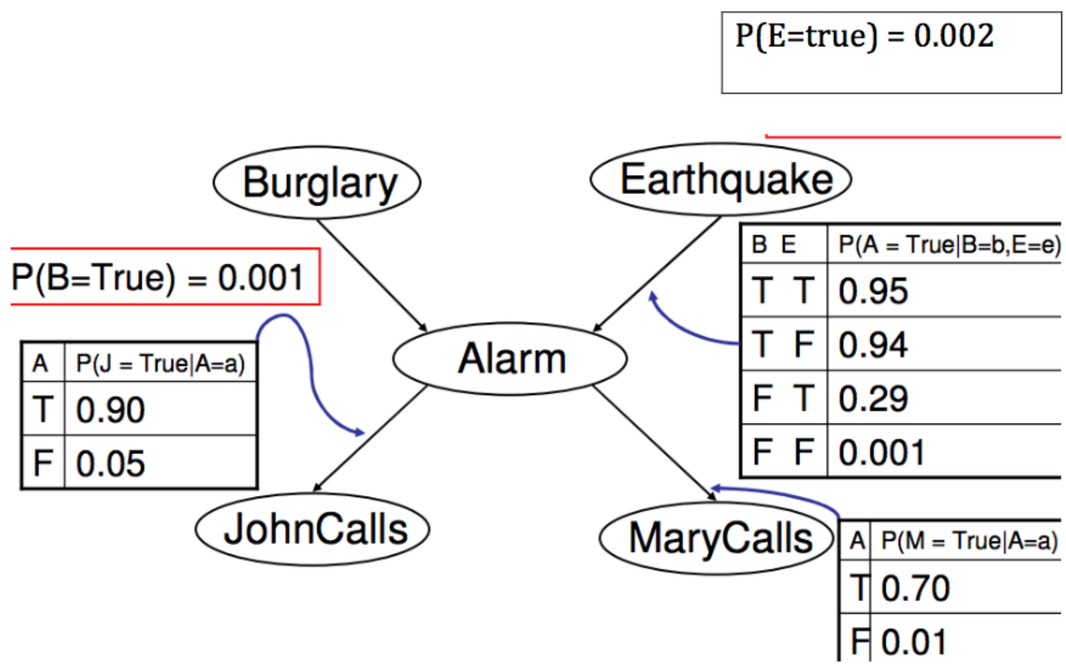
Earthquake and Burglary question (Bayes Net and conditional independent)

I am at work, and my neighbor John calls to say that the alarm in my house went off. My neighbor Mary also calls to say the alarm went off. Sometimes these neighbors tell the truth but sometimes they lie. The alarm might have been set off by a burglar, but sometimes it goes off because of minor earthquake. This system can be modeled with the following Bayes Net:



1.What's the probability that the alarm went off but there is neither an earthquake nor a burglar?

Solution:
We'll introduce random variables J, M, A, B and E to represent the events mentioned in the problem. Each of these is a binary random variable. We use the notation X^T to indicate that random variable X has value true, i.e. has happened, and X^F to indicate it is false.

Then compute the joint probability:

$$
\begin{aligned}
P(J^T, M^T, A^T, B^F, E^F) &= P(J^T | M^T, A^T, B^F, E^F) P(M^T, A^T, B^F, E^F) \\
&= P(J^T | A^T) P(M^T | A^T, B^F, E^F) P(A^T, B^F, E^F) \\
&= P(J^T | A^T) P(M^T | A^T) P(A^T | B^F, E^F) P(B^F, E^F) \\
&= P(J^T | A^T) P(M^T | A^T) P(A^T | B^F, E^F) P(B^F) P(E^F) \\
&= 0.90 * 0.70 * 0.001 * 0.999 * 0.998 = 0.0006
\end{aligned}
$$

2. Consider again the Bayes net above. For each of the following, indicate whether the conditional independence statement is true or false given the model:

a. B and E are independent
b. B and M are independent
c. J and M are independent
d. J and M are independent given B
e. J and M are independent given A
f. B and E are independent given A
g. B and E are independent given M

Solution:
a. yes
b. no
c. no
d. no
e. yes
f. no
g. no

Taken from Probabilistic Graphical Models: Principles and Techniques, by Koller and Friedman

