Artificial Intelligence
Exercises for Tutorial 3 on Probabilistic Inference and Bayesian Networks
Including Answers

Introduction

The following multiple choice questions are examples of typical questions one can expect on the AI exam. The questions on the AI exam are also multiple choice, but for this tutorial one has to explain the answers given. Moreover at the end one can find some open questions.

Exercises & Answers

1. Consider the following case of a car accident that involved a taxi.

   All taxis in town are blue or green. It is known that under dim lighting conditions discrimination between blue and green is 70% reliable; which means that $P(W = b|C = b)$ as well as $P(W = g|C = g)$ are 0.70, where C is the two valued variable with values g and b indicating the color of the taxi, $C = b$ means *"the taxi is blue"* and W is the two valued variable indicating the declaration of the witness, $W = b$ means *"the witness says the taxi is blue"*.

   Suppose that 7 out of 10 taxis are actually green.

   a: A witness declares that the taxi was blue. Given the declaration of our witness what is the probability that the taxi is indeed blue?

   b: Suppose two witnesses independently declare that the taxi was blue. Draw the Bayesian Network for this case and what is the probability that the taxi was indeed blue?

   c: Now assume that a third *independent* witness appears on the scene and declares that the taxi was green. What is the now probability that the taxi was indeed blue?

   Answer:

a: We must compute $P(C = b|W = b)$. Using Bayes Law this is equal to

$$P(C = b|W = b) = P(W = b|C = b)P(C = b)/P(W = b) = 0.7*0.3/P(W = b)$$

The denominator can be computed by marginalization:

$$\begin{aligned} P(W = b) &= P(W = b, C = b) + P(W = b, C = g) \\ &= P(W = b|C = g)P(C = g) + P(W = b|C = b)P(C = b) \\ &= 0.3 * 0.7 + 0.7 * 0.3 = 0.42 \end{aligned}$$

Hence $P(C = b|W = b) = 0.21/0.42 = 0.5$.

b: The Bayesian Network as three nodes (see slides lecture) now called C denoting the color of the car and a node W_1 for declaration of witness 1 and a node W_2 for the declaration of witness 2. Both CPTs are identical and follow directly from the text, entries are 0.7 and 0.3. Similar to part a we can compute:

$$\begin{aligned} &P(C = b|W_1 = b, W_2 = b) = P(W_1 = b, W_2 = b|C = b)P(C = b)/P(W_1 = b, W_2 = b) \\ &= P(W_1 = b|C = b)P(W_2 = b|C = b)P(C = b)/P(W_1 = b, W_2 = b) \\ &= 0.7 * 0.7 * 0.3/P(W_1 = b, W_2 = b) \end{aligned}$$

the third equality follows from the independency assumption and

$$\begin{aligned} &P(W_1 = b, W_2 = b) = \\ &P(W_1 = b|C = g)P(W_2 = b|C = g)P(C = g) + \\ &P(W_1 = b|C = b)P(W_2 = b|C = b)P(C = b) = \\ &0.3 * 0.3 * 0.7 + 0.7 * 0.7 * 0.3 = 0.063 + 0.147 = 0.200 \end{aligned}$$

Hence $P(C = b|W_1 = b, W_2 = b) = 0.142/0.200 = 0.710$ Hence $P(C = b|W_1 = b, W_2 = b) = 0.142/0.200 = 0.710$

c: For the three witnesses a similar computation gives that $P(C = b|W_1 = b, W_2 = b, W_3 = g) = 0.7 * 0.7 * 0.3 * 0.3/(0.3 * 0.3 * 0.7 * 0.7 + 0.7 * 0.7 * 0.3 * 0.3) = 0.5$

Notation. The use of the bold $\mathbf{P}$ in an expression like $\mathbf{P}(B|LB = t)$ is introduced in Russel and Norvigs' Artificial Intelligence book (Section 13.2.2.). It is the vector of probabilities for the values of a finite discrete random variable in a fixed order. For a boolean variable the vector with two components represents the probability distribution where the first number is the probability of the variable having value true, the second number is the value for false. Of course the sum of these two values should be 1. The constant α is used by Russel and Norvig for the normalization constant (see section 13.3). For example the product $\alpha < 0.8, 0.6 >$ equals $< \alpha \times 0.8, \alpha \times 0.6 >$. Since the sum is 1, we can compute α from $\alpha \times 1.4 = 1$.

