NATIONAL UNIVERSITY OF SINGAPORE

CS5340 - Uncertainty Modelling in AI

(Sem 2 AY2023/24)

SOLUTIONS

Time Allowed: 90 Minutes

- Write your student number only. Do not write your name.
- The assessment contains 5 multi-part problems. You have 90 minutes to earn 50 points.
- The assessment contains 26 pages, including this cover page.
- The assessment is open book. You may refer to any printed or handwritten material.
- You may not use your mobile phone, or any other electronic device, except for a simple calculator.
- **Write your answers on the Answer Sheet on page 25.**
- Don't panic. Questions often look harder than they actually are.
- Remember the strategies and techniques we covered in class and apply them.
- Good luck!

Student Number: A

Question	Points	Score
True or False	5	
D-Separation	10	
Estimators	9	
Bayes Net Inference	16	
MRF Inference	10	
Total:	50	

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)	$\frac{\lambda^x \exp(-\lambda)}{x!} \qquad \forall x \in \{1, 2, \dots\}$
Beta (α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
Gamma (α, β)	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad \forall x > 0$
Dirichlet $(\boldsymbol{\alpha})$	$\frac{\Gamma(\sum_k^K \alpha_k)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_K)} \prod_{k=1}^{K} x_k^{\alpha_k - 1}$
Multivariate Normal $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\frac{1}{(2\pi)^{D/2}\lvert\Sigma\rvert^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$
Uniform (a, b)	$\frac{1}{b-a}$

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

Problem 1. True or False (5 Points)

For each of the following statements, write **A** in the box provided if the statement is always true, and write **B** otherwise. Each statement is worth 1 point. **There is negative marking for this problem: each wrong answer will cost you 0.5 points.** If you are unsure, you can choose not to provide an answer (0 for skipped questions).

1.1. (1 point) Let $X_1, X_2 \sim \text{Bern}(0.5)$ be independent Bernoulli random variables and

$$Y = \begin{cases} 1, & \text{if there are an odd number of 1's in } X_1, X_2 \\ 0, & \text{otherwise.} \end{cases}$$

Then $X_1 \perp X_2 \mid Y$.

A. True

B. False

1.2. (1 point) Let $X_1, X_2, X_3 \sim \text{Bern}(0.5)$ be independent Bernoulli random variables and

$$Z = \begin{cases} 1, & \text{if there are an odd number of 1's in } X_1, X_2, X_3 \\ 0, & \text{otherwise.} \end{cases}$$

Then $X_1 \perp X_2 \mid Z$.

A. True

B. False

1.3. (1 point) Suppose a particular disease is present in 1 out of every 100,000 individuals. There exists a medical test that yields a positive result with a probability of 0.99 when the disease is present, and a negative result with a probability of 0.99 when the individual is disease-free. If your test result comes back positive, the probability that you have the disease is more than 0.5.

A. True

B. False

1.4. (1 point) Consider a Bernoulli random variable $X \sim \text{Bern}(\pi)$. Then, the associated log-likelihood is given by $\log p(X = x|\pi) = x \log(1 - \pi) + (1 - x) \log \pi$

A. True

B. False

1.5. (1 point) Assume a distribution P that factorizes according to a directed graph G. If $I(P) \subseteq I(G)$, then it must be the case that G is a perfect map for P.

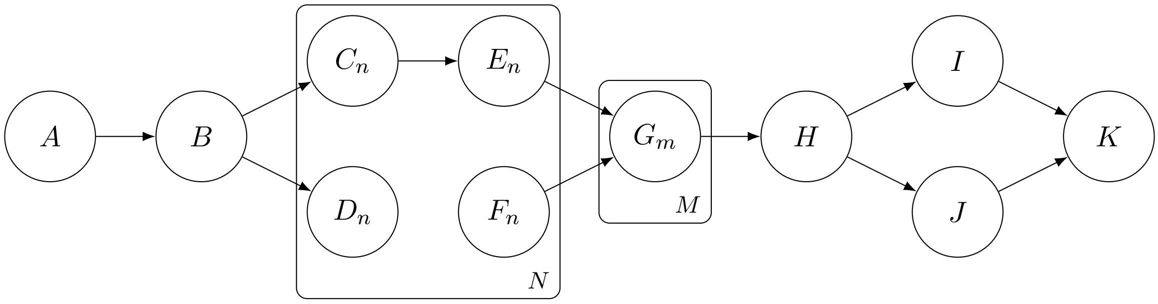
A. True

B. False

Problem 2. D-Separation (10 Points)

