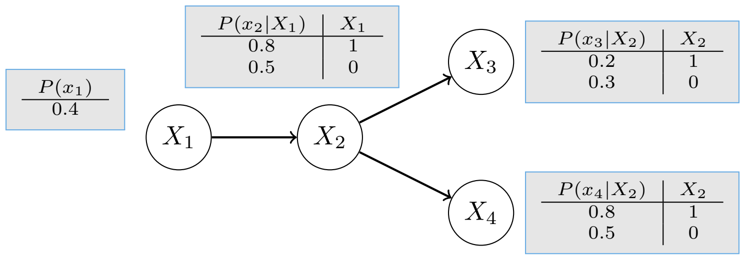
Question 1.

Consider the network below and compute

a) the marginal probability $P(X_3 = 0) = P(\neg x_3)$,

b) the conditional probability $P(X_2 = 1 \mid X_3 = 1) = P(x_2 \mid x_3)$.



Answer:

We can save much work by selecting a "smart" marginalization order. Some nodes may get eliminated immediately since they sum out to one (i.e., their value carries no information about the probability we are trying to compute). For the rest, we can cache intermediate results in so-called "factors".

a)

$$P(X_3 = 0) = \sum_{x_1=0}^{1} \sum_{x_2=0}^{1} \sum_{x_4=0}^{1} P(X_1 = x_1) \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_3 = 0 \mid X_2 = x_2) \cdot P(X_4 = x_4 \mid X_2 = x_2)$$

$$= \sum_{x_2=0}^{1} P(X_3 = 0 \mid X_2 = x_2) \overbrace{\sum_{x_1=0}^{1} P(X_1 = x_1)P(X_2 = x_2 \mid X_1 = x_1) \underbrace{\sum_{x_4=0}^{1} P(X_4 = x_4 \mid X_2 = x_2)}_{1}}^{G_{X_1}}$$

$$= \sum_{x_2=0}^{1} P(X_3 = 0 \mid X_2 = x_2) \cdot G_{X_1}(x_2) = G_{X_1,X_2}(X_3 = 0) = \mathbf{0.762}$$

$$G_{X_1}(X_2 = 0) = \sum_{x_1=0}^{1} P(X_1 = x_1)P(X_2 = 0 \mid X_1 = x_1) = 0.6 \cdot 0.5 + 0.4 \cdot 0.2 = 0.38$$

$$G_{X_1}(X_2 = 1) = \sum_{x_1=0}^{1} P(X_1 = x_1)P(X_2 = 1 \mid X_1 = x_1) = 0.6 \cdot 0.5 + 0.4 \cdot 0.8 = 0.62$$

$$G_{X_1,X_2}(X_3 = 0) = \sum_{x_2=0}^{1} P(X_3 = 0 \mid X_2 = x_2) \cdot G_{X_1}(x_2) = 0.7 \cdot 0.38 + 0.8 \cdot 0.62 = 0.762$$

$$G_{X_1,X_2}(X_3 = 1) = \sum_{x_2=0}^{1} P(X_3 = 1 \mid X_2 = x_2) \cdot G_{X_1}(x_2) = 0.3 \cdot 0.38 + 0.2 \cdot 0.62 = 0.238$$

After removing the node X_4, we only needed 6 multiplications and 3 additions (8 multiplications and 4 additions if the entire factor G_{X_1,X_2} was computed). As opposed to that, when computing the same value naively, we needed 24 multiplications and 7 additions.

