1. Bayes’ Rule

(a) We combine these using Bayes’ rule to find the joint probabilities and then marginalize over boxes to obtain the marginal probability of selecting an apple:

\[
P(\text{apple}) = P(\text{apple}, \text{red}) + P(\text{apple}, \text{green}) + P(\text{apple}, \text{blue})
= P(\text{apple}|\text{red})P(\text{red}) + P(\text{apple}|\text{green})P(\text{green})
+ P(\text{apple}|\text{blue})P(\text{blue})
= 0.3 \times 0.2 + 0.3 \times 0.6 + 0.5 \times 0.2 = 0.34
\]

(b) We want to compute \(P(\text{green}|\text{orange})\). We know from Bayes rule that

\[
P(\text{green}|\text{orange}) = \frac{P(\text{green}, \text{orange})}{P(\text{orange})}
\]

to use this formula we need \(P(\text{green}, \text{orange})\) and \(P(\text{orange})\). Now,

\[
P(\text{orange}) = P(\text{orange}|\text{red})P(\text{red}) + P(\text{orange}|\text{green})P(\text{green})
+ P(\text{orange}|\text{blue})P(\text{blue})
= 0.4 \times 0.2 + 0.3 \times 0.6 + 0.5 \times 0.2 = 0.36.
\]

Note that in the calculations above we found

\[
P(\text{orange}, \text{green}) = P(\text{orange}|\text{green})P(\text{green}) = 0.3 \times 0.6 = 0.18
\]

Putting all this together, we have

\[
P(\text{green}|\text{orange}) = \frac{P(\text{green}, \text{orange})}{P(\text{orange})} = \frac{0.18}{0.36} = 0.5
\]

2. First-order Markov Chain

Given the joint distribution \(p(X, Y, Z) = p(X)p(Y|X)p(Z|Y)\), we have that:

(a) \(p(X, Z|Y) = \frac{p(X)p(Y|X)p(Z|Y)}{p(Y)} = \frac{p(X, Y)}{p(Y)}p(Z|Y) = p(X|Y)p(Z|Y)\).

We have shown that the the conditional distribution \(p(X, Z|Y)\) factorizes as the product of conditionals \(p(X|Y)p(Z|Y)\). Hence \(X \perp Z \mid Y\).
(b) In general, for \( p(X, Y, Z) \) we require \( 2^3 - 1 \) parameters for the different settings of \( X, Y \) and \( Z \) as we can obtain the remaining probability by normalization. However, for the given distribution we exploit the conditional independencies as follows: For \( p(X) \) we require 1 parameter; for \( p(Y|X) \) we require 2 parameters for the different settings of \( X \); and for \( p(Z|Y) \) we also require 2 parameters. In total we require 5 parameters.

3. Inference in a Bayesian Network

(a) The joint distribution corresponding to the given Bayesian network is:

\[
\]

(b) To show that the distribution is correctly normalized we sum over all the variables and show that this sum is 1. This is true for any Bayesian network (as the individual distributions are correctly normalized).

\[
Z = \sum_{h, \ell, v, s, o, a, t, b} P(h)P(\ell)P(a|h, \ell)P(v|h, \ell)P(s|h, \ell)P(c|v)P(o|v, s)P(t|a)P(b|o, t)
\]

\[
= \sum_{h, \ell, v} P(h)P(\ell)P(v|h, \ell) \sum_{s} P(s|h, \ell) \sum_{c} P(c|v) \sum_{o} P(o|v, s)
\]

\[
= \sum_{h} P(h) \sum_{\ell} P(\ell) \sum_{v} P(v|h, \ell)
\]

\[
= 1.
\]

Note that the elimination order is important as one cannot marginalize a specific variable if there are other terms outside the sum containing this variable.

(c) Statistical independences:

i. All the paths between \( H \) and \( L \) have to go through either \( V \) or \( S \). In both cases we have head-to-head structures: \( \rightarrow V \leftarrow \) and \( \rightarrow S \leftarrow \), and neither \( V \) or \( S \) is observed. Therefore \( H \perp L \).

ii. All the paths between \( H \) and \( A \) have to go through \( B \), where we have a head-to-head structure: \( \rightarrow B \leftarrow \), where \( B \) is not observed. Hence \( H \perp A \).

iii. The path \( C \leftarrow V \leftarrow L \) is a head-to-tail structure but \( V \) is not observed. Hence this path is not blocked and \( C \perp L \).

iv. Observing \( B \) does not block the paths between \( C \) and \( L \) as \( B \) is as child of \( O \), which has a head-to-head structure. Therefore \( C \perp L | B \). What variable should we observe so that \( L \) and \( C \) are conditionally independent?

