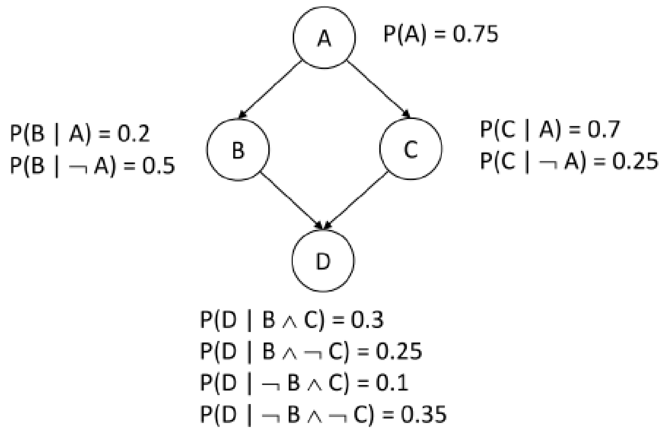
ROB311 – 12 October 2020

1. Consider the following Bayesian Network A, B, C and D are random Boolean variables. What is the probability of having D true if we know that A is true?



Calculate. P(D|A) =?

Solution:

$$P(D|A) = \frac{P(A,D)}{P(A)} = \frac{P(A,B,C,D) + P(A,B,\neg C,D) + P(A,\neg B,C,D) + P(A,\neg B,\neg C,D)}{P(A)}$$

$$= \begin{matrix} P(B|A)P(C|A)P(D|B,C) + P(B|A)P(\neg C|A)P(D|B,\neg C) + P(\neg B|A)P(C|A)P(D|\neg B,C) + \\ +P(\neg B|A)P(\neg C|A)P(D|\neg B,\neg C) \end{matrix}$$

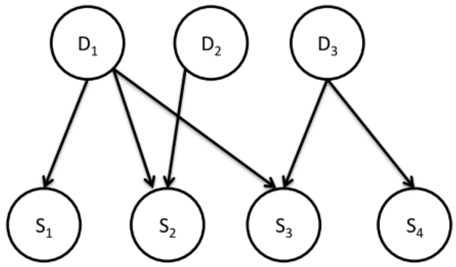
$$= 0.2 \times 0.7 \times 0.3 + 0.2 \times 0.3 \times 0.25 + 0.8 \times 0.7 \times 0.1 + 0.8 \times 0.3 \times 0.35$$

$$= 0.042 + 0.015 + 0.056 + 0.084 = 0.197$$

2. A patient goes to the doctor for a medical condition, the doctor suspects three diseases as the cause of the condition. The three diseases are D1, D2, D3, which are marginally independent from each other. There are four symptoms S1, S2, S3, and S4, which the doctor wants to check for presence in order to find the most probable cause of the condition. The symptoms are conditionally dependent to the three diseases as follows: S1 depends only on D1, S2 depends on D1 and D2. Symptom S3 depends on D1 and D3, and S4 depends only on D3. Assume all random variables are Boolean, they are either `true' or `false'.
   a) Draw the Bayesian network for this problem.
   b) Write down the expression for the joint probability distribution as a product of conditional probabilities.

Solution:

a)



b)

$$P(D_1, D_2, D_3, S_1, S_2, S_3) = P(D_1)P(D_2)P(D_3)P(S_1|D_1)P(S_2|D_1, D_2)P(S_3|D_1, D_3)P(S_4|D_3)$$

3. Disney Park hired an engineer. They want to predict when they will have a lot of visitors. They gathered a lot of data but they do not know how to process it.

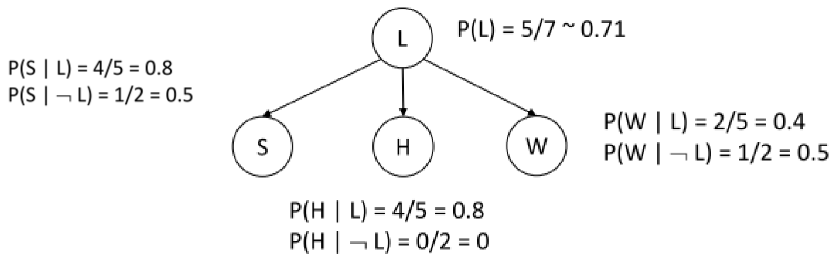
   They asked the engineer to do it.

