NATIONAL UNIVERSITY OF SINGAPORE

CS5340 - Uncertainty Modelling in AI

(Quiz 1, Semester 2 AY2021/22)

SOLUTIONS

Time Allowed: 1 hour

Instructions

- This is an open-book quiz. You may refer to any of the lecture slides and tutorials, or any of your own notes.
- You may *not* refer to any external online material or use any software to help you answer the questions. You may use a calculator.
- 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. **No justifications are needed**.

Grading Policy: 1 point for each correct answer.

Problem 1. [1 points] An example of epistemic uncertainty is the uncertainty over the probability that a biased coin will turn up heads.

Solution: True.

Problem 2. [1 points] Given two random variables X and Y that are *correlated*,

$$\mathbb{E}[2(X + 3Y)] = 2\mathbb{E}[X] + 6\mathbb{E}[Y]$$

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

Problem 3. [1 points] Given three *independent* random variables X, Y and Z,

$$\mathbb{V}[X + 2Y + Z] = \mathbb{V}[X] + 2\mathbb{V}[Y] + \mathbb{V}[Z]$$

Solution: False. The variance of $X + 2Y + Z$ is

$$\mathbb{V}[X] + 4\mathbb{V}[Y] + \mathbb{V}[Z]$$

Problem 4. [1 points] If the covariance of two random variables X and Y is non-zero, the two variables are *not* independent.

Solution: True.

Problem 5. [1 points] Given a model with a Gaussian likelihood with known mean and a Gamma prior over the variance parameter, then Bayesian inference over the variance parameter is intractable.

Solution: False.

Problem 6. [1 points] Consider two different MRFs comprising 3 random variables X, Y, and Z. In H_1, there are two edges: an edge between X and Y, and another edge between Y and Z. In

H_2, there is only one edge and it is between X and Y. Any distribution p that factorizes according to H_1 can be re-written to factorize according to H_2.

Solution: False.

Problem 7. [1 points] If two nodes X and Y are d-separated in a DGM given observed variables Z, we can conclude they are conditionally-independent given Z.

Solution: True.

Problem 8. [1 points] Define $Z = X + Y$. If two *independent* random variables Z and Y are normally distributed, then X is also normally distributed.

Solution: True.

Problem 9. [1 points] The computational complexity of the sum-product algorithm on a undirected tree with binary random variables is $O(n)$ where n is the number of nodes in the graph.

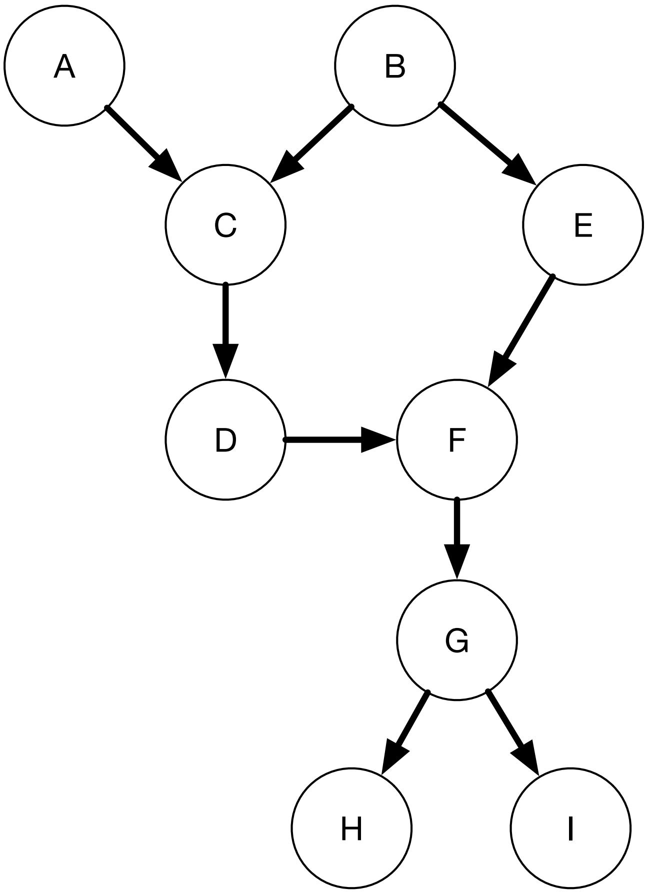
Solution: True.

