=A	G=B	G=C
I=lo,D=lo	0.3	0.4	0.3
I=lo,D=hi	0.05	0.25	0.7
I=hi, D=lo	0.9	0.08	0.02
I=hil,D=hi	0.5	0.3	0.2

	L=w	L=s
G=A	0.1	0.9
G=B	0.4	0.6
G=C	0.99	0.01

What is the probability of getting a strong letter if you are an intelligent student?

Review: inference in Bayes Nets

What is the probability of getting a strong letter if you are an intelligent student?

$$P(X = x \mid \mathbf{e}) = \frac{P(X = x \mid \mathbf{e})}{P(\mathbf{e})}$$

– We want to compute the probability of a set of **query variables** *X (= the letter)* given an event **e** (= *being intelligent)*

– An event = an assignment of values to a set of **evidence variables *E*** *(= intelligence)*

Review: conditional probability

$$P(X \mid Y) = \frac{P(X, Y)}{P(Y)}$$

$P(Y)$ can be computed by marginalization:

$$P(Y = y) = \sum_x P(X = x, Y = y)$$

If we can compute $P(X, Y)$, we can also compute $P(X \mid Y)$

Computing inferences in Bayes Nets

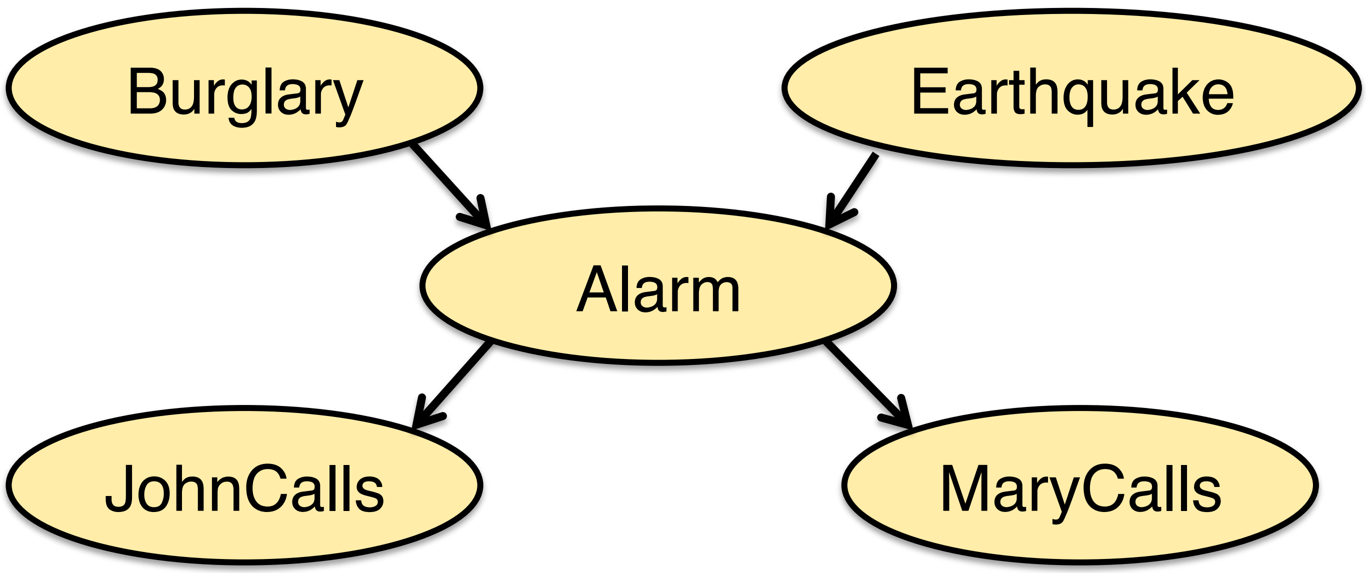
What is the probability of getting a strong letter if you are an intelligent student?

What about the other, hidden, variables **h** ?
We have to marginalize them out.

