NATIONAL UNIVERSITY OF SINGAPORE

CS5340 - Uncertainty Modelling in AI

(Quiz 1, Semester 2 AY2020/21)

SOLUTIONS

Time Allowed: 1 hour

Instructions

- This is an open-book quiz. You may refer to any of the lecture slides and tutorials.
- You may *not* refer to any external online material or use any software to help you answer the questions.
- Please do not cheat; your answers *must* be your own. Do *not* collaborate with anyone else.
- Please put all your answers in Luminus.
- Read each question *carefully*. Don't get stuck on any one problem. The questions are *not* in any particular order of difficulty.
- Don't panic. The problems often look more difficult than they really are.
- Good luck!

Student Number.: ______________________________

Common Probability Distributions

Distribution (Parameters)	PDF/PMF
Normal (μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
Bernoulli (r)	$r^x(1-r)^{(1-x)}$
Categorical (π)	$\prod_{k=1}^{K} \pi_k^{x_k}$
Binomial (μ, N)	$\binom{N}{x}\mu^x(1-\mu)^{N-x}$
Poisson (λ)	$\frac{\lambda^x \exp[-\lambda]}{x!}$
Beta (α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
Gamma (a, b)	$\frac{1}{\Gamma(a)} b^a x^{a-1} \exp[-bx]$
Dirichlet $(\boldsymbol{\alpha})$	$\frac{\Gamma(\sum_k^K \alpha_k)}{\Gamma(\alpha_1)...\Gamma(\alpha_K)} \prod_{k=1}^{K} x_k^{\alpha_k-1}$
Multivariate Normal $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\frac{1}{(2\pi)^{D/2}\|\Sigma\|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$
Uniform (a, b)	$\frac{1}{b-a}$

Note: $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ is the Gamma function.

1 True or False

For the following questions, please answer TRUE or FALSE. **Justifications are optional**.

Grading Policy: 1 point for each correct answer. For problem 10, we give everyone a point as the question is slightly ambiguous and the justifications provided by some students are valid.

Problem 1. [1 points] Epistemic uncertainty exists as a "natural" part of the system and cannot be decreased even if as more data is collected. A good example of epistemic uncertainty is whether it will rain today.

Solution: False. *Aleatoric uncertainty* exists as a "natural" part of the system and cannot be decreased even if more data was collected.

Problem 2. [1 points] Given two random variables X and Y, the mean of $X + Y$ is always equal to the sum of the mean of X and the mean of Y, i.e.,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Solution: True. This follows from the linearity of expectations.

Problem 3. [1 points] Given two random variables X and Y, the variance of $X + Y$ is always equal to the sum of the variance of X and the variance of Y, i.e.,

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$$

Solution: False. Refer to Tutorial 1 Problem 4.

Problem 4. [1 points] Given two independent random variables X and Y, the mean of $X \times Y$ is always equal to the product of the mean of X and the mean of Y, i.e.,

$$\mathbb{E}[X \times Y] = \mathbb{E}[X] \times \mathbb{E}[Y]$$

Solution: True.

Problem 5. [1 points] Two random variables X and Y are independent *if and only if* the covariance between them is zero, $\mathrm{Cov}[X, Y] = 0$.

Solution: False. This was a tutorial question.

Problem 6. [1 points] Given an arbitrary distribution P, we can always find a Bayesian network that captures all the conditional independencies in P.

Solution: False.

Problem 7. [1 points] Given an arbitrary Bayesian network graph (the DAG only) G, we can always find a distribution P that captures all the conditional independencies in G.

Solution: True.

Problem 8. [1 points] If there is a directed arrow from nodes X and Y in a Bayesian Network, then X causes Y.

Solution: False.

Problem 9. [1 points] Two random variables X and Y are independent if the random variable $X \times Y$ follows the distribution $p(X)p(Y)$.

Solution: False. X and Y are independent if the random variable $p(X, Y) = p(X)p(Y)$.

