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 40 points.
- The assessment contains 24 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 23**.
- 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	
Markov Chain	7	
Learning Parameters - Part I	9	
Learning Parameters - Part II	16	
Variational Bayes	3	
Total:	40	

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 $(\boldsymbol{\alpha})$	$\frac{\Gamma(\sum_k^K \alpha_k)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_K)} \prod_{k=1}^{K} x_k^{\alpha_k - 1}$
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) If X and Y are both univariate Gaussian random variables, then XY is also a Gaussian random variable.

A. True

B. False

1.2. (1 point) The sum product algorithm can be applied to any directed graphical model to obtain the marginals at every node.

A. True

B. False

1.3. (1 point) In hidden Markov model (HMM), the observed variables $\{X_n\}_n$ form a Markov chain.

A. True

B. False

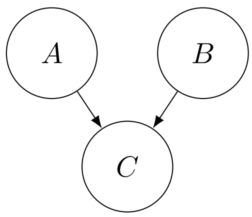
1.4. (1 point) In the expectation-maximization (EM) algorithm, given an arbitrary θ_{old}, the maximum likelihood estimator is

$$\theta_{\text{MLE}} = \arg\max_{\theta} \mathcal{Q}(\theta, \theta_{\text{old}}).$$

A. True

B. False

1.5. (1 point) Consider the following DGM $\mathcal{M}$.



If $A \sim \mathcal{N}(A|0,1)$, $B \sim \mathcal{N}(A|0,1)$, and $C \sim \mathcal{N}(C|AB,1)$, then $\mathcal{M}$ is a linear-Gaussian model.

A. True

B. False

Problem 2. Markov Chain (7 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).

2.1. (1 point) Let the transition matrix of a finite Markov chain[1] $\mathcal{M}$ be defined by

$$\begin{bmatrix} 0.1 & 0.9 & 0.0 \\ 0.0 & 0.0 & 1.0 \\ 1.0 & 0.0 & 0.0 \end{bmatrix}.$$

Then $\mathcal{M}$ is ergodic.

A. True

B. False

2.2. (1 point) Let the transition matrix of a finite Markov chain $\mathcal{M}$ be defined by

$$\begin{bmatrix} 0.0 & 0.5 & 0.5 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.5 & 0.5 & 0.0 \end{bmatrix}.$$

Then $\mathcal{M}$ is ergodic.

A. True

B. False

2.3. (1 point) Let $\mathcal{M}_1$ be an **irreducible but not aperiodic** finite Markov chain defined by the transition matrix $\mathbf{T}$ and let $\mathcal{M}_2$ be a finite Markov chain defined by the transition matrix $(\mathbf{I} + \mathbf{T})/2$ where $\mathbf{I}$ is the identity matrix. Then $\mathcal{M}_2$ is ergodic.

A. True

B. False

[1] A finite Markov chain is a Markov chain whose state space is finite.

2.4. (1 point) Let $\mathcal{M}_1$ be an **aperiodic but not irreducible** finite Markov chain defined by the transition matrix $\mathbf{T}$ and let $\mathcal{M}_2$ be a finite Markov chain defined by the transition matrix $(\mathbf{I} + \mathbf{T})/2$ where $\mathbf{I}$ is the identity matrix. Then $\mathcal{M}_2$ is ergodic.

A. True

B. False

2.5. (1 point) Let $\mathcal{M}_1$ be a finite Markov chain defined by the transition matrix $\mathbf{T}$ and $\mathcal{M}_2$ be a finite Markov chain defined by the transition matrix $\mathbf{T}^2 = \mathbf{TT}$. If $\mathcal{M}_2$ is ergodic, then $\mathcal{M}_1$ is ergodic.

A. True

B. False

2.6. (1 point) Let $\mathcal{M} = \{x_0, x_1, x_2 \ldots\}$ be a Markov chain whose state space is the set of real numbers $\mathbb{R}$. Suppose $\mathcal{M}$ always begins at 0 (i.e. $x_0 = 0$) and follows the transition kernel

$$\mathbf{T}(x_{t+1}|x_t) = \mathcal{N}(x_{t+1}|x_t, 1) \cdot \min\left\{1, \frac{\exp(-x_{t+1}^2) \cdot \mathcal{N}(x_{t+1}|x_t, 1)}{\exp(-x_t^2) \cdot \mathcal{N}(x_t|x_{t+1}, 1)}\right\}.$$

Then $\mathcal{M}$ is ergodic.

