Exercise 1 :

In the graphs in Figures 1 and 2, determine which variables are d-separated from A. Note that it is sufficient to find a single open path along which evidence can be transmitted; if such a path exists then the variables are *not* d-separated (instead they are said to be d-connected).

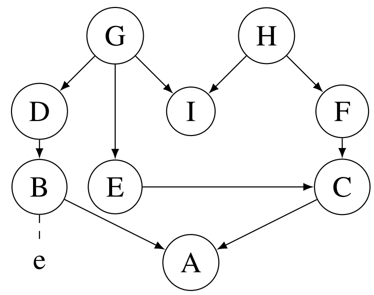


Figure 1: Figure for Exercise 1.

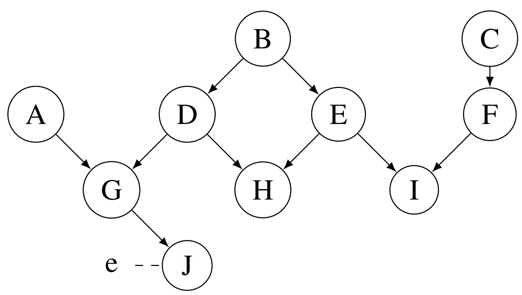


Figure 2: Figure for Exercise 1.

Solution:

In (a), all variables are d-connected to A. We can see that by checking whether there *exists* an open path between each pair of variables. Take A and D for instance. The path $A \leftarrow C \leftarrow F \leftarrow H \rightarrow I \leftarrow G \rightarrow D$ is closed, since it contains the connection ($H \rightarrow I \leftarrow G$), which is converging and there is no evidence on I (nor on any of its descendants for which it has none). However, the path $A \leftarrow C \leftarrow E \leftarrow G \rightarrow D$ is open since each triplet of nodes along the path (i.e., ($A \leftarrow C \leftarrow E$), ($C \leftarrow E \leftarrow G$), ($E \leftarrow G \rightarrow D$) defines an open connection according to the d-separation rules. With this path we have established the existence of an open path, and we therefore have that A and D are no d-separated given the evidence provided.

In (b), all variables except C and F are d-connected to A.

Exercise 2 :

Consider the network in Figure 3.

- What are the minimal set(s) of variables that we should have evidence on in order to d-separate C and E (that is, sets of variables for which no proper subset d-separates C and E)?

- What are the minimal set(s) of variables we should have evidence on in order to d-separate A and B?
- What are the maximal set(s) of variables that we can have evidence on and still d-separate C and E (that is, sets of variables for which no proper superset d-separates C and E)?
- What are the maximal set(s) of variables that we can have evidence on and still d-separate A and B?

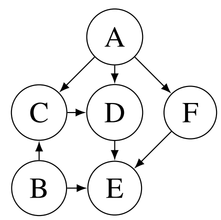


Figure 3: A causal network for Exercise 3.

Solution:

Minimal sets d-separating C and E: $\{B, D, F\}$ and $\{A, B, D\}$.

Minimal sets d-separating A and B: $\emptyset$.

Maximal set d-separating C and E: $\{A, B, D, F\}$.

Maximal set d-separating A and B: $\{F\}$.

Exercise 3 :

Consider the network in Figure 3. Which conditional probability tables must be specified to turn the graph into a Bayesian network?

Solution: The needed tables are $P(A)$, $P(B)$, $P(C \mid A, B)$, $P(D \mid A, C)$, $P(E \mid B, D, F)$, and $P(F \mid A)$.

Exercise 4 :

Peter is currently taking three courses on the topics of probability theory, linguistics, and algorithmics. At the end of the term he has to take an exam in two of the courses, but he has yet to be told which ones. Previously he has passed a mathematics and an English course, with good grades in the mathematics course and outstanding grades in the English course. At the moment, the workload from all three courses combined is getting too big, so Peter is considering dropping one of the courses, but he is unsure how this will affect his chances of getting good grades in the remaining ones. What are reasonable variables of interest in assessing Peter's situation? How do they group into information, hypothesis, and mediating variables?

Hint: The grades that Peter has already received constitute evidence about certain variables (i.e., variables for which we observe the states they are in).

