Bayesian Networks

Popular framework for working with probability and uncertain worlds. Very useful in artificial intelligence.

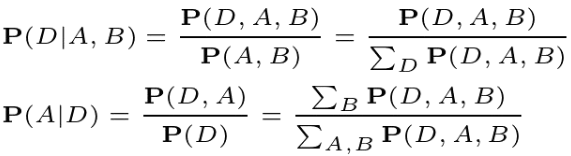
Joint Probability

In theory given the **full joint distribution** of a set of random variables, it is possible to answer any query by applying the rules of probability.

Fox example, if we know



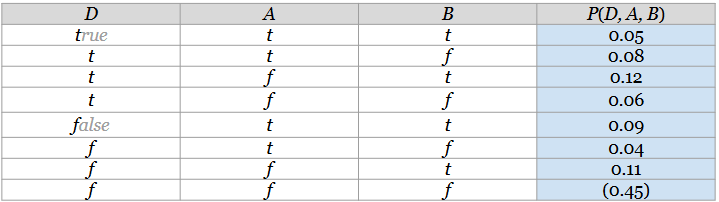
We can answer any query such as



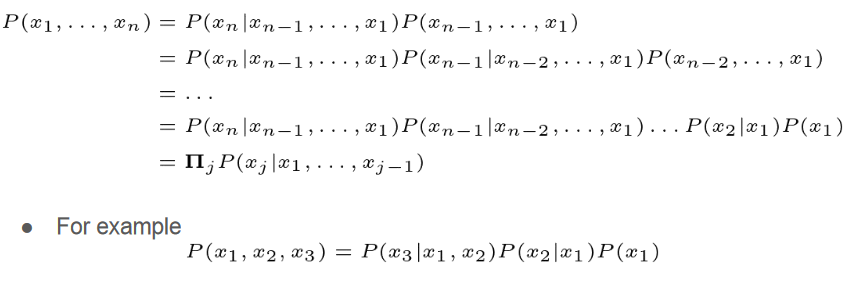
Limitations

However, the size of the joint distribution grows exponentially with number of variables!

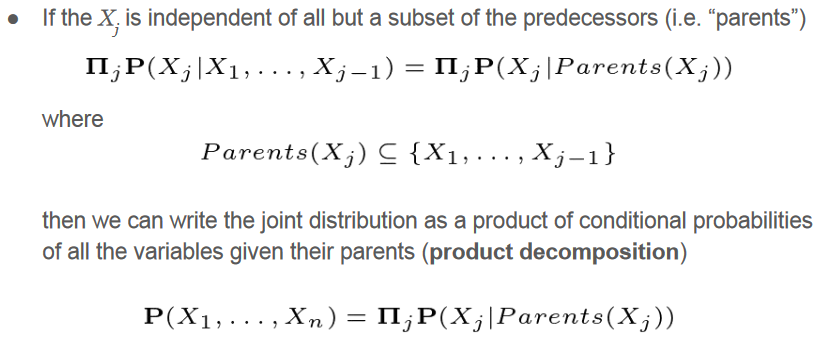
For the same example, assuming 3 Boolean variables, we need 2­3 ­= 8 values



