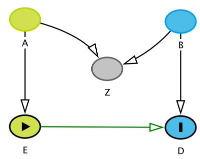
Radboud Universiteit

Exam



Probabilistic Graphical Models

January 24th, 2017

Johannes Textor, Perry Groot, Marcos Bueno
Institute for Computing and Information Sciences
Radboud Universiteit Nijmegen

Name ____________________

S-number ____________________

Question	1	2	3	4	5	6	7	Sum
maximum score	14	28	9	9	18	12	10	100
your score								

grade

This exam consists of 20 pages. Please verify the completeness of your copy.

Please note the following:

- You can write your answers in either English or Dutch language.
- You have 2.5 hours of time for completing this exam. Students who are eligible for extra time have 3 hours of time.
- Please think carefully before writing anything into the response boxes after each question. Should you require more space, please use the pages 17 to 20.
- This is an open-book exam. You are allowed to use any printed or hand-written material. The use of any electronic devices, such as smart-phones, is forbidden.

This page intentionally left blank.

Question 1 Bayesian networks basics, 14 points

Please cross out the box next to the correct statement or correct answer. Only one answer per question is correct.

(2 points) In the context of Bayesian networks, the abbreviation DAG stands for:

- ☐ Don't Actually Gnow
- ☐ Dirichlet-Anderson Gravity
- ☐ Directed Acyclic Graph

(2 points) A Bayesian network G encodes a factorization of a probability density in which each factor is a conditional probability density $p(y \mid \mathbf{X})$, where $\mathbf{X}$ is the set of

- ☐ children
- ☐ parents

of variable y in G.

(2 points) Each variable in a Bayesian network has at least one child.

- ☐ correct
- ☐ incorrect

(2 points) How does the number of terms in the factorization encoded by a Bayesian network change if we add an additional edge (arrow) to it?

- ☐ The number of terms increases.
- ☐ The number of terms decreases.
- ☐ The number of terms does not change.

(2 points) Suppose we have a Bayesian network containing n binary variables and an unknown amount of edges. Which of the three terms below is the smallest general upper bound on the number of parameters needed to encode the corresponding probability density factorization using conditional probability tables?

- ☐ $2+n$
- ☐ $2n$
- ☐ 2^n

(4 points) Consider the following Bayesian network:

$$
\begin{tikzcd}
 & W & \\
Z & & Y \arrow[ul] \\
 & X \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

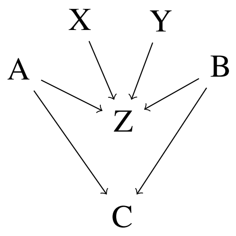
List all variables that match the description. (Remember that, by convention, each variable in a Bayesian network is a descendant and an ancestor of itself.)

Children of X:

Ancestors of Y:

Question 2 Binary Bayesian Networks, 28 points

A researcher has come up with the following Bayesian network for their problem domain. Throughout this question, we will refer to this network using the letter G.



(3 points) Assume that all variables are binary. How many parameters are needed to represent G using conditional probability tables? (In your answer, please mention how you encode each table.)

(3 points) How many parameters would you need for G if all variables were ternary instead of binary?

(6 points) List 6 different d-separation statements that are implied by G.

Consider the following partial table of a joint probability density on the variables A, B, C, X, Y, Z:

A	B	C	X	Y	Z	P
0	0	0	0	0	0	0.002
1	0	0	0	0	0	0.001
0	1	0	0	0	0	0.001
1	1	0	0	0	0	0.002
0	0	1	0	0	0	0.002
1	0	1	0	0	0	0.001

(2 points) Explain why this probability density is not consistent with the d-separation statement $A \perp\!\!\!\perp B \mid C, X, Y, Z$.

(2 points) Draw a Bayesian network that could have generated this probability table.

(3 points) Fill values into the last column of the probability table below that do not contradict consistency with the d-separation statement $A \perp\!\!\!\perp B \mid C, X, Y, Z$.

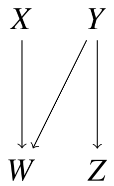
A	B	C	X	Y	Z	P
0	0	0	0	0	0	
1	0	0	0	0	0	
0	1	0	0	0	0	
1	1	0	0	0	0	
0	0	1	0	0	0	
1	0	1	0	0	0	

(6 points) Apply the "divorcing" technique to G (assuming binary variables) in order to reduce the number of parameters. Draw the modified network and determine the resulting number of parameters.

(3 points) Could you also use the "binning" technique to reduce the number of parameters? If so, explain how. If not, explain why not.

Question 3 Inference, 9 points

Consider the following Bayesian network G.

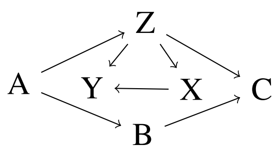


(3 points) Is this network a polytree? Explain.

(6 points) Draw the factor graph of G.