Solution:

There are two information variables, *English Grade* and *Math Grade* and three hypothesis variables, *Prob. Grade*, *Ling. Grade*, and *Alg. Grade*. Mediating variables could be *Mathematical Talent* and *Linguistic Talent*.

Exercise 5 :

Assume it is your responsibility to monitor volcanic eruptions. You receive data from two different stations (seismometers), S_1 and S_2. Each S_i is modeled as a Boolean variable where "true" stands for "I detected an eruption" and "false" stands for "I did not detect an eruption". The seismometers are not fully reliable, however; they may not detect an eruption even though there was one, and they may mistake an earthquake for an eruption of a volcano. We model this situation with two additional Boolean variables: V for volcanic eruption, and E for Earthquake.

Use the algorithm from the lecture to construct a Bayesian network for these 4 variables. Do so for the following two variable orders:

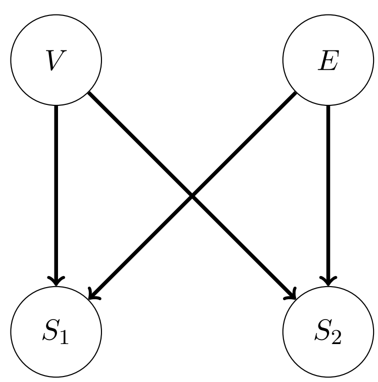
(a) $X_1 = V, X_2 = E, X_3 = S_1, X_4 = S_2$.

(b) $X_1 = S_1, X_2 = S_2, X_3 = E, X_4 = V$.

For each of these orders, draw the resulting Bayesian network. Justify your design, i.e., for each variable X_i added to the network explain why the set of parents you give X_i are needed, and why they are sufficient.

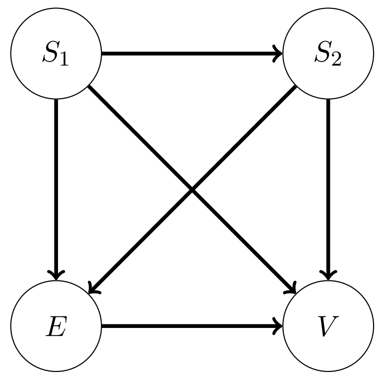
Solution:

(a) With this variable order, we get the following network:



$X_2 = E$ does not need $X_1 = V$ as a parent because Earthquakes are independent from volcano tests. $x_3 = S_1$ needs both $X_1 = V$ and $X_2 = E$ as parents because each of these may influence the measurement; same for $X_4 = S_2$, i.e., here we also need the parents $X_1 = V$ and $X_2 = E$. However, given the values of V and E, the measurements of $X_3 = S_1$ and $x_4 = S_2$ are independent. So $X_4 = S_2$ does not require the parent $X_3 = S_1$.

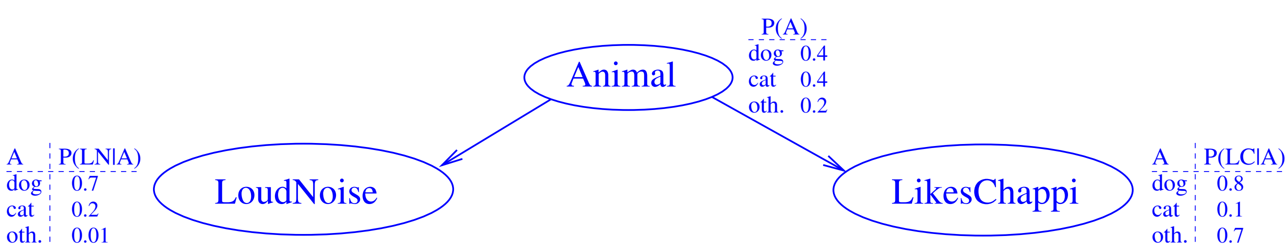
(b) With this variable order, we get the following network:



$X_2 = S_2$ needs $X_1 = S_1$ as a parent because, if S_1 detects a seismic phenomenon, then chances are higher S_2 will detect one as well. $X_3 = E$ needs each of $X_1 = S_1$ and $X_2 = S_2$ as parents because, if a station detects a seismic phenomenon, then chances are higher there was an earthquake; same for $X_4 = V$, i.e., here we also need the parents $X_1 = S_1$ and $X_2 = S_2$ because measurements indicate volcano tests as well. Finally, say we already know that S_1 and S_2 are true; then the value of E still has an influence on the value of V: If there was an earthquake, then there is a chance that the seismic measurements were caused by the earthquake rather than a volcano test. Thus V is *not* conditionally independent of E given S_1 and S_2, and we need $X_3 = E$ as a parent of $X_4 = V$ as well.

