MIDTERM EXAM - SOLUTIONS

STA 414/2104 WINTER 2026
University of Toronto

- Start: 18:15
- End: 20:15

Please check that your exam has 19 pages, including this one. The total possible number of points is 100.

Read the following instructions carefully:

1. Exam is closed book and internet. No calculators will be allowed during the midterm exam. You can use an optional **handwritten** aid sheet - A4 (or 8.5" × 11") double-sided.
2. If a question asks you to do some calculations, you must show your work for full credit.
3. Conceptual questions do not require long answers.
4. You will write your answers to each question in the space provided on the exam sheet. If you require additional paper, simply raise your hand.
5. After solving each question, you should write your answers immediately. Do not wait last minute to write them all at once.
6. Do not share the exam with anyone or in any platform!
7. Lastly, enjoy the problems!!!

1. Exponential families (11 points)

The lifetime of a piece of equipment can be modeled by a random variable X following a Weibull distribution $\mathcal{W}(\lambda, 1/2)$ with probability density function:

$$f_\lambda(x) = \frac{1}{2\lambda\sqrt{x}} \exp\left(-\frac{\sqrt{x}}{\lambda}\right) \mathbb{I}_{x>0}$$

where $\lambda > 0$ is a scale parameter.

(a) (4 points) Show that this is a valid probability density function. *Hint: Use the change of variable* $u = \sqrt{x}$.

Solution: First, $f_\lambda(x) \geqslant 0$ for all $x > 0$ since exponential and square root functions are positive. Next, we must show that $\int_0^\infty f_\lambda(x)dx = \int_0^\infty \frac{1}{2\lambda\sqrt{x}} \exp\left(-\frac{\sqrt{x}}{\lambda}\right) dx = 1$. Let $u = \sqrt{x}$, which implies $x = u^2$ and $dx = 2u\, du$. The limits remain 0 to ∞.

$$\int_0^\infty \frac{1}{2\lambda u} \exp\left(-\frac{u}{\lambda}\right) (2u\, du) = \int_0^\infty \frac{1}{\lambda} \exp\left(-\frac{u}{\lambda}\right) du = \left[-\exp\left(-\frac{u}{\lambda}\right)\right]_0^\infty = -(0-1) = 1.$$

(b) (3 points) Write the above distribution as an exponential family, and identify its sufficient statistics, natural parameter, and log-partition function.

Solution: We rewrite the pdf in the exponential family form $f(x) = h(x)\exp\{\eta T(x) - A(\eta)\}$:

$$\begin{aligned} f(x,\lambda) &= \mathbb{I}_{x>0}\frac{1}{\sqrt{x}} \exp\left\{\log\left(\frac{1}{2\lambda}\exp\left(-\frac{\sqrt{x}}{\lambda}\right)\right)\right\} \\ &= \mathbb{I}_{x>0}\frac{1}{\sqrt{x}} \exp\left\{-\log(2\lambda) - \frac{\sqrt{x}}{\lambda}\right\} \end{aligned}$$

Matching terms:

- **Sufficient Statistic:** $T(x) = \sqrt{x}$
- **Natural Parameter:** $\eta = -\frac{1}{\lambda}$ (which implies $\lambda = -\frac{1}{\eta}$)
- **Log-Partition Function:** $A(\eta) = \log(2\lambda) = \log\left(-\frac{2}{\eta}\right)$

(The term $\mathbb{I}_{x>0}\frac{1}{\sqrt{x}}$ is the base measure $h(x)$).

(c) (4 points) Assume that we observed $X_1, X_2, \ldots, X_n$ i.i.d. random variables from this Weibull distribution with an unknown parameter λ. Find the MLE for λ.

