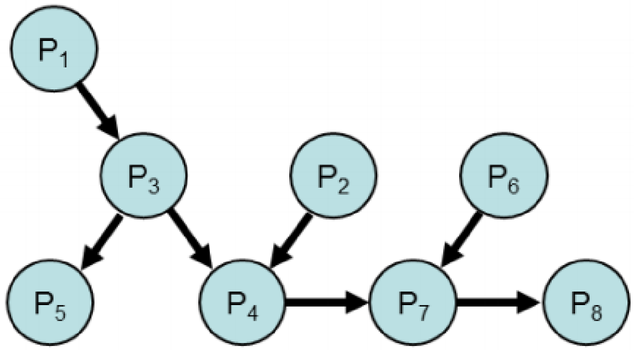
1 Bayesian Network Problems



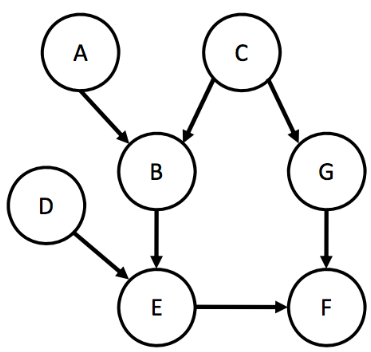
1. Given the Bayesian Network about, determine:

   (a) if P1 and P5 are independent of P6 given P8
   FALSE, the path through P3, P4 and P7 is not blocked; neither P1 and P6 or P5 and P6 are d-separated.

   (b) if P2 is independent of P6 given no information
   TRUE, the path is blocked by node P7.

   (c) if P1 is independent of P2 given P8
   FALSE, P1 and P2 converge on P4 and the path between them is un-blocked by P8.

   (d) if P1 is independent of P2 and P5 given P4
   FALSE, P4 unblocks the path of information from P2 and P3 is not blocked.

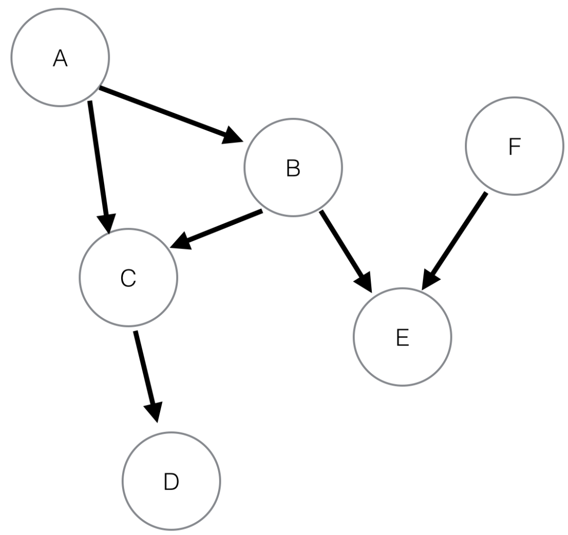


2. Given the Bayesian Network above, determine if:

   (a) A is independent of C given F.
   Answer: False. There is an unblocked (or not d-separated) path from A to B to E, and then thru F to G to C. Note that without information about F, the path from E to G is blocked.

   (b) G is independent of D given E.
   Answer: False. There is an unblocked (or not d-separated) path from D to E to B, and then to G.

(c) C is independent of D.
Answer: True. The fact that we have no information about E d-separates the path from D to B. No information about F d-separates E and G. So information about D is d-separated from paths to C both via F and E.



Given this network, calculate $P(B|D = false)$, $P(B|E = true)$ and $P(B|F = false)$.

Answer Part 1:

$P(B|D = false) =$
$\sum_A \sum_C \sum_E \sum_F P(A, B, C, D = false, E, F) / \sum_A \sum_B \sum_C \sum_E \sum_F P(A, B, C, D = false, E, F)$

$P(A, B, C, D = false, E, F) = P(A)P(B|A)P(C|A, B)P(D = false|C)P(E|F, B)P(F)$ so

$\sum_A \sum_C \sum_E \sum_F P(A, B, C, D = false, E, F)$
$= \sum_A \sum_C \sum_E \sum_F P(A)P(B|A)P(C|A, B)P(D = false|C)P(E|F, B)P(F)$
$= \sum_A P(A)P(B|A) \sum_C P(C|A, B)P(D = false|C) \sum_F P(F) \sum_E P(E|F, B)$

Note that $\sum_E P(E|F, B)$ results in a table of 1s and so multiplying this by $\sum_F P(F)$ will result in 1s. Both of E and F are not relevant to the query (and can be identified by the relevance algorithm presented in class). We therefore focus on eliminating C and A.

