Durham University

EXAMINATION PAPER

Examination Session:	Year:	Exam Code:
May/June	2026	MATH3421-WE01

Title:

Bayesian Computation and Modelling III

Time:	2 hours	
Additional Material provided:	Formula sheet	
Materials Permitted:	None	
Calculators Permitted:	Yes	Models Permitted: Casio FX83 series or FX85 series.

Instructions to Candidates:	Answer all questions. The indicative marks shown in brackets for the main parts of each question are given as a guide to the weighting the markers expect to apply. Write your answer in the white-covered answer booklet with barcodes. Begin your answer to each question on a new page.

Revision:	

ED01/2026

SECTION A

1. Consider a posterior density $\pi(\theta|\underline{x}) = \pi(\theta)f(\underline{x}|\theta)/k$ where $\theta \in \mathcal{S} \subseteq \mathbb{R}$, $\pi(\theta)$ is the prior density ascribed to θ, $f(\underline{x}|\theta)$ is the likelihood function and k is a normalising constant. Suppose that interest lies in estimating an expectation of the form

$$\mu_h = \mathbb{E}_{\pi(\theta|\underline{x})}[h(\theta)] = \int_{\mathcal{S}} h(\theta)\pi(\theta|\underline{x})d\theta$$

for some function $h(\cdot)$.

(a) Write down a self-normalised estimator $\tilde{\mu}_h$ of μ_h based on N independent draws $\theta_1, \ldots, \theta_N$ from the prior $\pi(\cdot)$. Give the (un-normalised) weight function, $w(\theta_i)$, explicitly. **[2]**

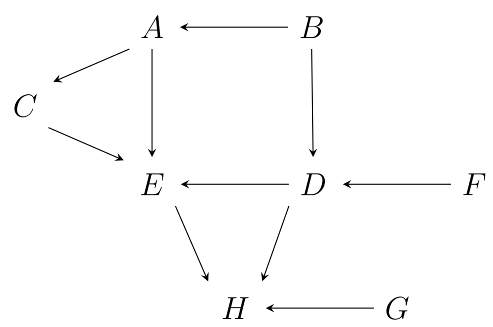
(b) What quantity does $\frac{1}{N}\sum_{i=1}^{N} w(\theta_i)$ estimate? Briefly explain your reasoning. **[2]**

(c) Show that $\tilde{\mu}_h$ is a consistent estimator of μ_h. **[4]**

(d) Briefly describe a potential drawback of this estimation procedure. **[2]**

---

2. Consider the following directed acyclic graph (DAG) that encodes the conditional independence relationships between eight variables.



(a) Consider the sub-DAG involving just D, H and G. Suppose that D, H and G are discrete variables with support on the sets $\mathcal{D}$, $\mathcal{H}$ and $\mathcal{G}$, respectively. In the joint distribution for D, H and G implied by this sub-DAG, show that D and G are marginally independent. **[3]**

(b) Consider the full DAG.

(i) Identify all the exogenous variables. **[1]**

(ii) Find an undirected path between A and F on which E is a collider. **[1]**

(iii) Are A and F d-separated by $\mathcal{S} = \{E\}$? Explain your answer. **[2]**

(iv) Can it be concluded that A and F are marginally independent? Explain your answer. **[3]**

SECTION B

3. (a) Consider the simulation of values from a Gamma(2, 1) distribution, truncated to the right at 1. We wish to sample this target using a Metropolis-Hastings independence sampler based on U(0, 1) proposals.

(i) Write down and simplify the acceptance probability for a move from a current value θ to a proposed value ϕ. **[2]**

(ii) If the Markov chain is currently at θ, find the probability that the chain will move (marginalised over the distribution of proposed values).
Hint: you may assume that the function xe^{-x} is increasing for $x \in (0, 1)$. **[4]**

(b) Suppose that you wish to generate samples from a posterior distribution with density $\pi(\theta|\underline{x}) \propto \pi(\theta)f(\underline{x}|\theta)$ and support $\mathcal{S} \subseteq \mathbb{R}$. Consider a Markov chain defined by an algorithm which performs the following steps for $j = 1, 2, \ldots$

Algorithm. At state $\theta^{(j-1)}$:

- **Step 1.** Draw $\theta^* \sim \pi(\cdot)$. (Sample the prior.)
- **Step 2.** With probability

$$\alpha_{\mathrm{B}}(\theta^*|\theta^{(j-1)}) = \frac{f(\underline{x}|\theta^*)}{f(\underline{x}|\theta^{(j-1)}) + f(\underline{x}|\theta^*)}$$

put $\theta^{(j)} = \theta^*$, otherwise put $\theta^{(j)} = \theta^{(j-1)}$.
- **Step 3.** Increment j and go to **Step 1**.