Solution: The log-likelihood is:

$$\ell(\lambda) = \sum_{i=1}^{n} \left(-\log(2\lambda) - \frac{1}{2}\log(X_i) - \frac{\sqrt{X_i}}{\lambda} \right)$$
$$= -n\log(2) - n\log(\lambda) - \frac{1}{2}\sum_{i=1}^{n}\log(X_i) - \frac{1}{\lambda}\sum_{i=1}^{n}\sqrt{X_i}$$

Taking the derivative w.r.t λ and setting to 0:

$$\frac{\partial \ell}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n}\sqrt{X_i} = 0$$

Multiply by λ^2:

$$-n\lambda + \sum_{i=1}^{n}\sqrt{X_i} = 0$$

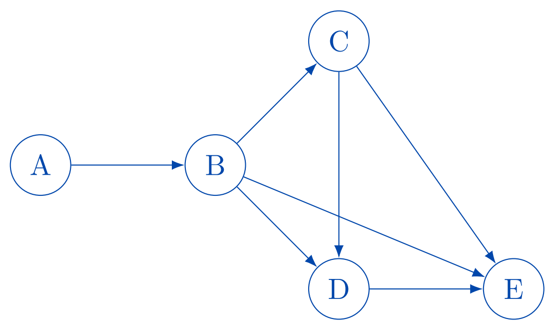
$$n\lambda = \sum_{i=1}^{n}\sqrt{X_i} \implies \hat{\lambda}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{X_i}$$

2. Graphical models (8 points)

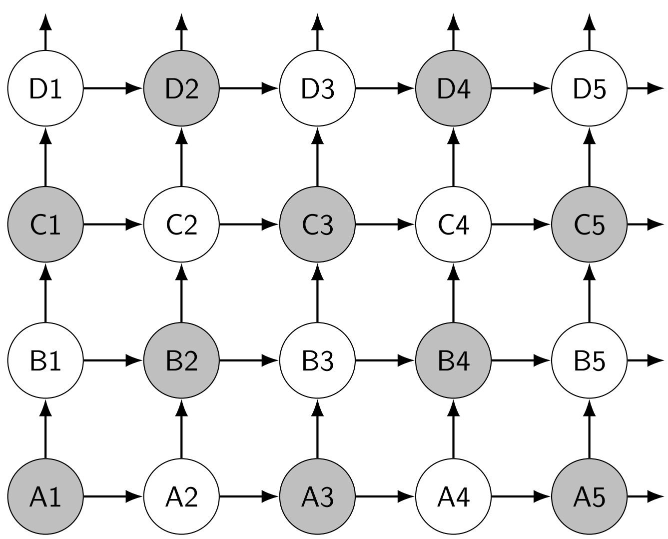
(a) (3 points) Draw the Directed Acyclic Graph corresponding to the following factorization of a joint distribution:

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

Solution:



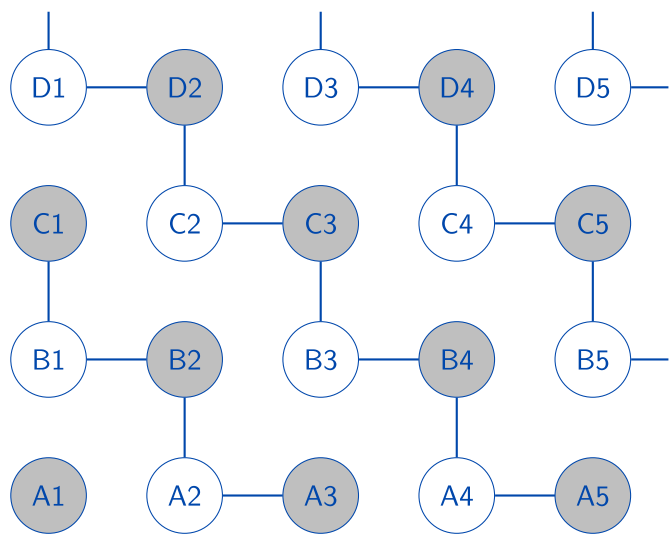
(b) (5 points) Consider the following "checkerboard" lattice of variables (assume it can be extended arbitrarily far up and to the right). The arrows point up and to the right.