For each of the following statements, write **A** in the box provided if the statement is always true, and write **B** otherwise. Each statement is worth 1 point. **There is negative marking for this problem: each wrong answer will cost you 0.5 points.** If you are unsure, you can choose not to provide an answer (0 for skipped questions).

You are given the following Bayesian Network with $N = M = 5$.



2.1. (1 point) $H \perp K \mid \{I, J\}$

A. True

B. False

2.2. (1 point) $I \perp J \mid H$

A. True

B. False

2.3. (1 point) $A \perp H \mid B$

A. True

B. False

2.4. (1 point) $A \perp H \mid G_1$

A. True

B. False

2.5. (1 point) $C_1 \perp C_5 \mid A$

A. True

B. False

2.6. (1 point) $C_1 \perp C_5 \mid \{A, B\}$

A. True

B. False

2.7. (1 point) $C_1 \perp C_5 \mid \{A, B, H\}$

A. True

B. False

2.8. (1 point) $D_1 \perp F_1 \mid \emptyset$

A. True

B. False

2.9. (1 point) $D_1 \perp F_1 \mid \{C_2, C_3, C_4, C_5, G_1\}$

A. True

B. False

2.10. (1 point) $D_1 \perp F_1 \mid \{C_2, C_3, C_4, C_5, G_2\}$

A. True

B. False

Problem 3. Estimators (9 Points)

The following sub-problems are multiple choice. Put your choice in the box provided in the Answer Sheet, e.g., if you believe an answer to be A, write "A" in the box. There is **no** negative marking for this problem.

Poisson Likelihood with Gamma Prior. Recall that the mass function of Poisson distribution with parameter $\lambda > 0$ is

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \forall x \in \{1, 2, \dots\}$$

and the density function of Gamma distribution with parameter $\alpha > 0$ and $\beta > 0$ is

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad \forall x > 0.$$

3.1. (1 point) Suppose $x_1, \dots, x_N$ are drawn from Poisson distribution with unknown parameter λ. Let $S = \sum_n x_n$. What is the maximum likelihood estimator λ_{MLE}?

A. S

B. S/N

C. $(S-1)/N$

D. N/S

E. $N/(S-1)$

F. None of the above.

3.2. (2 points) Suppose we model the prior $p(\lambda)$ using a Gamma distribution using some fixed α, β. What is the maximum a posteriori estimator λ_{MAP}?

A. $\lambda_{\text{MLE}} + \alpha$

B. $\lambda_{\text{MLE}} + \alpha - 1$

C. $\lambda_{\text{MLE}} + (\alpha - 1)/\beta$

D. $\lambda_{\text{MLE}} + \alpha/\beta$

E. $(S + \alpha)/(N + \beta)$

F. $(S + \alpha - 1)/(N + \beta)$

G. $(N + \beta)/(S + \alpha)$

H. $(N + \beta)/(S + \alpha - 1)$

I. None of the above.

Uniform Likelihood with Dirac Delta Prior. Recall that the density function of a continuous uniform distribution with parameter $a < b$ is

$$p(x) = \begin{cases} \frac{1}{b-a}, & \forall x \in [a, b] \\ 0, & \text{otherwise} \end{cases}$$

and the density function for a Dirac delta distribution with parameter γ is $p(x)$ such that the cumulative distribution function,

$$\int_{-\infty}^{u} p(x)\, dx = \begin{cases} 1, & u \geq \gamma \\ 0, & \text{otherwise.} \end{cases}$$

Intuitively, this is a distribution that places all its mass on a single point γ.