Problem 10. [1 points] To convert a DGM into a MRF, we perform moralization; for each node, we add an edge between its parents, and then replace the directed edge with an undirected edge.

Solution: True.

Problem 11. [1 points] Every Directed Graphical Model G can be converted into an Undirected Graphical Model H whilst preserving all the conditional independence assertions in G, i.e., $I(H) = I(G)$ where $I(A)$ is the set of conditional independence assertions associated with the graph A.

Solution: False. We have seen in class some couterexamples.

2 The Right Distribution

For each of the following subproblems, pick the *best* probability distribution to model the stated random variable among the choices given. **Provide a brief justification**. If multiple distributions are equally well-suited, pick any one of them.

Please refer to the online quiz for the specific questions.

Grading Policy: 2 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 0.5 point. If answer is correct, but justification is only partially correct, you get 1.5. If answer is wrong, but justification is reasonable(or partially correct), then you get 1 point. If answer is wrong, but justification is completely correct, then you get 1.5 points.

Problem 12. [2 points] You want to model the total amount of rainfall in Singapore on any given day.

A. Beta
B. Bernoulli
C. Gaussian
D. Gamma

Solution: Beta and Bernoulli are immediately out of question because they have support over $[0, 1]$ and $\{0, 1\}$ respectively. Among the choices, Gamma would be the best. Gaussian is not very appropriate here since the amount of rainfall should not be negative and a Gaussian distribution is symmetric and cannot account for skewness.

Problem 13. [2 points] You are designing an app for predicting travel times. What distribution is appropriate for capturing the number of busses arriving at COM2 every hour? You can assume that the arrival times between busses are independent of one another and the mean rate of bus arrivals is constant.

A. Poisson
B. Dirichlet
C. Exponential
D. Gaussian

Solution: The Poisson is an appropriate distribution to model the number of bus events per hour. All the other options have supports that are not suitable for the problem: They are all continuous distributions. Gaussian has negative support too and Dirichlet has support over a probability simplex.
Grading Policy: if answer is correct, but justification is trivial, you get 1 point for this question.

Problem 14. [2 points] You are an infectious disease expert interested in modeling the number of people that will be infected by each individual with COVID-19. Let's represent this random variable by X and you've assumed that X Bernoulli(n, p). Which of the following would be a good prior distribution for p?

A. Beta
B. Gamma
C. Gaussian
D. Poisson

Solution: A Beta distribution is most appropriate here given its support over $[0, 1]$ and it is conjugate to the Bernoulli.

Problem 15. [2 points] You work for a famous ride-hailing company called Super. The Super CEO wants you to model how long Super customers have to wait for their rides to arrive. You have a limited amount of data of waiting times (in seconds), which you can use to fit your parameters. Which of the following distributions would you pick?

A. Beta
B. Gamma
C. Gaussian
D. Categorical

Solution: Among the choices, the Gamma distribution would likely work best. The Categorical is a discrete distribution. Beta only supports over $[0, 1]$, and Gaussian has negative support.

Problem 16. [2 points] Bitcash's price continues to rise and you want to get in on the action. You've decided to model the inter-day price differences, i.e., the difference in the end-of-day/closing price from one day to the next. Which of the following distributions could you use?

A. Dirichlet
B. Poisson
C. Gamma
D. Gaussian

Solution: Among the choices, the Gaussian would work best here. The others are inappropriate given you are modeling the price differences (support over the entire real line).

3 Donald's Decisions

Donald's data analysis company—Data Unification and Contrived Knowledge (DUCK)—is a new startup. After finishing CS5340, they've recruited you to come in as their resident UAI expert.

Note: The questions below are intentionally open-ended. You will be graded on the *reasonableness* of your answers.

Please refer to the online quiz for the specific questions.