A. True

B. False

2.7. (1 point) Let $\mathcal{M} = \{x_0, x_1, x_2 \ldots\}$ be a Markov chain whose state space is the set of integers $\mathbb{Z}$. Suppose $\mathcal{M}$ always begins at 0 (i.e. $x_0 = 0$) and follows the transition kernel

$$\mathbf{T}(x_{t+1}|x_t) = \begin{cases} 0.5 & \text{if } x_{t+1} = x_t - 1 \\ 0.5 & \text{if } x_{t+1} = x_t + 1 \\ 0.0 & \text{otherwise.} \end{cases}$$

Then $\mathcal{M}$ is ergodic.

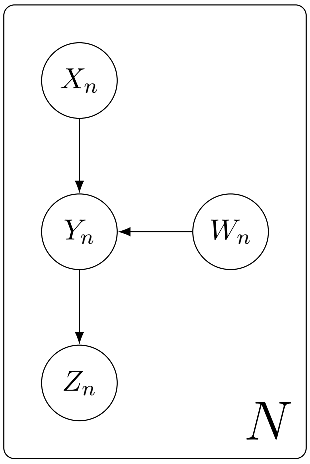
A. True

B. False

Problem 3. Learning Parameters - Part I (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.

You are given the following Bayesian Network.



The variables have the following distributions and relationships:

$$\begin{aligned} p(X_n) &= \mathrm{Cat}[\pi] \\ p(W_n) &= \mathcal{N}(W_n|\mu_w, \sigma_w^2) \\ Y_n &= X_n + W_n \\ p(Z_n|Y_n) &= \mathcal{N}(Z_n|\alpha Y_n, \sigma_z^2) \end{aligned}$$

Note that X_n takes on a value $\{1, 2, \ldots, K\}$ (it is *not* in 1-hot encoding). Let us denote $X_{1:N} = \{X_1, X_2, \ldots, X_N\}$ and similarly for $Y_{1:N}$, $W_{1:N}$, and $Z_{1:N}$. Assume σ_w^2 and σ_z^2 to be known, and denote the set of unknown parameters as $\theta = \{\pi, \mu_w, \alpha\}$. Suppose that we have a dataset $\mathcal{D} = \{(X_n, Z_n)\}_{n=1}^N$. We define

$$\mathbf{X} = \begin{bmatrix} X_1 & 1 \\ X_2 & 1 \\ \vdots & \vdots \\ X_N & 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix}$$

3.1. (1 point) Which of the following corresponds to the joint distribution $p(X_{1:N}, W_{1:N}, Y_{1:N}, Z_{1:N})$?

A. $\boldsymbol{\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | X_n + \mu_w, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \sigma_z^2))^{I[X_n=k]}}$

B. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | X_n + \mu_w, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \sigma_z^2 + \sigma_w^2))^{I[X_n=k]}$

C. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | X_n + \mu_w, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \alpha^2 \sigma_z^2))^{I[X_n=k]}$

D. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | X_n + W_n, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \sigma_z^2))^{I[X_n=k]}$

E. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | X_n + W_n, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \alpha^2 \sigma_z^2))^{I[X_n=k]}$

F. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | \mu_w, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \sigma_z^2))^{I[X_n=k]}$

G. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | X_n + \mu_w, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \alpha^2 \sigma_z^2))^{I[X_n=k]}$

H. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | \mu_w, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \alpha^2 \sigma_z^2))^{I[X_n=k]}$

I. $\prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(Y_n | W_n, \sigma_w^2) \mathcal{N}(Z_n | \alpha Y_n, \sigma_z^2))^{I[X_n=k]}$

J. None of the above.

3.2. (1 point) To perform MLE given $\mathcal{D}$, which of the following optimization problems do we solve?

