EXAMINATION PAPER

Examination Session:	Year:	Exam Code:
May/June	2022	MATH3341-WE01

Title:
Bayesian Statistics III

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 40% and Section B is worth 60%. Within each section, 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.

Q3 Let $y = (y_1,...,y_n)$ be a sequence of n observables sampled from a Geometric sampling distribution $y_i|\theta \overset{\text{iid}}{\sim} \mathrm{Ge}(\theta)$ with mass function

$$f\,(t|\theta) = \begin{cases} (1-\theta)^t\,\theta & \text{if}\, t \in \{0,1,2,...\} \\ 0 & \text{otherwise} \end{cases},$$

and unknown parameter $\theta \in [0,1]$. Assume that $n = 6$ and $y_* = \sum_{i=1}^{n} y_i = 3$. Assume that the (overall) prior distribution of θ is a Uniform distribution. Consider that you wish to assess whether $\theta \le 1/3$ or otherwise with a Bayesian hypothesis test.

(a) Partition the (overall) prior distribution of θ according to the pair of hypotheses under consideration.

(b) Perform the Bayesian hypothesis test under consideration according to Jeffreys' scale rule, and write down the strength of evidence.

Hint-1: The Beta distribution is denoted $\mathrm{Be}\,(a,b)$ and has pdf

$$f\,(x|a,b) = \frac{\Gamma\,(a+b)}{\Gamma\,(a)\,\Gamma\,(b)} x^{a-1}\,(1-x)^{b-1}\;1(x\in[0,1])\,.$$

Hint-2: Jeffreys' scale rule

B_{01}	**Strength of evidence**
$(1, +\infty)$	**H_0 is supported**
$(10^{-1/2}, 1)$	**Evidence against H_0: not worth more than a bare mention**
$(10^{-1}, 10^{-1/2})$	**Evidence against H_0: substantial**
$(10^{-3/2}, 10^{-1})$	**Evidence against H_0: strong**
$(10^{-2}, 10^{-3/2})$	**Evidence against H_0: very strong**
$(0, 10^{-2})$	**Evidence against H_0: decisive**

Q4 Let $f(\cdot)$ and $g(\cdot)$ be probability densities such that there is a constant $M > 1$ satisfying $f(x) \leq Mg(x)$ for all $x \in \mathbb{R}$. Consider the following recursive algorithm.

Algorithm:

step 1. Draw x from distribution with pdf $g(\cdot)$

step 2. Draw u from the continuous uniform distribution $\mathrm{U}(0, 1)$

step 3. If $\left(u \leq \frac{f(x)}{Mg(x)}\right)$ then

return x

else

go to step 1

(a) Prove that the algorithm above is valid for sampling from a distribution with pdf $f(\cdot)$.

(b) Prove that the acceptance rate of the above algorithm is

$$\mathrm{P}(\text{accept}) = \frac{1}{M}.$$

What does this result suggest about how to choose $g(\cdot)$?

SECTION B

Q5 Consider a sequence of observables $\{y_i\}_{i=1}^{n}$ drawn independently from a sampling distribution which admits density

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

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

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

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

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

(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

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

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

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

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

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

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

Q6 Consider the Bayesian model

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

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 $\mathrm{Pa}\left(L, a\right)$ and has pdf

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

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

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

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

(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 $\hat{\theta}_i^{\mathrm{EB}}$ for θ_i under the loss function

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

Q7 (a) Consider the Bayesian model

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

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

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

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

(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

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

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.

Q8 (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 $s(x,y) \le 1 + \frac{\pi(x)q(y|x)}{\pi(y)q(x|y)}$. 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:

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

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.