Problem 17. [4 points] DUCK has about a million Birdcoins. One of the DUCK analysts, Daphne, has been modeling the price of Birdcoin. From past data, she notices that price of Birdcoin B is positively correlated with the price of Catcoin C — in fact, the relationship appears to be $B = 10C + \epsilon$ where $\epsilon \sim \text{Normal}(0, 0.1)$. For example, when the price of Catcoin was S$0.10, the price of Birdcoin was S$1.00. The current price of Catcoin is S$1.00, and the price of Birdcoin is S$9.88.

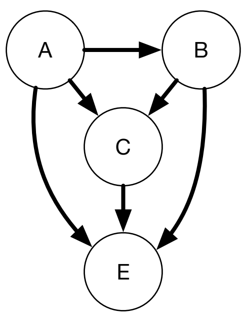
She's presented her results to Donald, who said, "*I have a GREAT idea! I'll buy lots of Catcoins to drive up its price! Since the price of Birdcoin will then rise by 10 times, we'll all be RICH!*"

Do you think Donald's idea is great? *State Yes or No and give a justification.*

Grading Policy: 4 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 1 point. If answer is correct, but justification is only partially correct, you get 3. If answer is wrong, but justification is completely correct, then you get 3 points. In this problem, your justifications are not required to be exactly the same as solution (e.g. some justifications like 'distribution of testing data may not be the same as training data' are regarded as correct) **Solution:** Not necessarily since the C and B may not be causally related.

Problem 18. [4 points]

One of the DUCK analysts, Mick, has been modeling how people make purchasing decisions. He has come up with the following Bayesian network structure:



where E is someone's purchasing choice, and the remaining variables are influencing factors (e.g., income level, age group). All the random variables are discrete and modeled by categorical distributions. Mick claims his model is simple and has sufficient conditional independence assumptions to ensure it doesn't overfit to data.

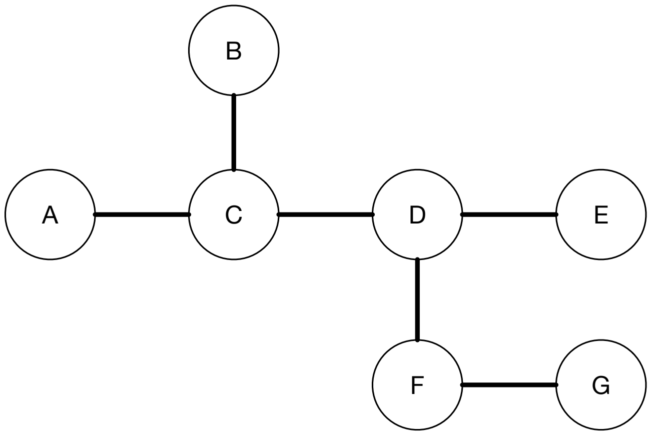
Is Mick correct about his Bayesian Network structure?.

Grading Policy: 4 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 1 point. If answer is correct, but justification is only partially correct, you get 3. If answer is wrong, but justification is completely correct, then you get 3 points. If your justifications are not relevant to question-'Mick claims his model is simple and has sufficient conditional independence assumptions to ensure it doesn't overfit to data' (e.g. Mick has not considered unobserved variables; Current purchasing may affect future, the model does not consider time evolution), they will be regarded as complete wrong justifications.

Solution: No, the graph is a fully connected network that has no conditional independencies.

4 The Most Likely Assignment

Your goal in this problem is to derive the most likely assignment to the variables given specific observations, i.e., the maximum a posteriori assignment of variables. Suppose you are given the following pairwise Markov Random Field:



which factorizes as:

$$p(A, B, \ldots, G) \propto p(A, C)p(B, C)p(C, D)p(D, E)p(D, F)p(F, G)$$

where each the following pairwise factors is:

x_i	x_j	$\psi(x_i, x_j)$
0	0	10
0	1	1
1	0	2
1	1	7

There are no other factors. For each of the following, compute the most likely assignment given the stated observations. If there is more than one MAP assignment, then please state just one of them.

