Exam Uncertainty In Artificial Intelligence

Name: ....

Student number: ....

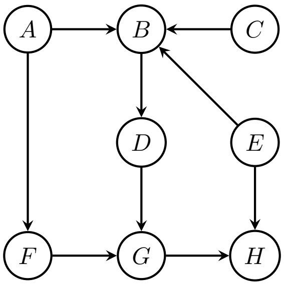
August 20, 2021

Instructions:

- Produce your answers on this exam copy. Use the available space (if needed you can use the back). Be crisp and to the point.
- Fill out your name and student number on EVERY page.
- Hand in ALL the pages of the exam, even if you did not fill on the questions on some pages.
- This is a closed book exam. You can only use the formularium without any additional notes. You may also use a calculator.
- For all questions with binary answers (YES/NO, TRUE/FALSE, . . .), points will be subtracted in case of a wrong answer. This is a 'correction for guessing'. Therefore, if you do not know the answer leave the question blank.
- The marks are indicated for each question. All questions together count for a total of 50 points.

Questions:

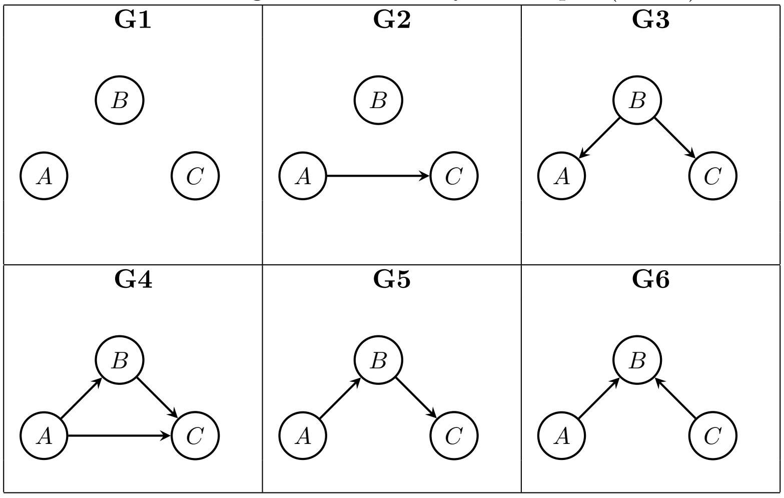
1. (2.5 pts) Independence
   Consider the Bayesian below. For each of the independency statements, judge the statement “The independency is encoded by the Bayesian network”. If it is not true, provide an unblocked/active path.



independency statement	**encoded by network**[1]	**active path if FALSE**
$A \perp\!\!\!\perp C$	**TRUE/FALSE**	
$A \perp\!\!\!\perp D \mid B, H$	**TRUE/FALSE**	A, F, G, H, G, D
$A \perp\!\!\!\perp E \mid F$	**TRUE/FALSE**	
$G \perp\!\!\!\perp E \mid B$	**TRUE/FALSE**	G, F, A, B, E
$F \perp\!\!\!\perp C \mid D$	**TRUE/FALSE**	F, A, B, O, C

[1]For all questions with binary answers (YES/NO, TRUE/FALSE, . . .), points will be subtracted in case of a wrong answer. This is a ‘correction for guessing’. Therefore, if you do not know the answer leave the question blank.

2. (3 pts) Maps
Consider the following six Directed Acyclic Graphs (DAGs):

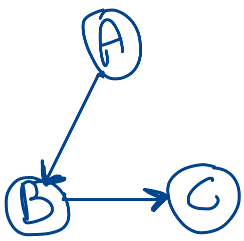


Below, consider the factorization of the joint distribution $P(A, B, C)$. Indicate for each of the six DAGs if the DAG is a perfect map, independence map, and/or dependence map for any distribution that can be factorized according the provided factorization.