A. $\arg\min_\theta \log p(X_{1:N} | \theta)$

B. $\arg\min_\theta \log p(Z_{1:N} | \theta)$

C. $\arg\min_\theta \log p(X_{1:N}, Z_{1:N} | \theta)$

D. $\arg\min_\theta \log p(Z_{1:N} | X_{1:N}, \theta)$

E. $\arg\min_\theta \log p(X_{1:N} | Z_{1:N}, \theta)$

F. $\arg\min_\theta \log p(\theta | X_{1:N}, Z_{1:N})$

G. $\arg\max_\theta \log p(X_{1:N} | \theta)$

H. $\arg\max_\theta \log p(Z_{1:N} | \theta)$

I. $\boldsymbol{\arg\max_\theta \log p(X_{1:N}, Z_{1:N} | \theta)}$

J. $\arg\max_\theta \log p(Z_{1:N} | X_{1:N}, \theta)$

K. $\arg\max_\theta \log p(X_{1:N} | Z_{1:N}, \theta)$

L. $\arg\max_\theta \log p(\theta | X_{1:N}, Z_{1:N})$

M. None of the above.

3.3. (2 points) Which of the following is equivalent to $\log p(X_{1:N}, Z_{1:N})$?

A. $\boldsymbol{\log \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(Z_n | \alpha(X_n + \mu_w), \alpha^2\sigma_w^2 + \sigma_z^2)\right)^{I[X_n=k]}}$

B. $\log \sum_{n=1}^{N} \sum_{k=1}^{K} \left(\pi_k \mathcal{N}(Z_n | \alpha(X_n + \mu_w), \alpha^2\sigma_w^2 + \sigma_z^2)\right)^{I[X_n=k]}$

C. $\sum_{n=1}^{N} \log \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(Z_n | X_n + \mu_w, \sigma_w^2 + \sigma_z^2)\right)^{I[X_n=k]}$

D. $\prod_{n=1}^{N} \log \sum_{k=1}^{K} \left(\pi_k \mathcal{N}(Z_n | \alpha(X_n + \mu_w), \sigma_w^2 + \sigma_z^2)\right)^{I[X_n=k]}$

E. $\log \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(Z_n | X_n + \mu_w, \alpha^2\sigma_w^2 + \sigma_z^2)\right)^{I[X_n=k]}$

F. $\log \prod_{n=1}^{N} \sum_{k=1}^{K} \left(\pi_k \mathcal{N}(Z_n | \alpha(X_n + \alpha\mu_w), \sigma_w^2 + \alpha^2\sigma_z^2)\right)^{I[X_n=k]}$

G. $\log \prod_{n=1}^{N} \left(\pi \mathcal{N}(Z_n | \alpha(X_n + \mu_w), \alpha\sigma_w^2 + \sigma_z^2)\right)$

H. $\log \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\pi_k \mathcal{N}(X_n | \mu_w, \sigma_w^2)\mathcal{N}(Z_n | \alpha Y_n, \sigma_z^2)\right)^{I[X_n=k]}$

I. The log-likelihood is intractable due to the latent variables. We can use the EM algorithm to learn the parameters.

J. None of the above.

3.4. (2 points) Let N_k be the number of samples where $X_n = k$. Which of the following is the MLE parameter π?

A. $\pi_{k,\text{MLE}} = N_k$.

B. $\boldsymbol{\pi_{k,\text{MLE}} = \frac{N_k}{N}}$**.**

C. $\pi_{k,\text{MLE}} = \frac{N}{N_k}$.

D. $\pi_{k,\text{MLE}} = \frac{N_k}{N^2}$.

E. $\pi_{k,\text{MLE}} = \frac{N + N_k}{N}$.

F. $\pi_{k,\text{MLE}} = \log\left(\frac{N_k}{N}\right)$.

G. $\pi_{k,\text{MLE}} = \sqrt{\frac{N_k}{N}}$.

H. $\pi_{k,\text{MLE}} = 1 - \frac{N_k}{N}$.

I. There is no closed-form solution and we can learn the parameters using EM.

J. None of the above.

3.5. (3 points) Let $\boldsymbol{\beta} = [\alpha, \alpha\mu_w]^\top$. Which of the following is the MLE for β?