Please refer to the online quiz for the specific questions.

Grading Policy: 2 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 0.5 points. If answer is correct, but justification is only partially correct, you get 1.5. If answer is wrong, but justification is reasonable(or partially correct), then you get 1 point. If your answer is partially correct, but justification is completely wrong or blank, you get 0 points. If your answer is wrong, but you answer with the correct potential value, and justification is completely correct, then you get 1.5 points.

Problem 19. [2 points]

You observe $C = 0$ and $D = 1$ and $F = 1$. What is the most likely assignment to all the other variables?

Solution: Since A is independent to other nodes given C

$$A = \text{argmax}_A \psi(A, C = 0) = 0 \tag{1}$$

Since B is independent to other nodes given C

$$B = \text{argmax}_B \psi(B, C = 0) = 0 \tag{2}$$

Since E is independent to other nodes given D

$$E = \text{argmax}_E \psi(D = 1, E) = 1 \tag{3}$$

Since G is independent to other nodes given F

$$G = \text{argmax}_G \psi(F = 1, G) = 1 \tag{4}$$

Problem 20. [2 points]

You observe $C = 1$ and $D = 0$ and $F = 1$. What is the most likely assignment to all the other variables?

Solution: Since A is independent to other nodes given C

$$A = \text{argmax}_A \psi(A, C = 1) = 1 \tag{5}$$

Since B is independent to other nodes given C

$$B = \text{argmax}_B \psi(B, C = 1) = 1 \tag{6}$$

Since E is independent to other nodes given D

$$E = \text{argmax}_E \psi(D = 0, E) = 0 \tag{7}$$

Since G is independent to other nodes given F

$$G = \text{argmax}_G \psi(F = 1, G) = 1 \tag{8}$$

Problem 21. [2 points]

You observe $A = 1, B = 0, C = 1, E = 1, F = 1$. What is the most likely assignment to all the other variables?

Solution: Since D is independent to other nodes given C, E, F

$$D = \text{argmax}_D \psi(D, F = 1)\psi(D, E = 1)\psi(C = 1, D) = 1 \tag{9}$$

Since G is independent to other nodes given F

$$D = \text{argmax}_D \psi(F = 1, G) = 1 \tag{10}$$

Problem 22. [2 points]

You observe $C = 1$. What is the most likely assignment for node A.

Solution: Since A is independent to other nodes given C

$$A = \text{argmax}_A \psi(A, C = 1) = 1 \tag{11}$$

Problem 23. [2 points]

You observe $F = 1$. What is the most likely assignment for node G.

Solution: Since G is independent to other nodes given F

$$G = \text{argmax}_G \psi(F = 1, G) = 1 \tag{12}$$

Problem 24. [2 points]

You observe $C = 0, D = 1, E = 0, F = 0$. What is the most likely assignment for nodes B.

Solution: Since B is independent to other nodes given C.

$$B = \text{argmax}_B \psi(B, C = 0) = 0 \tag{13}$$

Problem 25. [2 points]

You observe $B = 1$ and $D = 0$. What is the most likely assignment for node A.

Solution: There are two interpretations, the first is we choose A that maximizes the marginal distribution $p(A|B = 1, D = 0)$ and the second is that we jointly choose A, C that maximizes $p(A, C|B = 1, D = 0)$ (since other nodes are independent to A, C given D, so we just need to consider A, C). Both interpretations are correct.
We first look at the first interpretation.

