Durham University

EXAMINATION PAPER

Examination Session:	Year:	Exam Code:
May/June	2022	MATH4031-WE01

Title: Bayesian Statistics IV

Time:	3 hours	
Additional Material provided:		
Materials Permitted:		
Calculators Permitted:	Yes	Models Permitted: Casio FX83 series or FX85 series.

Instructions to Candidates:	Answer all questions. Section A is worth 20%, Section B is worth 60%, and Section C is worth 20%. Within Sections A and B, all questions carry equal marks. Students must use the mathematics specific answer book.

Revision:	

ED01/2022

SECTION A

Q1 Consider observables $\{x_i\}_{i=1}^{n}$ independently sampled from a sampling distribution with mass function

$$\mathrm{Pn}(x_i|\theta) = \frac{\theta^{x_i}\exp(-\theta)}{x_i!}\,1(x_i \in \{0,1,2,...\})\,,$$

with unknown parameter $\theta \in (0,\infty)$.

(a) Compute Jeffreys' prior for θ up to a multiplicative constant.

(b) Assume that you have only one observation $x_1 = 0$ in your sample. Can the Jeffreys' prior computed in part (a) be used to perform statistical inference?

Q2 In a Bayesian network, consider a directed acyclic graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$. Let $\mathcal{R} = \mathcal{V}\backslash\{w\}$, and let $\mathrm{Bl}\,(w)$ denote the Markov blanket of vertex w.

(a) Show that

$$f\,(w|\mathcal{R}) = \frac{\prod_{u\in\mathcal{V}} f\,(u|\mathrm{pa}\,(u))}{\int \prod_{u\in\mathcal{V}} f\,(u|\mathrm{pa}\,(u))\,\mathrm{d}w}$$

(b) Continue part (a) and prove that

$$f\,(w|\mathcal{R}) = f\,(w|\mathrm{Bl}\,(w))$$

Hint: You can use the factorization

$$\prod_{u\in\mathcal{V}} f\,(u|\mathrm{pa}\,(u)) = f\,(w|\mathrm{pa}\,(w)) \prod_{u\in\mathrm{ch}(w)} f\,(u|\mathrm{pa}\,(u)) \prod_{x\in\mathcal{X}} f\,(x|\mathrm{pa}\,(x))$$

where $\mathcal{X} = \mathcal{V}\backslash\{w,\mathrm{ch}(w)\}$ and $\mathrm{ch}(w)$ are the children of w.

SECTION B

Q3 Consider a sequence of observables $\{y_i\}_{i=1}^{n}$ drawn independently from a sampling distribution which admits density

$$f(t|\theta) = \begin{cases} \sqrt{\frac{2}{\pi}} t^2 \exp\left(-\frac{1}{2}\theta t^2 + \frac{3}{2}\log(\theta)\right) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases},$$

labelled by an unknown parameter $\theta \in (0, \infty)$.

Hint-1: The Gamma distribution is denoted $\text{Ga}(a, b)$ and has pdf

$$f(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1(x \geq 0)$$

with $a \in (0, \infty)$, $b \in (0, \infty)$, and mean $\text{E}(x|a, b) = \frac{a}{b}$.

(a) Show that the sampling distribution is an exponential family of distributions. Compute the minimal sufficient statistic.

(b) Derive a prior probability density function for θ which is conjugate to the likelihood. State the name of the distribution along with its parameters.

(c) Consider (without proof) that the posterior distribution of θ is

$$\theta|y \sim \text{Ga}(a^*, b^*)$$

with $a^* = \frac{3}{2}n + a$ and $b^* = \frac{1}{2}\sum_{i=1}^{n} y_i^2 + b$. Compute the posterior predictive density function of a future outcome $z = y_{n+1}$.

(d) Compute the Bayesian point estimator $\hat{\theta}$ for θ under the loss function

$$\ell(\delta, \theta) = w\frac{1}{n}\sum_{i=1}^{n}(y_i - \delta)^2 + (1 - w)(\delta - \theta)^2$$

where $w \in (0, 1)$ is a known constant. In particular, show that

$$\hat{\theta} = w\bar{y} + (1 - w)\frac{a^*}{b^*}$$

Hint-2: You may use without proof that $\sum_{i=1}^{n}(y_i - \delta)^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\delta - \bar{y})^2$, where $\bar{y}$ is the arithmetic average of $\{y_i\}$.