Problem 10. [1 points] Assume a positive distribution $P > 0$ over discrete variables X, Y and Z (where all probabilities are larger than zero), with a corresponding Bayesian network G. Assume P factorizes according to G. All conditional independencies in a distribution P can be found in the corresponding Bayesian Network G using d-separation.

Solution: False

2 D-Separation Test

You are given the following Bayesian Network.



Each node represents a binary random variable. For each of the following, state whether the conditional independence assertion is True or False given the graph.

Please refer to the online quiz for the specific questions.

Grading Policy: 2 points for correct answer.

Problem 11. [2 points] $(A \perp\!\!\!\perp B|\emptyset)$

Solution: True

Problem 12. [2 points] $(A \perp\!\!\!\perp B|C)$

Solution: False

Problem 13. [2 points] $(B \perp\!\!\!\perp F|E)$

Solution: False

Problem 14. [2 points] $(F \perp\!\!\!\perp H|G)$

Solution: True

Problem 15. [2 points] $(D \perp\!\!\!\perp E|I)$

Solution: False

Problem 16. [2 points] $(B \perp\!\!\!\perp D|C)$

Solution: True

Problem 17. [2 points] $(A \perp\!\!\!\perp E|G, B)$

Solution: False

Problem 18. [2 points] $(A \perp\!\!\!\perp E|H, C)$

Solution: False

Problem 19. [2 points] $(A \perp\!\!\!\perp D|C)$

Solution: True

Problem 20. [2 points] $(A \perp\!\!\!\perp D|C, E)$

Solution: True

Problem 21. [2 points] $(H \perp\!\!\!\perp I|F)$

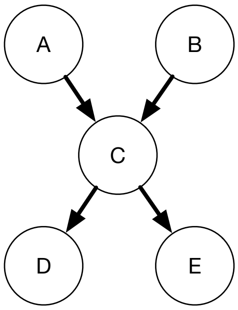
Solution: False

3 Factorizations

In class, we learnt that each graph structure (whether directed or undirected) can be associated with a probability distribution p over the same random variables. For each of the following graphs, select *all* factorizations of p where we can be certain that the G is an I-map for p, i.e., the independence set $I(G) \subseteq I(p)$. You may select none of them.

Grading Policy: 4 points if you select **all** the correct factorizations. If you selects a combination of correct and incorrect answers, they will get partial marks as follows: Full Marks * (Number of Selected Correct Answers - Number of Selected Incorrect Answers) / Total Number of Correct Answers.

Problem 22. [4 points]

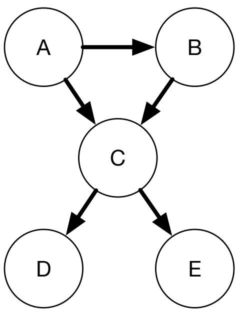


A. $p(A)p(B)p(C|A,B)p(D|C)p(E|C)$
B. $p(A,B,C)p(D)p(E|C)$
C. $p(A)p(B,C|A)p(D|C)p(E)$
D. $p(A)p(B)p(C)p(D)p(E)$
E. $p(A,B)p(C|A,B)p(D|C)p(E|C)$
F. None of the above

Solution: Factorizations such that $I(G) \subseteq I(p)$.

