T.61.5140 Machine Learning: Advanced Probablistic Methods
Hollmén, Raiko (Spring 2008)
Problem session, 1st of February, 2008
http://www.cis.hut.fi/Opinnot/T-61.5140/

1. Construct a causal network and follow the reasoning in the following story. Mr. Holmes is working in his office when he receives a phone call (C) from his neighbor, who tells him that Holmes' burglar alarm (A) has gone off. Convinced that a burglar has broken into his house (B), Holmes rushes to his car and heads for home. On his way, he listens to the radio, and in the news it is reported (R) that there has been a small earthquake (E) in the area. Knowing that earthquakes have a tendency to turn on burglar alarms, he returns to work.

Draw a causal network and write the joint probability for the random variables C,A,B,R,E. (pages 360–)

Solution:

A causal network (=Bayesian network) can be formed by connecting events that may cause each other with directed edges.

1. Burglar may cause the alarm ($B \rightarrow A$)
2. Earthquake may cause the alarm ($E \rightarrow A$)
3. Earthquake may cause a report ($E \rightarrow R$)
4. Alarm may cause phone call ($A \rightarrow C$)

The graphical model for the above is shown in Figure 1 (left). The joint probability can be written by conditioning each random variable by its parents:

$$P(C, A, B, R, E) = P(B)P(E)P(A \mid B, E)P(R \mid E)P(C \mid A). \tag{1}$$

(Earthquake might also tempt burglar to break in, but Mr. Holmes did not consider this possibility.)

2. Consider the network in Figure 1 (right). What is the Markov blanket of each variable? (pages 382–383)

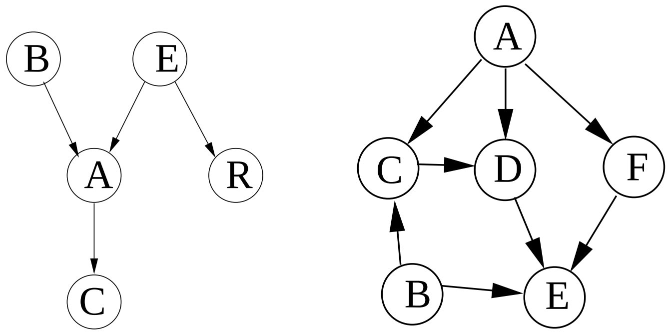

Figure 1: *Left:* Problem 1. *Right:* Problems 2 and 3.

Solution:

The Markov blanket includes the node's neighbours and co-parents, that is, other parents of the node's children.
Markov blanket for A: B, C, D, F
Markov blanket for B: A, C, D, E, F
Markov blanket for C: A, B, D
Markov blanket for D: A, B, C, E, F
Markov blanket for E: B, D, F
Markov blanket for F: A, B, D, E
Note: Problem 4 in the next weeks exercises will further deal with the Markov blanket.

3. From the network in Figure 1 (right), do the following conditional independencies follow? (D-separation, page 378)
(a) $A \perp B \mid C$
(b) $A \perp B \mid \emptyset$
(c) $C \perp E \mid B, D$
(d) $C \perp D \mid A, B$
(e) $B \perp F \mid A, C$
(f) $A \perp E \mid D, F$

Solution:

Conditional independency follows if all paths connecting the variables are blocked. A path is blocked in three cases: If the arrows meet at an observed node (1) head-to-tail or (2) tail-to-tail, or (3) the arrows meet head to head at the node, and neither the node, nor any of its descendants, are

observed. (Being observed is the same as being conditioned on, here.)
(a) No, the path through C is not blocked
(b) Yes, all paths are blocked
(c) No, the path through A and F is not blocked
(d) No, the direct path is not blocked
(e) Yes, all paths are blocked
(f) No, the path through C and B is not blocked (Note that D is an observed descendant of C)

4. Consider the Bayesian network defined by the following tables. Write a program that generates random realisations (samples) of the variables A, B, C, D. Based on 1000 samples, estimate $P(B = 1 \mid D = 1)$.

P(A)	
A=0	0.5
A=1	0.5

P(B)	
B=0	0.8
B=1	0.2

$P(C \mid A, B)$	A=0, B=0	A=0, B=1	A=1, B=0	A=1, B=1
C=0	0.8	0.7	0.6	0.3
C=1	0.2	0.3	0.4	0.7

$P(D \mid C)$	C=0	C=1
D=0	0.9	0.2
D=1	0.1	0.8

Solution:

A Bayesian network is a generative model, so generating random samples is rather easy. One needs to select the order in which to sample the variables such, that the parents are always sampled before the children. In this case, there are two such orders, A,B,C,D or B,A,C,D. With a proper order, you can read the probabilities directly from the given tables.

You can find the Matlab code of the solution on the course webpage.

The probability $P(B = 1 \mid D = 1)$ is computed accurately:

$$P(B = 1 \mid D = 1) = \frac{\sum_{A,C} P(D = 1 \mid C)P(C \mid A, B = 1)P(A)P(B = 1)}{\sum_{A,B,C} P(D = 1 \mid C)P(C \mid A, B)P(A)P(B)} \tag{2}$$

$$= \frac{0.09}{0.338} \approx 0.266 \tag{3}$$

and by sampling 1000 times from the joint distribution and taking the fraction of counts:

$$P(B = 1 \mid D = 1) = \frac{P(B = 1, D = 1)}{P(D = 1)} \approx \frac{93/1000}{317/1000} \approx 0.29 \tag{4}$$

The evolution of the sampled estimate is plotted in Figure 2.

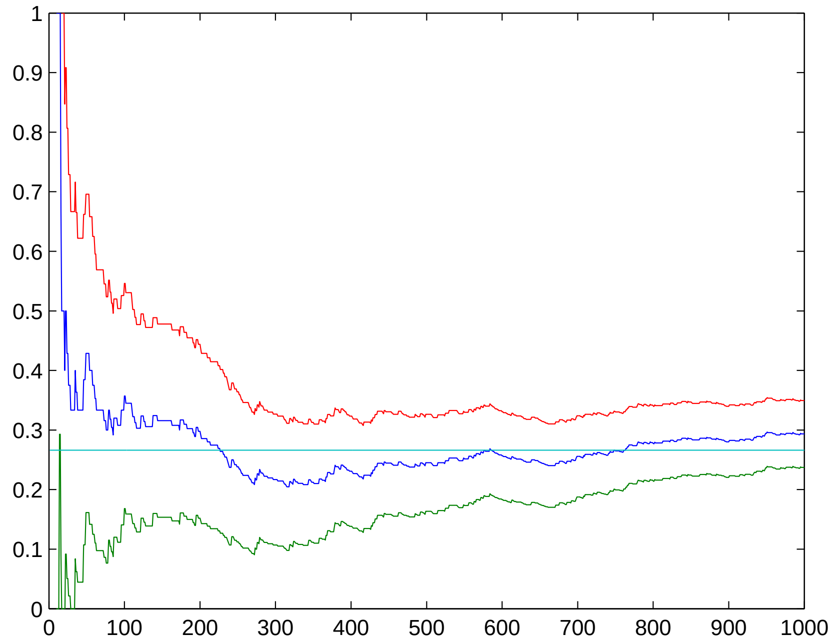

Figure 2: Problem 4. The accurate value of $P(B = 1 \mid D = 1)$ is shown as the horizontal line. The estimate with confidence intervals (twice the standard deviation), evolves as a function of the number of samples used.