Question 4 Markov Equivalence, 9 points

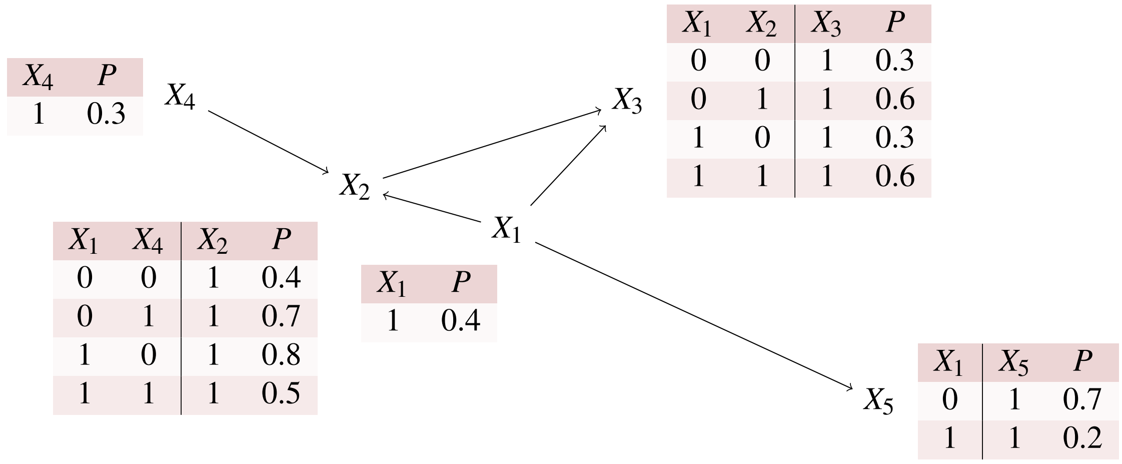
(6 points) Draw two different Bayesian networks that are Markov equivalent to the one below. Please clearly highlight the difference(s) to the original network.



(3 points) Draw a Bayesian network with 4 variables X, Y, Z, W whose Markov equivalence class contains only the network itself.

Question 5 Structure Learning, 18 points

Consider the following Bayesian network G.



X_4	P
1	0.3

X_1	X_4	X_2	P
0	0	1	0.4
0	1	1	0.7
1	0	1	0.8
1	1	1	0.5

X_1	P
1	0.4

X_1	X_2	X_3	P
0	0	1	0.3
0	1	1	0.6
1	0	1	0.3
1	1	1	0.6

X_1	X_5	P
0	1	0.7
1	1	0.2

(6 points) Explain why the graph of G is not faithful to the probability density p that it represents and draw a new graph that is faithful to p.

(12 points) Suppose we are given the following list of conditional independencies.

- $A \perp\!\!\!\perp Y \mid B, C, D$
- $B \perp\!\!\!\perp C \mid A$
- $B \perp\!\!\!\perp D \mid A$
- $C \perp\!\!\!\perp D \mid A$
- $X \perp\!\!\!\perp B \mid A$
- $X \perp\!\!\!\perp Y \mid B, C, D$
- $X \perp\!\!\!\perp Y \mid A, C, D$

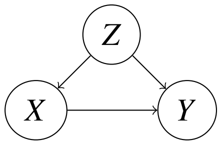
Perform the IC or PC algorithm on this list to infer a graph pattern from this list of conditional independencies. Explain which, if any, additional edges can be oriented after you have obtained your graph pattern.

Question 6 Causality, 12 points

Consider the following Bayesian network. The resulting joint distribution is given on the right-hand side.

Z	P
1	0.5

Z	X	P
0	1	0.1
1	1	0.5



Z	X	Y	P
0	0	1	0
0	1	1	0.5
1	0	1	0.5
1	1	1	1

Z	X	Y	P
0	0	0	0.45
0	0	1	0
0	1	0	0.025
0	1	1	0.025
1	0	0	0.125
1	0	1	0.125
1	1	0	0
1	1	1	0.25

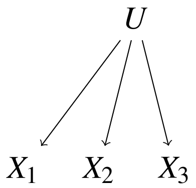
(3 points) Determine $p(Y = 1 \mid X = 1)$. (Show the calculation; you can give the result as a fraction of two numbers instead of a single number.)

(3 points) Determine $p(Y = 1 \mid \mathrm{do}(X = 1))$ (show the calculation). Is this smaller or larger than $p(Y = 1 \mid X = 1)$? How can you explain this difference?

(6 points) Determine $\sum_z p(Y = 1 \mid X = 1, z)p(z)$ (show the calculation). Explain why this gives the same result as $p(Y = 1 \mid \mathrm{do}(X = 1))$.

Question 7 Constraints with unobserved variables, 10 points

Consider the following Bayesian network G. All variables are assumed to be binary.



This network implies the constraint $X_i \perp\!\!\!\perp X_j \mid U$ for all $i \neq j \in \{1,2,3\}$. Therefore, not every probability density $p(x_1,x_2,x_3,u)$ is consistent to G. However, what happens if we treat U as an unobserved variable? Because we cannot condition on U anymore, the network then no longer implies any conditional independence. Does this mean that it can represent any probability density on the observed variables?

In more precise words, please determine whether the following statement is true:

For every probability density $p(x_1,x_2,x_3)$, there exists some other density $p'(x_1,x_2,x_3,u)$ that is consistent with G, such that $p(x_1,x_2,x_3) = \sum_u p'(x_1,x_2,x_3,u)$.

If the statement is true, prove it. If it is false, give a counterexample or explain how a counterexample could be constructed.

This page can be used for further writing.

This page can be used for further writing.

This page can be used for further writing.

This page can be used for further writing.