2. The Prosecution argument. The counsel for the prosecution argues as follows:

> *Ladies and gentlemen of the jury, the probability of the observed match between the sample at the scene of the crime and that of the suspect having arisen by innocent means is 1 in 10 million. This is an entirely negligible probability, and we must therefore conclude that with a probability overwhelmingly close to 1 that the suspect is guilty. You have no alternative but to convict.*

This argument is known as the Prosecutor's Fallacy. Explain the error in the counsel's reasoning.

Answer: The confusion is between $P(M|\neg G)$, the probability of a match given the suspect is not guilty, which is 1 in 10 million, and $P(\neg G|M)$. They are not the same. Bayes rule gives the relation. The values are the same when the priors $P(\neg G|B)$ and $P(G|B)$ (based on all other evidence B) are the same.

3. "Most car accidents are caused by people that do have a driver's licence." What is suggested by this statement? What are the relevant conditional probabilities?

   Answer: The suggestion is that having a driver's licence causes (or at least has a positive impact on) car accidents, more than not having a driver's licence. The relevant probabilities are: $P(Acc|\neg HasLisence)$ and $P(Acc|HasLisence)$. The statement is probably true since most people driving a car have a driver's license. But this doesn't imply that the first conditional probability is larger than the second one. You may construct two worlds in both worlds the statement is true, but in one world the suggestion is true, in the other world it is not true.

4. Make exercise 14.1 from the book of Russel and Norvig Artificial Intelligence (3rd edition).

   Solution (from R&N): see Figures 1 and 2.

14.1

a. With the random variable C denoting which coin $\{a, b, c\}$ we drew, the network has C at the root and X_1, X_2, and X_3 as children.

The CPT for C is:

C	$P(C)$
a	$1/3$
b	$1/3$
c	$1/3$

The CPT for X_i given C are the same, and equal to:

C	X_1	$P(C)$
a	$heads$	0.2
b	$heads$	0.6
c	$heads$	0.8

b. The coin most likely to have been drawn from the bag given this sequence is the value of C with greatest posterior probability $P(C|2 \text{ heads}, 1 \text{ tails})$. Now,

$$\begin{aligned} P(C|2 \text{ heads}, 1 \text{ tails}) &= P(2 \text{ heads}, 1 \text{ tails}|C)P(C)/P(2 \text{ heads}, 1 \text{ tails}) \\ &\propto P(2 \text{ heads}, 1 \text{ tails}|C)P(C) \\ &\propto P(2 \text{ heads}, 1 \text{ tails}|C) \end{aligned}$$

where in the second line we observe that the constant of proportionality $1/P(2 \text{ heads}, 1 \text{ tails})$ is independent of C, and in the last we observe that $P(C)$ is also independent of the value of C since it is, by hypothesis, equal to $1/3$.

From the Bayesian network we can see that X_1, X_2, and X_3 are conditionally independent given C, so for example

$$\begin{aligned} &P(X_1 = tails, X_2 = heads, X_3 = heads|C = a) \\ &= P(X_1 = tails|C = a)P(X_2 = heads|C = a)P(X_3 = heads|C = a) \\ &= 0.8 \times 0.2 \times 0.2 = 0.032 \end{aligned}$$

Note that since the CPTs for each coin are the same, we would get the same probability above for any ordering of 2 heads and 1 tails. Since there are three such orderings, we have

$$P(2heads, 1tails|C = a) = 3 \times 0.032 = 0.096.$$

Figure 1: Solution exercise 14.1 (first part))

Similar calculations to the above find that

$$\begin{aligned} P(2heads, 1tails|C = b) &= 0.432 \\ P(2heads, 1tails|C = c) &= 0.384 \end{aligned}$$

showing that coin b is most likely to have been drawn.

Alternatively, one could directly compute the value of $P(C|2 \text{ heads}, 1 \text{ tails})$.