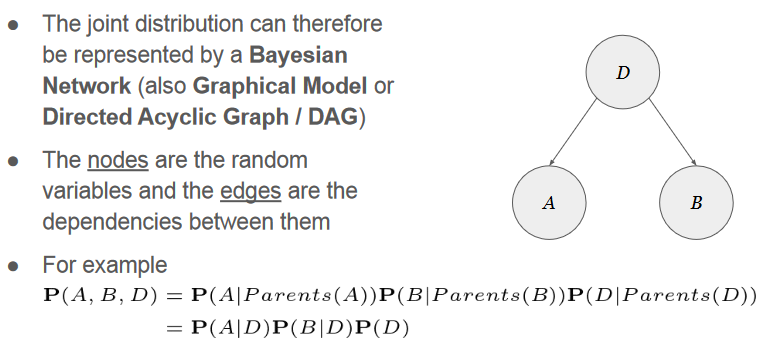
Chain Rule



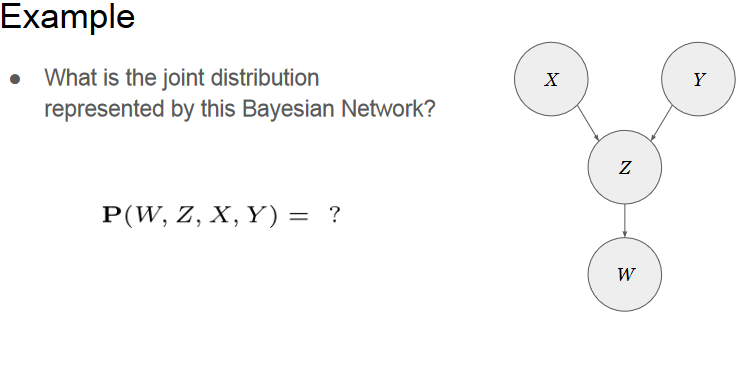
Product Decomposition

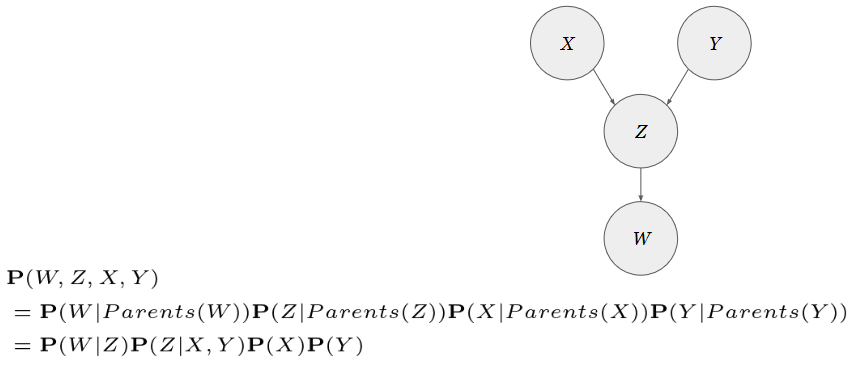


From joint Distribution to Bayesian Networks



Example



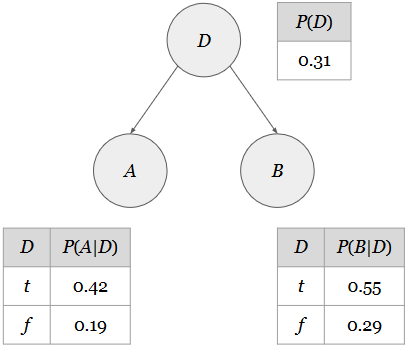


Conditional Probability Tables

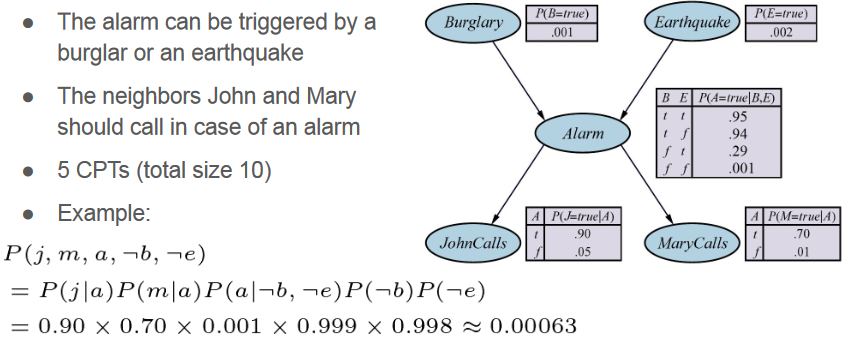
In case of discrete variables, a Conditional Probability Table (CPT) is associated to each node

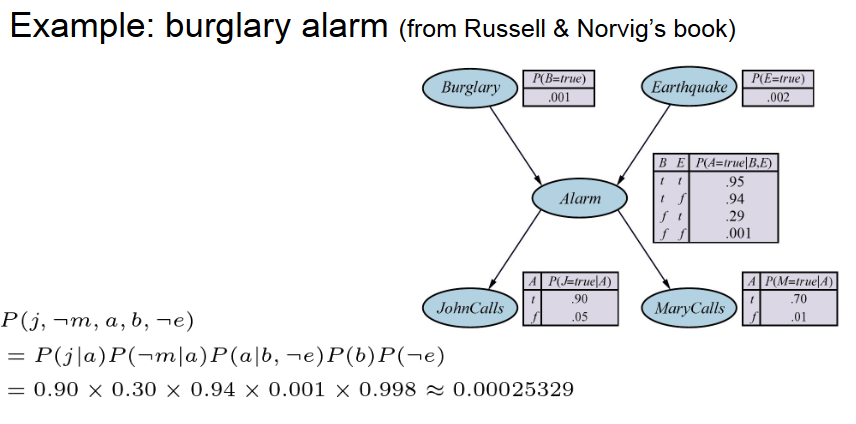
In the example, we now need to store 5 probability values instead of the 7 values for the full joint distribution (~30% less)

From these, we can derive all the probabilities of the joint distribution and therefore answer any query

