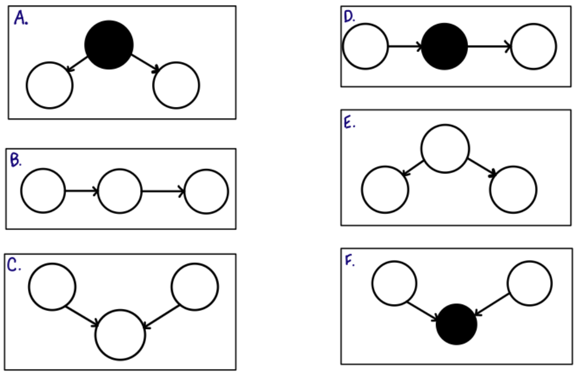
1 Warm Up

1. Bayes Ball Matching: Select the image that exemplifies each of the following Bayes Ball rules.



   (i) Active Causal Chain

   B

   (ii) Active Common Cause

   E

   (iii) Active Common Effect

   F

   (iv) Inactive Causal Chain

   D

   (v) Inactive Common Cause

   A

   (vi) Inactive Common Effect

   C

2. Are the following statements true or false? If it is false, briefly justify why not.

   (a) In a directed acyclic graph, each node can correspond to multiple random variables.

   False - each node will correspond to exactly one random variable.

   (b) Bayes Nets allow us to represent conditional independence relationships in a concise way.

   True

   (c) The number of entries in a conditional probability table in a Bayes Net is exponentially larger than in a joint probability table.

   False - The number of entries in a conditional probability table in a Bayes Net is exponentially smaller than in a joint probability table. A joint probability table will need to list out all combinations of possible domain values for the variables, which will be an exponential number of entries. In a Bayes Net, it only needs the different conditional probability relationships.

   (d) In a Bayes Net with no edges, the variables are all dependent on each other.

   False - the variables will all be independent of each other since there are no edges connecting them.

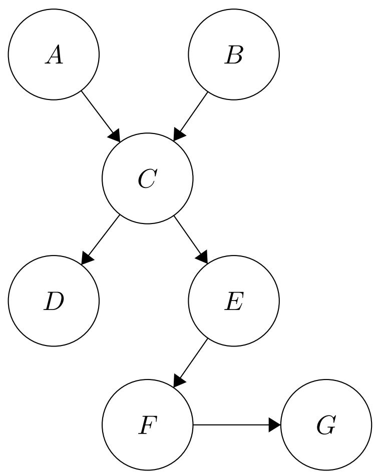
2 Bayes' Nets

Bayesian networks represent our beliefs about the conditional independence relationships between the variables in a system. The joint distribution is factored into conditional distributions, which are represented by the connections between nodes.

1. What would a Bayes net of variables which are all independent of each other look like?

   The BN would be a graph with no edges (no nodes are connected to any of the others).

2. Consider the following Bayes net.



   (a) Which of the following statements is true? (Hint: d-separation!)

   Recall that to determine whether $X \perp\!\!\!\perp Y \mid Z_1, Z_2, \ldots Z_n$ with d-separation, we must check the activeness of all paths between X and Y. A path is active if and only if each triple along the path is active:

   - Causal chain $A \rightarrow B \rightarrow C$ (or $A \leftarrow B \leftarrow C$) where B is unobserved
   - Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
   - Common effect $A \rightarrow B \leftarrow C$ where B or one of its descendants is observed

   If all paths are inactive, then $X \perp\!\!\!\perp Y \mid Z_1, Z_2, \ldots Z_n$.

   - $P(B \mid A) = P(B)$

     True: This statement is true if $A \perp\!\!\!\perp B$. The only path from A to B is ACB. For the triple ACB, the common effect C is unobserved (and none of its descendants are observed), so this triple is inactive. Therefore, path ACB is inactive.

   - $A \perp\!\!\!\perp B \mid D, F$

     False: Again, the only path is ACB. For triple ACB, descendants D and F of common effect C is observed, so this triple is active. Therefore, path ACB is active.
     Another way you could think of the "descendants" property is when considering paths between two nodes, we can "bounce" off nodes when valid. In this case, the path we can take is $ACDCB$, or $ACEFECB$.

   - $A \perp\!\!\!\perp B \mid C$

False: For triple ACB, common effect C is observed, so this triple is active. Therefore, path ACB is active.

- $D \perp\!\!\!\perp E$

  False: The only path from D to E is DCE. For triple DCE, the common cause C is unobserved, so this triple is active. Since all triples are active, path DCE is active.

- $P(D \mid C, E) = P(D \mid C)$

  True: This statement is true if $D \perp\!\!\!\perp E \mid C$. Again, the only path from D to E is DCE. For triple DCE, the common cause C is observed, so this triple is inactive. Therefore, path DCE is inactive.

- $D \perp\!\!\!\perp E \mid A, B$

  False: For triple DCE, the common cause C is unobserved, so this triple is active. Therefore, path DCE is active.

- $P(B \mid E) = P(B \mid G, E)$

  True: This statement is true if $B \perp\!\!\!\perp G \mid E$. The only path from B to G is $BCEFG$. The triple CEF is inactive, since it is a causal chain in which E is observed. Therefore, path $BCEFG$ is inactive.

(b) Suppose all variables have domains of size 2. How many entries are in the full joint probability table $P(A, B, C, D, E, F, G)$?

$2^7 = 128$ (assuming none have been implicitly omitted)

(c) How many entries are in the conditional probability table at each of the individual variables?