Condition on the shaded nodes (assume they are observed). Identify the set of all unobserved nodes that are **conditionally** independent of the variable B_3.

Solution: **Answer: All the nodes not in** $\{A_4, D_1, C_2\}$
Justification: Applying the pruning algorithm, we remark that there is no leaf in this graph, so we only remove the edges originating from the observed (shaded) nodes. We then remove the arrow directions to analyze connectivity.

This results in the following graph (with one more row and column to better see what is happening):



We can see that the only nodes that will be connected to B_3 are $\{A_4, D_1, C_2\}$.

3. Decision Theory (9 points)

Company A owns a tract of land that may contain oil. Due to this prospect, Company B has offered to purchase the land. Alternatively, Company A can drill for oil itself, or enter a joint venture. Our current belief is that the probability of finding oil is $X\%$.

$$P(\text{Oil}) = \frac{X}{100}, \quad P(\text{Dry}) = \frac{100 - X}{100}$$

The management has 3 possible actions:

- **A) Sell:** Sell the land to Company B for a guaranteed **$90,000**.
- **B) Drill:** Drill for oil at a cost of $100,000. If oil is found, the revenue is $800,000 (net profit $700,000). If the land is dry, the cost is a loss of $100,000.
- **C) Joint Venture:** Enter a partnership. Company A receives a signing bonus S which follows a uniform distribution between $60,000 and $100,000. Additionally, Company A retains 50% of the profits (or losses) from drilling.

$$S \sim U(60,000, 100,000)$$

1. (5 points) For what range of probabilities X would **Selling the land** (Option A) yield the highest expected profit?

   Solution: We calculate the expected value (in thousands) for each option:

$$\begin{aligned} \mathbb{E}[Sell] &= 90 \\ \mathbb{E}[Drill] &= \frac{X}{100}(700) + \frac{100 - X}{100}(-100) = 7X - (100 - X) = 8X - 100 \\ \mathbb{E}[JV] &= \mathbb{E}[S] + 0.5 \times \mathbb{E}[Drill] = 80 + 0.5(8X - 100) = 80 + 4X - 50 = 4X + 30 \end{aligned}$$

   We compare Selling vs Joint Venture:

$$90 \geqslant 4X + 30 \implies 60 \geqslant 4X \implies X \leqslant 15$$

   We also check Selling vs Drill:

$$90 \geqslant 8X - 100 \implies 190 \geqslant 8X \implies X \leqslant 23.75$$

   Thus, if $0 \leqslant X \leqslant 15$, Selling is the best option.

2. (4 points) For what range of probabilities X would the **Joint Venture** (Option C) yield the highest expected profit?

   Solution: We need $\mathbb{E}[JV] \geqslant \mathbb{E}[Sell]$ and $\mathbb{E}[JV] \geqslant \mathbb{E}[Drill]$. From the first question, we know $\mathbb{E}[JV] \geqslant \mathbb{E}[Sell]$ when $X \geqslant 15$. Now compare JV vs Drill:

   $$4X + 30 \geqslant 8X - 100$$

   $$130 \geqslant 4X$$

   $$32.5 \geqslant X$$

   Thus, if $15 \leqslant X \leqslant 32.5$, the Joint Venture is the best option. (Note: For $X > 32.5$, Drilling becomes the best option).

4. Simple Monte Carlo (17 points)

Imagine we have a dangerous pollution prediction model that outputs samples of

$$P(A_1, A_2, \dots, A_T | \text{measurements})$$

where each A_i is a Bernoulli random variable indicating whether the level of pollution is dangerous (1) or not (0) on the ith day ahead. We are given a set of N i.i.d. samples from this joint predictive distribution:

$$a_1^{(1)}, a_2^{(1)}, \dots, a_T^{(1)} \sim P(A_1, A_2, \dots, A_T | \text{measurements})$$
$$\vdots$$
$$a_1^{(N)}, a_2^{(N)}, \dots, a_T^{(N)} \sim P(A_1, A_2, \dots, A_T | \text{measurements})$$

A note on notation: In the questions below, you **won't** need to use the indicator function, you will simply use the random variables themselves to describe the events. For example, $a_1^{(1)} = 1$ means that there is a dangerous level of pollution on the first day.

