Tutorial 3 - Bayesian Networks

COMP9418 – Advanced Topics in Statistical Machine Learning

Lecturer: Gustavo Batista

Lecture: Bayesian Networks

Topic: Questions from lecture topics

Last revision: Sunday 27th September, 2020 at 22:04

Question 1

Consider the random variables X, Y, Z which have the following joint distribution:

$$P(X, Y, Z) = P(X)P(Y|X)P(Z|Y)$$

a. Show that X and Z are conditionally independent given Y.

b. If X, Y and Z are binary variables, how many parameters are needed to specify a distribution of this form?

Answer

a. According to the chain rule, $P(X, Y, Z)$ can be factorised in the following way

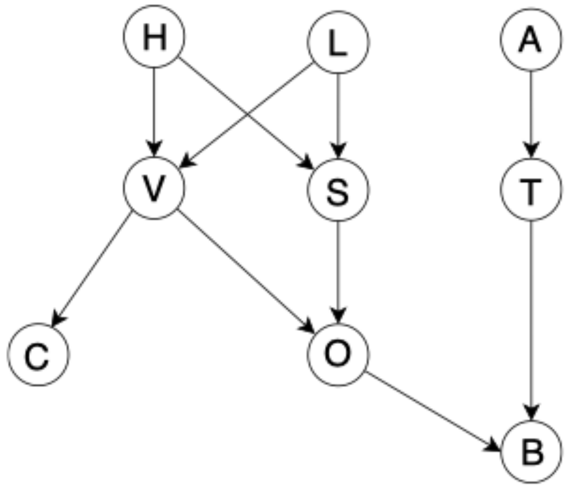
$$P(X, Y, Z) = P(X)P(Y|X)P(Z|X, Y)$$

Remember that this factorisation do not assume any independencies between variables. Notice that, to make $P(Z|X, Y)$ (that occurs in the chain rule) equal to $P(Z|Y)$ (that occurs in the exercise), we need assume that $X \perp\!\!\!\perp Z|Y$.

b. In general, for $P(X, Y, Z)$ we require $2^3 - 1 = 7$ 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.

Question 2

The Bayesian network shown below is a greatly simplified version of a network used for medical diagnosis in an intensive care unit. The diagnostic variables are the hypovolemia (H), the left ventricular failure (L) and the anaphylaxis (A). The intermediate variables are the left ventricular endiastolic volume (V), the stroke volume (S) and the total peripheral resistance (T). The measurement variables are the central venous pressure (C), the cardiac output (O) and the blood pressure (B). H, L and A take on the values $\{true, false\}$; V, S and T take on the values $\{low, high\}$; and C, O and B take on the values $\{low, medium, high\}$.



The Bayesian network is fully specified by its CPTs. We have:

$P(H = true) = 0.2$	
$P(L = true) = 0.05$	$P(A = true) = 0.01$
$P(V = low \vert H = false, L = false) = 0.05$	$P(S = low \vert H = false, L = false) = 0.05$
$P(V = low \vert H = false, L = true) = 0.01$	$P(S = low \vert H = false, L = true) = 0.95$
$P(V = low \vert H = true, L = false) = 0.98$	$P(S = low \vert H = true, L = false) = 0.5$
$P(V = low \vert H = true, L = true) = 0.95$	$P(S = low \vert H = true, L = true) = 0.98$
$P(T = low \vert A = false) = 0.3$	$P(T = low \vert A = true) = 0.98$
$P(C = low \vert V = low) = 0.95$	$P(C = medium \vert V = low) = 0.04$
$P(C = low \vert V = high) = 0.01$	$P(C = medium \vert V = high) = 0.29$
$P(O = low \vert V = low, S = low) = 0.98$	$P(O = medium \vert V = low, S = low) = 0.01$
$P(O = low \vert V = low, S = high) = 0.3$	$P(O = medium \vert V = low, S = high) = 0.69$
$P(O = low \vert V = high, S = low) = 0.8$	$P(O = medium \vert V = high, S = low) = 0.19$
$P(O = low \vert V = high, S = high) = 0.01$	$P(O = medium \vert V = high, S = high) = 0.01$
$P(B = low \vert O = low, T = low) = 0.98$	$P(B = medium \vert O = low, T = low) = 0.01$
$P(B = low \vert O = low, T = high) = 0.3$	$P(B = medium \vert O = low, T = high) = 0.6$
$P(B = low \vert O = medium, T = low) = 0.98$	$P(B = medium \vert O = medium, T = low) = 0.01$
$P(B = low \vert O = medium, T = high) = 0.05$	$P(B = medium \vert O = medium, T = high) = 0.4$
$P(B = low \vert O = high, T = low) = 0.9$	$P(B = medium \vert O = high, T = low) = 0.09$
$P(B = low \vert O = high, T = high) = 0.01$	$P(B = medium \vert O = high, T = high) = 0.09$

a. Write down the factorised joint distribution defined by the Bayesian network.

b. Show that the above factorised joint distribution is correctly normalized, using the rules of probability.

c. Let $X \perp\!\!\!\perp Y$ denote that X and Y are marginally independent and $X \perp\!\!\!\perp Y|Z$ denote that X and Y are conditionally independent given Z. Using the concept of d-separation, show or refute the following independence statements:

   i. $H \perp\!\!\!\perp L$

   ii. $H \perp\!\!\!\perp A$

   iii. $C \perp\!\!\!\perp L$

   iv. $V \perp\!\!\!\perp A|B$

d. For the cases where independence holds in item c, prove these independences using the rules of probability.