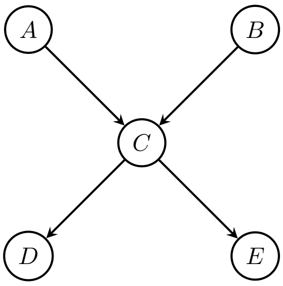
Factorization: $P(C|B)P(B|A)P(A)$

	G1	**G2**	**G3**	**G4**	**G5**	**G6**
perfect map:	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**
independence map:	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**
dependence map:	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**	**YES/NO**

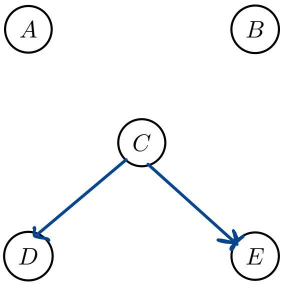
$A \perp\!\!\!\perp C \mid B$



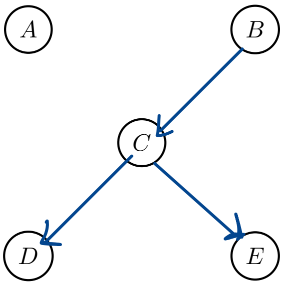
3. (4.5 pts) Independencies
Consider a Bayesian network over the random variables A, B, D, D, and E with the structure shown below, with full joint distribution $P(A, B, C, D, E)$.



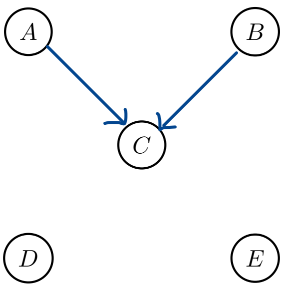
(a) On the diagram below, add arrows between nodes such that the resulting Bayesian network most faithfully encodes the independences of $P(A, B, D, E) = \sum_C P(A, B, C, D, E)$
If no arrows are needed write "no arrows needed". So I think we want an independence map



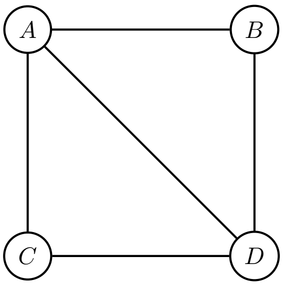
(b) On the diagram below, add arrows between nodes such that the resulting Bayesian network most faithfully encodes the independences of $P(B, C, D, E|A = a)$.
If no arrows are needed write "no arrows needed".



(c) On the diagram below, add arrows between nodes such that the resulting Bayesian network most faithfully encodes the independences of $P(A, B, C|D = d, E = e)$.
If no arrows are needed write "no arrows needed".



4. (3 pts) Proving
Consider the following Markov Network with random variables A, B, C, D and with the structure below, with full joint distribution $P(A, B, C, D)$.



Prove that B and C are conditionally independent given A and D by using the factorization and the rules of probability only.

Answer:

$$P(A,B,C,D) = \frac{1}{Z}\,\phi(A,D,C)\,\phi(A,B,D)$$

$$= P(A)P(C|A,D)P(D|A)P(B|A,D)$$ using factorization of markov network

$$= P(A)P(D|A)\underline{P(C|A,D)P(B|A,D)}$$

$$P(A,B,C,D) = P(A)P(D|A)\underline{P(B,C|A,D)}$$ chain rule

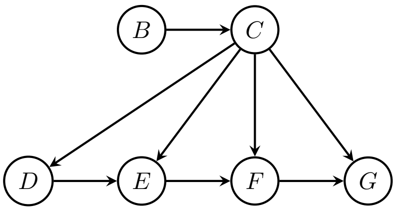
which means that

$$P(C,B|A,D) = P(C|A,D)P(B|A,D)$$

$$\Rightarrow$$

$$C \perp\!\!\!\perp B \mid A,D$$

5. (5 pts) Inference 1:
Consider the Bayesian network with binary variables drawn below.



$P(B,C,D,E,F,G)$
$= P(B)P(C|B)P(D|C)\,P(E|D,C)P(F|E,C)$
$P(G|F,C)$

- (3 pts) Use bucket elimination to answer the query $P(B, D|F = 1)$. Complete the scheme below with the given order.

variable	bucket	message sent (+indicate with an arrow where the message is sent to)
C	$P(C|B)P(D|C)P(E|D,C)$ $P(F|E,C)P(G|F,C)$	$\gamma_C(B,D,E,F,G) =$ $P(C|B)P(D|C)P(E|D,C)P(F|E,C)P(G|F,C)$
E	$\gamma_C(B,D,E,F,G)$	$\gamma_E(B,D,F,G) =$ $\sum_E \gamma_C(B,D,E,F,G)$
G	$\gamma_E(B,D,F,G)$	$\gamma_G(B,D,F) = \sum_G \gamma_E(B,D,F,G)$

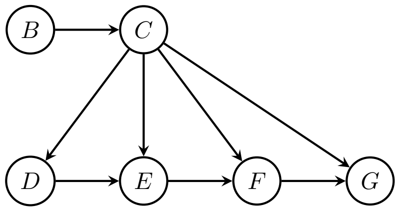
- (1 pt) Now show how to compute $P(B, D|F = 1)$ from this.

$$P(B,D|F=1) = \frac{\sum_{CEG} P(B,C,D,E,F,G)}{\sum_{CEGBD} P(B,C,D,E,F,G)} = \frac{\gamma_G(B,D,F)}{\sum_{B,D}\gamma_G(B,D,F)}$$

- (1 pt) Is there a more efficient order possible? Yes/No, circle your answer. If yes, specify that order and shortly explain why this order is more efficient.