Figure 2: Solution exercise 14.1 (continuation)

5. Make exercise 14.4 from the book of Russel and Norvig Artificial Intelligence (3rd edition).

   Solution (from R&N): see Figure 3

a. Yes. Numerically one can compute that $P(B, E) = P(B)P(E)$. Topologically B and E are d-separated by A.

b. We check whether $P(B, E|a) = P(B|a)P(E|a)$. First computing $P(B, E|a)$

$$P(B, E|a) = \alpha P(a|B, E)P(B, E)$$

$$= \alpha \begin{cases} .95 \times 0.001 \times 0.002 & \text{if } B = b \text{ and } E = e \\ .94 \times 0.001 \times 0.998 & \text{if } B = b \text{ and } E = \neg e \\ .29 \times 0.999 \times 0.002 & \text{if } B = \neg b \text{ and } E = e \\ .001 \times 0.999 \times 0.998 & \text{if } B = \neg b \text{ and } E = \neg e \end{cases}$$

$$= \begin{cases} 0.0008 & \text{if } B = b \text{ and } E = e \\ 0.3728 & \text{if } B = b \text{ and } E = \neg e \\ 0.2303 & \text{if } B = \neg b \text{ and } E = e \\ 0.3962 & \text{if } B = \neg b \text{ and } E = \neg e \end{cases}$$

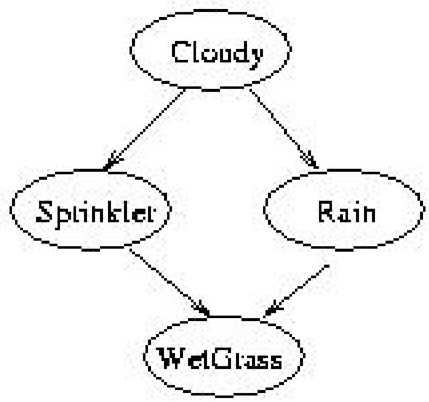
where α is a normalization constant. Checking whether $P(b, e|a) = P(b|a)P(e|a)$ we find

$$P(b, e|a) = 0.0008 \neq 0.0863 = 0.3736 \times 0.2311 = P(b|a)P(e|a)$$

showing that B and E are not conditionally independent given A.

Figure 3: Solution exercise 14.4

P(C=F)	P(C=T)
0.5	0.5



C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

Figure 4: The Sprinkler Bayesian network

6. Given the Sprinkler network shown in Figure 4. What is the best approximation of the value of $P(S = True|W = True)$ (the probability that the **S**prinkler was on given that the grass is **W**et)? Use the enumeration method. Indicate where you use the "conditional independency" relation represented by the BN.

   (a) 0.2781

   (b) 0.6471

   (c) 0.1945

   (d) 0.4298

   We compute the distribution **P** and use normalization.

$\mathbf{P}(S|W = True)$

$=$

$\alpha\mathbf{P}(S, W = True)$

$=$

$\alpha \sum_c \sum_r \mathbf{P}(C = c, S, R = r, W = True)$

$=$

$\alpha \sum_c \sum_r \mathbf{P}(C = c)\mathbf{P}(S|C = c)\mathbf{P}(R = r|C = c)\mathbf{P}(W = True|S, R = r)$

$=$ (factoring out $P(C = c)$ and $P(S|C = c)$ of the summation over r:)

$\alpha \sum_c \mathbf{P}(C = c)\mathbf{P}(S|C = c) \sum_r \mathbf{P}(R = r|C = c)\mathbf{P}(W = True|S, R = r)$

We compute this expression for each value of S (i.e. for both S=True and S=False). to obtain both values of the distribution $\mathbf{P}(S|W = True)$.

We start with the conditional probability value for $S = True$.

$P(S = True|W = True)$

$=$

$\alpha \sum_c P(C = c)P(S = True|C = c) \sum_r P(R = r|C = c)P(W = True|S = True, R = r)$ (where α stands for the constant $P(W = True)$)

$=$

$\alpha P(C = True).P(S = True|C = True) \sum_r P(R = r|C = True)P(W = True|S = True, R = r)$

$+$

$\alpha P(C = False).P(S = True|C = False) \sum_r P(R = r|C = False)P(W = True|S = True, R = r)$.