$$\begin{aligned} A &= \text{argmax}_A p(A|B = 1, D = 0) && (14) \\ &= \text{argmax}_A \sum_C \psi(B = 1, C) \cdot \psi(C, D = 0) \cdot \psi(A, C) && (15) \end{aligned}$$

if $A = 0$

$$\begin{aligned} \sum_C \psi(B = 1, C) \cdot \psi(C, D = 0) \cdot \psi(A, C) &= \sum_C \psi(B = 1, C) \cdot \psi(C, D = 0) \cdot \psi(A = 0, C) && (16) \\ &= 2 \cdot 10 \cdot 10 + 7 \cdot 2 \cdot 1 && (17) \\ &= 200 + 14 = 214 && (18) \end{aligned}$$

if $A = 1$

$$\begin{aligned} \sum_C \psi(B = 1, C) \cdot \psi(C, D = 0) \cdot \psi(A, C) &= \sum_C \psi(B = 1, C) \cdot \psi(C, D = 0) \cdot \psi(A = 1, C) && (19) \\ &= 2 \cdot 10 \cdot 2 + 7 \cdot 2 \cdot 7 && (20) \\ &= 40 + 98 = 138 && (21) \end{aligned}$$

Since $214 > 138$, $A = \text{argmax}_A p(A|B = 1, D = 0) = 0$
Then, we look at the second interpretation,

$$A, C = \text{argmax}_{A,C} p(A, C|B = 1, D = 0) \tag{22}$$

Message from B to C:

$$m_{BC}(C = 0) = 2$$

$$m_{BC}(C = 1) = 7$$

.
Message from D to C

$$m_{DC}(C = 0) = 10$$

$$m_{DC}(C = 1) = 2$$

.
Message from C to A:

$$m_{CA}(A = 0) = \max_C \psi(A = 0, C) m_{BC}(C) m_{DC}(C) = 200$$

$$m_{CA}(A = 1) = \max_C \psi(A = 1, C) m_{BC}(C) m_{DC}(C) = 98$$

Since $200 > 98$, MAP is $A = 0$.

Problem 26. [2 points]

You observe $D = 1$ and $F = 0$. What is the most likely assignment for nodes A and E.

Solution: There are two interpretations, the first is we choose A, E that maximizes the marginal distribution $p(A, E|D = 1, F = 0)$ and the second is that we jointly choose A, B, C, E that maximizes $p(A, E|D = 1, F = 0)$ (since other nodes are independent to A, B, C, E given D, so we just need to consider A, B, C, E). Both interpretations are correct.
We first look at the first interpretation.
Since A and E are conditional independent given D, we can look at them separately.

$$\begin{aligned} E &= \text{argmax}_A p(E|D = 1) && (23)\\ &= \text{argmax}_A \psi(D = 1, E) = 1 && (24) \end{aligned}$$

$$\begin{aligned} A &= \text{argmax}_A p(A|D = 1) && (25)\\ &= \text{argmax}_A \sum_{B,C} \psi(A, C) \cdot \psi(B, C) \cdot \psi(C, D = 1) && (26) \end{aligned}$$

if $A = 0$

$$\begin{aligned} \sum_{B,C} \psi(A, C) \cdot \psi(B, C) \cdot \psi(C, D = 1) &= \sum_{B,C} \psi(A = 0, C) \cdot \psi(B, C) \cdot \psi(C, D = 1) && (27)\\ &= 10 \cdot 10 \cdot 1 + 10 \cdot 2 \cdot 1 + 1 \cdot 1 \cdot 7 + 1 \cdot 7 \cdot 7 && (28)\\ &= 100 + 2 + 7 + 49 = 176 && (29) \end{aligned}$$

if $A = 1$

$$\begin{aligned} \sum_{B,C} \psi(A, C) \cdot \psi(B, C) \cdot \psi(C, D = 1) &= \sum_{B,C} \psi(A = 1, C) \cdot \psi(B, C) \cdot \psi(C, D = 1) && (30)\\ &= 2 \cdot 10 \cdot 1 + 2 \cdot 2 \cdot 1 + 7 \cdot 1 \cdot 7 + 7 \cdot 7 \cdot 7 && (31)\\ &= 20 + 4 + 49 + 343 = 416 && (32) \end{aligned}$$

Since $416 > 176$, $A = \text{argmax}_A p(A|D = 1) = 1$.