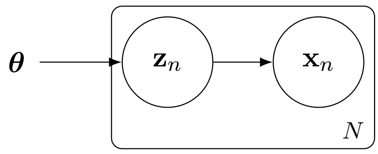
3.3. (2 points) Suppose $x_1, \ldots, x_N$ are drawn from Uniform distribution with $a = 0$ and unknown parameter b. What is the maximum likelihood estimator b_{MLE}?

A. $\sum_n x_n$
B. $\frac{1}{N} \sum_n x_n$
C. $\min\{x_1, \ldots, x_N\}$
D. $\min\{x_1, \ldots, x_N\} - 1$
E. $0.5 \min\{x_1, \ldots, x_N\}$
F. $\max\{x_1, \ldots, x_N\}$
G. $\max\{x_1, \ldots, x_N\} + 1$
H. $2 \max\{x_1, \ldots, x_N\}$
I. $\text{median}\{x_1, \ldots, x_N\}$
J. $\text{median}\{x_1, \ldots, x_N\} - 1$
K. $\text{median}\{x_1, \ldots, x_N\} + 1$
L. None of the above.

3.4. (2 points) Let $0 \leq x_n \leq 10$ for all $n \in \{1, \ldots, N\}$. Suppose we model the prior $p(b)$ as a Dirac delta distribution with parameter $\gamma = 20$. What is the maximum a posteriori estimator b_{MAP}?

A. $\sum_n x_n$
B. $\frac{1}{N} \sum_n x_n$
C. $\min\{x_1, \ldots, x_N\}$
D. $\min\{x_1, \ldots, x_N\} - 1$
E. $0.5 \min\{x_1, \ldots, x_N\}$
F. $\max\{x_1, \ldots, x_N\}$
G. $\max\{x_1, \ldots, x_N\} + 1$
H. $2 \max\{x_1, \ldots, x_N\}$
I. $\text{median}\{x_1, \ldots, x_N\}$
J. $\text{median}\{x_1, \ldots, x_N\} - 1$
K. $\text{median}\{x_1, \ldots, x_N\} + 1$
L. None of the above.

Likelihood from DGM. You are given the following Bayesian Network.



3.5. (2 points) Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are drawn from the DGM with unknown parameter $\boldsymbol{\theta}$. Note that we do not observe $\mathbf{z}_n$. What loss function should we minimize to obtain the maximum likelihood estimator $\boldsymbol{\theta}_{\text{MLE}}$?

A. $\sum_{n=1}^{N} \left[\ln p(\mathbf{z}_n|\boldsymbol{\theta}) \right]$

B. $\sum_{n=1}^{N} \left[\ln p(\mathbf{z}_n|\boldsymbol{\theta}) + \ln p(\mathbf{x}_n|\mathbf{z}_n) \right]$

C. $\sum_{n=1}^{N} \left[\int \ln p(\mathbf{z}_n|\boldsymbol{\theta}) \, d\mathbf{z}_n \right]$

D. $\sum_{n=1}^{N} \left[\int \ln p(\mathbf{z}_n|\boldsymbol{\theta}) p(\mathbf{x}_n|\mathbf{z}_n) \, d\mathbf{z}_n \right]$

E. $\sum_{n=1}^{N} \left[\ln \int p(\mathbf{z}_n|\boldsymbol{\theta}) \, d\mathbf{z}_n \right]$

F. $\sum_{n=1}^{N} \left[\ln \int p(\mathbf{z}_n|\boldsymbol{\theta}) p(\mathbf{x}_n|\mathbf{z}_n) \, d\mathbf{z}_n \right]$