Note, that the factors do not have to sum up to one! For one, we have the factor $G_{X_4}(x_2)$ which sums up to two. Also, if we eliminated X_2 before X_1, we would obtain

$$G_{X_2}(X_1 = 0) = \sum_{x_2=0}^{1} P(X_2 = x_2 \mid X_1 = 0) \cdot P(X_3 = 0 \mid X_2 = x_2) = 0.5 \cdot 0.7 + 0.5 \cdot 0.8 = 0.75$$

$$G_{X_2}(X_1 = 1) = \sum_{x_2=0}^{1} P(X_2 = x_2 \mid X_1 = 1) \cdot P(X_3 = 0 \mid X_2 = x_2) = 0.2 \cdot 0.7 + 0.8 \cdot 0.8 = 0.78$$

b)

$$\begin{aligned}
P(X_2 = 1, X_3 = 1) &= \sum_{x_1=0}^{1} \sum_{x_4=0}^{1} P(X_1 = x_1) \cdot P(X_2 = 1 \mid X_1 = x_1) \cdot P(X_3 = 1 \mid X_2 = 1) \cdot P(X_4 = x_4 \mid X_2 = 1) \\
&= P(X_3 = 1 \mid X_2 = 1) \overbrace{\sum_{x_1=0}^{1} P(X_1 = x_1) P(X_2 = 1 \mid X_1 = x_1) \underbrace{\sum_{x_4=0}^{1} P(X_4 = x_4 \mid X_2 = 1)}_{1}}^{G_{X_1}(X_2=1)} \\
&= 0.2 \cdot 0.62 = 0.124 \\
P(X_2 = 1 \mid X_3 = 1) &= \frac{0.124}{0.238} \approx \mathbf{0.5210}
\end{aligned}$$

Question 2.

Consider the same network as above.

Assume that the sequence $\{r_i\}_{i=1}^{20}$ was generated at random uniformly from the interval $(0; 1)$. Use the sequence to

a) approximate $P(x_3)$ using a suitable sampling method,

b) approximate $P(x_1 \mid x_2, \neg x_3)$ using *rejection sampling* and *likelihood weighting*.

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}
0.2551	0.5060	0.6991	0.8909	0.9593	0.5472	0.1386	0.1493	0.1975	0.8407
r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}
0.0827	0.9060	0.7612	0.1423	0.5888	0.6330	0.5030	0.8003	0.0155	0.6917

Answer:

First, let us notice that when estimating either of the probabilities below, we can marginalize over X_4. Thus, we will not be sampling values for X_4.

Once we obtain samples from the distribution, we estimate the probability using the *Monte Carlo method*.

For all sampling methods below, we require a topological ordering of the nodes, i.e., the random variables. We will use

$$X_1 < X_2 < X_3 < X_4.$$

a) For this task, we can use the *forward sampling* algorithm.

	X_1	X_2	X_3
s^1	$P(x_1) > r_1 \rightarrow 1$	$P(x_2 \mid x_1) > r_2 \rightarrow 1$	$P(x_3 \mid x_2) < r_3 \rightarrow 0$
s^2	$P(x_1) < r_4 \rightarrow 0$	$P(x_2 \mid \neg x_1) < r_5 \rightarrow 0$	$P(x_3 \mid \neg x_2) < r_3 \rightarrow 0$
s^3	1	1	1
s^4	0	1	0
s^5	0	1	0
s^6	0	0	0

Hence, we estimate $P(X_3 = 1) = \frac{|\{s^i \,:\, s^i_{X_3} = 1\}|}{6} = \frac{1}{6}$.

b) (1) Let us use *rejection sampling* first. Rejection sampling iteratively employs the forward sampling algorithm, rejecting all samples inconsistent with the evidence.

	X_1	X_2	X_3	
s^1	1	1	0	✓
s^2	0	0	?	**reject**
s^3	0	1	1	**reject**
s^4	1	0	?	**reject**
s^5	1	0	?	**reject**
s^6	0	1	0	✓
s^7	0	0	?	**reject**
s^8	0	1	0	✓

Hence, we estimate $P(X_1 = 1 \mid X_2 = 1, X_3 = 0) = \frac{1}{3}$.

(2) Now, let us try *likelihood weighting*, which does not reject any samples. Instead, it computes a *weight* (likelihood) for each sample.