A. $\boldsymbol{\beta}_{\text{MLE}} = \mathbf{X}^\top\mathbf{X}\mathbf{X}^\top\mathbf{Z}$

B. $\boldsymbol{\beta}_{\text{MLE}} = (\mathbf{X}^\top\mathbf{Z})^{-1}\mathbf{X}^\top\mathbf{X}$

C. $\boldsymbol{\beta}_{\text{MLE}} = \mathbf{X}\mathbf{X}^\top(\mathbf{Z}\mathbf{Z}^\top)^{-1}$

D. $\boldsymbol{\beta}_{\text{MLE}} = \mathbf{Z}^\top\mathbf{Z}(\mathbf{X}^\top\mathbf{X})^{-1}$

E. $\boldsymbol{\beta}_{\text{MLE}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Z}$.

F. $\boldsymbol{\beta}_{\text{MLE}} = \mathbf{X}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{Z}$

G. $\boldsymbol{\beta}_{\text{MLE}} = \mathbf{X}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}$

H. $\boldsymbol{\beta}_{\text{MLE}} = \sqrt{(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Z}}$

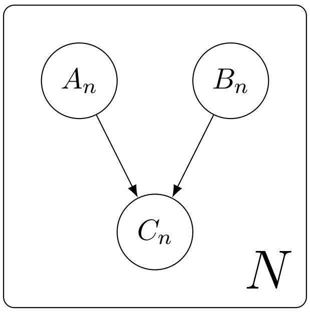
I. There is no closed-form solution and we can learn the parameters using EM.

J. None of the above.

Problem 4. Learning Parameters - Part II (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.



The variables have the following distributions and relationships:

$$p(A_n) = \text{Bern}[\pi_A]$$
$$p(B_n) = \text{Bern}[\pi_B]$$
$$p(C_n|A_n, B_n) = \text{Poi}[2^{A_n+B_n}\alpha]$$

where $p(C_n|A_n, B_n)$ is Poisson distributed with rate $2^{A_n+B_n}\alpha$. For example, when $A_n = 1$ and $B_n = 0$, we have $C_n \sim \text{Poi}(2\alpha)$.

Let us denote $A_{1:N} = \{A_1, A_2, \ldots, A_N\}$ and similarly for $B_{1:N}$ and $C_{1:N}$. Let $\theta = \{\pi_A, \pi_B, \alpha\}$. Suppose that we have a dataset $\mathcal{D} = \{(A_n, C_n)\}_{n=1}^N$. We shall find the MLE for π_A directly and use the EM algorithm for π_B, α.

4.1. (1 point) According to the DGM, how does the joint $p(A_{1:N}, B_{1:N}, C_{1:N})$ factorize?

A. $\prod_{n=1}^N p(A_n)p(B_n)p(C_n)$

B. $\prod_{n=1}^N p(A_n)p(B_n)p(C_n|A_n, B_n)$

C. $\prod_{n=1}^N p(A_n, B_n|C_n)p(C_n)$

D. $\prod_{n=1}^N p(A_n|C_n)p(B_n|C_n)p(C_n)$

E. $\prod_{n=1}^N p(A_n)p(B_n|A_n)p(C_n|A_n, B_n)$

F. $\sum_{n=1}^N p(A_n)p(B_n)p(C_n)$

G. $\sum_{n=1}^N p(A_n)p(B_n)p(C_n|A_n, B_n)$

H. $\sum_{n=1}^N p(A_n, B_n|C_n)p(C_n)$

I. $\sum_{n=1}^N p(A_n|C_n)p(B_n|C_n)p(C_n)$

J. $\sum_{n=1}^N p(A_n)p(B_n|A_n)p(C_n|A_n, B_n)$

4.2. (2 points) The posterior $p(B_n = 1|A_n, C_n) = \widehat{\pi}_n$. What is $\widehat{\pi}_n$?
Hint: Consider the definition of conditional probability or apply Bayes rule.