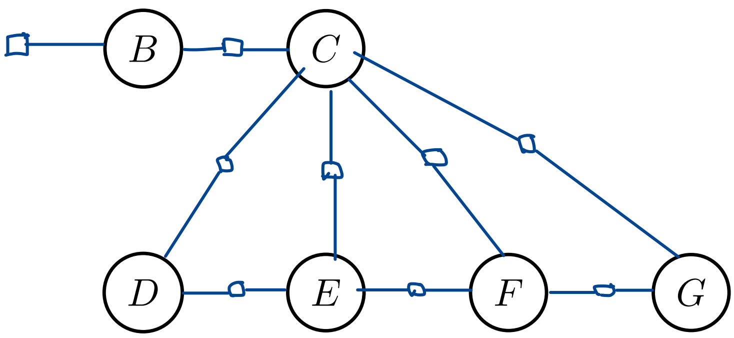
Yes/No	More efficient order: G, E, C

6. (5 pts) Inference 2:
Now let's use message passing on the Bayesian Network of the previous question, repeated below.



- Answer the questions below considering you would use belief propagation and possibly loop cut conditioning to answer $P(G|F=1)$. Indicate if you need loop cut yes/no, and if yes which variables are in the loop cut. Draw the factor graph both before and after loop cut conditioning (if necessary), clearly define all factors, define the root, define the message order (by writing numbering the arrows in the factor graph used for message passing). Try to make the procedure as efficient as possible. There is no need to elaborate the other steps typically involved in message passing.

Factor graph before possible loop cut + define factors!:



Loop cut needed:
Yes / No

If loop cut needed: variables for most efficient cut set:

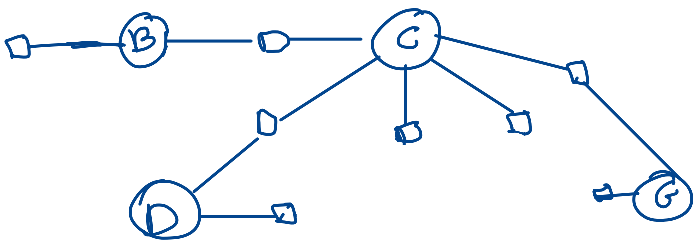
$\{E\}$

↳ not F since already in evidence

Root:

G

Factor graph after possible loop cut + define factors:



Wondering how this works then

7. (6 pts) Probabilistic reasoning, modelling, and learning
Consider the problem of predicting a coin flip on the basis of several previous observations. Imagine that you observed three coin flips, and they all came up heads. Make a prediction about the probability that the fourth coin flip will be heads as well.

(a) (1pt) Assume that you know the coin is fair (probability that outcome is heads is equal to probability that outcome is tails). What is the probability that the fourth coin flip will be heads, given the three heads observations?

Answer: $1/2$

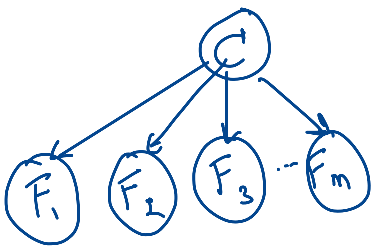
(b) (1pt) Now assume that you had no prior knowledge on what is on the sides of the coin before you flipped it. What is the probability that the fourth coin flip will be heads, given the three heads observations?

Answer: 1

ML estimation

(c) (4 pts) Now assume that you know your coin is one of three types: fair, two-headed (they always come up heads), and two-tailed (they always come up tails). Moreover, assume that you have a prior belief or knowledge that these three types show up with probability 0.5 for fair coins, and two-headed or two-tailed with probability 0.25 each.

i. (2 pts) Define the Bayesian network (graph + probability tables) that represents this knowledge of the coin and the four coin flips.



C	P(C)
fair	1/2
2t	1/4
2h	1/4

$P(H_i \mid C)$ H_i	f	2h	2t
0	1/2	0	1
1	1/2	1	0

ii. (1 pt) What is the posterior probability that the coin is fair given the three heads observations?