(i) Write down the transition kernel, $p(\phi|\theta)$, of the Markov chain defined by this algorithm. **[3]**

(ii) Show that the Markov chain defined by this algorithm is reversible with stationary density $\pi(\theta|\underline{x})$.
Hint: show that the detailed balance equation $\pi(\theta|\underline{x})p(\phi|\theta) = \pi(\phi|\underline{x})p(\theta|\phi)$ holds. **[3]**

(iii) The Metropolis-Hastings algorithm uses

$$\alpha_{\mathrm{MH}}(\theta^*|\theta^{(j-1)}) = \min\left\{1, \frac{f(\underline{x}|\theta^*)}{f(\underline{x}|\theta^{(j-1)})}\right\}$$

in Step 2 (but all other steps are unchanged). By comparing $\alpha_{\mathrm{B}}(\theta^*|\theta^{(j-1)})$ and $\alpha_{\mathrm{MH}}(\theta^*|\theta^{(j-1)})$, briefly explain why the Metropolis-Hastings algorithm is preferred. **[3]**

4. (a) Consider a logistic regression model, $\mathcal{M}_1$, for n independent pairs of observations, $(x_1, y_1), \ldots, (x_n, y_n)$:

$$\begin{aligned} Y_i|\alpha, \beta, x_i &\sim \text{Bernoulli}\left\{\frac{1}{1+\exp(-\alpha-\beta x_i)}\right\}, \\ \alpha &\sim \text{N}(0, \sigma_\alpha^2), \\ \beta &\sim \text{N}(0, \sigma_\beta^2). \end{aligned}$$

(i) Prior to modelling, the predictor variables, $x_1, \ldots, x_n$, have been standardised such that $\gamma = 1/\{1+\exp(-\alpha)\}$ can be interpreted as the probability that $Y = 1$ when the predictor variable is equal to its mean value. Calculate the prior density function for γ, $\pi(\gamma)$. **[3]**

(ii) Show that an equation satisfied by all the turning points, γ^*, of $\pi(\gamma)$ is

$$\text{logit}(\gamma^*) = \sigma_\alpha^2(2\gamma^* - 1),$$

where $\text{logit}(\gamma^*) = \log\{\gamma^*/(1-\gamma^*)\}$ is the logit function. Hence show that $\gamma^* = 1/2$ is a turning point for all values of σ_α^2. **[3]**

(iii) Before seeing any data, you believe that the most likely value for γ is $1/2$. Suggest a sensible upper bound for the prior standard deviation σ_α of α, giving justification for your answer. **Hint**: you may find it useful to consider the second derivative of $\log\{\pi(\gamma)\}$. **[5]**

(b) A second model, $\mathcal{M}_2$, is also considered for the same data:

$$\begin{aligned} Y_i|\alpha, \beta, \delta, x_i &\sim \text{Bernoulli}\left\{\frac{1}{1+\exp(-\alpha-\beta x_i - \delta x_i^2)}\right\}, \\ \alpha &\sim \text{N}(0, \sigma_\alpha^2), \\ \beta &\sim \text{N}(0, \sigma_\beta^2), \\ \delta &\sim \text{N}(0, 1). \end{aligned}$$

(i) Explain why the Savage-Dickey density ratio can be used to calculate the Bayes factor to compare models $\mathcal{M}_1$ and $\mathcal{M}_2$. **[2]**

(ii) Before seeing the data, you believe that $\mathcal{M}_2$ is twice as likely as $\mathcal{M}_1$. Find the set of values of $\pi(\delta = 0|\mathbf{x}, \mathbf{y}, \mathcal{M}_2)$ for which the posterior odds for $\mathcal{M}_1$ against $\mathcal{M}_2$ exceed one. **[2]**
