Figure 1: Bayesian Network

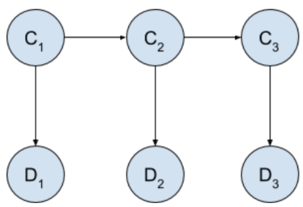


Figure 2: After removing non-ancestor nodes

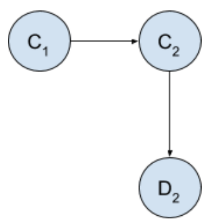


Figure 3: Converting to factor graph

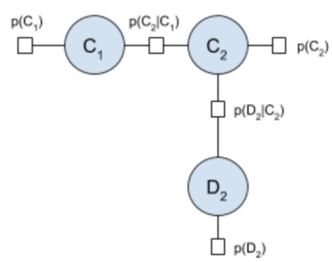
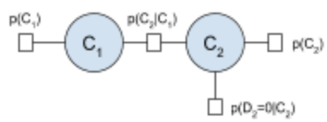


Figure 4: Condition on $D_2 = 0$



Remove any disconnected node is a no-op. Hence, we proceed to the step to "Do the work!"

$$\mathbb{P}(C_2|D_2 = 0) = p(C_2)p(D_2 = 0|C_2)f(C_2)$$

where

$$f(C_2) \propto \sum_{C_1} p(C_1)p(C_2|C_1)$$

Let us first evaluate $p(C_2)$:

$$\begin{aligned}\mathbb{P}(C_2) &= \sum_{C_1} p(C_1)p(C_2|C_1) \\ &= \frac{1}{2}(\epsilon + 1 - \epsilon) \\ &= \frac{1}{2}\end{aligned}$$

Overall, taking α as the constant of proportion:

$$\begin{aligned}\mathbb{P}(C_2|D_2 = 0) &= \alpha p(C_2)p(D_2 = 0|C_2)\sum_{C_1} p(C_1)p(C_2|C_1) \\ &= \alpha\frac{1}{2}p(D_2 = 0|C_2) * \frac{1}{2}(\epsilon + 1 - \epsilon) \\ &= \frac{\alpha}{4}p(D_2 = 0|C_2)\end{aligned}$$

$$\begin{aligned}\mathbb{P}(C_2 = 0|D_2 = 0) &= \frac{\alpha}{4}p(D_2 = 0|C_2 = 0) \\ &= \frac{\alpha}{4}(1 - \eta)\end{aligned}$$

$$\begin{aligned}\mathbb{P}(C_2 = 1|D_2 = 0) &= \frac{\alpha}{4}p(D_2 = 0|C_2 = 1) \\ &= \frac{\alpha}{4}\eta\end{aligned}$$

Since the two should sum to 1, we have:

$$\begin{aligned}1 &= \frac{\alpha}{4}(1 - \eta) + \frac{\alpha}{4}\eta \\ &= \frac{\alpha}{4}(1 - \eta + \eta)\end{aligned}$$

Hence:

$$\mathbb{P}(C_2 = 1|D_2 = 0) = \eta$$

Figure 5: Bayesian Network

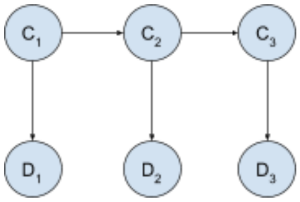


Figure 6: After removing non-ancestor nodes

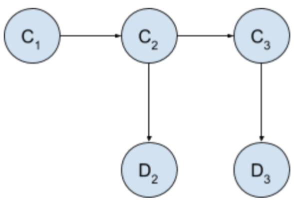


Figure 7: Converting to factor graph

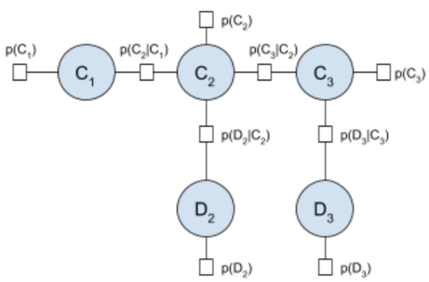
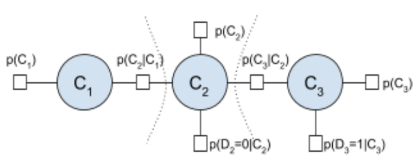