Answer: $1/5$

$$P(C=\text{fair} \mid F_1=H, F_2=H, F_3=H, F_4=H)$$

$$= \frac{P(F_1=H, F_2, F_3, F_4 \mid C=\text{fair})\, P(C=\text{fair})}{\sum_c \dots} = P(F_1=H \mid C=\text{fair}) \cdot \dots$$

$$= \frac{1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2}{(1/2)^4 + 0 + (1 \cdot 1/4)} = \frac{(1/2)^4}{(1/2)^4 + 1/4} = 1/5$$

iii. (1 pt) What is the probability that the fourth coin flip will be heads?

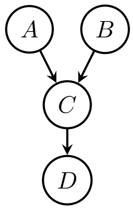
Answer: $9/10$

$$P(H_4 = H \mid H_1 = H, H_2 = H, H_3 = H) = \sum_c P(H_4 = H \mid c) P(C \mid D)$$

$$= \left(\frac{1}{5} \cdot \frac{1}{2}\right) + \left(0 \right) + \left(0{,}8 \cdot 1 \right)$$

$$= 0{,}9 = 9/10$$

8. (6 pts) **Learning.** Consider the following Bayesian network with binary variables A, B, C, and D.



A	B	C	D
1	0	1	1
0	1	0	?
?	1	0	0
0	1	1	1
1	1	0	0
0	1	1	1
1	0	1	1
1	1	1	1
1	1	0	0

$P(A)P(B)P(C|A,B)P(D|C)$

(a) (2 pts) How many parameters are needed fully define the CPTs of the Bayesian network? Define these parameters. Define them as follows: $\theta_{X10} = p(X = 1|Y = 1, Z = 0)$, $\theta_{Y0} = p(Y = 1|Z = 0)$, and $\theta_Z = p(IZ1)$, and use an alphabetic order for the variables in the conditioning set.

Number of parameters = $1 + 1 + 4 + 2 = 8$

Definition parameters:

$\theta_A = P(A)$

$\theta_B = P(B)$

$\theta_{C00} = P(C|A=0, B=0)$

$\theta_{C01} = P(C|A=0, B=1)$

$\theta_{C10} = P(C|A=1, B=0)$

$\theta_{C11} = P(C|A=1, B=1)$

$\theta_{D0} = P(D|C=0)$

$\theta_{D1} = P(D|C=1)$

(b) (4 pts) Consider now the dataset with missing observations. Assume all network probabilities are initialized to take the value 1 with probability 0.7. Use Expectation-Maximization (EM) to estimate $P(C = 1|A = 1, B = 1)$. Peform ONE iteration of EM. Use the tables below for your answers. For the E-step fill in the completed data table, where the completed values are printed bold. For the M-step provide formula and value for the first iteration.

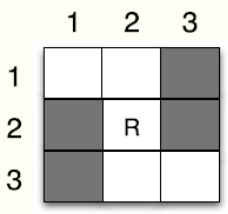
E-step: Completed data table: (fill only for the examples you need to compute the new estimate)

	A	B	C	D	**weight formula** using parameters of question a	**weighted value** 1st iteration
	1	0	1	1		
	0	1	0	**1**	$q^2(D = 1) =$	
	0	1	0	**0**	$q^2(D = 0) =$	
x	**1**	1	0	0	$q^3(A = 1) = P(A=1 \mid B=1, C=0, D=0) = \frac{P(A=1, B=1, C=0, D=0)}{P(A=0, B=1, C=0, D=0) + P(A=1, B=1, C=0, D=0)}$ $= \frac{P(A=1)\cancel{P(B=1)}P(C=0 \mid A=1, B=1)P(D=0 \mid A=1)}{\dots + P(A=0)\cancel{P(B=1)}P(C=0 \mid A=0, B=1)P(D=0 \mid A=0)}$	$\frac{0{,}7 \cdot 0{,}3 \cdot 0{,}7}{0{,}7 \cdot 0{,}3 \cdot 0{,}7 + 0{,}3 \cdot 0{,}3 \cdot 0{,}7}$
	0	1	0	0	$q^3(A = 0) =$ $= \frac{\theta_A (1-\theta_{C11})\theta_{D1}}{\theta_A(1-\theta_{C11})\theta_{D1} + (1-\theta_A)(1-\theta_{C01})\theta_{D0}}$	$= 0{,}7$
	0	1	1	1		
x	1	1	0	0		
	0	1	1	1		
	1	0	1	1		
x	1	1	1	1		
x	1	1	0	0		