A. $\widehat{\pi}_n = \dfrac{\pi_B \mathrm{Poi}[2\alpha]^{A_n} \mathrm{Poi}[4\alpha]^{1-A_n}}{\pi_B \mathrm{Poi}[2\alpha]^{A_n} \mathrm{Poi}[4\alpha]^{1-A_n} + (1-\pi_B)\mathrm{Poi}[\alpha]^{A_n} \mathrm{Poi}[2\alpha]^{1-A_n}}$

B. $\widehat{\pi}_n = \dfrac{(1-\pi_B)\mathrm{Poi}[4\alpha]^{A_n} \mathrm{Poi}[2\alpha]^{1-A_n}}{\pi_B \mathrm{Poi}[2\alpha]^{A_n} \mathrm{Poi}[\alpha]^{1-A_n} + (1-\pi_B)\mathrm{Poi}[4\alpha]^{A_n} \mathrm{Poi}[2\alpha]^{1-A_n}}$

C. $\widehat{\pi}_n = \dfrac{\mathrm{Poi}[2\alpha]^{A_n} \mathrm{Poi}[\alpha]^{1-A_n}}{\mathrm{Poi}[4\alpha]^{A_n} \mathrm{Poi}[2\alpha]^{1-A_n}}$

D. $\widehat{\pi}_n = \pi_B \dfrac{\mathrm{Poi}[4\alpha]^{A_n} \mathrm{Poi}[2\alpha]^{1-A_n}}{\mathrm{Poi}[2\alpha]^{A_n} \mathrm{Poi}[\alpha]^{1-A_n}}$

E. $\widehat{\pi}_n = \dfrac{\pi_A \mathrm{Poi}[4\alpha]^{B_n} \mathrm{Poi}[2\alpha]^{1-B_n}}{\pi_A \mathrm{Poi}[4\alpha]^{B_n} \mathrm{Poi}[2\alpha]^{1-B_n} + (1-\pi_A)\mathrm{Poi}[2\alpha]^{B_n} \mathrm{Poi}[\alpha]^{1-B_n}}$

F. $\boldsymbol{\widehat{\pi}_n = \dfrac{\pi_B \mathrm{Poi}[4\alpha]^{A_n} \mathrm{Poi}[2\alpha]^{1-A_n}}{\pi_B \mathrm{Poi}[4\alpha]^{A_n} \mathrm{Poi}[2\alpha]^{1-A_n} + (1-\pi_B)\mathrm{Poi}[2\alpha]^{A_n} \mathrm{Poi}[\alpha]^{1-A_n}}}$

G. $\widehat{\pi}_n = \dfrac{\pi_B \mathrm{Poi}[4\alpha]}{\pi_B \mathrm{Poi}[4\alpha] + (1-\pi_B)\mathrm{Poi}[2\alpha]}$

H. $\widehat{\pi}_n = \dfrac{\pi_B \mathrm{Poi}[2\alpha] + (1-\pi_B)\mathrm{Poi}[\alpha]}{\pi_B \mathrm{Poi}[4\alpha] + (1-\pi_B)\mathrm{Poi}[2\alpha]}$

I. None of the above.

4.3. (2 points) What is $\log p(A_n, B_n, C_n)$?

A. $A_n \log \pi_A + B_n \log(1-\pi_A) + B_n \log \pi_B +$
$C_n \log(1-\pi_B) + C_n \log(\alpha 2^{A_n+B_n}) - \alpha 2^{A_n+B_n}$

B. $A_n \log \pi_A + (1-A_n)\log(1-\pi_A) + B_n \log \pi_B + (1-B_n)\log(1-\pi_B) +$
$C_n \log(\alpha) - \alpha 2^{A_n+B_n} - \log(C_n!)$

C. $A_n \log \pi_A + (1-A_n)\log \pi_B + B_n \log \pi_B + (1-B_n)\log \pi_A +$
$C_n \log(\alpha 2^{A_n+B_n}) - \alpha 2^{A_n+B_n} - \log(C_n!)$

D. $\boldsymbol{A_n \log \pi_A + (1-A_n)\log(1-\pi_A) + B_n \log \pi_B + (1-B_n)\log(1-\pi_B) +}$
$\boldsymbol{C_n \log(\alpha 2^{A_n+B_n}) - \alpha 2^{A_n+B_n} - \log(C_n!)}$

E. $A_n \log \pi_A + (1-A_n)\log(1-\pi_A) + B_n \log \pi_B + (1-B_n)\log(1-\pi_B) -$
$C_n \log(\alpha 2^{A_n+B_n}) + \alpha 2^{A_n+B_n} - \log(C_n!)$