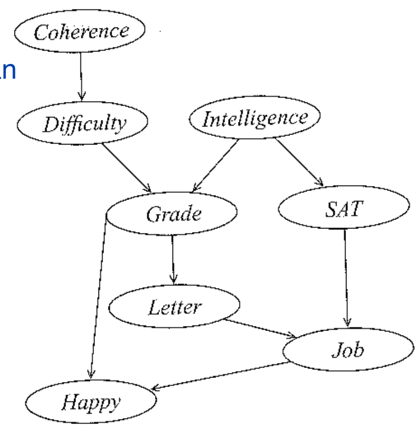


Figure 9.8 **The** Extended-Student **Bayesian network**

run the elimination algorithm. As for Bayesian networks, we then apply the variable elimination algorithm to the set $\boldsymbol{Z} = \mathcal{X} - \boldsymbol{Y}$. The procedure returns an *unnormalized* factor over the query variables $\boldsymbol{Y}$. The distribution over $\boldsymbol{Y}$ can be obtained by normalizing the factor; the partition function is simply the normalizing constant.

Example 9.1

Let us demonstrate the procedure on a nontrivial example. Consider the network demonstrated in figure 9.8, which is an extension of our Student *network. The chain rule for this network asserts that*

$$
\begin{aligned}
P(C, D, I, G, S, L, J, H) &= P(C)P(D \mid C)P(I)P(G \mid I, D)P(S \mid I) \\
&\quad P(L \mid G)P(J \mid L, S)P(H \mid G, J) \\
&= \phi_C(C)\phi_D(D, C)\phi_I(I)\phi_G(G, I, D)\phi_S(S, I) \\
&\quad \phi_L(L, G)\phi_J(J, L, S)\phi_H(H, G, J).
\end{aligned}
$$

We will now apply the VE algorithm to compute $P(J)$. We will use the elimination ordering: C, D, I, H, G, S, L:

1. *Eliminating C: We compute the factors*

$$
\begin{aligned}
\psi_1(C, D) &= \phi_C(C) \cdot \phi_D(D, C) \\
\tau_1(D) &= \sum_C \psi_1.
\end{aligned}
$$

2. *Eliminating D: Note that we have already eliminated one of the original factors that involve D — $\phi_D(D, C) = P(D \mid C)$. On the other hand, we introduced the factor $\tau_1(D)$ that involves*

Taken from Probabilistic Graphical Models: Principles and Techniques by Koller and Friedman

D. Hence, we now compute:

$$\begin{aligned}\psi_2(G, I, D) &= \phi_G(G, I, D) \cdot \tau_1(D) \\ \tau_2(G, I) &= \sum_D \psi_2(G, I, D).\end{aligned}$$

3. *Eliminating I: We compute the factors*

$$\begin{aligned}\psi_3(G, I, S) &= \phi_I(I) \cdot \phi_S(S, I) \cdot \tau_2(G, I) \\ \tau_3(G, S) &= \sum_I \psi_3(G, I, S).\end{aligned}$$

4. *Eliminating H: We compute the factors*

$$\begin{aligned}\psi_4(G, J, H) &= \phi_H(H, G, J) \\ \tau_4(G, J) &= \sum_H \psi_4(G, J, H).\end{aligned}$$

Note that $\tau_4 \equiv 1$ (all of its entries are exactly 1): we are simply computing $\sum_H P(H \mid G, J)$, which is a probability distribution for every G, J, and hence sums to 1. A naive execution of this algorithm will end up generating this factor, which has no value. Generating it has no impact on the final answer, but it does complicate the algorithm. In particular, the existence of this factor complicates our computation in the next step.

5. *Eliminating G: We compute the factors*

$$\begin{aligned}\psi_5(G, J, L, S) &= \tau_4(G, J) \cdot \tau_3(G, S) \cdot \phi_L(L, G) \\ \tau_5(J, L, S) &= \sum_G \psi_5(G, J, L, S).\end{aligned}$$

Note that, without the factor $\tau_4(G, J)$, the results of this step would not have involved J.