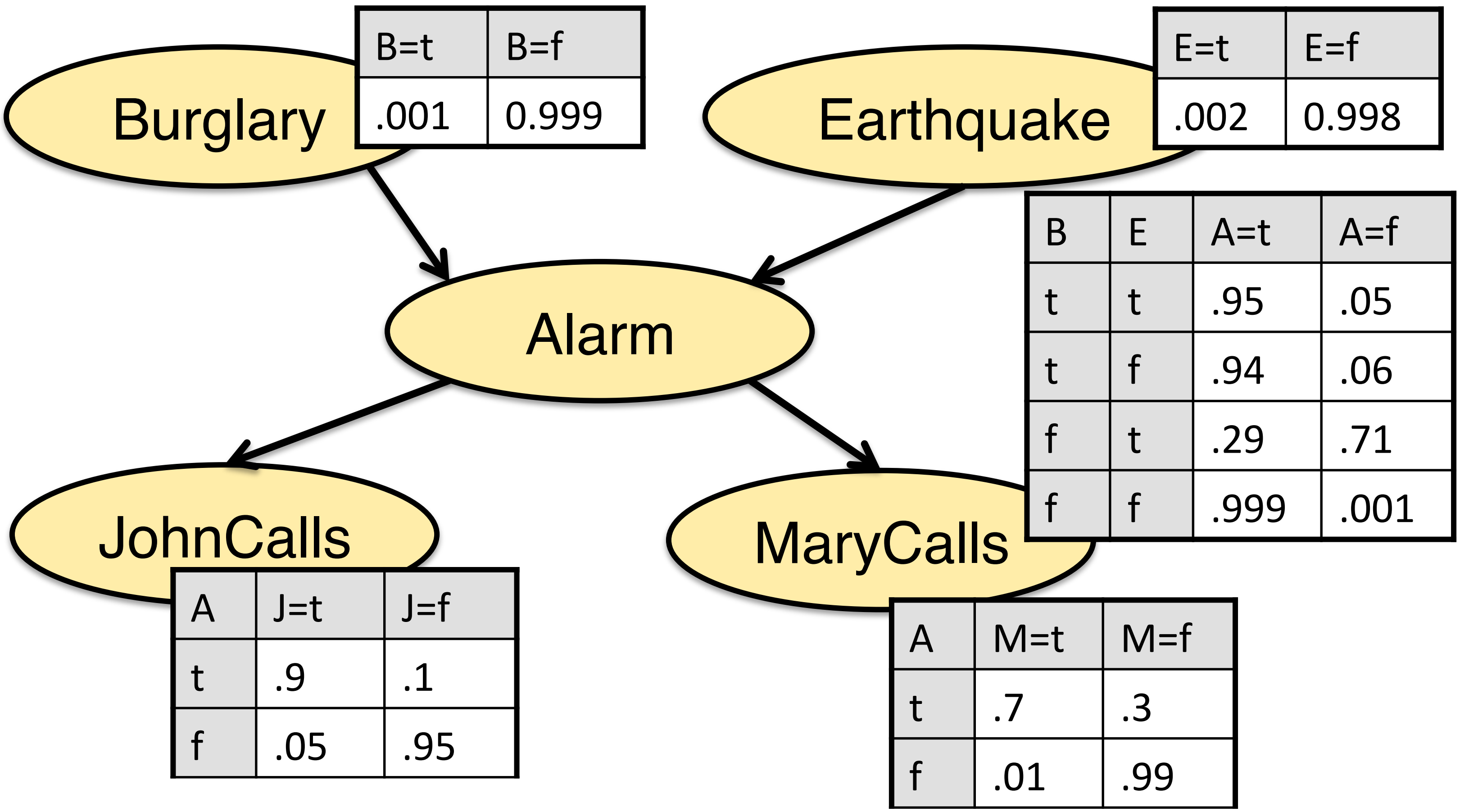
$$P(X = x, \mathbf{E} = \mathbf{e}) = \sum_{H=\mathbf{h}} P(x, \mathbf{h}, \mathbf{e})$$

The *Burglary* example

The alarm can go off if there is a burglary or if there is an earthquake.
If the alarm goes off, John might call.
If the alarm goes off, Mary might call.