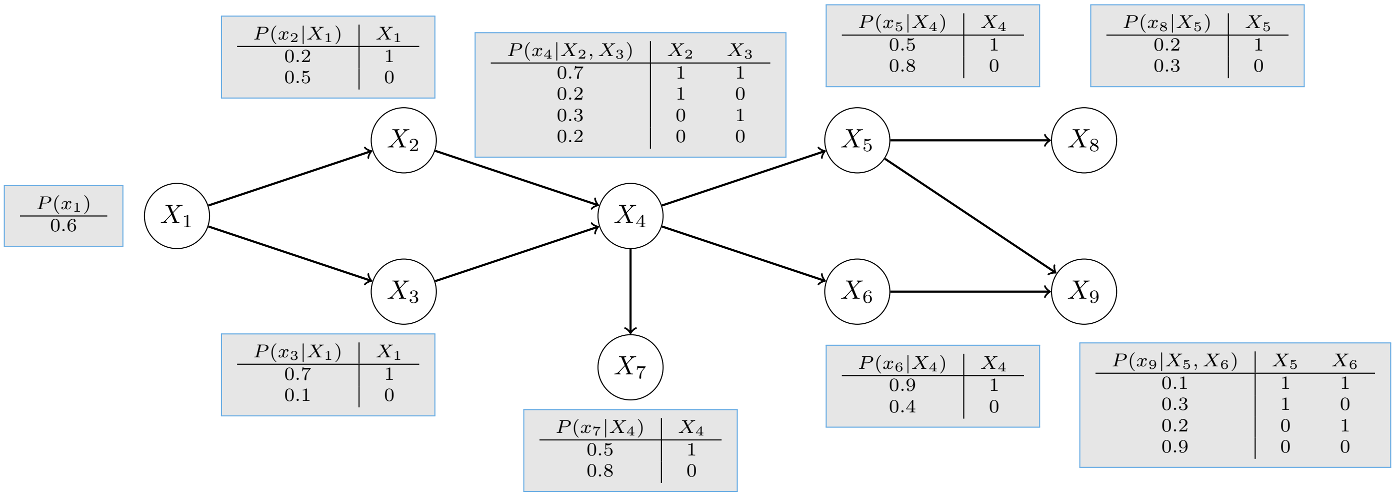
In likelihood weighting, evidence remains fixed. Since we also summed out X_4, we only need to determine the value of X_1. Hence, for each random number r_i, we obtain a sample s^i.

	X_1	X_2	X_3	w
s^1	$P(x_1) > r_1 \rightarrow 1$	1	0	$P(x_2 \mid x_1)P(\neg x_3 \mid x_2) = 0.8 \cdot 0.8 = 0.64$
s^2	$P(x_1) < r_2 \rightarrow 0$	1	0	$P(x_2 \mid \neg x_1)P(\neg x_3 \mid x_2) = 0.5 \cdot 0.8 = 0.4$
s^3	0	1	0	0.4
s^4	0	1	0	0.4
s^5	0	1	0	0.4
s^6	0	1	0	0.4
s^7	1	1	0	0.64
s^8	1	1	0	0.64
s^9	1	1	0	0.64
s^{10}	0	1	0	0.4
s^{11}	1	1	0	0.64
s^{12}	0	1	0	0.4
s^{13}	0	1	0	0.4
s^{14}	1	1	0	0.64
s^{15}	0	1	0	0.4
s^{16}	0	1	0	0.4
s^{17}	0	1	0	0.4
s^{18}	0	1	0	0.4
s^{19}	1	1	0	0.64
s^{20}	0	1	0	0.4

We estimate $P(X_1 = 1 \mid X_2 = 1, X_3 = 0) = \frac{7 \cdot 0.64}{7 \cdot 0.64 + 13 \cdot 0.4} = \frac{4.48}{9.68} \approx 0.4628$.

Question 3.

