

EXAMINATION PAPER

Examination Session: May/June

2020

Year:

Exam Code:

MATH3341-WE01

Title:

Bayesian Statistics III

Time (for guidance only):	3 hours	
Additional Material provided:		
Materials Permitted:		
Calculators Permitted:	Yes	Models Permitted: There is no restriction on the model of calculator which may be used.

Instructions to Candidates:	Credit will be given for your answers to all questions. All questions carry the same marks.		
	Please start each question on a new page. Please write your CIS username at the top of each page.		
	Show your working and explain your reasoning.		

Revision:



Q1 1.1 Let θ denote the average recovery time (in days) for patients with a disease. You are interested in specifying a suitable prior distribution for θ , and you wish to elicit it based on expert knowledge. Your expert judges that $\phi = \theta^{\alpha}$, for some $\alpha > 0$, has an exponential distribution with cumulative distribution function

$$F(\phi) = \begin{cases} 1 - \exp(-\beta\phi) & \text{for } \phi \in (0, +\infty) \\ 0 & \text{otherwise} \end{cases}$$

for some $\beta > 0$. In addition, from his graduate school studies, he judges that there is 25% probability that $\theta < 8$ days and 25% probability that $\theta > 15$ days. Specify the prior distribution for θ that you can use in your analysis by stating its probability density function and computing its hyper-parameter values.

1.2 Suppose that $y_1, ..., y_n$ are a random sample from a uniform distribution on the range $(0, \theta)$. Assume that $\gamma = \log(\theta)$ admits a Normal prior distribution with known mean μ and variance σ^2 . Calculate the posterior distribution of γ given the observations, including an explicit formula for the normalising constant.

Hint: $n\gamma + \frac{1}{2\sigma^2}(\gamma - \mu)^2 = \frac{1}{2\sigma^2}(\gamma - \mu + n\sigma^2)^2 + n\mu - \frac{1}{2}n^2\sigma^2$

- **1.3** An important theorem in relation to Bayesian networks states that: If A, B and S are collections of random quantities in a Bayesian network, then A and B are conditionally independent given S if A and B are separated by S in the moral graph of an ancestral subgraph containing $A \cup B \cup S$ of the directed acyclic graph of the Bayesian network.
 - (i) Consider a joint distribution for 7 random quantities which can be written in the form

 $p(a, b, c, d, e, f, g) = p(a)p(b)p(c \mid a, b)p(d \mid b)p(e \mid c, d)p(f \mid e, g)p(g \mid e).$

Draw the DAG representing this factorization.

- (ii) Illustrate the application of the theorem by enumerating all the sets of quantities V for which the theorem implies that $c \perp d \mid V$.
- 1.4 Use Laplace's method to derive an approximation for the normalizing constant of the following non-normalized posterior distribution of parameter $\theta \in \mathbb{R}$

$$f(\theta|y) \propto e^{(\alpha+y)\theta} e^{-\beta e^{\theta}},$$

where $y \in \mathbb{R}$, $\alpha > 0$ and $\beta > 0$.

- Exam code MATH3341-WE01
- **Q2** Let $y = (y_1, ..., y_n)$ be a sequence of *n* observables assumed to be iid according to a Log-Normal sampling distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, +\infty)$; i.e.

 $y_i | \mu, \sigma^2 \stackrel{\text{iid}}{\sim} \text{LN}(\mu, \sigma^2)$, i = 1, ..., n

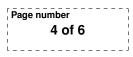
where μ is an unknown parameter, and σ^2 is assumed known.

Hint: The Log-Normal distribution denoted by $LN(\mu, \sigma^2)$ has density function

$$f(x|\mu,\sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(-\frac{1}{2} \frac{(\log(x)-\mu)^2}{\sigma^2}\right) &, \text{ if } x \in (0,+\infty)\\ 0 &, \text{ otherwise} \end{cases}$$

- **2.1** Show that the LN distribution with known σ^2 is an exponential family of distributions.
- **2.2** Compute the likelihood of y given (μ, σ^2) and its sufficient statistic.
- **2.3** Derive the prior distribution which is conjugate to the likelihood function of the problem.
- **2.4** Using the conjugate prior, construct the (1 a)100% HPD posterior credible interval for μ , and show your working. Compute the bounds of the 95% HPD posterior credible interval for μ , when there is available a dataset y = (0.05, 0.36, 0.13, 0.22, 0.60) of size n = 5 observations; $\sigma^2 = 1$; the prior mean of μ is 0; and the prior variance of μ is 10.

Hint: The 0.975-quantile of the standard Normal distribution is $z_{0.975}^* = 1.959964$.