1. (2 points) Write an estimator for the probability that day 2 does not have a dangerously high level of pollution.

   Solution:

   $$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} (1 - a_2^{(i)})$$

2. (3 points) For each of the following estimators for the probability of observing zero days with dangerous pollution levels, state whether the estimator is unbiased. No justification is required.

   (a) Estimator 1: $\frac{1}{N} \sum_{i=1}^{N} \prod_{t=1}^{T} (1 - a_t^{(i)})$

   (b) Estimator 2: 0

   (c) Estimator 3: $\prod_{t=1}^{T} (1 - a_t^{(1)})$

   Solution: only estimators 1 and 3 are unbiased.

3. (3 points) Write an estimator for the probability that there is a dangerously high level of pollution on day 1 but not on day 3.

   Solution:

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} a_1^{(1)} (1 - a_3^{(i)})$$

4. (5 points) Write an estimator for the probability that day 1 has a dangerously high level of pollution given day 2 did not.

   Solution:

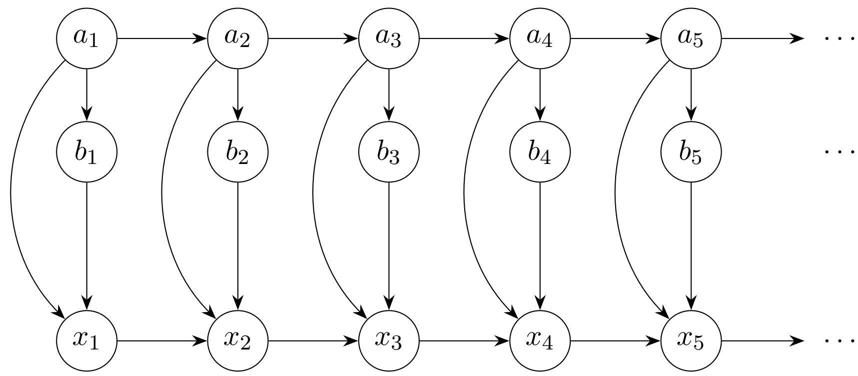
$$\frac{\sum_{i=1}^{N} a_1^{(1)} (1 - a_2^{(i)})}{\sum_{i=1}^{N} (1 - a_2^{(i)})}$$

5. (4 points) What is the variance of your estimator in question 1. as a function of N?

   Solution: If $p = P(A_2 = 0 \mid \text{measurements})$ is the true probability, the variance is $\frac{p(1-p)}{N}$.

5. Hidden Markov Models (17 points)

Given the following directed acyclic graphical model:



1. (2 points) Write the factorized joint distribution implied by this DAG (up to index T).

   Solution:

$$
\begin{aligned}
&p(a_1, \ldots, a_T, b_1, \ldots, b_T, x_1, \ldots, x_T) \\
&= p(a_1) \left[\prod_{t=2}^{T} p(a_t|a_{t-1}) \right] \left[\prod_{t=2}^{T} p(b_t|a_t) \right] p(x_1|a_1, b_1) \left[\prod_{t=2}^{T} p(x_t|x_{t-1}, a_t, b_t) \right]
\end{aligned}
$$

2. Assuming we stop the DAG at index T, each variable a_i can take one of K_a states, each variable b_i can take one of K_b states, and each variable x_i can take one of K_x states:

   - (1 points) How many states can this set of variables take on?
     Solution: $(K_a K_b K_x)^T$, where T is the length of the chain

- (2 points) How many parameters are required to parameterize the joint distribution?