$=$

$\alpha 0.5 \times 0.1 \times (P(R = True|C = True)P(W = True|S = True, R = True) + P(R = False|C = True)P(W = True|S = True, R = False))$

$+$

$\alpha 0.5 \times 0.5 \times (P(R = True|C = False)P(W = True|S = True, R = True) + P(R = False|C = False)P(W = True|S = True, R = False))$

$=$

$\alpha 0.5 \times 0.1 \times (0.8 \times 0.99 + 0.2 \times 0.9) + \alpha 0.5 \times 0.5 \times (0.2 \times 0.99 + 0.8 \times 0.9)$

$=$

$\alpha 0.2781$

And, for $S = False$:

$P(S = False|W = True)$

$=$

$\alpha \sum_c P(C = c) P(S = False|C = c) \sum_r P(R = r|C = c) P(W = True|S = False, R = r)$

$=$

$\alpha P(C = True).P(S = False|C = True) \sum_r P(R = r|C = True) P(W = True|S = False, R = r)$

$+$

$\alpha P(C = False).P(S = False|C = False) \sum_r P(R = r|C = False) P(W = True|S = False, R = r)$.

$=$

$\alpha 0.5 \times 0.9 \times (P(R = True|C = True) P(W = True|S = False, R = True) + P(R = False|C = True) P(W = True|S = False, R = False))$

$+$

$\alpha 0.5 \times 0.5 \times (P(R = True|C = False) P(W = True|S = False, R = True) + P(R = False|C = False) P(W = True|S = False, R = False))$

$=$

$\alpha 0.5 \times 0.9 \times (0.8 \times 0.9 + 0.2 \times 0.0) + \alpha 0.5 \times 0.5 \times (0.2 \times 0.9 + 0.8 \times 0.0)$

$=$

$\alpha 0.369$

Thus: $\mathbf{P}(S|W = True) = \alpha < 0.2781, 0.369 >$.

After normalization we obtain: $\mathbf{P}(S|W = True) =< 0.2781/0.2781+0.369, 0.369/0.2781+0.369 >=< 0.4296, 0.5702 >$.

Conclusion: answer d) (0.4298) is the best approximation of $P(S = True|W = True)$.

7. Given the Sprinkler network shown in Figure 4. What is the best approximation of the value of $P(S = True|W = True, R = True)$ (the probability that the **S**prinkler was on given that the grass is **W**et and that it was **R**aining)?

   (a) 0.2781
   (b) 0.6471
   (c) 0.1945
   (d) 0.4298

   Using the same technique as in the previous exercise, we compute:

   $\mathbf{P}(S|W = True, R = True)$

$=$

$\alpha \mathbf{P}(S, W = True, R = True)$

$=$

$\alpha \sum_c \mathbf{P}(C = c, S, R = True, W = True)$

$=$

$\alpha \sum_c \mathbf{P}(C = c)\mathbf{P}(S|C = c)\mathbf{P}(R = True|C = c)\mathbf{P}(W = True|S, R = True)$

We first compute the value for $S = True$:

$P(S = True|W = True, R = True)$

$=$

$\alpha \sum_c P(C = c)P(S = True|C = c)P(R = True|C = c)P(W = True|S = True, R = True)$

$=$

$\alpha P(C = True)P(S = True|C = True)P(R = True|C = True)P(W = True|S = True, R = True)$

$+$

$\alpha P(C = False)P(S = True|C = False)P(R = True|C = False)P(W = True|S = True, R = True)$

$=$

$\alpha(0.5 \times 0.1 \times 0.8 \times 0.99 + 0.5 \times 0.5 \times 0.2 \times 0.99)$

$=$

$\alpha 0.0891$

For $S = False$ we compute:

$P(S = False|W = True, R = True)$

$=$

$\alpha \sum_c P(C = c)P(S = False|C = c)P(R = True|C = c)P(W = True|S = False, R = True)$

$=$

$\alpha P(C = True)P(S = False|C = True)P(R = True|C = True)P(W = True|S = False, R = True)$

$+$

$\alpha P(C = False)P(S = False|C = False)P(R = True|C = False)P(W = True|S = False, R = True)$

$=$

$\alpha(0.5 \times 0.9 \times 0.8 \times 0.9 + 0.5 \times 0.5 \times 0.2 \times 0.9)$

$=$

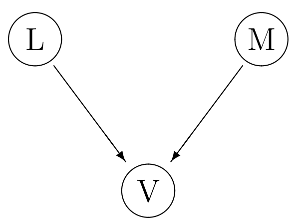
$\alpha 0.369$

Normalization: $\mathbf{P}(S|W = True, R = True) = \alpha < 0.0891, 0.369 >=< 0.0891/(0.0891 + 0.369), 0.369/(0.0891 + 0.369) >=< 0, 1945, 0, 8055 >$.