Consider the Bayes net below:



a) Compute the marginal probability distribution $P(X_7)$ using *variable elimination* with the elimination order

$$X_1, X_8, X_9, X_5, X_6, X_2, X_3, X_4.$$

b) Compute $P(x_8 \mid \neg x_4)$ however you see fit.

Answer:

a) When eliminating X_i, we first collect all factors containing X_i and compute their product $\psi^{(i)}$. Then, we compute $G_{X_i} = \sum_{x_i} \psi^{(i)}$, remove all the collected factors and add G_{X_i} instead.

$$P(X_7 = x_7) = \sum_{x_1=0}^{1} \dots \sum_{x_6=0}^{1} \sum_{x_8=0}^{1} \sum_{x_9=0}^{1} P(X_1 = x_1) \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_3 = x_3 \mid X_1 = x_1) \cdot$$
$$P(X_4 = x_4 \mid X_2 = x_2, X_3 = x_3) \cdot P(X_5 = x_5 \mid X_4 = x_4) \cdot P(X_6 = x_6 \mid X_4 = x_4) \cdot$$
$$P(X_7 = x_7 \mid X_4 = x_4) \cdot P(X_8 = x_8 \mid X_5 = x_5) \cdot P(X_9 = x_9 \mid X_5 = x_5, X_6 = x_6)$$

(1) Eliminate X_1:

$$G_{X_1}(x_2, x_3) = \sum_{x_1=0}^{1} P(X_1 = x_1) \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_3 = x_3 \mid X_1 = x_1)$$

X_2	X_3	G_{X_1}
0	0	$0.6 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.5 \cdot 0.9 = \mathbf{0.324}$
0	1	$0.4 \cdot 0.5 \cdot 0.1 + 0.6 \cdot 0.8 \cdot 0.7 = \mathbf{0.356}$
1	0	$0.4 \cdot 0.5 \cdot 0.9 + 0.6 \cdot 0.2 \cdot 0.3 = \mathbf{0.216}$
1	1	$0.4 \cdot 0.5 \cdot 0.1 + 0.6 \cdot 0.2 \cdot 0.7 = \mathbf{0.104}$

(2) Eliminate X_8:

$$G_{X_8}(x_5) = \sum_{x_8=0}^{1} P(X_8 = x_8 \mid X_5 = x_5) = 1$$

(3) Eliminate X_9:

$$G_{X_9}(x_5, x_6) = \sum_{x_9=0}^{1} P(X_9 = x_9 \mid X_5 = x_5, X_6 = x_6) = 1$$

Now, the overall sum is given as

$$P(X_7 = x_7) = \sum_{x_2=0}^{1}\sum_{x_3=0}^{1}\sum_{x_4=0}^{1}\sum_{x_5=0}^{1}\sum_{x_6=0}^{1} P(X_4 = x_4 \mid X_2 = x_2, X_3 = x_3) \cdot P(X_5 = x_5 \mid X_4 = x_4) \cdot$$
$$P(X_6 = x_6 \mid X_4 = x_4) \cdot P(X_7 = x_7 \mid X_4 = x_4) \cdot \underbrace{G_{X_9}(x_5, x_6)}_{1} \cdot \underbrace{G_{X_8}(x_5)}_{1} \cdot G_{X_1}(x_2, x_3)$$

(4) Eliminate X_5:

$$G_{X_5}(x_4) = G_{X_9,X_5}(x_4, x_6) = \sum_{x_5=0}^{1} P(X_5 = x_5 \mid X_4 = x_4) \cdot G_{X_9}(x_4, x_6) = 1$$

(5) Eliminate X_6:

$$G_{X_6}(x_4) = G_{X_9,X_5,X_6}(x_4) = \sum_{x_6=0}^{1} P(X_6 = x_6 \mid X_4 = x_4) \cdot G_{X_5}(x_4) = 1$$

(6) Eliminate X_2:

$$G_{X_1,X_2}(x_3, x_4) = \sum_{x_2=0}^{1} P(X_4 = x_4 \mid X_2 = x_2, X_3 = x_3) \cdot G_{X_1}(x_2, x_3)$$