  Solution:

  - a_1: $K_a - 1$
  - $a_t|a_{t-1}$: $(T-1) \times K_a(K_a - 1)$
  - $b_t|a_t$: $T \times K_a(K_b - 1)$
  - $x_1|a_1, b_1$: $K_a K_b(K_x - 1)$
  - $x_t|x_{t-1}, a_t, b_t$: $(T-1) \times K_x K_a K_b(K_x - 1)$

  This gives

  $$(K_a-1)+(T-1)\times K_a(K_a-1)+T\times K_a(K_b-1)+K_aK_b(K_x-1)+(T-1)\times K_xK_aK_b(K_x-1)$$

3.
- (2 points) Is $b_1 \perp b_2 \mid a_1$?

  Solution: Yes.
- (2 points) Is $b_1 \perp b_2 \mid a_2$?

  Solution: Yes.
- (2 points) Is $b_1 \perp b_2 \mid x_1$?

  Solution: No.
- (2 points) Is $x_2 \perp b_3 \mid x_3$?

  Solution: No.
- (2 points) Is $x_1 \perp x_3 \mid x_2, a_2$?

  Solution: Yes.
- (2 points) Is $b_1 \perp b_3 \mid a_2, b_2$?

  Solution: Yes.

6. Markov chains and their stationary distributions (16 points)

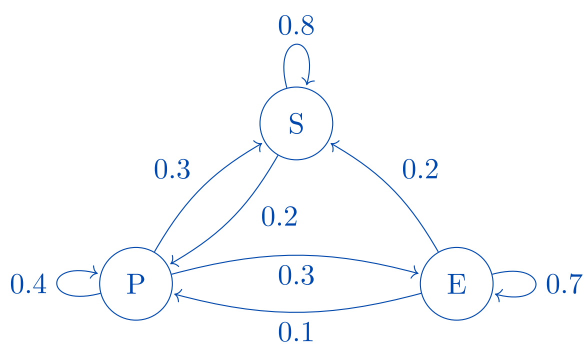
Consider a Markov chain $X_0, X_1, X_2, \dots$ representing the activity of a house cat. The state space is $S = \{S, E, P\}$, where the states represent Sleeping (S), Eating (E), and Playing (P). The transition matrix is given by:

$$P = \begin{pmatrix} 0.8 & 0 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

Note: The rows and columns are ordered S, E, P. For example, $P_{1,3} = 0.2$ is the probability that the cat goes from Sleeping to Playing.

1. (2 points) Draw the state diagram for this chain, labelling all the nodes and edges with their transition probabilities.

   Solution:



2. (3 points) Given the starting distribution $\pi_0 = \begin{pmatrix} 1/2 & 0 & 1/2 \end{pmatrix}$ (meaning the cat starts by Sleeping or Playing with equal probability), find the probability distribution at time $t = 1$.

   Solution: We compute the vector-matrix product $\pi_1 = \pi_0 P$:

$$\begin{pmatrix} 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0.8 & 0 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.55 & 0.15 & 0.3 \end{pmatrix}$$

   The distribution at $t = 1$ is 55% Sleeping, 15% Eating, and 30% Playing.

3. (6 points) Find the stationary distribution π of this Markov chain. The stationary distribution is given as the solution to the vector equation $\pi P = \pi$ (where π is a row vector).

Solution: Let $\pi = (\pi_S, \pi_E, \pi_P)$. We solve the system $\pi P = \pi$ subject to $\sum \pi_i = 1$:

$$1) \quad 0.8\pi_S + 0.2\pi_E + 0.3\pi_P = \pi_S$$
$$2) \quad 0\pi_S + 0.7\pi_E + 0.3\pi_P = \pi_E$$
$$3) \quad 0.2\pi_S + 0.1\pi_E + 0.4\pi_P = \pi_P$$

From (2): $0.3\pi_P = 0.3\pi_E \implies \pi_P = \pi_E$.