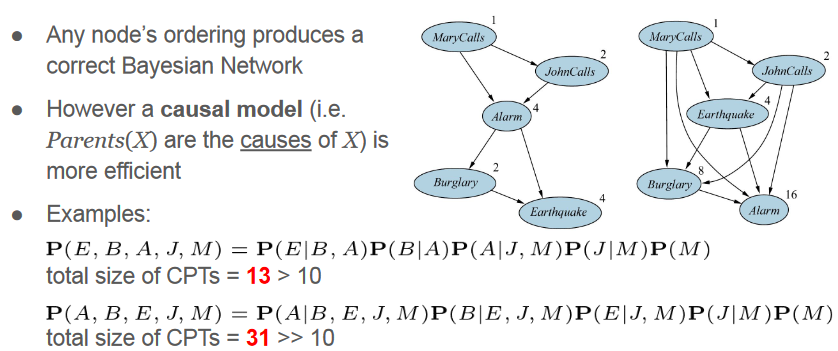


Example: Burglary alarm (from Russell & Norvig’s book)





Network Construction



Advantages of Product Decomposition

Besides storage space, the product decomposition represented by a Bayesian Network improves also the accuracy of the estimation

Example:

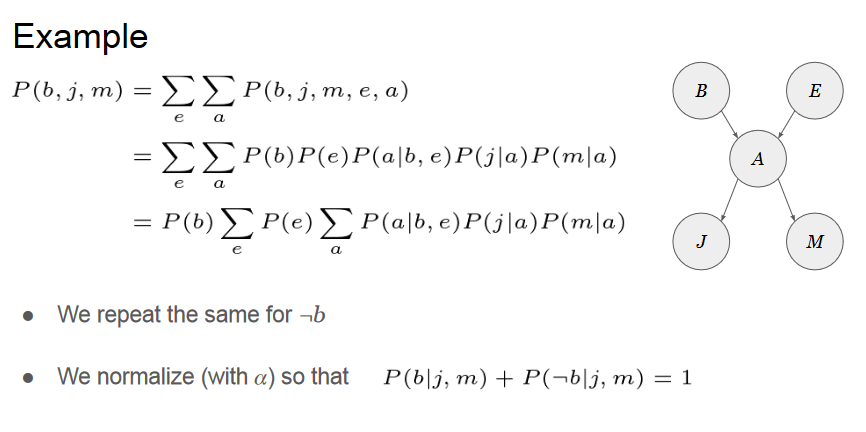
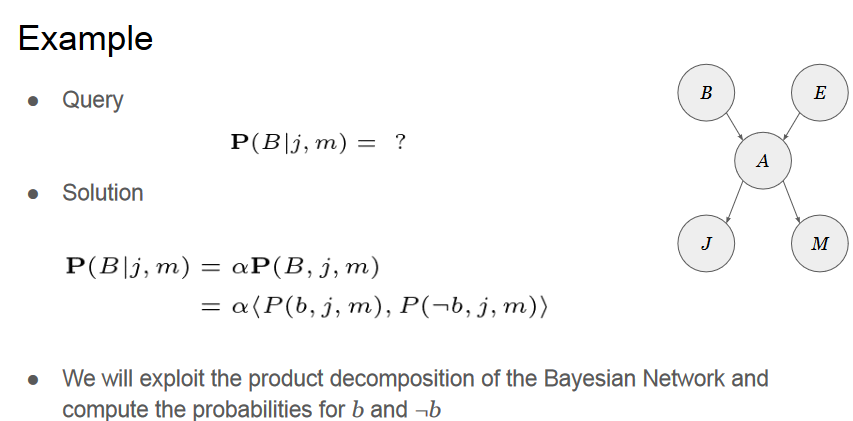
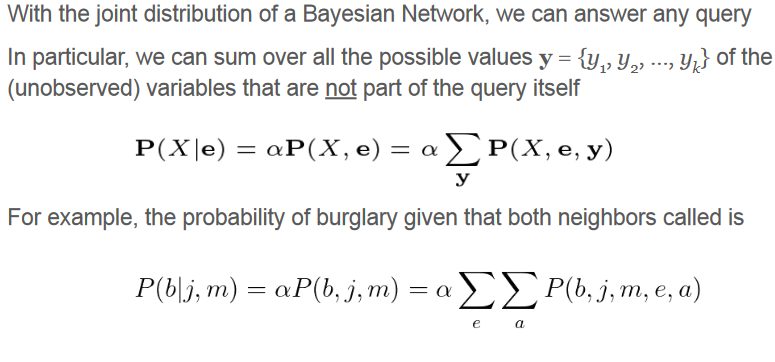
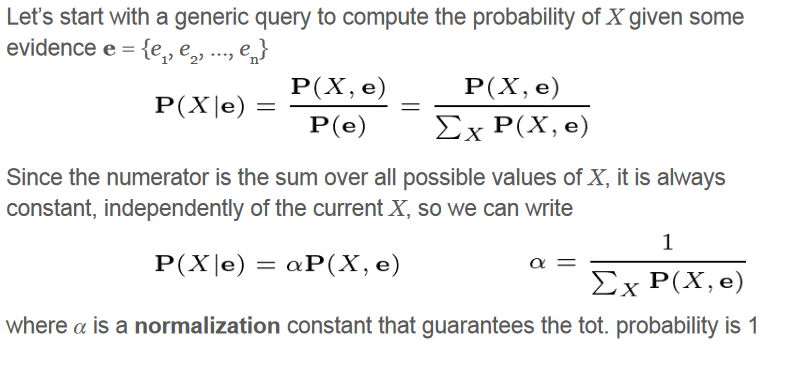
We want to estimate the probability of P(A,B,D) from a dataset with only 25 samples