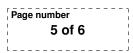
- **Q3** 3.1 Consider random variables $\theta \sim N(0, 1)$ and $y \mid \theta \sim N(\theta, 1)$, and a loss function $\ell(\theta, d) = (\theta d)^2$ for all $d \in \mathbb{R}$ and $\theta \in \mathbb{R}$. We denote as $N(\mu, \sigma^2)$, the Normal distribution with mean μ and variance σ^2 .
 - (i) Consider a decision rule δ_a , where $\delta_a(y) = ay$ for all $y \in \mathbb{R}$, and where $a \in \mathbb{R}$ is an arbitrary constant. Show that the (Frequentist) risk function for decision δ_a is

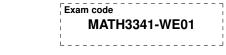
$$R(\theta, \delta_a) = (1-a)^2 \theta^2 + a^2.$$

- (ii) Show that δ_a is inadmissible when a < 0 or a > 1.
- (iii) Compute the Bayesian point estimator of θ under the aforesaid loss function and Bayesian model. State if this estimator is admissible and justify your answer.
- **3.2** Assume a 1-dimensional random quantity $x \sim Q(x|y)$, with unimodal density q(x|y). Show that the (1-a)-credible interval $C_a = [L, U]$ for x as a Bayesian rule C_a under the loss function

$$\ell(x, C_a; L, U) = k(U - L) - 1(x \in [L, U]), \quad \text{with} \quad k \in (0, \max_{\forall x \in \mathbb{R}} (q(x|y)))$$

is given by q(L) = q(U) = k, and $\mathsf{P}_Q(x \in [L, U]|y) = 1 - a$. Discuss known properties of the derived credible interval.





Q4 4.1 Consider the following probability density function, which is known up to a constant of proportionality

$$f(x) = \frac{e^x}{c},$$

where $x \in [0, 1]$. Assuming the numbers below are a sequence of independent random numbers uniformly distributed on [0, 1], generate 3 values from f(x) using inverse sampling.

$$0.156 \quad 0.579 \quad 0.936$$

4.2 The Kumaraswamy distribution is a flexible alternative to the beta distribution. The probability density function of this distribution is given by

$$f(x) = \alpha \beta x^{(\alpha-1)} (1 - x^{\alpha})^{(\beta-1)},$$

where $x \in (0, 1)$, $\alpha > 0$ and $\beta > 0$. Assuming that $\alpha = \beta = 2$, perform 3 iterations of rejection sampling from f(x) using a uniform distribution on [0, 1] as proposal distribution. State whether the generated values are accepted or not. Base your calculations on the following sequence of uniformly distributed random numbers between 0 and 1:

$$0.046 \quad 0.495 \quad 0.307 \quad 0.138 \quad 0.645 \quad 0.515$$

4.3 We want to use Gibbs sampling to sample from the joint distribution of A and B, which has probabilities proportional to the table below.

			B	
		1	2	3
	4	0.4	0.5	0.8
A	5	0.4	0.6	0.5
	6	$0.4 \\ 0.4 \\ 0.5$	0.7	0.9

Generate the output of the Gibbs sampler assuming availability of the sequence of uniform random numbers in [0, 1] given below and using as initial value $B^{(0)} = 2$.

 $0.963 \quad 0.801 \quad 0.526 \quad 0.039 \quad 0.101 \quad 0.675$



- **Q5** A colony of insects is studied with a view to quantifying variation in the number of eggs laid and in the rate at which eggs successfully develop. Let N_i denote the number of eggs laid by insect *i*, for i = 1, ..., n, and E_i the number of eggs which develop. Eggs laid by insect *i* develop independently with probability of success p_i . Expert judgment is that the variability between insects in numbers of eggs laid should be modelled well by a Poisson distribution with fixed rate λ . Furthermore, it is thought that a beta distribution with fixed parameters α and β would adequately describe variability between insects to p_i .
 - **5.1** Draw a directed acyclic graph using plate notation for the Bayesian network describing the joint distribution of the $\{N_i\}$, $\{E_i\}$ and $\{p_i\}$ based on the above, and add vertices for the parameters λ , α and β .
 - **5.2** Specify the distributions for the vertices $\{N_i\}$, $\{E_i\}$ and $\{p_i\}$ given their respective parents.
 - **5.3** On closer inspection, the experts are uneasy about the choice of fixed values of the parameters. Therefore, the model is modified by assigning exponential prior distributions with rate 1 for the parameters α and β of the beta distribution and an improper prior proportional to $1/\lambda$ for λ . Derive (up to multiplicative constants) all the conditional distributions required for Gibbs sampling. Which of the conditional distributions are known distributions and what are their parameters?
 - 5.4 Describe an efficient approach to generating values from each of the conditional distributions in part 5.3. You may refer to standard functions in R or name standard algorithms. For any algorithm you name, show that preconditions (if any) for its application are met. You may use the fact that the third derivative of $\log \Gamma(x)$ is negative for all positive x, where $\Gamma(\cdot)$ is the gamma function.

Hint-1: The Poisson distribution for $x \in \{0, 1, ...\}$ with parameter λ takes the form

$$P(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

Hint-2: The pdf of a beta-distributed random quantity $x \in (0, 1)$ with parameters a and b is

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

Hint-3: The exponential distribution for $x \in [0, \infty)$ with parameter θ takes the form

$$f(x|\theta) = \theta e^{-\theta x}$$