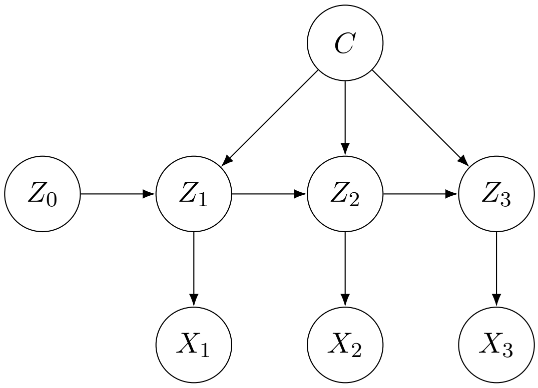
G. $\sum_{n=1}^{N} \left[\ln \int p(\mathbf{z}_n|\boldsymbol{\theta}) \, d\mathbf{z}_n + \ln \int p(\mathbf{x}_n|\mathbf{z}_n) \, d\mathbf{z}_n \right]$

H. None of the above.

Problem 4. Bayes Net Inference (16 Points)

The following sub-problems are multiple choice. Put your choice in the box provided in the Answer Sheet, e.g., if you believe an answer to be A, write "A" in the box. There is **no** negative marking for this problem.

You are given the following Bayesian Network.



Z_0	$p(Z_0)$
0	0.5
1	0.5

C	$p(C)$
0	0.8
1	0.2

X_i	Z_i	$p(X_i \mid Z_i)$
0	0	0.8
0	1	0.2
1	0	0.2
1	1	0.8

Z_{i+1}	Z_i	C	$p(Z_{i+1} \mid Z_i, C)$
0	0	0	0.1
0	1	0	0.9
1	0	0	0.9
1	1	0	0.1
0	0	1	0.9
0	1	1	0.1
1	0	1	0.1
1	1	1	0.9

Part I: In the following, identify the **Markov Boundary** for specific nodes. The Markov Boundary is a *minimal* Markov Blanket; if you remove a node from the Markov Boundary, it is no longer a Markov Blanket.

4.1. (1 point) What is the Markov Boundary for Z_0?

A. $\{Z_1\}$

B. $\{Z_2\}$

C. $\{Z_3\}$

D. $\{Z_1, C\}$

E. $\{Z_1, X_1, C, X_2\}$

F. $\{Z_1, Z_1, Z_3, C\}$

G. $\Phi = \{\}$ (Empty Set)

H. None of the above.

4.2. (1 point) What is the Markov Boundary for Z_2?

A. $\{Z_1\}$

B. $\{Z_3\}$

C. $\{X_2\}$

D. $\{Z_1, Z_3, C\}$

E. $\{X_2, Z_1, Z_3, C\}$

F. $\{X_1, X_2, X_3, Z_1, Z_3, C\}$

G. $\Phi = \{\}$ (Empty Set)

H. None of the above.

4.3. (1 point) What is the Markov Boundary for X_1?

A. $\{Z_1\}$

B. $\{Z_2\}$

C. $\{Z_3\}$

D. $\{X_2, X_3\}$

E. $\{Z_0, Z_1, C\}$

F. $\{Z_0, Z_1, Z_2, C_3, C\}$

G. $\Phi = \{\}$ (Empty Set)

H. None of the above.

4.4. (1 point) What is the Markov Boundary for X_3?

A. $\{Z_1\}$

B. $\{Z_2\}$

C. $\{Z_3\}$

D. $\{X_1, X_2\}$

E. $\{X_2, Z_1, Z_3, C\}$

F. $\{X_2, Z_3, C\}$

G. $\Phi = \{\}$ (Empty Set)

H. None of the above.

Part II. Compute the conditional probabilities for each of the following. Answers are rounded to two decimal places. Select the best answer.