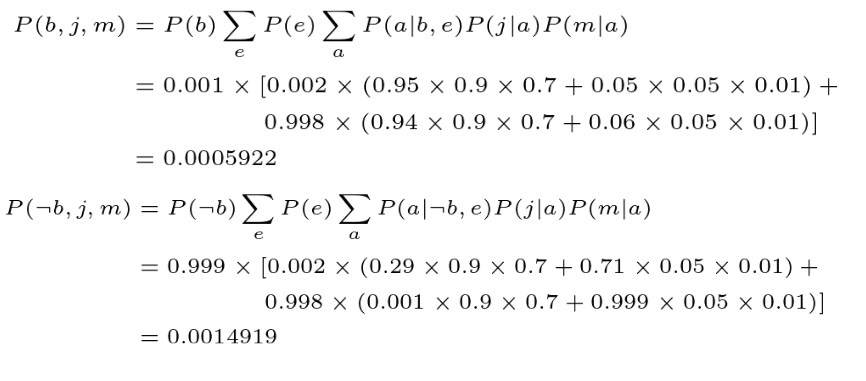
Since there are 8 possible cases (A,B,D) cases, we will have 25/8 = ~3 samples in average for each case, but in reality some will have many samples, some none!

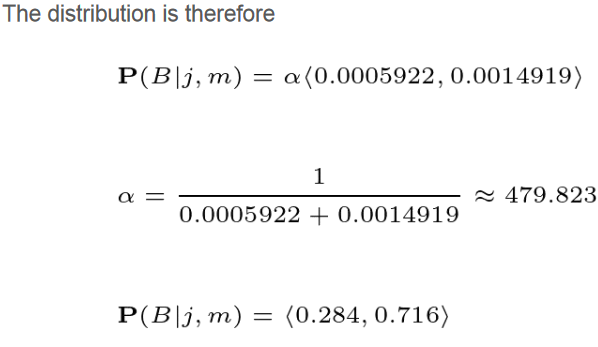
If we estimate P(A|D), P(B|D), and P(D), instead it’s more likely to have enough samples

25/4 ≈ 6 samples in average for P(A|D)  
25/4 ≈ 6 samples in average for P(B|D)  
25/2 ≈ 12 samples in average for P(D)

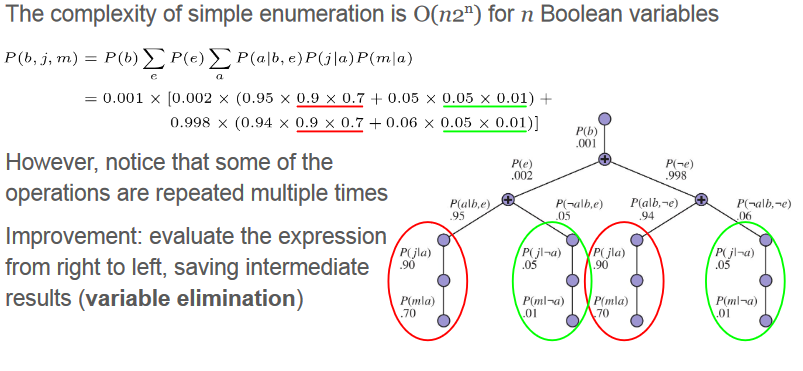
Inference by Enumeration







Variable Elimination



Reading

● Pearl et al. “Causal Inference in Statistics -- A Primer”  
○ Sec 1.4  
○ Sec 1.5.2  
● Russell & Norvig “Artificial Intelligence -- A Modern Approach”  
○ Ch 14 (until Sec 14.4)