Figure 8: Condition on $D_2 = 0$



Remove any disconnected node is a no-op. Hence, we proceed to the step to "Do the work!". Final required probability is:

$$\mathbb{P}(C_2|D_2 = 0, D_3 = 1) = p(C_2)p(D_2 = 0|C_2)f_1(C_2)f_2(C_2)$$

Where:

$$f_1(C_2) \propto \sum_{C_1} p(C_1)p(C_2|C_1)$$
$$f_2(C_2) \propto \sum C_3 p(C_3)p(D_3 = 1|C_3)p(C_3|C_2)$$

Hence, taking α as the constant of proportion, we have:

$$\mathbb{P}(C_2|D_2 = 0, D_3 = 1) = \alpha p(C_2)p(D_2 = 0|C_2)(\sum_{C_1} p(C_1)p(C_2|C_1))(\sum_{C_3} p(C_3)p(D_3 = 1|C_3)p(C_3|C_2))$$

We know that $p(C_1) = p(C_2) = \frac{1}{2}$. Also, with the same calculation:

$$\mathbb{P}(C_3) = \sum_{C_2} p(C_2)p(C_3|C_2) = \frac{1}{2}(1 - \epsilon + \epsilon) = \frac{1}{2}$$

So, we have:

$$\begin{aligned}
\mathbb{P}(C_2|D_2 = 0, D_3 = 1) &= \alpha\frac{1}{2}p(D_2 = 0|C_2)(\sum_{C_1} \frac{1}{2}p(C_2|C_1))(\sum_{C_3} \frac{1}{2}p(D_3 = 1|C_3)p(C_3|C_2)) \\
&= \frac{\alpha}{8}p(D_2 = 0|C_2)(\epsilon + 1 - \epsilon)(\sum_{C_3} p(D_3 = 1|C_3)p(C_3|C_2)) \\
&= \frac{\alpha}{8}p(D_2 = 0|C_2)\sum_{C_3} p(D_3 = 1|C_3)p(C_3|C_2)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(C_2 = 0|D_2 = 0, D_3 = 1) &= \frac{\alpha}{8}\mathbb{P}(D_2 = 0|C_2 = 0)\sum_{C_3} p(D_3 = 1|C_3)p(C_3|C_2 = 0) \\
&= \frac{\alpha}{8}(1 - \eta)(\eta(1 - \epsilon) + (1 - \eta)\epsilon) \\
&= \frac{\alpha(\eta - 3\eta\epsilon + \epsilon - \eta^2 + 2\eta^2\epsilon)}{8}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(C_2 = 1|D_2 = 0, D_3 = 1) &= \frac{\alpha}{8}p(D_2 = 0|C_2 = 1)\sum_{C_3} p(D_3 = 1|C_3)p(C_3|C_2 = 1) \\
&= \frac{\alpha}{8}\eta(\eta\epsilon + (1 - \eta)(1 - \epsilon)) \\
&= \frac{\alpha(\eta - \eta^2 - \eta\epsilon + 2\eta^2\epsilon)}{8}
\end{aligned}$$

Since the sum of the two probabilities should be 1, we have:

$$\begin{aligned}
1 &= \frac{\alpha(\eta - 3\eta\epsilon + \epsilon - \eta^2 + 2\eta^2\epsilon)}{8}) + \frac{\alpha(\eta - \eta^2 - \eta\epsilon + 2\eta^2\epsilon)}{8} \\
&= \frac{\alpha(2\eta - 4\eta\epsilon + \epsilon - 2\eta^2 + 4\eta^2\epsilon)}{8} \\
\alpha &= \frac{8}{2\eta - 4\eta\epsilon + \epsilon - 2\eta^2 + 4\eta^2\epsilon}
\end{aligned}$$