M-step:

formula for M-step using q's of E-step	**value** 1st iteration
$P(C = 1 \mid A = 1, B = 1) = \frac{1}{3 + q^3(A=1)}$	$\frac{1}{3 + 0{,}7}$ $= 0{,}27$

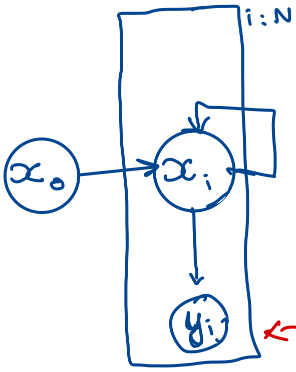
9. (6 pts) Hidden Markov Model



Suppose a robot (R) navigates a maze with 9 grid cells (see picture), where some of the cells are free (white) and some are blocked (grey). At each time step, the robot is occupying one of the free cells. The robot is equipped with sensors that provide noisy observations on whether the neighbouring cells or blocked (grey) or not (white): (y_U, y_D, y_L, y_R) of the four cells adjacent to the robot's current position (U=up, D=down, L=left, and R=right). An observation outside the grid can be considered equal to an observation of a blocked cell (grey). Each observation y_i is accurate 80% of the time. Imagine that at each time step the robot moves randomly to the up, down, left, or right. If the robot tries to move to an adjacent blocked cell, it will fail to do so and stay in its current position.

(a) (2 pts) Draw the Hidden Markov Model corresponding to the problem description and define the probability tables associated with this network **where the middle cell position (2,2) is involved.**.

HMM:



N = amount of steps taken

incorrect, we need 1 evidence per cell

Definition of probability tables with middle cell position (2,2) involved:

x_{i-1}	$P(x_i=(1,1) \mid x_{i-1})$
(1,1)	0
(1,2)	1/4
(2,2)	2/4
(3,2)	1/4
(3,3)	0

x_{i+1}	$P(x_{i+1} \mid x_i=(2,2))$
(1,1)	0
(1,2)	1/4
(2,2)	2/4
(3,2)	1/4
(3,3)	0

$P(y_i=(w,w,G,G) \mid x_i=(2,2)) = 0,8$

$P(\neg y_i=(w,w,G,G) \mid x_i=(2,2)) = 0,2$

(b) (1 pt) Suppose you start in the central cell (2,2). One time step passes and you are now in a new, possibly different state and your sensors indicate (free,blocked,blocked,blocked). Which states have a non-zero probability of being your new position?

Answer: (2,2) , (1,2), (3,2)

(c) (1 pt) Calculate the posterior probability distribution over your new position after one time step (situation of previous subquestion).

Answer:

$P((2,2) \mid \text{free}, \text{blocked}, \text{blocked}, \text{blocked}, (2,2)) = \underline{P(\vec{v} \mid (2,2), (2,2))\, P((2,2) \mid (2,2))}$

$= \dfrac{0{,}8 \cdot 0{,}2 \cdot 0{,}8 \cdot 0{,}8 \cdot 0{,}5}{Z} = \dfrac{32}{49}$

$Z = (0{,}8 \cdot 0{,}2 \cdot 0{,}8 \cdot 0{,}8 \cdot 0{,}5) + (0{,}2 \cdot 0{,}2 \cdot 0{,}2 \cdot 0{,}8) + (0{,}8 \cdot 0{,}8 \cdot 0{,}8 \cdot 0{,}2)$

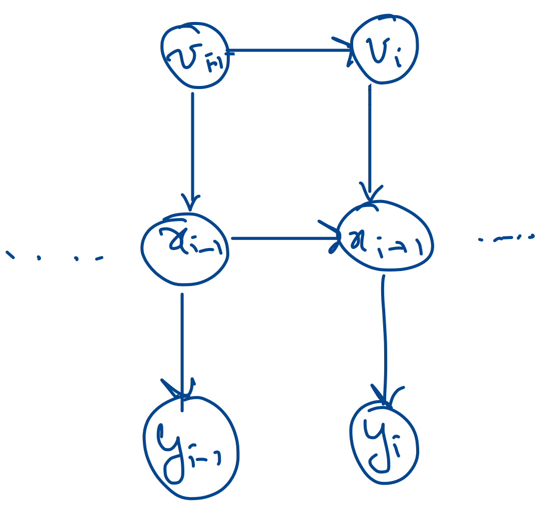
$= 0{,}1552$

(d) (1 pt) Which algorithms that we learned in our course could be used to estimate the most likely position of the robot after five time steps and five observations? Which one would you use and why?