Substitute into (1): $0.8\pi_S + 0.2\pi_E + 0.3\pi_E = \pi_S \implies 0.5\pi_E = 0.2\pi_S \implies \pi_S = 2.5\pi_E = \frac{5}{2}\pi_E$. (we would obtain the same result substituting in (3)).

Normalize: $\pi_S + \pi_E + \pi_P = 1 \implies \frac{5}{2}\pi_E + \pi_E + \pi_E = 1 \implies 4.5\pi_E = 1$.

$$\pi_E = \frac{1}{4.5} = \frac{2}{9}, \quad \pi_P = \frac{2}{9}, \quad \pi_S = \frac{5}{9}$$
$$\pi = \left(\frac{5}{9}, \frac{2}{9}, \frac{2}{9}\right)$$

4. (5 points) Describe how to use the Metropolis-Hastings algorithm that uses this Markov chain (as the proposal distribution) to generate draws from the **uniform distribution** on $\{S, E, P\}$. Explicitly state the acceptance probability formula using the values from matrix P. What are the moves (given by sequences of the form $x \to x'$) that are always accepted with probability 1?

Solution: We want to sample from the target distribution $\pi_{target}(x) = \frac{1}{3}$ for all $x \in \{S, E, P\}$. The proposal distribution is $q(x'|x) = P_{x,x'}$.

The steps of the Metropolis-Hastings algorithm are:

(a) Initialize the state $x^{(0)} \in \{S, E, P\}$ (or select it randomly).

(b) For $t = 0, 1, 2, \ldots$:

- Propose a candidate state x' from the transition matrix $q(x'|x^{(t)})$.
- Calculate the acceptance probability:

$$A(x^{(t)}, x') = \min\left(1, \frac{\pi_{target}(x')q(x^{(t)}|x')}{\pi_{target}(x^{(t)})q(x'|x^{(t)})}\right) = \min\left(1, \frac{P_{x',x^{(t)}}}{P_{x^{(t)},x'}}\right)$$

- Accept the proposal with probability $A(x^{(t)}, x')$ and set $x^{(t+1)} = x'$. Otherwise, reject the proposal and set $x^{(t+1)} = x^{(t)}$.

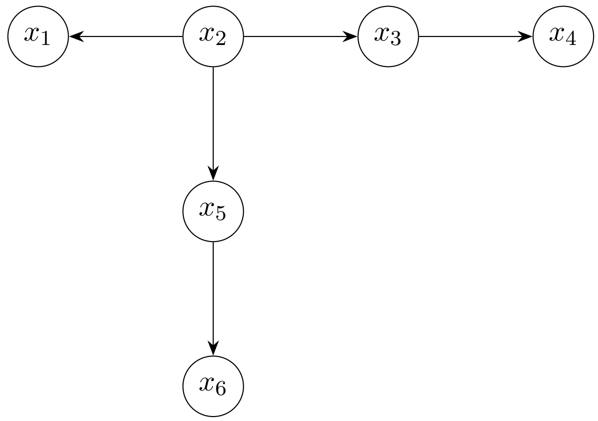
The moves $x \to x'$ that are always accepted with probability 1 are those where $P_{x',x} \geqslant P_{x,x'}$. Looking at the matrix P, these moves are:

- $S \to P$: $A(S, P) = \min(1, 0.3/0.2) = 1$
- $E \to P$: $A(E, P) = \min(1, 0.3/0.1) = 1$
- Self-transitions ($S \to S$, $E \to E$, $P \to P$) are also accepted with probability 1.

We also accept $S \to E$ though it never happens in the proposal step.