Eliminate C: $f_1(A, B) = \sum_C P(C|A, B)P(D = false|C)$

$f_1(a, b) = P(c|a, b)P(D = false|c) + P(-c|a, b)P(D = false| - c)$
$f_1(a, -b) = P(c|a, -b)P(D = false|c) + P(-c|a, -b)P(D = false| - c)$
$f_1(-a, b) = P(c| - a, b)P(D = false|c) + P(-c| - a, b)P(D = false| - c)$
$f_1(-a, -b) = P(c| - a, -b)P(D = false|c) + P(-c| - a, -b)P(D = false| - c)$

Eliminate A: $f_2(B) = \sum_A P(A)P(B|A)f_1(A, B)$

$f_2(b) = P(a)P(b|a)f_1(a, b) + P(-a)P(b| - a)f_1(-a, b)$

$f_2(-b) = P(a)P(-b|a)f_1(a, -b) + P(-a)P(-b|-a)f_1(-a, -b)$

Normalize: $P(B|D = false) = normalize(f_2(B)) = f_2(B)/(f_2(b) + f_2(-b))$

Answer Part 2:

$P(B|E = true) =$
$\sum_A \sum_C \sum_D \sum_F P(A, B, C, D, E = true, F)/ \sum_A \sum_B \sum_C \sum_D \sum_F P(A, B, C, D, E = true, F)$

$P(A, B, C, D, E = true, F) = P(A)P(B|A)P(C|A, B)P(D|C)P(E = true|F, B)P(F)$ so

$\sum_A \sum_C \sum_D \sum_F P(A, B, C, D, E = true, F)$
$= \sum_A \sum_F \sum_C \sum_D P(A)P(B|A)P(C|A, B)P(D|C)P(E = true|F, B)P(F)$
$= \sum_A P(A)P(B|A) \sum_F P(E = true|F, B)P(F) \sum_C P(C|A, B) \sum_D P(D|C)$

Note that $\sum_D P(D|C)$ results in a table of 1s and $\sum_C P(C|A, B)$ is also a table of 1s. Both of C and D are not relevant to the query (and can be identified by the relevance algorithm presented in class). We therefore focus on eliminating A and F.

Eliminate F: $f_1(B) = \sum_F P(E = true|F, B)P(F)$

$f_1(b) = P(e|f, b)P(f) + P(e|-f, b)P(-f)$
$f_1(-b) = P(e|f, -b)P(f) + P(e|-f, -b)P(-f)$

Eliminate A: $f_2(B) = \sum_A P(A)P(B|A)$

$f_2(b) = P(b|a)P(a) + P(b|-a)P(-a)$
$f_2(-b) = P(-b|a)P(a) + P(-b|-a)P(-a)$

Normalize: $P(B|E = true) = normalize(f_1(B)f_2(B)) = f_1(B)f_2(B)/(f_1(b)f_2(b) + f_1(-b)f_2(-b))$

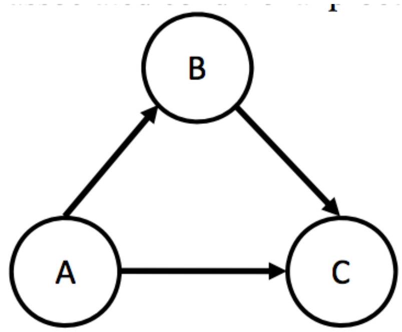
Answer Part 3:

$P(B|F = false) =$
$\sum_A \sum_C \sum_D \sum_E P(A, B, C, D, E, F = false)/ \sum_A \sum_B \sum_C \sum_D \sum_E P(A, B, C, D, E, F = false)$

Note however that F is d-separated from B by E. This means $P(B|F) = P(B)$ and can be computed directly from input CPTs.

$P(B) = \sum_A P(A, B) = \sum_A P(B|A) * P(A)$

3. Given the Bayesian Network Below:



$P(A = true) = 0.75$	$P(C = true \mid A = true, B = true) = 0.8$ $P(C = true \mid A = true, B = false) = 0.8$ $P(C = true \mid A = false, B = true) = 0.25$ $P(C = true \mid A = false, B = false) = 0.25$
$P(B = true \mid A = true) = 0.9$ $P(B = true \mid A = false) = 0.8$	

(a) Are any variables in the graph conditionally independent of each other? Why or why not?

Answer: Even tho there is a line between C and B, C and B are independent given A. This is because $P(C|A, B) = P(C|A)$ for all combinations of A,B and C. This should tell you that while lack of a line can indicate independences (or conditional independences) between variables, presence of a line does not necessarily indicate independences (or conditional independences).

(b) Calculate $P(A = true|B = true, C = true)$

Answer:

$P(A = true|B = true, C = true) = P(A = true, B = true, C = true)/P(B = true, C = true)$

$= P(A = true, B = true, C = true)/\sum_A P(A, B = true, C = true)$

$= P(C = true|A = true, B = true) * P(B = true|A = true) * P(A = true)/\sum_A P(C = true|A, B = true) * P(B = true|A) * P(A)$

$= (0.75 * 0.9 * 0.8)/((0.75 * 0.9 * 0.8) + (0.25 * 0.8 * 0.2)) = 0.92$

Note that this can be simplified if you substitute $P(C|A, B) = P(C|A)$ in the equations above.