Exercise 6 :

Consider the following Bayesian network BN:



Use this network to compute the following probabilities:

(a) $P(loudnoise, dog, likeschappi)$.

(b) $P(loudnoise, other, likeschappi)$.

(c) $P(\neg loudnoise, dog, likeschappi)$.

Include intermediate steps at a level of granularity as in the lecture slides examples. In particular, for each of the probabilities P demanded in (a) – (c), write down *which* probabilities provided in BN can be combined *how* to obtain P.

Solution:

In all cases, we use the method from Chapter 16 slides 19 and 20. As the variable ordering consistent with BN, we choose $X_1 = \mathit{Animal}$, $X_2 = \mathit{LoudNoise}$, $X_3 = \mathit{LikesChappi}$.

(a) $P(\mathit{loudnoise}, \mathit{dog}, \mathit{likeschappi}) = P(\mathit{likeschappi} \mid \mathit{loudnoise}, \mathit{dog}) * P(\mathit{loudnoise} \mid \mathit{dog}) * P(\mathit{dog}) = P(\mathit{likeschappi} \mid \mathit{dog}) * P(\mathit{loudnoise} \mid \mathit{dog}) * P(\mathit{dog}) = 0.8 * 0.7 * 0.4 = 0.224$.

(b) $P(\mathit{loudnoise}, \mathit{other}, \mathit{likeschappi}) = P(\mathit{likeschappi} \mid \mathit{loudnoise}, \mathit{other}) * P(\mathit{loudnoise} \mid \mathit{other}) * P(\mathit{other}) = P(\mathit{likeschappi} \mid \mathit{other}) * P(\mathit{loudnoise} \mid \mathit{other}) * P(\mathit{other}) = 0.7 * 0.01 * 0.2 = 0.0014$.

(c) $P(\neg \mathit{loudnoise}, \mathit{dog}, \mathit{likeschappi}) = P(\mathit{likeschappi} \mid \neg \mathit{loudnoise}, \mathit{dog}) * P(\neg \mathit{loudnoise} \mid \mathit{dog}) * P(\mathit{dog}) = P(\mathit{likeschappi} \mid \mathit{dog}) * P(\neg \mathit{loudnoise} \mid \mathit{dog}) * P(\mathit{dog}) = 0.8 * 0.3 * 0.4 = 0.096$. Note here that $P(\neg \mathit{loudnoise} \mid \mathit{dog})$ is not explicitly given in the picture, but can be calculated as $1 - P(\mathit{loudnoise} \mid \mathit{dog})$; compare Chapter 16 slide 12.

Exercise 7 :

We want to construct a Bayesian network for the following random variables (all with domain *true*,*false*):

Mexico	Person X has recently travelled to Mexico
Svine_flu_infection	Person X has been infected with the svine flu virus
Vaccination	Person X has been vaccinated against svine flu
Svine_flu_sick	Person X is sick from svine flu
Fever	Person X has fever

a. Use the above ordering of the variables to determine a Baysian network structure based on the chain rule and conditional independence relations.

b. Repeat the construction with the alternative variable ordering:

Vaccination,*Mexico*,*Svine_flu_sick*,*Fever*,*Svine_flu_infection*

Solution:

a. We write the joint distribution of the variables using the chain rule:

$$\begin{aligned} &P(M, S\_f\_i, V, S\_f\_s, F) = \\ &\quad P(M)\cdot \\ &\quad P(S\_f\_i \mid M)\cdot \\ &\quad P(V \mid M, S\_f\_i)\cdot \\ &\quad P(S\_f\_s \mid M, S\_f\_i, V)\cdot \\ &\quad P(F \mid M, S\_f\_i, V, S\_f\_s)\cdot) \end{aligned}$$