	Feature 1 Sunny?	Feature 2 High Temperature?	Feature 3 Weekend?	Class Lots of Visitors?
Day 1	yes	yes	yes	yes
Day 2	yes	no	yes	yes
Day 3	no	yes	no	yes
Day 4	yes	yes	no	yes
Day 5	yes	yes	no	yes
Day 6	yes	no	no	no
Day 7	no data since you were on business travel			
Day 8	no	no	yes	no

   a) Draw the Bayesian network based on what we can learn from the data and show the conditional probabilities
   b) Based on the learned Bayesian network, what's the probability of receiving many visitors at Disney Park on a cloudy and hot weekend day?

Solution:

a)



b)

$$P(L \mid \neg S, H, W) = P(L, \neg S, H, W)/P(\neg S, H, W)$$
$$= 0.045/0.045 = 1$$
with
$$P(L, \neg S, H, W) = P(L)\ P(\neg S \mid L)\ P(H \mid L)\ P(W \mid L)$$
$$= 0.71 \times 0.2 \times 0.8 \times 0.4 \sim 0.045$$
$$P(\neg L, \neg S, H, W) = P(\neg L)\ P(\neg S \mid \neg L)\ P(H \mid \neg L)\ P(W \mid \neg L)$$
$$= 0.29 \times 0.5 \times 0 \times 0.5 = 0$$
$$P(\neg S, H, W) = P(L, \neg S, H, W) + P(\neg L, \neg S, H, W) = 0.045 + 0 = 0.045$$

4. Medical diagnosis

A hospital uses a support system for detecting lung problems. The system is designed to help in the diagnosis of tuberculosis, cancer, and bronchitis.
The system will use previous data from the hospital gathered from previous consultations.

At the registration in the hospital, a new patient it is asked to fill in a questionnaire and answer 2 questions: "Have you recently visited Asia? " and " Are you a smoker? ".

The data shows that:

- among all the patients, 10% have recently visited Asia, and 30% are smokers;
- Tuberculosis is present in Asia, and a patient who recently visited Asia has 10% of having tuberculosis and a patient who have not been recently to Asia has only 1% of having tuberculosis;
- Patients that smoke and complain of lung problems have 20% of having cancer (against only 2% for patients that do not smoke);
- Patients that do not smoke are suffering in 80% of cases of only a bronchitis (against only 60% for people that smoke).

The doctor proposes only 2 tests:

- The doctor auscultates the patient's lungs with a stethoscope. A bronchitis or a lung cancer can be detected in 60% of cases. When the patient has none of these two diseases, the doctor will detect it with a probability of 99%;
- The doctor orders an X-Ray. With the X-Ray the tuberculosis or lung cancer are detected in 70% of cases. If the patient has none of these two diseases, nothing will be observed on the X-Ray with a probability of 98%.

Questions:

- Model this problem using a Bayesian network;
- If the patient is not smoking and has not recently visited Asia, can you infer with disease?
- According to the disease inferred in Point 2, the doctor decides to auscultate the patient's lungs with a stethoscope? Why?
  The stethoscope test is negative. What is the new inferred diagnosis?
- The doctor orders an X-Ray. The X-Ray test is positive. What is the new inferred diagnosis?
- Was the X-Ray needed?

Solution:

1.

A=Asia
S=smoker
T= tuberculosis,
C=cancer
B=bronchitis
X=x-Ray
M=stethoscope/medical device