Therefore:

$$\mathbb{P}(C_2 = 1 | D_2 = 0, D_3 = 1) = \frac{\eta - \eta^2 - \eta\epsilon + 2\eta^2\epsilon}{2\eta - 4\eta\epsilon + \epsilon - 2\eta^2 + 4\eta^2\epsilon}$$

i)

$$\mathbb{P}(C_2 = 1 | D_2 = 0) = \eta = 0.2$$

$$\begin{aligned}\mathbb{P}(C_2 = 1 | D_2 = 0, D_3 = 1) &= \frac{0.2 - 0.04 - 0.02 + 0.008}{0.4 - 0.08 + 0.1 - 0.08 + 0.016} \\ &= \frac{0.148}{0.356} &= 0.4157\end{aligned}$$

ii) Addiing the second reading $D_3 = 1$ increases the probability of C_2 being 1. In the first case, the probability is small since expected C_2 and observed D_2 differ which has a very small probability.

Later, we observe D_3 to be 1, which means that C_3 has a high probability of being 1. This can be used to also say that C_2 has a high probability of being 1. So, with good probability at this point we can say that the sensor reading D_2 was probably flawed and and the probability of C_2 being 1 increases.

iii) Putting $\eta = 0.2$ and using the two equations:

$$\begin{aligned}\eta &= \frac{\eta - \eta^2 - \eta\epsilon + 2\eta^2\epsilon}{2\eta - 4\eta\epsilon + \epsilon - 2\eta^2 + 4\eta^2\epsilon} \\ 0.2 &= \frac{0.2 - 0.04 - 0.2\epsilon + 0.08\epsilon}{0.4 - 0.8\epsilon + \epsilon - 0.08 + 0.16\epsilon} \\ 0.2 &= \frac{0.16 - 0.12\epsilon}{0.32 + 0.36\epsilon} \\ 0.064 + 0.072\epsilon &= 0.16 - 0.12\epsilon \\ 0.192\epsilon &= 0.096 \\ \epsilon &= 96/192 = 0.5\end{aligned}$$

At the value of 0.5, we can say that $p(c_t | c_{t-1}) = 0.5$. This essentially means that the new location of the car is not dependent on the location during the previous timestamp. This can also be rephrased as the new location of the car cannot provide any information of the location of the car at the previous timestamp (if it was unknown).

Hence, we can disconnect the path from C_2 to C_3, which essentially means that C_2 and D_3 won't be connected either and information about D_3 will not provide any value in determining C_2. This explains why providing D_3 does not affect the posterior probability of C_2.

$$\begin{aligned}
\mathbb{P}(C_{11} = c_{11}, C_{12} = c_{12} | E_1 = e_1) &\propto p(c_{11})p(c_{12})p(e_1|c_{11}, c_{12}) \\
&\propto p(c_{11})p(c_{12})\frac{1}{2}(p(e_{11}|c_{11})p(e_{12}|c_{12}) + p(e_{12}|c_{11})p(e_{11}|c_{12})) \\
&\propto p(c_{11})p(c_{12})\frac{1}{2}(p_{\mathcal{N}}(e_{11}; ||a_1 - c_{11}||; \sigma^2)p_{\mathcal{N}}(e_{12}; ||a_1 - c_{12}||; \sigma^2) \\
&+ p_{\mathcal{N}}(e_{12}; ||a_1 - c_{11}||; \sigma^2)p_{\mathcal{N}}(e_{11}; ||a_1 - c_{12}||; \sigma^2))
\end{aligned}$$

$\mathbb{P}(C_{11} = c_{11}, \ldots, C_{1K} = c_{1K} \mid E_1 = e_1)$ will have at least one maxima. Since the prior distributions $p(c_{1i})$ are same for K cars, and the sensor differentiate between the cars either, all the cars can be treated as indistinguishable and can be permuted in any order.

Given that we are assuming that the car locations that maximize this probability are unique, we can have $K!$ permutations.

In the bayesian network, since the sensor reading E_i depends on all the K car positions, converting it to factor graph will yield a factor of size K+1. Hence, Markov blanket of any node in this graph has size at least K.

It is also easy to find an elimination strategy with which we always remove nodes with arity of K. We can remove cars one by one for each time stamp and all removed nodes will have a parity of at max K. Hence, the tree width is K.