We now simplify the conditional distributions by making some conditional independence assumptions (these are reasonable assumptions based on our knowledge of the domain – but not necessarily provably correct).

$$P(V \mid M, S\_f\_i) = P(V \mid M)$$

Justification: whether or not a person gets a vaccination may depend on whether he plans to travel to mexico. His decision to get a vaccination (generally) takes place before he the potential time of infection, so the decision is independent of whether he will catch the virus.

$$P(S\_f\_s \mid M, S\_f\_i, V) = P(S\_f\_s \mid S\_f\_i, V)$$

Justification: Given that we know whether a person has been infected with the virus, and whether or not he has a vaccination, it is no longer relevant to know where he may have contracted the virus, especially whether or not he traveled to Mexico. Note that this does not mean that $P(S\_f\_s \mid M) = P(S\_f\_s)$: as long as we don't know about S_f_s and V, the information of whether the person traveled to Mexico is still very relevant for the probability of S_f_s.

$$P(F \mid M, S\_f\_i, V, S\_f\_s) = P(F \mid S\_f\_s)$$

Justification: the fever is a direct consequence of the sickness. Once we know whether a person is sick, it gives us no additional useful information (regarding the probability of fever) to know whether the person was in Mexico, or got a vaccination.

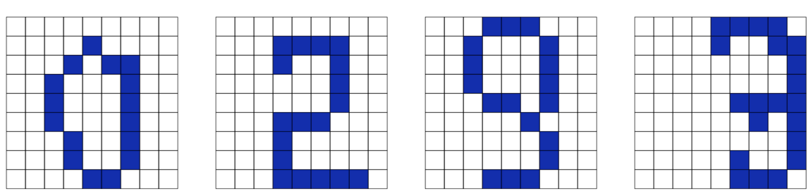
The representation of the joint distribution as the product of the simplified conditional distributions corresponds to the Bayesian network structure

b.

Going through the same procedure for the alternative ordering leads to fewer simplifications, and a more complicated network structure. The basic reason for this is that the second ordering does not respect the cause-effect relationships between the variables (i.e., taking cause variables before their effects).

Exercise 8 :

Design a Bayesian network that can be used to recognize handwritten digits 0,1,2,. . . ,9 from scanned, pixelated images like these:



- What are hypothesis and information variables?
- Could there be any useful mediating variables (consider e.g. the last image above)?
- How could you design a network structure
  - so that the conditional independencies are (approximately) reasonable
  - so that specification and inference complexity remain feasible
- How do you fill in the conditional probability tables?

Solution:

- The hypothesis variable(s) must enable us to make the intended inferences by querying these variables. Here we are interested in predicting which digit is actually represented by the image. Thus, we should have a variable
$$\textit{digit} \in \{0, 1, \ldots, 9\}$$
The information variables must allow us to enter into the Bayesian networks the relevant observations which we make. Here we observe $9 \cdot 9 = 81$ pixels, each of which can be black or white:
$$\textit{pixel\_i} \in \{b, w\} \quad i = 1, \ldots, 81$$
- Mediating variables: see below
- Writing a given digit can be seen as a cause for the pixels to become black or white. Just inserting edges for these direct cause-effect relationships gives us the structure According to this structure, any two pixels are conditionally independent given the digit (this type of structure is also called a *Naive Bayes Model*). This is not entirely realistic: for example consider pixels 27 and 36: According to the network structure:
$$P(\textit{pixel\_27} = b \mid \textit{digit} = 3, \textit{pixel\_36}{=}b) = P(\textit{pixel\_27} = b \mid \textit{digit} = 3),$$
i.e. knowing that the actual digit is a 3, the information on pixel 36 carries no relevant information for pixel 27. However, if in addition to *digit*=3 we know that *pixel_36*=b, this will indicate that the digit is written flush right in the box, and should thereby increase the probability of *pixel_27*=b.