Q4 Consider the Bayesian model

$$\begin{cases} y_i|\theta_i & \overset{\text{ind.}}{\sim} \text{Pa}(L, \theta_i), \quad i = 1, ..., n \\ \theta_i & \overset{\text{iid}}{\sim} \text{Ga}(a, b) \end{cases},$$

where $L \in (0, \infty)$, $a \in (0, \infty)$ and $b \in (0, \infty)$ are unknown hyper-parameters.

Hint-1: The Pareto distribution is denoted $\text{Pa}(L, a)$ and has pdf

$$f(x|L, a) = \frac{aL^a}{x^{a+1}} 1(x \geq L).$$

Hint-2: The Gamma distribution is denoted $\text{Ga}(a, b)$ and has pdf

$$f(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1(x \geq 0).$$

Hint-3: If $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log(\Gamma(x))$, then $\psi(x+1) = \psi(x) + \frac{x}{2}$.

(a) Derive the equations that are satisfied by the values $\hat{a}$ and $\hat{b}$ of a and b and which can be computed by the method ML-II. The system of these equations should contain one equation that can be solved analytically and one which cannot.

(b) Assume that the values $\hat{a}$ and $\hat{b}$ from part (a) are available. Compute the Empirical Bayes estimator $\hat{\theta}_i^{\text{EB}}$ for θ_i under the loss function

$$\ell(\delta, \theta) = \frac{\delta}{\theta} - \log\left(\frac{\delta}{\theta}\right) - 1.$$

Q5 (a) Consider the Bayesian model

$$\begin{cases} y_i|\theta & \overset{\text{ind}}{\sim} \text{Exp}\left(\frac{1}{\theta}\right), \quad i = 1, ..., n \\ \theta & \sim \mathrm{d}\Pi(\theta) \propto \frac{1}{\theta} \mathrm{d}\theta \end{cases}$$

Compute the Laplace approximation to the posterior expectation $\text{E}(\theta|y)$, where $y = (y_1, ..., y_n)$.

Hint: The Exponential distribution is denoted $\text{Exp}(\lambda)$ and has pdf

$$f(x|\lambda) = \lambda \exp(-\lambda x) 1(x \geq 0)$$

(b) Assume that the only available random number generator is that for simulating a uniform distribution in the interval $[0, 1]$. Design a random number generator that simulates from a continuous probability distribution with density function

$$f(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) 1(x \in \mathbb{R}),$$

where $b > 0$ and $\mu \in \mathbb{R}$ are known constants. This random number generator should require the generation of no more than two uniform random variates per call. Write down the pseudo-algorithm of your random number generator. Explain your working and include your calculations in the solutions.

Q6 (a) Consider the following Bayesian model

$$\begin{cases} y_i|\gamma,k \overset{\text{ind}}{\sim} \text{Poi}(\gamma) & \text{for } i=1,2,...,k \\ y_i|p,k \overset{\text{ind}}{\sim} \text{Geo}(p) & \text{for } i=k+1,k+2,...,n \\ \gamma \sim \text{Ga}(a,b) \\ p \sim \text{Be}(c,d) \\ k \sim \pi(k) \propto 1(k \in \{1,2,3,...,n-1\}) \end{cases}$$

Let $y=(y_1,...,y_n)$ be a vector of observables. Assume that $a \in (0,\infty)$, $b \in (0,\infty)$, $c \in (0,\infty)$, $d \in (0,\infty)$ and $n \in \mathbb{N}-\{0\}$ are fixed constants. Design a Gibbs sampler that targets the posterior distribution as stationary distribution.

Hint-1 $\text{Poi}(\lambda)$ has mass function $f(x|\lambda) = \frac{\lambda^x \exp(-\lambda)}{x!} 1(x \in \{0,1,2,...\})$.
Hint-2 $\text{Geo}(p)$ has mass function $f(x|p) = p(1-p)^x \; 1(x \in \{0,1,2,...\})$.
Hint-3 $\text{Ga}(a,b)$ has density function $f(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \; 1(x>0)$.
Hint-4 $\text{Be}(a,b)$ has density function $f(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \; 1(x \in [0,1])$.

(b) Let $\pi(\cdot)$ and $q(\cdot|\cdot)$ be density functions. Assume $s(x,y)$ is any symmetric positive function satisfying $s(x,y) \leq 1 + \frac{\pi(x)q(y|x)}{\pi(y)q(x|y)}$. Prove that the following recursive algorithm simulates a reversible Markov chain whose stationary distribution admits density $\pi(\cdot)$.