Conclusion: c is the correct answer.

Remark: the two exercises show an example of **explaining away**: the presence of one of the possible causes of an observed event makes the other less probable. The fact that the Sprinkler is on makes it less probable that there has been Rain.

8. In the Bayesian Network below with three boolean variables the probabilities for P and M are: $P(M = true) = 0,1$ and $P(L = true) = 0.7$ and the conditional probabilities for variable V are as shown in the table.



L	M	$P(V = true \mid L, M)$
true	true	0,9
true	false	0,5
false	true	0,3
false	false	0,05

What is the value of $P(V = true \mid L = true)$?

(a) 0.72

(b) 0.54

(c) 0.46

(d) 0.28

Answer: b (0.54)

Following the same strategy via computing the full joint probability and using the definition of conditional probability: $P(X|Y) = P(X, Y)/P(Y)$:

$P(V = t|L = t)$

$=$ (summing out M)

$\sum_m P(V = t, M = m|L = t)$

$=$ (definition cond. prob)

$\frac{\sum_m P(V=t,M=m,L=t)}{\sum_m \sum_v P(V=v,M=m,L=t)}$

$=$ (use BN)

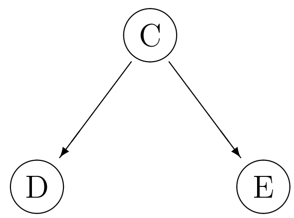


Figure 5: A Bayesian Network.

$\frac{\sum_m P(L=t)P(M=m)P(V=t|L=t,M=m)}{\sum_m \sum_v P(L=t)P(M=m)P(V=v|L=t,M=m)}$

= (fill in from the CPT's in the BN)

0.54.

A short route:

$P(V|L) = \sum_m P(V|L,M).P(M) = 0.9 \times 0.1 + 0.5 \times 0.9 = 0.54$

So, the correct answer is b)

9. Consider the Bayesian Network in Figure 5.

   All three nodes represent boolean variables. The probability distributions for the nodes of the network are as follows.

   For node C: $P(C = true) = 0.4$.

   For node D: $P(D = true|C = true) = 0.8$ and $P(D = true|C = false) = 0.3$.

   For node E: $P(E = true|C = true) = 0.9$ and $P(E = true|C = false) = 0.2$.

   What is the value of $P(D = true)$?

   (a) 0.50
   (b) 0.32
   (c) 0.18
   (d) 0.90

   Answer: $P(D = t) = \sum_c \sum_e P(C = c, E = e, D = t) = \sum_c \sum_e P(C = c).P(E = e|C = c).P(D = t|C = c)$ (using the BN semantics). By factoring out and simplifying because $\sum_e P(E = e|C = f) = 1$, as well as $\sum_e P(E = e|C = t) = 1$ we obtain: $P(C = t).P(D = t|C = t) + P(C = f).P(D = t|C = f) = 04. \times 0.8 + 0.6 \times 0.3 = 0.5$.

10. Consider again the Bayesian Network in Figure 5 with the probability distributions as given in the exercise above. One of the following statements is true. Which one?

(a) $P(D = true|E = true) > P(D = true)$

(b) $P(D = true|E = true) = P(D = true)$

(c) $P(D = true|E = true) < P(D = true)$

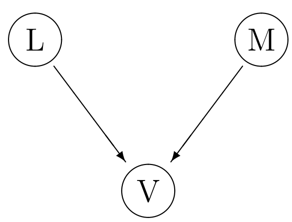
(d) There is not enough information to compute $P(D = true|E = true)$.

Answer: (a) $P\,(D = true|E = true) > P(D = true)$. We already know that P(D=true) = 0.5, so we need to compute $P\,(D\textbf{=}t|\ E\textbf{=}t) =$ P(D=t, E=t) **/** P(E=t). Computing the numerator and denominator one by one we obtain:

P(D=t, E=t) = P(D=t, E=t, C=t) + P(D=t, E=t, C=f) = (from Bayes) P(D=t **|** E=t, C=t) P(E=t, C=t) + P(D=t **|** E=t, C=f) P(E=t, C=f) = P(D=t **|** E=t, C=t) P(E=t **|** C=t) P(C=t) + P(D=t **|** E=t, C=f) P(E=t **|** C=f) P(C=f) = (D and E conditionally independent) P(D=t **|** C=t) P(E=t **|** C=t) P(C=t) + P(D=t **|** C=f) P(E=t **|** C=f) P(C=f) = 0.8 x 0.9 x 0.4 + 0.3 x 0.2 x 0.6 = 0.324