Variable	A	B	C	D	E	F	G
# Entries							

Variable	A	B	C	D	E	F	G
# Rows	2	2	8	4	4	4	4

(d) How many CPT entries are needed to represent this Bayes net in total?

Summing the number of entries in the table above, we get 28.

(e) How do we represent the full joint probability using the conditional probability tables of each variable?

Joint probability = product of all conditional probability distributions.
$P(A, B, C, D, E, F, G) = P(A)P(B)P(C \mid A, B)P(D \mid C)P(E \mid C)P(F \mid E)P(G \mid F)$

Now let's look at some common types of queries we might be interested in answering using a Bayes' Net.

Let C = some variable we care about, D = some variable we don't care about (and we also don't know the value of), E = some variable we've observed to have the value e (i.e., current knowledge/evidence). (Suppose for now C, D, E are the only nodes in our BN.)

3. Express each of the following generic queries using probability notation, then write how we would compute it using the full joint distribution, $P(C, D, E)$. The first expression has been written for you as an example to get started.

(a) What's the probability of the outcome $C = c$ given what I know?

$$P(c \mid e) = \frac{P(c, e)}{P(e)} =$$

$$P(c \mid e) = \frac{P(c,e)}{P(e)} = \frac{\sum_d P(c,d,e)}{\sum_{c'} \sum_d P(c',d,e)}$$

(b) What are the probabilities of all possible outcomes of C given what I know?

$$P(C \mid e) = \frac{P(C,e)}{P(e)} = \frac{\sum_d P(C,d,e)}{\sum_c \sum_d P(c,d,e)}$$

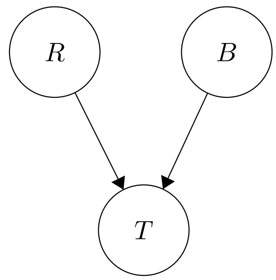
(c) Which outcome of event C is the least likely outcome given what I know?

$$\operatorname*{argmin}_c P(c \mid e) = \operatorname*{argmin}_c \frac{P(c,e)}{P(e)} = \operatorname*{argmin}_c \frac{\sum_d P(c,d,e)}{\sum_{c'} \sum_d P(c',d,e)}$$

Note that for parts (a) and (c) above, we use c' in the denominator as the variable being summed over to differentiate it from c itself, which is already in the expression.

3 Common Effect

Recall the common effect example from lecture: both raining (R) and a ballgame (B) can cause traffic (T).



In this example $P(R = 1) = 0.7, P(B = 1) = 0.2$, and the probability table of T given R and B is below:

R	B	$P(T = 1\|R, B)$
1	1	0.9
1	0	0.5
0	1	0.5
0	0	0.1

We will now answer a few queries on this Bayes Net to determine the probability of a ballgame given different sets of knowledge.

1. What is the factorization of the joint distribution? What is the probability that it is not raining and there is no ballgame but there is still traffic?

   Based on the Bayes net, the joint distribution $P(R, B, T) = P(R)P(B)P(T|R, B)$.
   So $P(R = 0, B = 0, T = 1) = P(R = 0)P(B = 0)P(T = 1|R = 0, B = 0) = .3 * .8 * .1 = .024$.

2. What is the probability of a ballgame if we have not observed any of the variables?

   Since we have not observed any variables, the probability is simply $P(B = 1) = 0.2$.

3. What is the probability of a ballgame given that there is traffic? How does it compare to the probability from the previous question? Is ballgame independent of traffic?

   $P(B = 1|T = 1) = \frac{P(B=1,T=1)}{P(T=1)}$.
   To find $P(B = 1, T = 1)$, we sum the joint distribution over the domain of R:
   $\sum_r P(R = r, B = 1, T = 1) = P(R = 0, B = 1, T = 1) + P(R = 1, B = 1, T = 1) = .3*.2*.5+.7*.2*.9 = .156$.
   To find $P(T = 1)$, we sum the joint distribution over the domains of R and B:
   $\sum_{r,b} P(R = r, B = b, T = 1) = P(R = 0, B = 0, T = 1) + P(R = 0, B = 1, T = 1) + P(R = 1, B = 0, T = 1) + P(R = 1, B = 1, T = 1) = .3 * .8 * .1 + .3 * .2 * .5 + .7 * .8 * .5 + .7 * .2 * .9 = .46$.
   So therefore $P(B = 1|T = 1) = \frac{P(B=1,T=1)}{P(T=1)} = \frac{.156}{.46} = .339$.
   The probability of ballgame given traffic is greater than the probability of a ballgame without observations, so ballgame is not independent of traffic.

4. What is the probability of a ballgame given that there is traffic but no rain? How does it compare to the probability from the previous questions? Is ballgame conditionally independent of rain given traffic?

$P(B = 1|R = 0, T = 1) = \frac{P(R=0,B=1,T=1)}{P(R=0,T=1)}$.
$P(R = 0, B = 1, T = 1) = .3 * .2 * .5 = .03$.
To find $P(R = 0, T = 1)$, we sum the joint distribution over the domain of B:
$\sum_b P(R = 0, B = b, T = 1) = P(R = 0, B = 0, T = 1) + P(R = 0, B = 1, T = 1) = .3*.8*.1+.3*.2*.5 = .054$.
So therefore $P(B = 1|R = 0, T = 1) = \frac{P(R=0,B=1,T=1)}{P(R=0,T=1)} = \frac{.03}{.054} = .556$.
The probability of a ballgame given traffic and no rain is even greater than the probability of a ballgame given just traffic. The absence of rain increases the probability of a ballgame as the cause of traffic. Because the probability changed, ballgame is not conditionally independent of rain given traffic.