Solution: Then, we look at the second interpretation,

$$E = \text{argmax}_A p(E|D=1) \tag{33}$$
$$= \text{argmax}_A \psi(D=1, E) = 1 \tag{34}$$

$$A, B, C = \text{argmax}_{A,B,C} p(A, B, C|D=1) \tag{35}$$

Message from D to C

$$m_{DC}(C=0) = 1$$
$$m_{DC}(C=1) = 7$$

Message from B to C

$$m_{BC}(C=0) = \max_B \psi(B, C=0) = 10$$
$$m_{BC}(C=1) = \max_B \psi(B, C=1) = 7$$

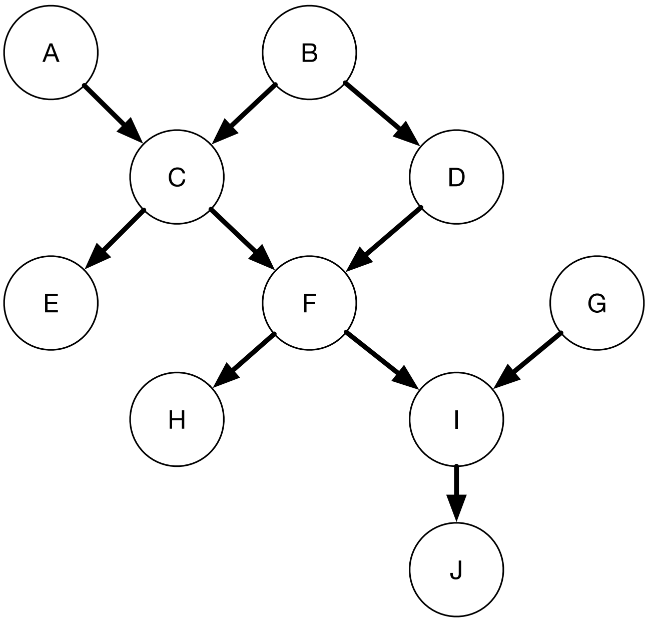
Message from C to A

$$m_{CA}(A=0) = \max_C \psi(A=0, C) m_{BC}(C) m_{DC}(C) = 100$$
$$m_{CA}(A=1) = \max_C \psi(A=1, C) m_{BC}(C) m_{DC}(C) = 343$$

Since $343 > 100$, MAP is $A = 1$

5 Interpreting Network Structure

You are given the following Bayesian Network.



Each node represents a binary random variable. Provide your answers to the solutions below. Justifications are optional.

Please refer to the online quiz for the specific questions.

Grading Policy: 2 points for correct answer, regardless of justifications. If answer is wrong, but justification is completely correct, then you get 1.5 points.

Problem 27. [2 points] What is the Markov Blanket of node A?

Solution: C, B

Problem 28. [2 points] What is the Markov Blanket of node B?

Solution: A, C, D

Problem 29. [2 points] What is the Markov Blanket of node C?

Solution: A, B, D, F, E

Problem 30. [2 points] What is the Markov Blanket of node D?

Solution: F, C, B

Problem 31. [2 points] What is the Markov Blanket of node F?

Solution: C, D, I, G, H

Problem 32. [2 points] What is the Markov Blanket of node G?

Solution: I, F

The following questions test d-separation ability. For the following answer True or False with a short justification.

Problem 33. [2 points] Is it true that $p(A|C,B) = p(A|B)$

Solution: False. Since ACB forms a head-to-head structure and , A and B are not conditional independent given C.

Problem 34. [2 points] Is it true that $p(A|C,D,F) = p(A|C,D)$

Solution: True. Since ACF and BDF are head to tail structures, the paths from A to F are blocked.

Problem 35. [2 points] Is it true that $p(H|I,F) = p(H|F)$

Solution: True. Since HFI is a tail-to-tail structure, F block the path from H to I.

Problem 36. [2 points] Is it true that $p(C|D,J) = p(C|J)$

Solution: False. The path CBD is not blocked by J.