7. Belief propagation (13 points)

Given the following graph of binary variables:



Select x_2 as root. Suppose we observe $\bar{x}_1 = 0$ (first state) and $\bar{x}_3 = 1$ (second state). The node potentials are $\psi_6(x_6) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, and $\psi_i(x_i) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ for all other $i \neq 6$. The edge potentials depend on the specific edge:

- For edges connected to the root (edges $2 \to 1$, $2 \to 3$, $2 \to 5$):

$$\psi_{root}(x_2, x_i) = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$$

- For the bottom edge ($5 \to 6$, $3 \to 4$):

$$\psi_{bottom} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

1. (3 points) Calculate the message from the observed node 3 to the root 2: $m_{3\to 2}(x_2)$.

   Solution: The edge is $x_2 \to x_3$, so the potential is $\psi(x_2, x_3)$. Since x_3 is observed as state 1 (the second index), we must take the **second column** of the root potential.

$$m_{3\to 2}(x_2) = \psi_i(\bar{x}_3)\psi_{root}(x_2, \bar{x}_3 = 1) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

2. (5 points) Calculate the message $m_{5\to 2}(x_2)$.

Solution:

Step 1: Message $6 \to 5$. Arrow is $x_5 \to x_6$, so potential is $\psi(x_5, x_6)$. We sum over x_6.

$$m_{6\to 5}(x_5) = \sum_{x_6} \psi_{bottom}(x_5, x_6)\psi_6(x_6)$$
$$= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 * 2 + 1 * 1 \\ 1 * 2 + 1 * 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

The message is uniform and does not bring any information, which is expected as we just eliminated an unobserved leaf variable. So, we also accept answers starting directly with Step 2.

Step 2: Message $5 \to 2$. Arrow is $x_2 \to x_5$, so potential is $\psi_{root}(x_2, x_5)$. We sum over x_5.

$$m_{5\to 2}(x_2) = \sum_{x_5} \psi_{root}(x_2, x_5)\psi_5(x_5)m_{6\to 5}(x_5)$$
$$= \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \left[\begin{pmatrix} 1 \\ 1 \end{pmatrix} \odot \begin{pmatrix} 3 \\ 3 \end{pmatrix} \right] = \begin{pmatrix} 2 * 3 + 1 * 3 \\ 1 * 3 + 3 * 3 \end{pmatrix} = \begin{pmatrix} 9 \\ 12 \end{pmatrix}$$

3. (5 points) Calculate the normalized posterior $p(x_2|\bar{x}_1, \bar{x}_3)$.

Solution: We need to combine messages from all neighbors (x_1, x_3, x_5).

1. From observed $x_1 = 0$ (first state): Take the **first column** of ψ_{root}.

$$m_{1\to 2}(x_2) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

2. From observed x_3: Calculated in Q1.

$$m_{3\to 2}(x_2) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

3. From branch x_5: Calculated in Q2.

$$m_{5\to 2}(x_2) = \begin{pmatrix} 9 \\ 12 \end{pmatrix}$$

Unnormalized Belief:

$$\tilde{p}(x_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \odot \begin{pmatrix} 2 \\ 1 \end{pmatrix} \odot \begin{pmatrix} 1 \\ 3 \end{pmatrix} \odot \begin{pmatrix} 9 \\ 12 \end{pmatrix} = \begin{pmatrix} 2 \cdot 1 \cdot 9 \\ 1 \cdot 3 \cdot 12 \end{pmatrix} = \begin{pmatrix} 18 \\ 36 \end{pmatrix}$$

Sum $= 18 + 36 = 54$.