4.5. (2 points) Compute $p(Z_0 = 1 | Z_1 = 0, C = 1)$

A. 0.06

B. 0.10

C. 0.20

D. 0.28

E. 0.50

F. 0.72

G. 0.80

H. 0.90

I. 0.94

J. None of the above.

4.6. (2 points) Compute $p(Z_2 = 1 | X_2 = 1, Z_1 = 0, Z_3 = 0, C = 0)$

A. 0.06

B. 0.10

C. 0.20

D. 0.28

E. 0.50

F. 0.72

G. 0.80

H. 0.90

I. 0.94

J. None of the above.

4.7. (2 points) Compute $p(X_3 = 1 | Z_1 = 0, Z_2 = 1, Z_3 = 0)$

A. 0.06

B. 0.10

C. 0.20

D. 0.28

E. 0.50

F. 0.72

G. 0.80

H. 0.90

I. 0.94

J. None of the above.

Part III. For the next three questions, assume that we observe $Z_0 = 0, X_1 = 0, X_3 = 1$, and $C = 1$. Answers are rounded to two decimal places. Select the best answer. Compute the following conditionals:

4.8. (2 points) Compute $p(Z_1 = 0 | Z_0 = 0, X_1 = 0, X_3 = 1, C = 1)$

A. 0.06

B. 0.10

C. 0.20

D. 0.28

E. 0.50

F. 0.72

G. 0.80

H. 0.90

I. 0.94

J. None of the above.

4.9. (2 points) Compute $p(Z_2 = 0|Z_0 = 0, X_1 = 0, X_3 = 1, C = 1)$

A. 0.06

B. 0.10

C. 0.20

D. 0.28

E. 0.50

F. 0.72

G. 0.80

H. 0.90

I. 0.94

J. None of the above.

4.10. (2 points) Compute $p(Z_3 = 0|Z_0 = 0, X_1 = 0, X_3 = 1, C = 1)$

A. 0.06

B. 0.10

C. 0.20

D. 0.28

E. 0.50

F. 0.72

G. 0.80

H. 0.90

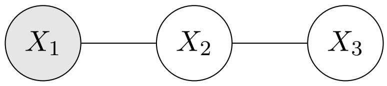
I. 0.94

J. None of the above.

Problem 5. MRF Inference (10 Points)

The following sub-problems are multiple choice. Put your choice in the box provided in the Answer Sheet, e.g., if you believe an answer to be A, write "A" in the box. There is **no** negative marking for this problem.

Consider the following Markov Random Field.



Each random variable X_i can take on a value $x_i \in \mathbb{R}^+$. In other words, x_i is a positive real number. The MRF is parameterized using the following unary and pairwise potentials:

$$\psi(X_i = x_i) = \exp(-vx_i)$$

$$\psi(X_i = x_i, X_j = x_j) = \exp(-w_1 x_i + -w_2 x_j) \quad \text{where } i < j$$

where $w_1, w_2, v > 0$, that is, the parameters w_1, w_2, and v are all positive real numbers. Hence, the joint probability is given by,

$$p(X_1, X_2, X_3) = \frac{1}{Z}\psi(X_1)\psi(X_2)\psi(X_3)\psi(X_1, X_2)\psi(X_2, X_3)$$

We will perform belief propagation / sum-product on this MRF. **Assume that X_3 is the root node and $X_1 = 1$ is observed.**

The following identity may be useful:

$$\int_0^\infty \exp(-ax)dx = \frac{1}{a} \qquad \text{for } a > 0$$

5.1. (2 points) What is the message sent from X_1 to X_2? Note that $X_1 = 1$ is observed.

A. $m_{1\to 2} = 1$

B. $m_{1\to 2} = \exp(-v)$

C. $m_{1\to 2}(x_2) = \exp(-v - w_1 x_1 - w_2)$

D. $m_{1\to 2}(x_2) = \exp(-v - w_2 x_2)$

E. $m_{1\to 2}(x_2) = \exp(-v - w_1 - w_2 x_2)$

F. $m_{1\to 2}(x_2) = \dfrac{\exp(-v - w_1 - w_2 x_2)}{w_1 + v}$

G. $m_{1\to 2}(x_2) = \dfrac{\exp(-v - w_1 - w_2 x_2)}{w_2 + v}$

H. $m_{1\to 2}(x_2) = \dfrac{\exp(-v - w_1 - w_2 x_2)}{w_1 + w_2 + v}$

I. None of the above.

5.2. (2 points) What is the message sent from X_2 to X_3? Note that $X_1 = 1$ is observed.