F. $A_n \log \pi_A + B_n \log \pi_B +$
$C_n \log(\alpha 2^{A_n+B_n}) + \alpha 2^{A_n+B_n} - \log(C_n!)$

G. $A_n \pi_A + (1-A_n)(1-\pi_A) + B_n \pi_B + (1-B_n)(1-\pi_B) +$
$C_n(\alpha 2^{A_n+B_n}) - \alpha 2^{A_n+B_n} - C_n!$

H. $\log\left[\dfrac{\pi_A^{A_n} \pi_B^{B_n} (\alpha 2^{A_n+B_n})^{C_n} \exp(-\alpha 2^{A_n+B_n})}{C_n!}\right]$

I. None of the above.

4.4. (2 points) We shall find the MLE for π_A directly. Let:

- N_A is the count of samples where $A_n = 1$
- N_B is the count of samples where $\widehat{\pi}_n > \frac{1}{2}$
- N_C is the count of samples where $C_n > 0$

What is the MLE for π_A?

A. $\pi_A = \dfrac{N - N_A}{N}$

B. $\pi_A = \dfrac{N_A}{N_B}$

C. $\pi_A = \dfrac{N_B}{N}$

D. $\boldsymbol{\pi_A = \dfrac{N_A}{N}}$

E. $\pi_A = \dfrac{N}{N_A}$

F. $\pi_A = \dfrac{\sum_{n=1}^{N} A_n^2}{N}$

G. $\pi_A = \dfrac{N_C}{N}$

H. $\pi_A = \dfrac{\sum_{n=1}^{N} A_n B_n}{N}$

I. $\pi_A = \dfrac{\sum_{n=1}^{N} C_n}{N_A}$

J. None of the above.

4.5. (2 points) We shall find the MLE for π_B using the EM algorithm. Let $\widehat{\pi} = \sum_{n=1}^{N} \widehat{\pi}_n$. How should we update π_B in the M-step?

A. $\pi_B^{\text{new}} = \widehat{\pi}$

B. $\pi_B^{\text{new}} = \dfrac{\widehat{\pi}}{N}$

C. $\pi_B^{\text{new}} = \widehat{\pi} + \dfrac{\sum_{n=1}^{N} \alpha}{N}$

D. $\pi_B^{\text{new}} = \dfrac{\widehat{\pi}}{N} + \dfrac{\sum_{n=1}^{N} \alpha}{N}$

E. $\pi_B^{\text{new}} = \dfrac{\widehat{\pi}}{N + \sum_{n=1}^{N} \alpha}$

F. $\pi_B^{\text{new}} = \dfrac{N}{\widehat{\pi} + \sum_{n=1}^{N} \alpha}$

G. $\pi_B^{\text{new}} = \dfrac{\sum_{n=1}^{N} \alpha}{\widehat{\pi}}$

H. $\pi_B^{\text{new}} = \dfrac{\widehat{\pi} + N}{2N}$

I. None of the above.

4.6. (2 points) Recall that the posterior $p(B_n = 1|A_n, C_n) = \widehat{\pi}_n$. What is $\mathbb{E}_{p(B_n|A_n,C_n)}[2^{B_n}]$?
Hint: Consider the definition of expectation.

A. $\widehat{\pi}_n$

B. $1 + \widehat{\pi}_n$

C. $1 - \widehat{\pi}_n$

D. $2 + \widehat{\pi}_n$

E. $2 - \widehat{\pi}_n$

F. $2^{\widehat{\pi}_n}$

G. None of the above.

4.7. (2 points) We shall find the MLE for π_B using the EM algorithm. How should we update α in the M-step?

A. $\alpha^{\text{new}} = \frac{\sum_{n=1}^{N} C_n}{N}$

B. $\alpha^{\text{new}} = \frac{\sum_{n=1}^{N} C_n}{\sum_{n=1}^{N} 2^{A_n}(1+\widehat{\pi}_n)}$

C. $\alpha^{\text{new}} = \frac{\sum_{n=1}^{N} C_n}{\sum_{n=1}^{N} 2^{A_n}(1-\widehat{\pi}_n)}$

D. $\alpha^{\text{new}} = \frac{\prod_{n=1}^{N} C_n}{N}$

E. $\alpha^{\text{new}} = \frac{\prod_{n=1}^{N} C_n}{\prod_{n=1}^{N} 2^{A_n}(1+\widehat{\pi}_n)}$

F. $\alpha^{\text{new}} = \frac{\prod_{n=1}^{N} C_n}{\prod_{n=1}^{N} 2^{A_n}(1-\widehat{\pi}_n)}$

G. There is no closed form update for α and we need to use numerical optimization.

H. None of the above.

4.8. (3 points) We would now like to learn the MLE parameters of the model using variational inference (similar to the VAE). Suppose we define the variational distribution as

$$q(B_{1:N}) = \prod_{n=1}^{N} q_n(B_n)$$

where $q_n(B_n) = \text{Bern}[v_n]$. Which of the following is the ELBO?