  We can improve the model by introducing a mediating variable
$$\textit{alignment} \in \{l, r, c\}$$
(for left, right, and center alignment) and add it to the network like this: Now two pixels are only independent given both *digit* and *alignment* – which still will not make this a perfectly accurate model, but already better than the first one.

Exercise 9 :

You are confronted with three doors, A, B, and C. Behind exactly one of the doors there is $10,000. When you have pointed at a door, an official will open another door with nothing behind it. After he has done so, you are allowed to alter your choice. Should you do that (i.e., will altering your choice improve your chances of winning the prize)?

Solution: See the Hugin network.

Exercise 10 :

For 10000 emails in your inbox you determine the values of the following three boolean variables:

Spam	the email is spam
Caps	the subject line is in all capital letters
Pills	Body of the message contains the word "pills"

You obtain the following counts:

	Caps			
	yes		*no*	
	Pills		*Pills*	
Spam	*yes*	*no*	*yes*	*no*
yes	150	850	600	3400
no	1	99	49	4851

Are

- *Spam* and *Caps* independent?
- *Pills* and *Caps* independent?
- *Pills* and *Caps* independent given *Spam*?
- *Spam* and *Caps* independent given *Pills*?

Solution:

To determine whether *Spam* and *Caps* are independent, we consider the joint and marginal distribution of these two variables:

	Caps		
Spam	*yes*	*no*	
yes	0.1	0.4	0.5
no	0.01	0.49	0.5
	0.11	0.89	

The entries are obtained by summing over the *Pills* variable, and normalizing by dividing by 10000. E.g. $0.4 = \frac{600+3400}{10000}$. Since, for example $P(\textit{Spam} = \textit{no}, \textit{Caps} = \textit{yes}) = 0.01 \neq P(\textit{Spam} = \textit{no})P(\textit{Caps} = \textit{yes}) = 0.5 \cdot 0.11 = 0.055$, we see that the two variables are not independent.

To see whether *Spam* and *Caps* are independent given *Pills*, we first determine the conditional distribution of *Spam* and *Caps* given *Pills*=*yes*:

	Caps		
Spam	*yes*	*no*	
yes	0.1875	0.75	0.9375
no	0.00125	0.06125	0.0625
	0.18875	0.81125	

Here, e.g. $0.1875 = \frac{150}{800}$ (800 is the total number of cases with *Pills*=*yes*).

Again, we find that the joint distribution is not the product of the marginals, e.g. $0.0625 \cdot 0.18875 = 0.0118 \neq 0.00125$. Thus, *Spam* and *Caps* are not independent given *Pills*.

For the (conditional) independence of *Pills* and *Caps* one proceeds in the same manner. Now the findings should be: *Pills* and *Caps* are not independent, but *Pills* and *Caps* are independent given *Spam*. For the last result one has to check both the conditional distribution given *Spam=yes* and given *Spam=no* (above it was enough to consider the conditional distribution of *Spam* and *Caps* given *Pills=yes*, because there we already found that *Spam* and *Caps* are not conditionally independent).

Exercise 11 (Optional):

Use Hugin to solve this exercise The following relations hold for the Boolean variables A, B, C, D, E, and F:

$(A \vee \neg B \vee C) \wedge (B \vee C \vee \neg D) \wedge (\neg C \vee E \vee \neg F) \wedge (\neg A \vee D \vee F) \wedge$
$(A \vee B \vee \neg C) \wedge (\neg B \vee \neg C \vee D) \wedge (C \vee \neg E \vee \neg F) \wedge (A \vee \neg D \vee F)$.

(i) Is there a truth value assignment to the variables making the expression true? (Hint: Represent the expression as a Bayesian network.)

(ii) We receive the evidence that A is false and B is true. Is there a truth value assignment to the other variables making the expression true?

Solution:

(1) See the Hugin network here. As the probability of $Result = y$ is positive, there are assignments of truth values making the expression true.

(ii) Insert $A = n$ and $B = n$ as evidence and propagate. As $P(Result = y) > 0$, there are assignments of the remaining variables making the expression true. If you insert "$Result = y$" and propagate, you see that the assignments must be $C = y, D = y, E = y, F = y$.