6. *Eliminating S: We compute the factors*

$$\begin{aligned}\psi_6(J, L, S) &= \tau_5(J, L, S) \cdot \phi_J(J, L, S) \\ \tau_6(J, L) &= \sum_S \psi_6(J, L, S).\end{aligned}$$

7. *Eliminating L: We compute the factors*

$$\begin{aligned}\psi_7(J, L) &= \tau_6(J, L) \\ \tau_7(J) &= \sum_L \psi_7(J, L).\end{aligned}$$

We summarize these steps in table 9.1.

Note that we can use any elimination ordering. For example, consider eliminating variables in the order G, I, S, L, H, C, D. We would then get the behavior of table 9.2. The result, as before, is precisely $P(J)$. However, note that this elimination ordering introduces factors with much larger scope. We return to this point later on. ■

Taken from Probabilistic Graphical Models:
Principles and Techniques
by Koller and Friedman

Step	Variable eliminated	Factors used	Variables involved	New factor
1	C	$\phi_C(C)$, $\phi_D(D,C)$	C,D	$\tau_1(D)$
2	D	$\phi_G(G,I,D)$, $\tau_1(D)$	G,I,D	$\tau_2(G,I)$
3	I	$\phi_I(I)$, $\phi_S(S,I)$, $\tau_2(G,I)$	G,S,I	$\tau_3(G,S)$
4	H	$\phi_H(H,G,J)$	H,G,J	$\tau_4(G,J)$
5	G	$\tau_4(G,J)$, $\tau_3(G,S)$, $\phi_L(L,G)$	G,J,L,S	$\tau_5(J,L,S)$
6	S	$\tau_5(J,L,S)$, $\phi_J(J,L,S)$	J,L,S	$\tau_6(J,L)$
7	L	$\tau_6(J,L)$	J,L	$\tau_7(J)$

Table 9.1 **A run of variable elimination for the query** $P(J)$

Step	Variable eliminated	Factors used	Variables involved	New factor
1	G	$\phi_G(G,I,D)$, $\phi_L(L,G)$, $\phi_H(H,G,J)$	G,I,D,L,J,H	$\tau_1(I,D,L,J,H)$
2	I	$\phi_I(I)$, $\phi_S(S,I)$, $\tau_1(I,D,L,S,J,H)$	S,I,D,L,J,H	$\tau_2(D,L,S,J,H)$
3	S	$\phi_J(J,L,S)$, $\tau_2(D,L,S,J,H)$	D,L,S,J,H	$\tau_3(D,L,J,H)$
4	L	$\tau_3(D,L,J,H)$	D,L,J,H	$\tau_4(D,J,H)$
5	H	$\tau_4(D,J,H)$	D,J,H	$\tau_5(D,J)$
6	C	$\tau_5(D,J)$, $\phi_C(C)$, $\phi_D(D,C)$	D,J,C	$\tau_6(D,J)$
7	D	$\tau_6(D,J)$	D,J	$\tau_7(J)$

Table 9.2 **A different run of variable elimination for the query** $P(J)$

9.3.1.3 Semantics of Factors

It is interesting to consider the semantics of the intermediate factors generated as part of this computation. In many of the examples we have given, they correspond to marginal or conditional probabilities in the network. However, although these factors often correspond to such probabilities, this is not always the case. Consider, for example, the network of figure 9.9a. The result of eliminating the variable X is a factor

$$\tau(A,B,C) = \sum_X P(X) \cdot P(A \mid X) \cdot P(C \mid B, X).$$

This factor does not correspond to any probability or conditional probability in this network. To understand why, consider the various options for the meaning of this factor. Clearly, it cannot be a conditional distribution where B is on the left hand side of the conditioning bar (for example, $P(A,B,C)$), as $P(B \mid A)$ has not yet been multiplied in. The most obvious candidate is $P(A,C \mid B)$. However, this conjecture is also false. The probability $P(A \mid B)$ relies heavily on the properties of the CPD $P(B \mid A)$; for example, if B is deterministically equal to A, $P(A \mid B)$ has a very different form than if B depends only very weakly on A. Since the CPD $P(B \mid A)$ was not taken into consideration when computing $\tau(A,B,C)$, it cannot represent the conditional probability $P(A,C \mid B)$. In general, we can verify that this factor