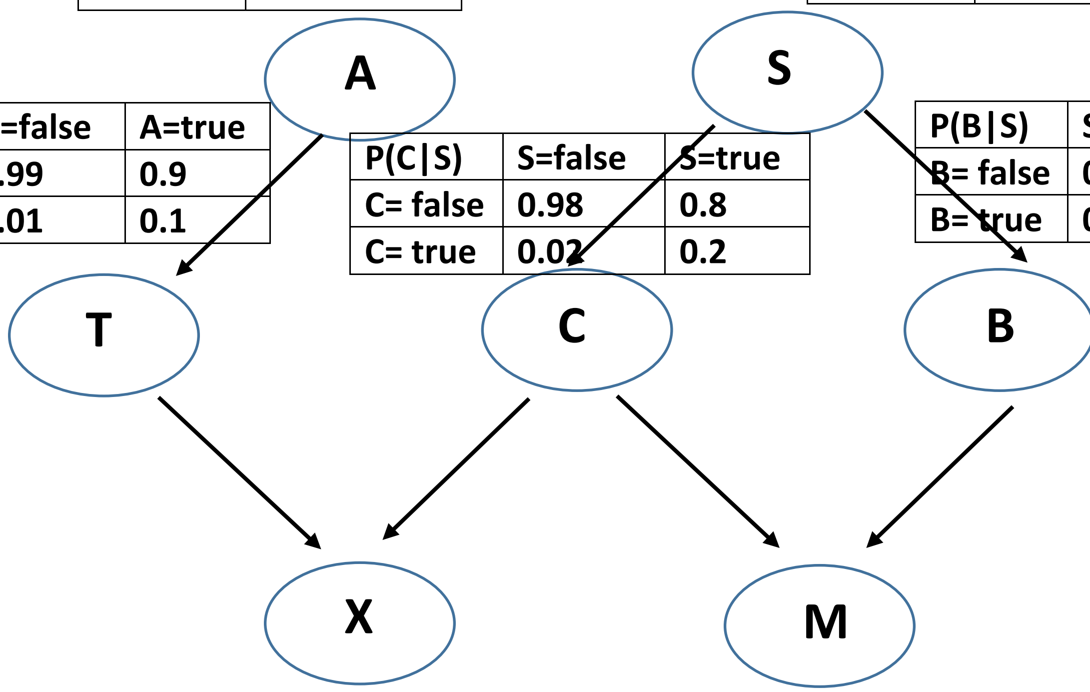
P(A)	
A= false	0.9
A= true	0.1

P(S)	
S= false	0.7
S= true	0.3

P(T|A)	A=false	A=true
T= false	0.99	0.9
T= true	0.01	0.1

P(C|S)	S=false	S=true
C= false	0.98	0.8
C= true	0.02	0.2

P(B|S)	S=false	S=true
B= false	0.2	0.4
B= true	0.8	0.6



P(X|T,C)	T=false	T=false	T=true	T=true
	C=false	C=true	C=false	C=true
X= false	0.98	0.3	0.3	0.3
X= true	0.02	0.7	0.7	0.7

P(M|C,B)	C=false	C=false	C=true	C=true
	B=false	B=true	B=false	B=true
M= false	0.99	0.4	0.4	0.4
M= true	0.01	0.6	0.6	0.6

2. P(T|A=false, S=false) = α x P(T,A=false, S=false)

= α x $\sum_{C,B,X,M} P(A = false)xP(S = false)\ x\ P(T|A = false)xP(C|S = false)xP(B|S = false)xP(X|T,C)xP(M\,|B,C) =$

= α x $P(A = false)xP(S = false)\ x\ P(T|A = false)x(\sum_C P(C|S = false)x(\sum_B P(B|S = false)x(\sum_X P(X|T,C)x(\sum_M P(M|B,C)$

= α x $P(A = false)xP(S = false)\ x\ P(T|A = false)x(1x\big(1x(1x1)\big))$

= α x $P(A = false)xP(S = false)\ x\ P(T|A = false)$

= α” x 0.01

In the same manner we calculate

P(C|A=false, S=false) = α x P(C,A=false, S=false)

= α x $P(A = false)xP(S = false)\ x\ P(C|S = false)$

= α” x 0.02

P(B|A=false, S=false) = α x P(B,A=false, S=false)

= α x $P(A = false)xP(S = false)\ x\ P(B|S = false)$

= α” x 0.8

After the first diagnosis we can infer that the patient has bronchitis

3. The doctor wants to confirm the diagnosis inferred at point 2.

   P(T|A=false, S=false, M=false) = α x P(T,A=false, S=false, M=false)

   = α x $\sum_{C,B,X} P(A = false)xP(S = false)\ x\ P(T|A = false)xP(C|S = false)xP(B|S = false)xP(X|T,C)xP(M = false\,|B,C) =$

   = α x $P(A = false)xP(S = false)\ x\ P(T|A = false)x(\sum_C P(C|S = false)x(\sum_B P(B|S = false)x\ P(M = false|B,C)\ x\ (\sum_X P(X|T,C)))$

   = α x $\boldsymbol{P(A = false)xP(S = false)\ x\ P(T|A = false)x(\sum_C P(C|S = false)x(\sum_B P(B|S = false)x\ P(M = false|B,C)}$