The *Burglary* example



B=t	B=f
.001	0.999

E=t	E=f
.002	0.998

B	E	A=t	A=f
t	t	.95	.05
t	f	.94	.06
f	t	.29	.71
f	f	.999	.001

A	J=t	J=f
t	.9	.1
f	.05	.95

A	M=t	M=f
t	.7	.3
f	.01	.99

What is the probability of a burglary if John and Mary call?

Inference by enumeration

What is the probability of a burglary if John and Mary call?

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

What is the joint probability $P(b,j,m)$?

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

Inference by enumeration

What is the joint probability $P(b,j,m,e,a)$?

$$P(b,j,m,a,e)$$
$$= P(b)P(e)P(a \mid b,e)P(j \mid a)P(m \mid a)$$

So,

$$P(b,j,m) = \underbrace{\sum_{e}\sum_{a}}_{\text{4 summands = 20 computations}} \underbrace{P(e)P(b)P(a \mid b,e)P(j \mid a)P(m \mid a)}_{\text{5 terms}}$$

Inference by enumeration

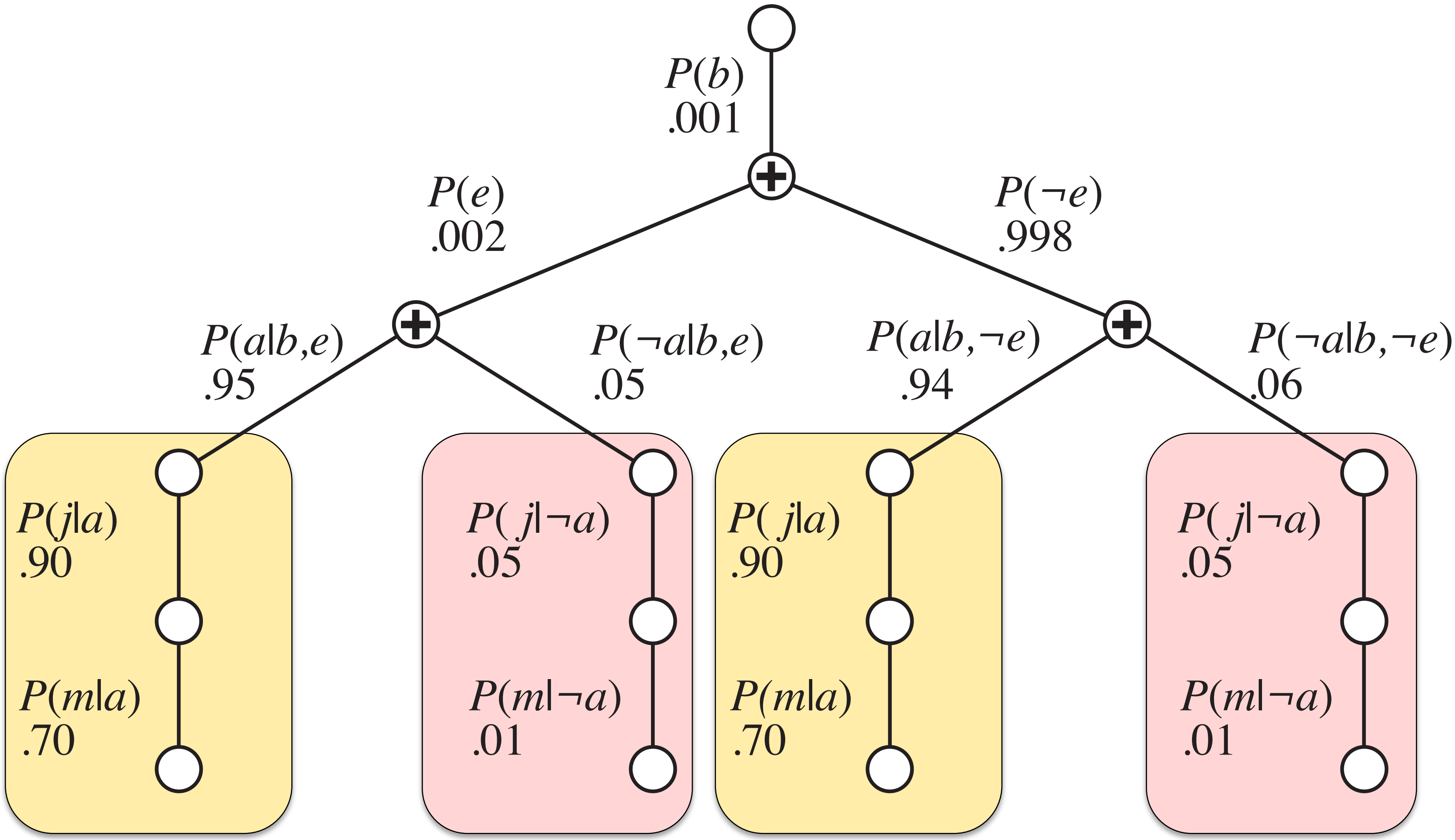
$$\sum_e \sum_a P(e)P(b)P(a \mid b,e)P(j \mid a)P(m \mid a)$$