(d) To prove statistical independence we show that the corresponding distributions factorize:
(e) In order to calculate $P(L|C)$, we first apply Bayes rule:

$$P(L|C) = \frac{P(L, C)}{P(C)} \tag{1}$$

where the computation of $P(L, C)$ can be done by summing over all the states of those variables not included in the marginal required:

$$P(L, C) = \sum_{H, V, S, O, A, T, B} P(H, L, V, S, O, A, T, B)$$

$$= \sum_{H, V, S, O} P(H)P(L)P(V|H, L)P(S|H, L)P(C|V)P(O|V, S) \left\{ \sum_A P(A) \left\{ \sum_T P(T|A) \left\{ \sum_B P(B|O, T) \right\} \right\} \right\}$$

$$= \sum_{H, V} P(H)P(L)P(V|H, L)P(C|V) \left\{ \sum_S P(S|H, L) \left\{ \sum_O P(O|V, S) \right\} \right\}$$

$$= P(L) \sum_{H, V} P(H)P(V|H, L)P(C|V),$$

where it is clear that $\sum_X P(X|Y) = 1$. Note, however, that the elimination order is important as (for example) we could not have eliminated $A$ without previously eliminating $T$. Simplifying we obtain:

$$P(L, C) = P(L) \sum_V P(C|V) \sum_H P(V|H, L)P(H).$$

Defining $m_H(V, L) = \sum_H P(V|H, L)P(H)$ and realizing that $m_H(V, L) = P(V|L)$ we have:

$$m_H(V = low, L = true) = P(V = low|H = true, L = true)P(H = true)$$

$$+ P(V = low|H = false, L = true)P(H = false)$$

$$= (0.95)(0.2) + (0.01)(0.8) = 0.198$$

$$m_H(V = high, L = true) = 1 - m_H(V = low, L = true) = 0.802.$$
Similarly:

\[ m_H(V = \text{low}, L = \text{false}) = P(V = \text{low}|H = \text{true}, L = \text{false})P(H = \text{true}) + P(V = \text{low}|H = \text{false}, L = \text{false})P(H = \text{false}) \]

\[ = (0.98)(0.2) + (0.05)(0.8) = 0.236 \]

\[ m_H(V = \text{high}, L = \text{false}) = 1 - m_H(V = \text{low}, L = \text{false}) = 0.764. \]

Hence,

\[ P(L = \text{true}, C = \text{high}) = P(L = \text{true}) \sum_V P(C = \text{high}|V)m_H(V, L = \text{true}) \]

\[ = P(L = \text{true})[P(C = \text{high}|V = \text{low})m_H(V = \text{low}, L = \text{true}) + P(C = \text{high}|V = \text{high})m_H(V = \text{high}, L = \text{true})] \]

\[ = 0.05[(0.01)(0.198) + (0.7)(0.802)] \]

\[ = 0.05(0.56338) = 0.028169. \]

Similarly,

\[ P(L = \text{false}, C = \text{high}) = P(L = \text{false}) \sum_V P(C = \text{high}|V)m_H(V, L = \text{false}) \]

\[ = P(L = \text{false})[P(C = \text{high}|V = \text{low})m_H(V = \text{low}, L = \text{false}) + P(C = \text{high}|V = \text{high})m_H(V = \text{high}, L = \text{false})] \]

\[ = 0.95[(0.01)(0.236) + (0.7)(0.764)] \]

\[ = 0.95(0.53716) = 0.510302. \]

It is clear that \( P(C = \text{high}) = \sum_L P(L, C = \text{high}) = 0.028169 + 0.510302 = 0.538471. \) Therefore, from equation 1 we have that:

\[ P(L = \text{true}|C = \text{high}) = \frac{P(L = \text{true}, C = \text{high})}{P(C = \text{high})} = \frac{0.028169}{0.538471} \approx 0.05231 \]

\[ P(L = \text{false}|C = \text{high}) = 0.94769. \]