- $p(A)p(B)p(C|A,B)p(D|C)p(E|C)$
- $p(A)p(B,C|A)p(D|C)p(E)$
- $p(A)p(B)p(C)p(D)p(E)$

Problem 23. [4 points]

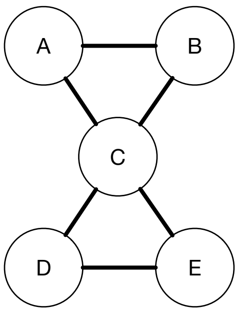


A. $p(A)p(B)p(C|A,B)p(D|C)p(E|C)$
B. $p(A,B,C)p(D)p(E|C)$
C. $p(A)p(B,C|A)p(D|C)p(E)$
D. $p(A)p(B)p(C)p(D)p(E)$
E. $p(A,B)p(C|A,B)p(D|C)p(E|C,D)$
F. None of the above

Solution: Factorizations such that $I(G) \subseteq I(p)$.

- $p(A)p(B)p(C|A,B)p(D|C)p(E|C)$
- $p(A,B,C)p(D)p(E|C)$
- $p(A)p(B,C|A)p(D|C)p(E)$
- $p(A)p(B)p(C)p(D)p(E)$

Problem 24. [4 points]

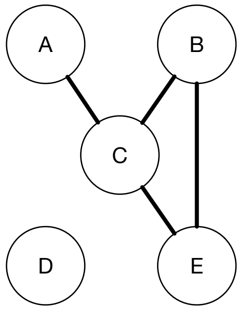


A. $p(A)p(B)p(C|A,B)p(D|C)p(E|C)$
B. $p(A,B,C)p(D)p(E|C)$
C. $p(A)p(B,C|A)p(D|C)p(E)$
D. $p(A)p(B)p(C)p(D|A)p(E)$
E. $p(A,B)p(C|A,B)p(D|C)p(E|C,D)$
F. None of the above

Solution: Factorizations such that $I(G) \subseteq I(p)$.

- $p(A)p(B)p(C|A,B)p(D|C)p(E|C)$
- $p(A,B,C)p(D)p(E|C)$
- $p(A)p(B,C|A)p(D|C)p(E)$
- $p(A,B)p(C|A,B)p(D|C)p(E|C,D)$

Problem 25. [4 points]

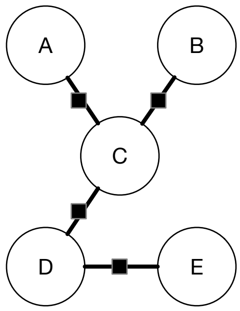


A. $p(A)p(B)p(C|A,B)p(D|C)p(E|C)$
B. $p(A,B,C,D)p(E|C)$
C. $p(A)p(B|A)p(C|A)p(D)p(E|C)$
D. $p(A)p(B)p(C|A)p(D|C)p(E)$
E. $p(A,B)p(C|A,B)p(D|C)p(E|C,D)$
F. None of the above

Solution: No factorizations are such that $I(G) \subseteq I(p)$.

4 Factor Graphs (Part 1)

Consider the following factor graph:



along with following factor table which is shared by all the factors in the graph.

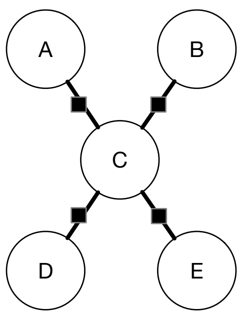
x_i	x_j	$\psi(x_i, x_j)$
0	0	1
0	1	2
1	0	2
1	1	4

Problem 26. [1 points] What is the factorization given the graph above? Select all factorizations that apply.

A. $\psi(A, C)\psi(B, C)\psi(C, D)\psi(D, E)$
B. $\psi(A, B, C)\psi(D)\psi(E, C)$
C. $\psi(A, B)\psi(B)\psi(C, A, B)\psi(D, C)\psi(E, C, D)$
D. $\psi(A)\psi(B)\psi(C)\psi(D)\psi(E)$
E. $\psi(A, C)\psi(B, C)\psi(C, D)\psi(C, E)$

Solution: The correct factorization is $\psi(A, C)\psi(B, C)\psi(C, D)\psi(D, E)$

Consider the following factor graph:



along with following factor table which is shared by all the factors in the graph.