$$= P(b)\sum_e \sum_a P(e)P(a \mid b,e)P(j \mid a)P(m \mid a)$$

$$= P(b)\underbrace{\sum_e}_{\text{2 summands: 14 computations}} \underbrace{P(e)\underbrace{\sum_a}_{\text{2 summands: 6 computations}} \underbrace{P(a \mid b,e)P(j \mid a)P(m \mid a)}_{\text{3 terms}}}_{\text{2 terms: 7 = 1 + 6 computations}}$$

2 terms: 15 = 1+ 14 computations

Or, visually….:



P(b)
.001
P(e)
.002
P(¬e)
.998
P(a|b,e)
.95
P(¬a|b,e)
.05
P(a|b,¬e)
.94
P(¬a|b,¬e)
.06
P(j|a)
.90
P(m|a)
.70
P(j|¬a)
.05
P(m|¬a)
.01
P(j|a)
.90
P(m|a)
.70
P(j|¬a)
.05
P(m|¬a)
.01

Enumeration is inefficient…

… it repeatedly evaluates the same subexpressions.

Time complexity for n Boolean variables: $O(2^n)$

Insights:

… it's better to evaluate each subexpression only once and save the result

… often, the number of subexpressions will be relatively small

Variable elimination

We want to compute

$$P(b, j, m, a, e)$$

$$= P(b)\sum_e P(e)\sum_a P(a \mid b, e)P(j \mid a)P(m \mid a)$$

- We have ordered the terms so that when we sum over a variable, we sum only over terms that depend on this variable.
- Now we evaluate this equation in right-to-left order and store intermediate results