X_3	X_4	G_{X_1,X_2}
0	0	$0.8 \cdot 0.324 + 0.8 \cdot 0.216 = \mathbf{0.432}$
0	1	$0.2 \cdot 0.324 + 0.2 \cdot 0.216 = \mathbf{0.108}$
1	0	$0.7 \cdot 0.356 + 0.3 \cdot 0.104 = \mathbf{0.2804}$
1	1	$0.3 \cdot 0.356 + 0.7 \cdot 0.104 = \mathbf{0.1796}$

Now, the overall sum is given as

$$P(X_7 = x_7) = \sum_{x_4=0}^{1} P(X_7 = x_7 \mid X_4 = x_4) \sum_{x_3=0}^{1} G_{X_1,X_2}(x_3, x_4)$$

(7) Eliminate X_3:

$$G_{X_1,X_2,X_3}(x_4) = \sum_{x_3=0}^{1} G_{X_1,X_2}(x_3, x_4)$$

X_4	G_{X_1,X_2,X_3}
0	$0.432 + 0.2804 = \mathbf{0.7124}$
1	$0.108 + 0.1796 = \mathbf{0.2876}$

(8) Eliminate X_4:

$$H(x_7) = \sum_{x_4=0}^{1} P(X_7 = x_7 \mid X_4 = x_4) \cdot G_{X_1,X_2,X_3}(x_4)$$

X_7	H
0	$0.2 \cdot 0.7124 + 0.5 \cdot 0.2876 = \mathbf{0.28628}$
1	$0.8 \cdot 0.7124 + 0.5 \cdot 0.2876 = \mathbf{0.71372}$

Finally, we have $P(X_7 = x_7) = H(x_7)$.

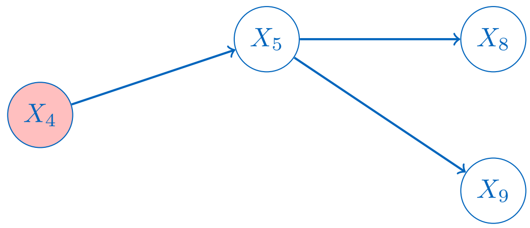
b) First, let us simplify the network using the conditional independence. It holds

$$X_8 \perp\!\!\!\perp \{X_1, X_2, X_3, X_6, X_7\} \mid X_4.$$

All paths between X_1, X_2, X_3 and X_8 are blocked by X_4 (causal chain).
All paths between X_7 and X_8 are blocked by X_4 (common cause).
Path $\langle X_8, X_5, X_4, X_6 \rangle$ is blocked by observed X_4 (common cause).
Path $\langle X_8, X_5, X_9, X_6 \rangle$ is blocked by unobserved X_9 (common effect).
Hence, we can simplify the network to

$P(x_5 \mid X_4)$	X_4
0.5	1
0.8	0

$P(x_8 \mid X_5)$	X_5
0.2	1
0.3	0



$P(x_9 \mid X_5, X_6)$	X_5	X_6
0.1	1	1
0.3	1	0
0.2	0	1
0.9	0	0

$$\begin{aligned}
P(X_8 = 1 \mid X_4 = 0) &= \sum_{x_5=0}^{1} \sum_{x_9=0}^{1} P(X_5 = x_5 \mid X_4 = 0) \cdot P(X_8 = 1 \mid X_5 = x_5) \cdot P(X_9 = x_9 \mid X_5 = x_5) \\
&= \sum_{x_5=0}^{1} P(X_5 = x_5 \mid X_4 = 0) \cdot P(X_8 = 1 \mid X_5 = x_5) \underbrace{\sum_{x_9=0}^{1} P(X_9 = x_9 \mid X_5 = x_5)}_{1} \\
&= 0.2 \cdot 0.3 + 0.8 \cdot 0.2 = \mathbf{0.22}
\end{aligned}$$