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

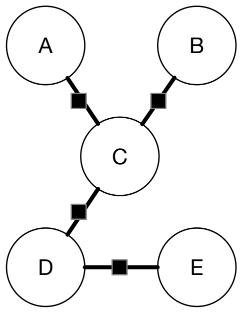
Problem 27. [1 points] What is the factorization given the graph above? Select all factorizations that apply.

A. $\psi(A, C)\psi(B, C)\psi(C, D)\psi(D, E)$
B. $\psi(A, B, C)\psi(D)\psi(E, C)$
C. $\psi(A, B)\psi(B)\psi(C, A, B)\psi(D, C)\psi(E, C, D)$
D. $\psi(A)\psi(B)\psi(C)\psi(D)\psi(E)$
E. $\psi(A, C)\psi(B, C)\psi(C, D)\psi(C, E)$

Solution: The correct factorization is $\psi(A, C)\psi(B, C)\psi(C, D)\psi(C, E)$

5 Factor Graphs (Part 2)

Consider the following factor graph:



along with following factor table which is shared by all the factors in the graph.

x_i	x_j	$\psi(x_i, x_j)$
0	0	1
0	1	2
1	0	2
1	1	4

Problem 28. [2 points] Given you observe $C = 0, D = 0$, what value of A and E maximizes the joint probability?

A. $A = 0, E = 0$
B. $A = 1, E = 0$
C. $A = 0, E = 1$
D. $A = 1, E = 1$
E. Cannot be computed from the above information.

Solution: $A = 1, E = 1$

Problem 29. [2 points] Given you observe $C = 1, D = 0$, what value of A and E maximizes the joint probability?

A. $A = 0, E = 0$
B. $A = 1, E = 0$
C. $A = 0, E = 1$
D. $A = 1, E = 1$
E. Cannot be computed from the above information.

Solution: $A = 1, E = 1$

Problem 30. [2 points] Given you observe $B = 1, D = 0$, what value of A and E maximizes the joint probability?

A. $A = 0, E = 0$
B. $A = 1, E = 0$
C. $A = 0, E = 1$
D. $A = 1, E = 1$
E. Cannot be computed from the above information.

Solution: $A = 1, E = 1$

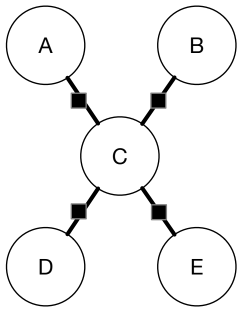
Problem 31. [2 points] Given you observe $B = 0, C = 1, D = 1$, what value of A and E maximizes the joint probability?

A. $A = 0, E = 0$
B. $A = 1, E = 0$
C. $A = 0, E = 1$
D. $A = 1, E = 1$
E. Cannot be computed from the above information.

Solution: $A = 1, E = 1$

6 Factor Graphs (Part 3)

Consider the following factor graph:



along with following factor table which is shared by all the factors in the graph.

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

Problem 32. [4 points] What is $p(A = 0|E = 0)$

A. 17/37
B. 20/37
C. 13/18
D. 5/18
E. 1/6
F. 5/6
G. 3/10
H. None of the above is correct.

Solution: 13/18

Problem 33. [4 points] What is $p(B = 1|E = 0)$

A. 17/37
B. 20/37
C. 13/18
D. 5/18
E. 1/6
F. 5/6
G. 3/10

H. None of the above is correct.

Solution: 5/18

Problem 34. [4 points] What is $p(D = 0|E = 0)$

A. 17/37
B. 20/37
C. 13/18
D. 5/18
E. 1/6
F. 5/6
G. 3/10
H. None of the above is correct.

Solution: 13/18

7 Factor Graphs (Part 4)

Problem 35. [4 points] What is $p(C = 1|D = 0)$

A. 17/37
B. 20/37
C. 13/18
D. 5/18
E. 1/6
F. 5/6
G. 3/10
H. None of the above is correct.

Solution: 5/6

8 The Right Distribution

For each of the following subproblems, pick the *best* probability distribution for each of the scenarios below. **Provide a brief justification**.