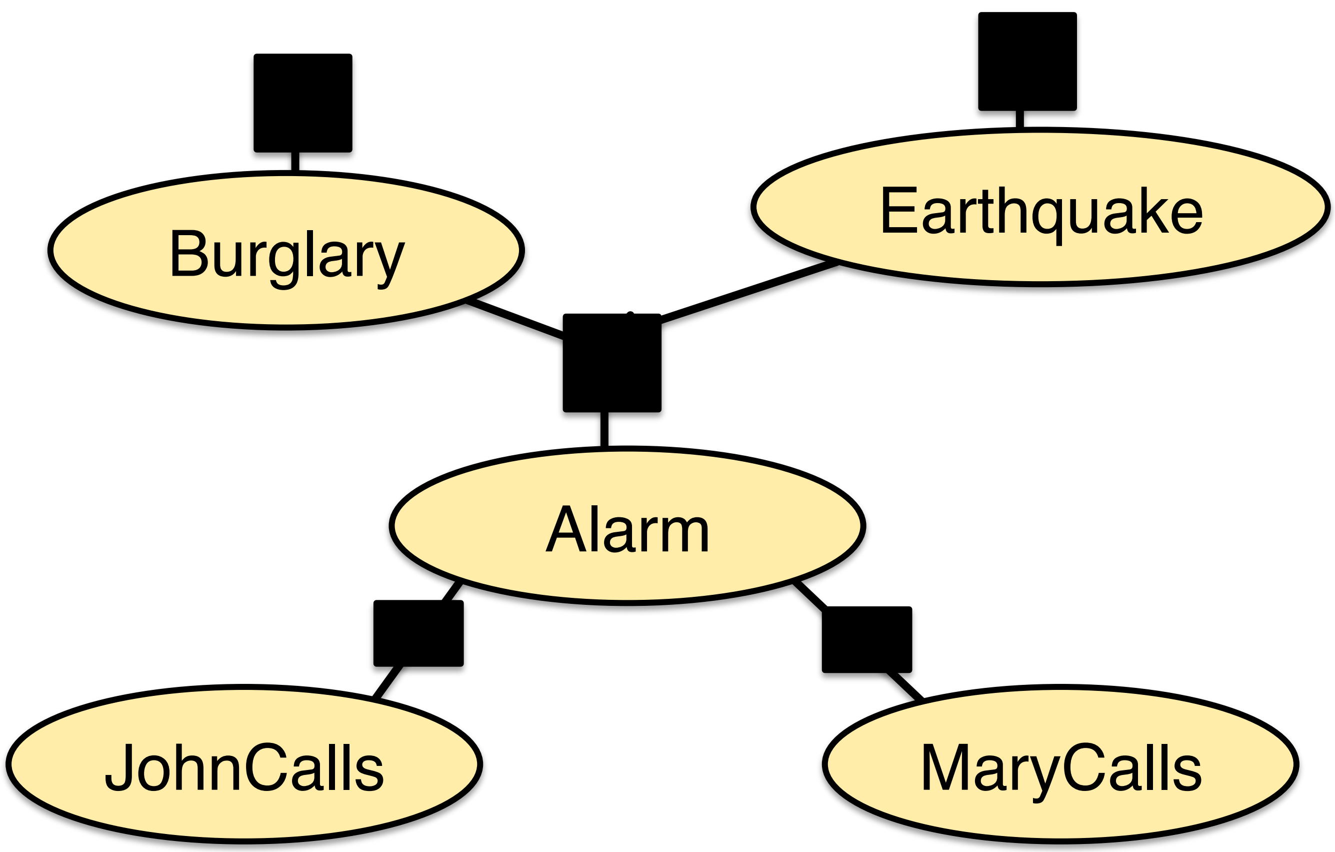
Factors

Variable elimination stores intermediate results in factors.

A factor is a table of probability distributions that is indexed by the variables of interest.