Answer: filtering, discrete dynamical model
=> forward algorithm

what we have also seen: general Bayes filter, Kalman filter, particle filter
but that's for continuous dynamical models

(e) (1 pt) Now suppose the robot does not move randomly but is programmed to have a more fluent behaviour that ensures there is a higher probability that the robot keeps on moving in the same direction that it was already heading. Encode this knowledge in the directed acyclic graph (no need to try to define the probability tables, just give the graph).



10. (3pts) ProbLog: Consider the following ProbLog program:

```
0.5:: q.
0.3:: t1.
0.3:: t2.

p :- q.
p :- s.
s :- t1.
s :- t2.
```

(a) What is the probability of `p` ?

Answer: 0,9

$P(p) = P(q) + P(t_1) + P(t_2) - P(q \wedge t_1) - P(q \wedge t_2) - P(t_1 \wedge t_2)$
$+ P(q \wedge t_1 \wedge t_2)$

$= 0,5 + 0,3 + 0,2 - 0,15 - 0,15 - 0,25 + 0,45$

$= 0,9$

(b) What is the probability of `P(p | t1)` ?

Answer: 1

(c) What is the probability of `P(p | ¬t1)` ?

Answer: 0,65

$P(p \mid \neg t_1) = P(q) + P(t_1) - P(q \wedge t_1)$

$= 0,5 + 0,3 - (0,5 \cdot 0,3)$

$= 0,65$

11. (6 pts) Sampling: Consider the following Bayesian network with binary variables.

$$A \rightarrow B \rightarrow C \rightarrow D$$

(a) (3 pts) Assume you want to obtain, as efficient as possible, ONE sample for each of the queries below using **ancestral sampling**, assuming you have a list of random and independent samples from a uniform distribution between 0 and 1 available.

Indicate for each of the queries a) if you would need rejection or not, b) list the (un)conditional probabilities available in the network that have the be sampled from, in the order they have to be sampled from, and c) order the queries according to efficiency (assuming each (un)conditional probability of the network is equally efficient to sample from).

query	**rejection needed: Yes/No**	**ordered list of (un)conditional network probabilities**	**computational efficiency (1=most efficient, 3=least efficient)**
$p(C)$	No	P(A), P(B|A), P(C|B)	2
$p(C\|B)$	Yes	P(C|B)	1
$p(C\|D)$	No	P(A), P(B|A), P(C|B), P(D|C)	3

(b) (1.5 pt) Assume you have the 11 samples available given in the table below that were generated by ancestral sampling on the given Bayesian network (without evidence). Use these samples to answer the queries below.

sample	A	B	C	D
sample 1	0	1	0	0
sample 2	1	1	0	1
sample 3	1	1	0	1
sample 4	1	0	1	0
sample 5	0	0	0	1
sample 6	1	0	0	0
sample 7	0	1	1	0
sample 8	0	0	0	1
sample 9	0	0	0	1
sample 10	1	1	1	1
sample 11	0	1	0	0

query	answer to query
$p(C)$	2/11
$p(C = 1 \mid B = 1)$	2/6
$p(C = 1 \mid D = 1)$	1/6

(c) (1.5 pt) Indicate for each of the sets of three subsequent samples below if they could have been generated by performing Gibbs sampling on the given Bayesian network?

sequence	could have been generated by Gibbs sampling
$\text{sample}_i = (A = 1, B = 0, C = 0, D = 1)$ $\text{sample}_{i+1} = (A = 1, B = 0, C = 0, D = 1)$ $\text{sample}_{i+2} = (A = 1, B = 0, C = 1, D = 1)$	Yes/No
$\text{sample}_i = (A = 1, B = 0, C = 0, D = 1)$ $\text{sample}_{i+1} = (A = 1, B = 0, C = 0, D = 0)$ $\text{sample}_{i+2} = (A = 0, B = 0, C = 0, D = 1)$	Yes/No
$\text{sample}_i = (A = 1, B = 0, C = 0, D = 1)$ $\text{sample}_{i+1} = (A = 1, B = 0, C = 0, D = 0)$ $\text{sample}_{i+2} = (A = 1, B = 0, C = 0, D = 0)$	Yes/No