Please refer to the online quiz for the specific questions.

Grading Policy: The justification and the choice are both important in these set of problems. The objective was to test your understanding of the distributions. You need both choice and justification to be correct to gain 2 points. If you get the choice correct but your rationale is incorrect or not meaningful, you get 0 points. If you have a partially correct rationale, you get 1 pt.

Problem 36. [2 points] You want to model number of votes a fixed set of candidates receives over a certain pre-defined number of elections. The votes are anonymous and you can only observe the counts.

A. Beta
B. Bernoulli
C. Gaussian
D. Multinomial
E. Categorical

Solution: Multinomial. The Multinomial distribution models the probability of counts for n independent trials.

Grading Policy: If you choose Multinomial and the justification includes that the distribution can be used to model the votes during multiple trials, we give full marks. Otherwise, we give 0 mark.

Problem 37. [2 points] You are designing a random number generator for picking fairly among 10 distinct choices. All choices should have an equal chance of being picked.

A. Poisson
B. Categorical
C. Discrete Uniform
D. Gaussian
E. Bernoulli

Solution: Discrete Uniform

Grading Policy: If Discrete Uniform is chosen and justification states it models equal chance of success for each category, we give full marks. If Categorical is chosen and justification only states it is a discrete distribution, we give 1 mark (the Categorical does not enforce equal probability for choice). If Categorical is chosen and is stated that the probabilities for the 10 choices are set to be equal, we give full marks. Otherwise, we give 0 mark.

Problem 38. [2 points] Your friend is playing a board game with you but you begin to suspect he is cheating. Specifically, you think the dice he is using is loaded (unfair). Denote the outcome of a die roll as X. You can observe the outcome of each roll. Which of the following distributions could model your uncertainty over the distribution of X?

A. Categorical
B. Multinomial
C. Gaussian
D. Poisson
E. Dirichlet

Solution: Dirichlet. Since we are modeling the uncertainty over the distribution of X (categorical distribution), a suitable distribution could be the conjugate prior of categorical distribution, which is Dirichlet.

Grading Policy: If you choose Dirichlet and the justification mentioned that it can model the uncertainty over the categorical distribution, we give full marks, otherwise, 0 mark

Problem 39. [2 points] You want to model the quiz score distribution of the students in CS5340. Assume that the scores are integers that lie between 0 and 100 inclusive. You believe that the distribution might turn out to be multi-modal (have 2 or more peaks).

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

Solution: Categorical. The categorical is a discrete distribution, which can comprise multiple peaks. We use categorical to model the score distribution over 101 different categories.
Multinomial is also correct since it models the probability of counts for each category. When using the multinomial, we treat multiple students as independent trials and each student can score 1 out of the 100 possible scores.

Grading Policy: If Categorical is chosen and the justification mentions that categorical is a discrete distribution that allows multiple peaks. If Multinomial is chosen and the multinomial is a discrete distribution that allows multiple peaks. The latter should state that the n different students are regarded as n independent trials.

Problem 40. [2 points] Consider the random variable A which represents whether you will get an A in CS5340 (or not). What distribution is A drawn from?

A. Normal
B. Bernoulli
C. Gamma
D. Gaussian
E. Exponential

Solution: Bernoulli. The random variable only takes 2 possible values, so it should be Bernoulli.

Grading Policy: If you choose Bernoulli and the justification states that it is a discrete binary variable, we give full marks, otherwise, 0 mark.