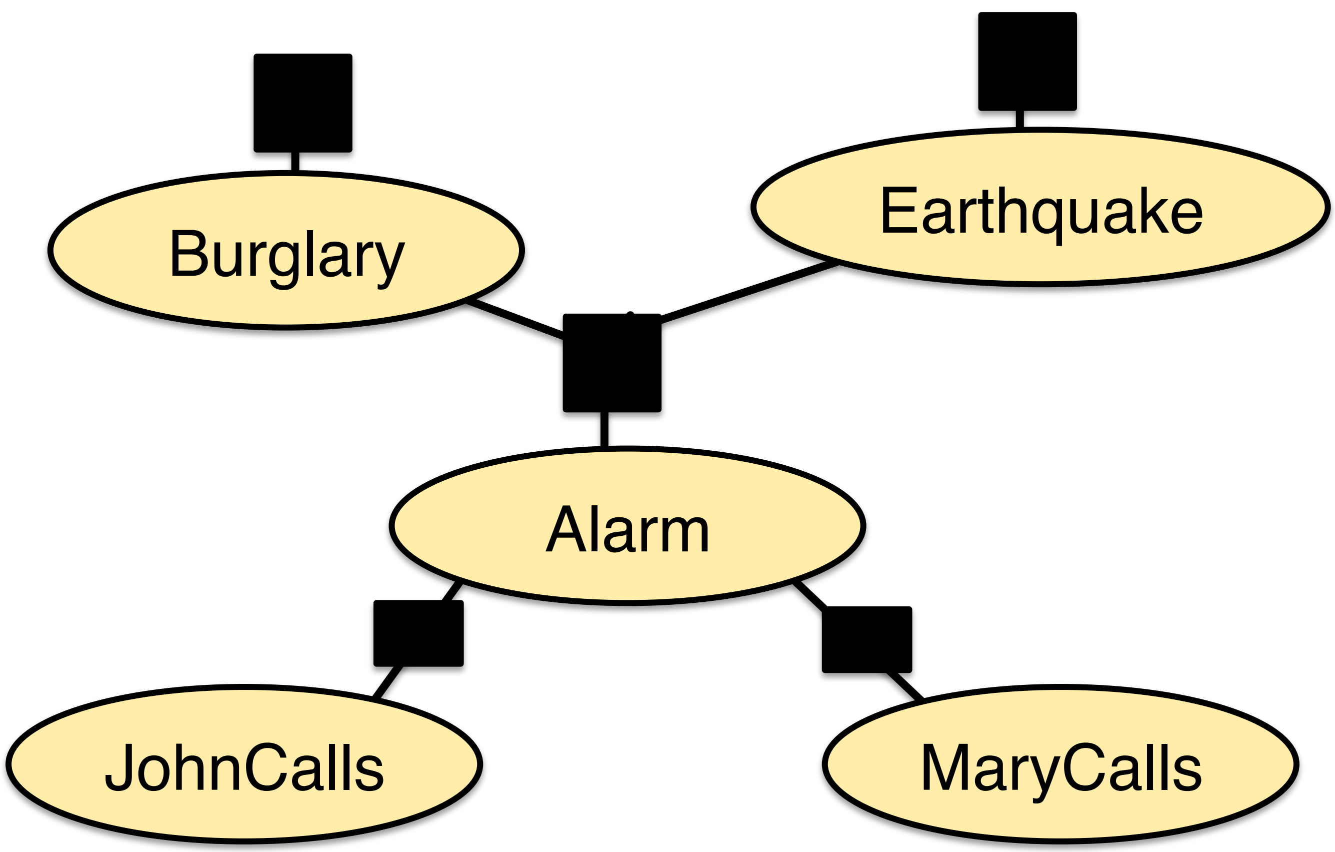
Factor graphs

We can also draw each Bayes Net as a factor graph:



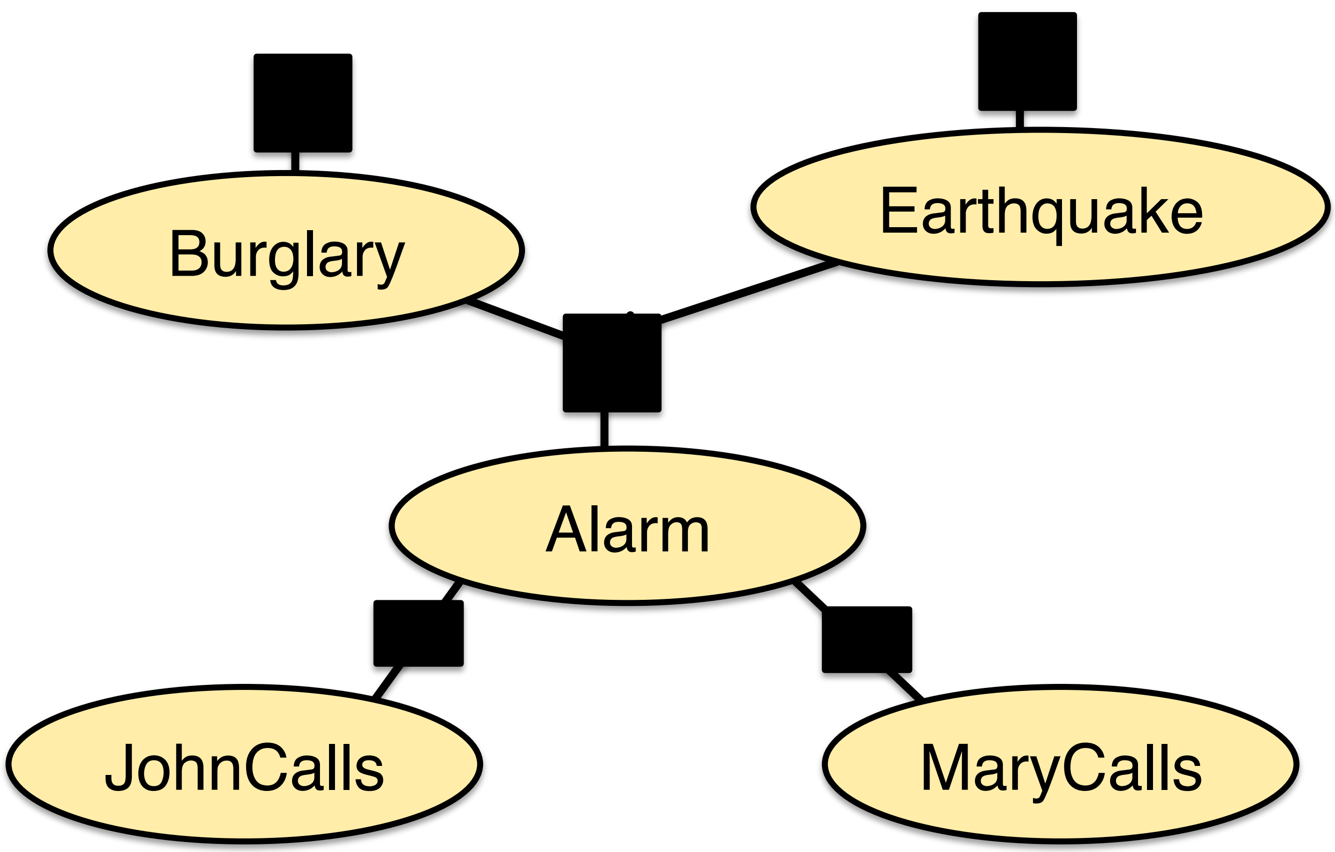
Factor graphs

Each CPT is one factor (= black box)



Factor graphs

The factor graph is a hypergraph



As factors

$$P(b, j, m, a, e)$$

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

$$= f1(B)\sum_{e} f2(E)\sum_{a} f3(A, B, E)f4(A)f5(A)$$

Evaluate the innermost sum

$$\text{f6}(B,E) = \sum_{a} \text{f3}(A,B,E) \times \text{f4}(A) \times \text{f5}(A)$$

$$= \text{f3}(a,B,E) \times \text{f4}(a) \times \text{f5}(a)$$

$$+ \text{f3}(\neg a,B,E) \times \text{f4}(\neg a) \times \text{f5}(\neg a)$$

$$P(b,j,m,a,e)$$

$$= \text{f1}(B) \times \sum_{e} \text{f2}(E) \times \sum_{a} \text{f3}(A,B,E) \times \text{f4}(A) \times \text{f5}(A)$$

$$= \text{f1}(B) \times \sum_{e} \text{f2}(E) \times \text{f6}(B,E)$$

So…

$$
\begin{aligned}
&P(b, j, m, a, e) \\
&= \mathrm{f1(B)} \times \sum_{\mathrm{e}} \mathrm{f2(E)} \times \sum_{a} \mathrm{f3(A,B,E)} \times \mathrm{f4(A)} \times \mathrm{f5(A)} \\
&= \mathrm{f1(B)} \times \sum_{\mathrm{e}} \mathrm{f2(E)} \times \mathrm{f6(B,E)}
\end{aligned}
$$

Evaluate the next sum

$$\text{f7(B)} = \sum_{\text{e}} \text{f2(E)} \times \text{f6(B,E)}$$

$$P(b, j, m, a, e)$$

$$= \text{f1(B)} \times \sum_{\text{e}} \text{f2(E)} \times \sum_{a} \text{f3(A,B,E)} \times \text{f4(A)} \times \text{f5(A)}$$

$$= \text{f1(B)} \times \sum_{\text{e}} \text{f2(E)} \times \text{f6(B,E)}$$

$$= \text{f1(B)} \times \text{f7(B)}$$

Complexity of variable elimination

Polytree networks:
only one (undirected) path between each pair of nodes.

Complexity is linear in the size of the network (number of CPT entries)

Complexity of variable elimination

Multiply connected networks:
multiple (undirected) paths between some pair of nodes.

Complexity is **exponential** in the size of the network (number of CPT entries)

– (Similar to finding number of models in propositional logic)