A. $\sum_n A_n \log \pi_A + (1-A_n)\log(1-\pi_A) + v_n \log \pi_B + (1-v_n)\log(1-\pi_B) + C_n \log \alpha + C_n(A_n+v_n)\log 2 - \alpha 2^{A_n}(1+v_n) - \log(C_n!)$

B. $\sum_n A_n \log \pi_A + (1-A_n)\log(1-\pi_A) + v_n \log \pi_B + (1-v_n)\log(1-\pi_B) + C_n \log \alpha + C_n(A_n+v_n) - \alpha 2^{A_n}(1+v_n) - \log(C_n!)$

C. $\sum_n A_n \log \pi_A + v_n \log \pi_B + C_n \log \alpha + C_n(A_n+v_n)\log 2 - \alpha 2^{A_n}(1+v_n) - \log(C_n!)$

D. $\sum_n A_n \log \pi_A + v_n \log \pi_B + C_n \log \alpha + C_n(A_n+v_n) - \alpha 2^{A_n}(1+v_n) - \log(C_n!)$

E. $\sum_n A_n v_n \log(\pi_A \pi_B) + C_n \log \alpha + C_n(A_n+v_n)\log 2 - \alpha 2^{A_n}(1+v_n) - \log(C_n!)$

F. $\sum_n A_n v_n \log(\pi_A \pi_B) + C_n \log \alpha + C_n(A_n+v_n) - \alpha 2^{A_n}(1+v_n) - \log(C_n!)$

G. $\sum_n A_n \log \pi_A + (1-A_n)\log(1-\pi_A) + v_n \log \pi_B + (1-v_n)\log(1-\pi_B) + C_n \log \alpha + C_n(A_n+v_n) - \alpha 2^{A_n}(1-v_n) - \log(C_n!)$

H. $\sum_n A_n \log \pi_A + v_n \log \pi_B + C_n \log \alpha + C_n(A_n+v_n)\log 2 - \alpha 2^{A_n}(1-v_n) - \log(C_n!)$

I. $\sum_n A_n \log \pi_A + v_n \log \pi_B + C_n \log \alpha + C_n(A_n+v_n) - \alpha 2^{A_n}(1-v_n) - \log(C_n!)$

J. None of the above.

Problem 5. Variational Bayes (3 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.

5.1. (1 point) What is the *main* objective of variational inference?

A. To compute the maximum aposteriori estimate of the parameters.

B. To maximize the log-likelihood of the data.

C. To find a simpler distribution that closely approximates the true posterior distribution.

D. To maximize the posterior of the latent variables.

E. To compute expectations.

F. None of the above.

5.2. (1 point) Which of the following statements is true about the Kullback-Leilbler (KL) divergence $D_{\mathrm{KL}}[p\|q]$ between distributions p and q?

A. It is symmetric, $D_{\mathrm{KL}}[p\|q] = D_{\mathrm{KL}}[q\|p]$.

B. It is non-negative.

C. It is only well defined if for all x, $q(x) = 0 \implies p(x) = 0$.

D. It is equivalent to the ELBO.

E. A, B, and C

F. B and C

G. C and D

H. B, C, and D.

I. All of the above.

J. None of the above.

5.3. (1 point) Why is the reparameterization trick is used in the training of the VAE?

A. It allows for gradient-based optimization methods to be used.

B. It reduces the variance of the gradient estimates.

C. It enables us to compute the ELBO.

D. It is used to get the MLE of the variational parameters.

E. A, B, and C

F. B and C

G. C and D

H. B, C, and D.

I. All of the above.

J. 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.

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			
1.5		2.5		3.5		4.5			
		2.6				4.6			
		2.7				4.7			
						4.8			

END OF PAPER