P(E=t) = P(E=t, C=t) + P(E=t, C=f) = (from Bayes) P(E=t **|** C=t) P(C=t) +P (E=t **|** C=f) P(C=f) = 0.9 x 0.4 + 0.2 x 0.6 (we know that P(C=f) = 1 - P(C=t)) = 0.36 + 0.12 =0.48

The last step is to compute $(D\textbf{=}t|E\textbf{=}t)$ = P(D=t, E=t) **/** P(E=t) =0.324 **/** 0.48 = 0.675.



L	M	$P(V = true \mid L, M)$
true	true	0,9
true	false	0,5
false	true	0,3
false	false	0,05

11. In the Bayesian Network above with three boolean variables the probabilities for P and M are: $P\,(M = true) = 0,2$ and $P\,(L = true) = 0.7$ and the conditional probabilities for variable V are as shown in the table.

What is the value of $P(V = false \mid L = false)$?

(a) 0.3

(b) 0.7

(c) 0.9

(d) 0.1

Answer: c (0.9). In a similar manner to how we computed the answer for the previous exercise, we know that P(V = f **|** L = f) = P(V = f, L = f) **/** P(L = f). Additionally, we know P(M = f) = 1 - P(M = t), P(L = f) = 1 - P(L = t) and P(V = f **|** M, L) = 1 - P(V = t **|** M, L). We will consider the numerator separately:

P(V = f, L = f) = P(V = f, L =f, M = t) + P(V = f, L =f, M = f) = (from Bayes) P (V = f **|** L =f, M = t) P(L = f, M = t) + (V = f **|** L =f, M = f) P(L = f, M = f) = (L and M independent) (V = f **|** L =f, M = t) P(L = f) P(M = t) + (V = f **|** L =f, M = f) P(L = f) P(M = f) = 0.7 x 0.3 x 0.2 + 0.95 x 0.3 x 0.8 = 0.27.

Then, P(V = f **|** L = f) = 0.27 **/** 0.3 = 0.9

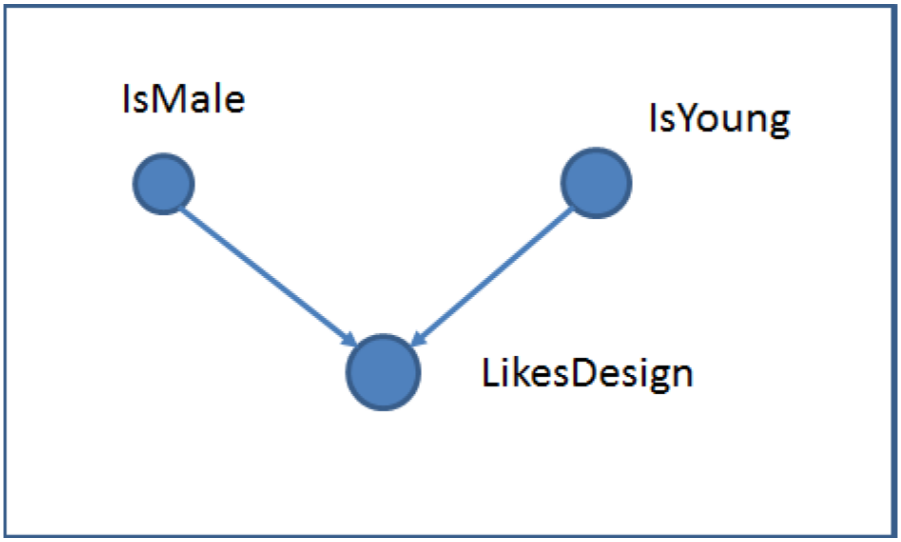


Figure 6: The dependencies for the user tests of my design

12. The Bayesian Network structure given in Figure 6 models the dependencies between three properties related to my design. IsMale is true when the user is male, IsYoung is true when the user is young, LikesDesign is true when the user likes my design. Which of the following statements is true? Give a proof or counter example.

    (a) IsYoung and IsMale are independent

    (b) IsYoung and IsMale are independent given LikesDesign

    Answer: a) is true, b) is false.