Algorithm: At state $x^{(t-1)}$,
step 1 Draw $y \sim q\left(y|x^{(t-1)}\right)$
step 2 Draw $u \sim \text{U}(0,1)$; that is the Uniform distribution
step 3 Set:

$$x^{(t)} = \begin{cases} y & a^{\text{B}}\left(y|x^{(t-1)}\right) \geq u \\ x^{(t-1)} & \text{otherwise} \end{cases},$$

where

$$a^{\text{B}}(y|x) = \frac{s(x,y)}{1 + \frac{\pi(x)q(y|x)}{\pi(y)q(x|y)}}.$$

(c) Show that the acceptance probability $a^{\text{B}}(y|x)$ of the Algorithm in part (b) with $s(\cdot,\cdot)=1$ is less than or equal to the acceptance probability $a^{\text{MH}}(y|x)$ of the Metropolis-Hastings algorithm targeting distribution $\pi(\cdot)$ with proposal distribution $q(\cdot|\cdot)$, for any x,y.

SECTION C

Q7 Let $p(x, z)$ be the joint density of a Bayesian model of interest with posterior density $p(z|x)$ and evidence $p(x)$. Let $q(z)$ be another density.

(a) Show that the following decomposition is valid

$$\log (p(x)) = \mathcal{L}(q) + \mathrm{KL}\left(q\middle|\middle|p\right)$$

where

$$\mathcal{L}(q) = \int q(z) \log\left(\frac{p(x, z)}{q(z)}\right) \mathrm{d}z$$
$$\mathrm{KL}\left(q\middle|\middle|p\right) = -\int q(z) \log\left(\frac{p(z|x)}{q(z)}\right) \mathrm{d}z$$

(b) Suppose that z is partitioned into M disjoint components $z = (z_1, z_2, ..., z_M)$ and assume that $q(\cdot)$ admits factorization

$$q(z) = \prod_{i=1}^{M} q_i(z_i)$$

Prove that $\mathcal{L}(q)$ can be expressed as

$$\mathcal{L}(q) = \int q_j(z_j) \log(\tilde{p}(x, z_j))\, \mathrm{d}z_j - \int q_j(z_j) \log(q_j(z_j))\, \mathrm{d}z_j + \text{const} \quad (1)$$

for any $j = 1, ..., M$, and compute the quantity $\log(\tilde{p}(x, z_j))$.

(c) Show that the optimal solution $q_j^{\star}(z_j)$ in Equation 1 according to the Variational Inference method for Bayesian inference, is

$$q_j^{\star}(z_j) = \frac{\exp(\mathrm{E}_{i\neq j}(\log(p(x, z))))}{\int \exp(\mathrm{E}_{i\neq j}(\log(p(x, z))))\, \mathrm{d}z_j}$$

where $\mathrm{E}_{i\neq j}(\cdot)$ denotes the expected value under the marginal distribution $\prod_{\{\forall i: i\neq j\}} q_i(z_i)$.

(d) Consider the Bayesian model

$$y_i|\mu, \tau \overset{\text{iid}}{\sim} \mathrm{N}\left(\mu, \tau^{-1}\right), \quad i = 1, ..., N$$
$$(\mu, \tau) \sim \mathrm{d}\pi(\mu, \tau) \propto \tau^{-1} \mathrm{d}\mu\, \mathrm{d}\tau$$

Assume a variational approximation density to the posterior of the form

$$q(\mu, \tau) = q_{\mu}(\mu)\, q_{\tau}(\tau)$$

Compute the optimal solutions for $q_{\mu}(\mu)$ and $q_{\tau}(\tau)$ as Normal and Gamma distribution densities respectively, and compute the associated parameters of these densities.

Hint-1 You may use without proof that $\sum_i \frac{1}{y_i}(\theta - x_i)^2 = \frac{1}{\hat{y}}(\theta - \hat{x})^2 + \text{const}$ where $\hat{y} = \left(\sum_i y_i^{-1}\right)^{-1}$ and $\hat{x} = \hat{y}\left(\sum_i y_i^{-1} x_i\right)$.

Hint-2 The Gamma distribution is denoted $\mathrm{Ga}(a, b)$ and has pdf

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)\ 1(x > 0)$$

and expected value $\mathrm{E}(x|a, b) = a/b$.