Normalized:

$$p(x_2|\bar{x}_1, \bar{x}_3) = \frac{18}{54} = 1/3, \quad p(x_2|\bar{x}_1, \bar{x}_3) = \frac{36}{54} = 2/3$$

8. Rejection Sampling (9 points)

Let X be a random variable following a Gamma distribution with parameter α, where $0 < \alpha < 1$. Its probability density function is given by:

$$f_\alpha(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \mathbb{1}_{\{x>0\}}$$

We wish to simulate draws from this distribution using the Rejection Sampling method. Consider the auxiliary (proposal) density $g_\alpha(x)$ defined as:

$$g_\alpha(x) = \frac{\alpha e}{\alpha + e} \left(x^{\alpha-1} \mathbb{1}_{\{0<x<1\}} + e^{-x} \mathbb{1}_{\{x \geqslant 1\}} \right)$$

1. (4 points) Find the smallest constant C_α such that $f_\alpha(x) \leqslant C_\alpha g_\alpha(x)$ for all $x > 0$. *Hint: Analyze the ratio or bound the term e^{-x} on the interval $(0, 1)$ and the term $x^{\alpha-1}$ on the interval $[1, \infty)$.*

   Solution: We check the inequality on the two intervals defined by the indicator functions in $g_\alpha(x)$.

   Case 1: $0 < x < 1$. In this interval, $g_\alpha(x) \propto x^{\alpha-1}$. We know that $e^{-x} \leqslant 1$ for $x > 0$. Thus:

   $$x^{\alpha-1} e^{-x} \leqslant x^{\alpha-1}$$

   Case 2: $x \geqslant 1$. In this interval, $g_\alpha(x) \propto e^{-x}$. Since $0 < \alpha < 1$, the term $\alpha - 1$ is negative, so $x^{\alpha-1}$ is a decreasing function. Thus, for $x \geqslant 1$, $x^{\alpha-1} \leqslant 1$. Thus:

   $$x^{\alpha-1} e^{-x} \leqslant e^{-x}$$

   Combining these bounds, we can write:

   $$x^{\alpha-1} e^{-x} \leqslant x^{\alpha-1} \mathbb{1}_{\{0<x<1\}} + e^{-x} \mathbb{1}_{\{x \geqslant 1\}}$$

   The density $f_\alpha(x)$ is $\frac{1}{\Gamma(\alpha)}$ times the LHS. The density $g_\alpha(x)$ is $\frac{\alpha e}{\alpha+e}$ times the RHS. To satisfy $f_\alpha(x) \leqslant C_\alpha g_\alpha(x)$, we substitute the densities:

   $$\frac{1}{\Gamma(\alpha)} \left(x^{\alpha-1} e^{-x} \right) \leqslant C_\alpha \frac{\alpha e}{\alpha + e} \left(x^{\alpha-1} \mathbb{1}_{\{0<x<1\}} + e^{-x} \mathbb{1}_{\{x \geqslant 1\}} \right)$$

   Using our derived bound, it is sufficient that:

   $$\frac{1}{\Gamma(\alpha)} \leqslant C_\alpha \frac{\alpha e}{\alpha + e}$$

   Solving for C_α, we can select:

   $$C_\alpha = \frac{\alpha + e}{\alpha e \Gamma(\alpha)}$$

2. (3 points) Recall the Rejection Sampling algorithm. Describe explicitly the steps to generate a single realization $X \sim f_\alpha$ using independent draws from g_α and a Uniform distribution.

Solution:

(a) Generate Y from the proposal distribution g_α.

(b) Generate $U \sim \mathcal{U}[0, 1]$ independent of Y.

(c) Accept Y as a realization of X if:

$$U \leqslant \frac{f_\alpha(Y)}{C_\alpha g_\alpha(Y)}$$

(d) If the condition is not met, reject Y and return to Step 1 (repeat until acceptance).

3. (2 points) Let τ be the number of draws (iterations) before we accept the first sample. What is the distribution of τ and what is its expected value $\mathbb{E}[\tau]$ as a function of α?

Solution: The number of trials τ follows a **Geometric distribution** with success parameter $p = \frac{1}{C_\alpha}$.

The expected number of iterations is:

$$\mathbb{E}[\tau] = \frac{1}{p} = C_\alpha = \frac{\alpha + e}{\alpha e \Gamma(\alpha)}$$

SCRAP PAPER – WILL NOT BE GRADED.

SCRAP PAPER – WILL NOT BE GRADED.