Problem 37. [2 points] Is it true that $p(H|B,C,D) = p(H|C,D)$

Solution: True. Since FCB and FDB are head-to-tail structure, the paths from H to B are blocked.

Problem 38. [2 points] Is it true that $p(A|F,B,C) = p(A|F,C)$

Solution: False. Since ACB is a head-to-head structure, A and B are not conditional independent given C.

6 Final 4 Questions

Problem 39. [4 points] How many networks are I-equivalent to the directed chain $X_1 \to X_2 \to X_3 \to \cdots \to X_N$? *Hint:* Two graphs that are I-equivalent must the have the same skeleton (the same graph where the edges are undirected instead of directed) and the same set of immoralities. An immorality is a v-structure $X \to Z \leftarrow Y$ if there is no directed edge between X and Y.

A. 2^N
B. $2^{(N-1)}$
C. $N - 1$
D. $2(N - 1)$
E. N^2

Grading Policy: 4 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 1 point. If answer is correct, but justification is only partially correct, you get 3. If answer is wrong, but justification is completely correct, then you get 3 points.

Solution: $N - 1$ (N is also acceptable). Since the conditional independence assertions of networks should be the same as the original directed chain, we can not have head-to-head structure or additional edges in new graph. Therefore, we can only change the direction of one edge in each network, which ends up with $N - 1$ networks.

Problem 40. [3 points] If $(X \perp Y|Z)$ and $(Y \perp A|Z)$, then $(X \perp A|Z)$.

A. True
B. False
C. Impossible to determine given the setup.

Grading Policy: 3 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 0.75 point. If answer is correct, but justification is only partially correct, you get 2.25. If answer is wrong, but justification is completely correct, then you get 2.25 points. If you do not provide clear counterexamples in your justifications, it will be regarded as completely wrong. **Solution:** False. Consider $X \to A \to Z \to Y$.

Problem 41. [2 points] Suppose $(A \perp B|C) \in I(G)$ where G is a Bayesian Network structure. Suppose P factorizes according to G. Is it true that $(A \perp B|C) \in I(P)$

A. Yes
B. No
C. Impossible to determine given the setup.

Grading Policy: 3 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 0.75 point. If answer is correct, but justification is only partially correct, you get 2.25. If answer is wrong, but justification is completely correct, then you get 2.25 points. Your justifications are regarded as correct, if they include either one of the following points: 'If P factorizes according to G, then any conditional independence in G exists in P' or 'Bayesian Networks are correct/sound'.

Solution: Yes. Because Bayesian Network is correct/sound. If a distribution P factorizes according to a Bayesian Network G, then any conditional independencies in G exist in P.

Problem 42. [2 points] Suppose we have a Bayesian Network with N binary random variables, $X_1, X_2, \ldots, X_N$. Suppose that $X_1 = 0$. What is the output of the variable elimination algorithm on this Bayesian network if we eliminate all of the variables?

A. 1
B. $p(X_1, X_2, ..., X_N)$
C. $p(X_1 = 0)$
D. $1 - p(X_1 = 0)$
E. None of the above.
F. Impossible to determine given the information provided.

Grading Policy: 2 points for correct answer and valid justification. If answer is correct, but justification is completely wrong or blank, then you get 0.5 point. If answer is correct, but justification is only partially correct, you get 1.5. If answer is wrong, but justification is completely correct, then you get 1.5 points.

Solution: $p(X_1 = 0)$. The variable elimination algorithm will eliminate $p(X_1, \ldots, X_N)\delta(X_1 = 0)$, where $\delta(X_1) = 1$ if $X_1 = 0$ and $\delta(X_1) = 0$ else. Therefore,

$$\sum_{X_1,\ldots,X_N} p(X_1, \ldots, X_N)\delta(X_1 = 0) = \sum_{X_1} p(X_1)\delta(X_1 = 0) \tag{36}$$

$$= p(X_1 = 0) \tag{37}$$