   = α” x

   $P(T|A = false)x(P(C|S = false)x\big(P(B = false|S = false)x\ P(M = false|B = false, C) + P(B = true|S = false)x\ P(M = false|B = true, C)\big) + P(C = false|S = false)x\big(P(B = false|S = false)xP(S = false|B = false, C = false) + (B = true|S = false)xP(S = false|B = true, C = false)\big)\big)$

   = α” x 0.01x(0.02 x (0.2x0.4+0.8x0.4)+0.98x(0.2x0.99+0.8x0.4))

   = α” x 0.0051564

   We will calculate in the same manner

   P(B|A=false, S=false, M=false) = α x P(B,A=false, S=false, M=false)=

= α” x 0.32

P(C|A=false, S=false, M=false) = α x P(C,A=false, S=false, M=false)=
= α” x 0.008

The doctor trues to confirm his first diagnosis with a respiratory test with the stethoscope.
The test with the stethoscope should have been positive.
The stethoscope test is negative
However, the doctor still believes the patient has a bronchitis but his confidence in the first diagnosis decreased

4. P(T|A=false, S=false, M=false, X=true) = α x P(T,A=false, S=false, M=false, X=true)
= α x $\sum_{C,B} P(A = false) x P(S = false)\ x\ P(T|A = false) x P(C|S = false) x P(B|S = false) x P(X = true|T, C) x P(M = false\ |B, C)$
= α” x
$P(T = true|A = false) x\ \sum_{C}\ P(C|S = false) x P(X = true|T, C) x (\sum_{B}\ P(B|S = false) x P(S = false|B, C)))$
= α” x $P(T = true|A = false) x$
(P(C=true|S=false)xP(X=true|T=true,C=true)x(P(B=false|S=false)x P(M=false|B=false,C=true)+ P(B=true|S=false)x P(M=false|B=true,C=true))) + P(C=false|S=false)xP(X=true|T=true,C=false)x(P(B=false|S=false)x P(M=false|B=false,C=false)+ P(B=true|S=false)x P(M=false|B=true,C=false))=
= α” x (0.01x(0.98 x 0.7 x (0.2 x 0.99 + 0.8 x 0.4) + 0.2 x 0.7 x (0.2 x 0.4 ° 0.8 x 0.4))
= α” x 0.00360948

We calculate similarly
P(C|A=false, S=false, M=false, X=true) = α x P(C,A=false, S=false, M=false, X=true)
= α” x 0.0056

P(B|A=false, S=false, M=false, X=true) = α x P(B,A=false, S=false, M=false, X=true)
= α” x 0.01288448

The doctor trues to confirm his first diagnosis with a x-ray.
A priori the respiratory test with the stethoscope should be positive and the x-ray should be negative – that confirms a bronchiolitis.

The respiratory test with the stethoscope was negative and the x-ray positive , however the doctor continues to think that the patient has a bronchiolitis, however his confidence decreased again.

5. The X-ray was not needed. No matter what the result from the X-ray is the doctor will continue to think that the patient has a bronchiolitis.
   (we can show that by calculating P(T|A=false, S=false, M=false, X=false); P(C|A=false, S=false, M=false, X=false), P(B|A=false, S=false, M=false, X=false))

   The confidence given to the 2 tests (60%-stethoscope and 70%-x-ray) are low, which make the result not change the first diagnosis of the doctor.

5. We are in a bank. Let C be a random Boolean variable indicating if a person is a criminal that is doing bank robberies ($c = 1$) or not ($c = 0$) and A, be a random Boolean variable indicating an arrest.
   A criminal shall be arrested with probability $P(A = 1|C = 1) = 0.98$, a non-criminal with probability $P(A = 1|C = 0) = 0.001$. One in 100 000 is a criminal, $P(C = 1) = 0.00001$.

   What is the probability that an arrested person actually is a criminal?

   $P(C=1|A=1) = P(A,C)/P(A)= (P(A=1|C=1) * P(C=1))/P(A=1) =$
   $= (P(A=1|C=1) * P(C=1))/(P(A=1/C=0)P(C=0)+ P(A=1/C=1)P(C=1))$
   $= (0.98*0.00001)/(0.001*0.99999+0.98*0.00001)$
   $= 0.0000098 / 0.00100979$
   $= 0.01$