A. $m_{2\to3}(x_3) = \exp(-w_2x_3)$

B. $m_{2\to3}(x_3) = \exp(-v - w_2x_3)$

C. $m_{2\to3}(x_3) = \exp(-vx_3 - w_2x_3)$

D. $m_{2\to3}(x_3) = \exp(-v - w_1 - w_2)$

E. $m_{2\to3}(x_3) = \exp(-v - w_1 - w_2x_3)$

F. $m_{2\to3}(x_3) = \dfrac{\exp(-v - w_1 - w_2x_3)}{w_1 + v}$

G. $m_{2\to3}(x_3) = \dfrac{\exp(-v - w_1 - w_2x_3)}{w_2 + v}$

H. $\boldsymbol{m_{2\to3}(x_3) = \dfrac{\exp(-v - w_1 - w_2x_3)}{w_1 + w_2 + v}}$

I. None of the above.

5.3. (2 points) What is message sent from X_3 to X_2? Note that $X_1 = 1$ is observed.

A. $m_{3\to2}(x_2) = \exp(-w_1x_2)$

B. $m_{3\to2}(x_2) = \exp(-v - w_1x_2)$

C. $m_{3\to2}(x_2) = \exp(-v - w_1 - w_2)$

D. $m_{3\to2}(x_2) = \exp(-v - w_1x_2 - w_2)$

E. $m_{3\to2}(x_2) = \dfrac{\exp(-v - w_1 - w_2)}{w_1 + v}$

F. $m_{3\to2}(x_2) = \dfrac{\exp(-v - w_1x_2 - w_2)}{w_2 + v}$

G. $\boldsymbol{m_{3\to2}(x_2) = \dfrac{\exp(-w_1x_2)}{w_2 + v}}$

H. $m_{3\to2}(x_2) = \dfrac{\exp(-w_1x_2)}{w_1 + w_2 + v}$

I. None of the above.

5.4. (2 points) What is $p(X_2 = x_2|X_1 = 1)$?

A. $\exp(-(v + w_1)x_2)$

B. $\exp(-(v + w_1 + w_2)x_2)$

C. $(v + w_1)\exp(-(v + w_1)x_2)$

D. $(v + w_2)\exp(-(v + w_1)x_2)$

E. $(v + w_1 + w_2)\exp(-(v + w_1)x_2)$

F. $(v + w_1)\exp(-(v + w_1 + w_2)x_2)$

G. $(v + w_2)\exp(-(v + w_1 + w_2)x_2)$

H. $(v + w_1 + w_2)\exp(-(v + w_1 + w_2)x_2)$

I. None of the above.

5.5. (2 points) What is $p(X_3 = x_3|X_1 = 1)$?

A. $\exp(-(v + w_2)x_3)$

B. $\exp(-(v + w_1 + w_2)x_3)$

C. $(v + w_1)\exp(-(v + w_2)x_3)$

D. $(v + w_2)\exp(-(v + w_2)x_3)$

E. $(v + w_1 + w_2)\exp(-(v + w_2)x_3)$

F. $(v + w_1)\exp(-(v + w_1 + w_2)x_3)$

G. $(v + w_2)\exp(-(v + w_1 + w_2)x_3)$

H. $(v + w_1 + w_2)\exp(-(v + w_1 + w_2))x_3$

I. None of the above.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

This page is intentionally left blank.

Answer Sheet

Fill in your student number. Put your answer for each question into each of the answer boxes below. If you decide to skip a question, leave the box blank.

Student Number: A

1.1		2.1		3.1		4.1		5.1	
1.2		2.2		3.2		4.2		5.2	
1.3		2.3		3.3		4.3		5.3	
1.4		2.4		3.4		4.4		5.4	
1.5		2.5		3.5		4.5		5.5	
		2.6				4.6			
		2.7				4.7			
		2.8				4.8			
		2.9				4.9			
		2.10				4.10			

END OF PAPER