Answer

a. $P(H, L, V, S, C, O, A, T, B) = P(H)P(L)P(A)P(V|H, L)P(S|H, L)P(C|V)P(O|V, S)P(T|A)P(B|O, T)$.

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).

$$Z = \sum_{h,l,v,s,c,o,a,t,b} P(h)P(l)P(a)P(v|h, l)P(s|h, l)P(c|v)P(o|v, s)P(t|a)P(b|o, t)$$

$$Z = \sum_{h,l,v,s,c,o} P(h)P(l)P(v|h, l)P(s|h, l)P(c|v)P(o|v, s) \underbrace{\sum_a P(a) \underbrace{\sum_t P(t|a) \underbrace{\sum_b P(b|o, t)}_{1}}_{1}}_{1}$$

$$Z = \sum_{h,l,v} P(h)P(l)P(v|h, l) \underbrace{\sum_s P(s|h, l) \underbrace{\sum_c P(c|v) \underbrace{\sum_o P(o|v, s)}_{1}}_{1}}_{1}$$

$$Z = \sum_h P(h) \underbrace{\sum_l P(l) \underbrace{\sum_v P(v|h, l)}_{1}}_{1}$$

$$Z = 1$$

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. i. All the paths between H and L have to go through either V or S. In both cases we have a convergent structure, and neither V or S is observed. Therefore $H \perp\!\!\!\perp L$.

ii. All the paths between H and A have to go through B, where we have a convergent structure, where B is not observed. Hence, $H \perp\!\!\!\perp A$.

iii. The path $C - V - L$ is a causal chain, and V is not observed. Hence, this path is not blocked and C is not d-separated of L. If C is independent of L will depend on the parametrisation of the Bayesian network.

iv. The paths between V and A have to go through B that is a convergent node. As B is observed, these paths are not blocked. Hence, V is not d-separated of A given B. Again, the independence between V and A given B will depend on the network parametrisation.

d. To prove statistical independence we show that the corresponding distributions factorize:

i. $$P(H, L) = \sum_{v,s,c,o,a,t,b} P(H)P(L)P(a)P(v|H, L)P(s|H, L)P(c|v)P(o|v, s)P(t|a)P(b|o, t)$$

$$= P(H)P(L) \sum_{a,v,s,c} P(a)P(v|H, L)P(s|H, L)P(c|v) \underbrace{P(o|v, s) \underbrace{P(t|a) \underbrace{P(b|o, t)}_{1}}_{1}}_{1}$$

$$= P(H)P(L) \sum_v P(v|H, L) \sum_s P(s|H, L) \sum_c P(c|v) \sum_a P(a)$$

$$= P(H)P(L)$$

ii. $P(H,A) = \sum_{l,v,s,c,o,t,b} P(H)P(l)P(A)P(v|H,l)P(s|H,l)P(c|v)P(o|v,s)P(t|A)P(b|o,t)$

$$= P(H)P(A) \sum_{l,v,s} P(H)P(l)P(A)P(v|H,l)P(s|H,l) \underbrace{\sum_c P(c|v) \sum_o P(o|v,s) \sum_t P(t|A) \sum_b P(b|o,t)}_{1}$$

$$= P(H)P(A) \underbrace{\sum_l P(l) \sum_v P(v|H,l) \sum_s P(s|H,l)}_{1}$$

$$= P(H)P(A)$$

Question 3

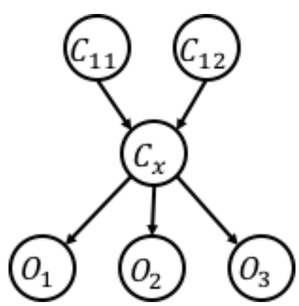
Jack has three coins C_1, C_2 and C_3 with p_1, p_2 and p_3 as their corresponding probabilities of landing heads. Jack flips coin C_1 twice and then decides, based on the outcome, whether to flip coin C_2 or C_3 next. In particular, if the two C_1 flips come out the same, Jack flips coin C_2 three times next. However, if the C_1 flips come out different, he flips coin C_3 three times next. Given the outcome of Jack's last three flips, we want to know whether his first two flips came out the same.

a. Show a Bayesian network structure (graph) for this problem.
b. Show the network conditional probability tables (CPTs) for all variables. If you have parameters that are shared among variables, define the CPT once and indicate which variables use that CPT.
c. Provide a query that solves this problem.

Answer

a. This is a modelling exercise. We can create two variables C_{11} and C_{12} to register the outcome of the two flips of coin C_1. An intermeddiate variable Cx does a XOR operation on these two variables to determine if their values are equal or not. Finally, the variables O_1, O_2 and O_3 have the final three flips. Those flips will depend on Cx as well as the outcome of the flips of the coins C_2 or C_3, represented by the variables $C_{21}, \ldots, C_{23}$ and $C_{31}, \ldots, C_{33}$, respectively.

   The following figure shows the Bayesian network.



b. Variables C_{11} and C_{12} have the following CPT

C_i	$P(C_i)$
heads	p_1
tails	$1-p_1$

   Variable Cx has the following CPT

C_{11}	C_{12}	Cx	$P(Cx\|C_{11}, C_{12})$
heads	heads	0	1
heads	heads	1	0
heads	tails	0	0
heads	tails	1	1
tails	heads	0	0
tails	heads	1	1
tails	tails	0	1
tails	tails	1	0

Variables O_1, O_2, and O_3 have the following CPT

Cx	O_i	$P(O_i\|Cx)$
0	heads	$p2$
0	tails	$1 - p_2$
1	heads	p_3
1	tails	$1 - p_3$

c. $P(Cx|O_1 = o_1, O_2 = o_2, O_3 = o_3)$