Main Examination period 2022 – January – Semester A

MTH6102: Bayesian Statistical Methods

You should attempt ALL questions. Marks available are shown next to the questions.

In completing this assessment:

- **You may use books and notes.**
- **You may use calculators and computers, but you must show your working for any calculations you do.**
- **You may use the Internet as a resource, but not to ask for the solution to an exam question or to copy any solution you find.**
- **You must not seek or obtain help from anyone else.**

All work should be **handwritten** and should **include your student number**.

The exam is available for a period of **24 hours**. Upon accessing the exam, you will have **3 hours** in which to complete and submit this assessment.

When you have finished:

- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;
- with your e-mail, include a photograph of the first page of your work together with either yourself or your student ID card.

Please try to upload your work well before the end of the submission window, in case you experience computer problems. **Only one attempt is allowed – once you have submitted your work, it is final**.

Examiners: J. Griffin, D. Stark

 Continue to next page

Question 1 [14 marks]. A six-sided die has an unknown number of faces marked with a six. Let k be this unknown number, which we would like to estimate. Our prior distribution for k is

$$P(k=j) = \begin{cases} 5/8, \ j=1 \\ 1/16, \ j=0,2,3,4,5,6. \end{cases}$$

When the die is thrown each face has an equal chance of showing. The observed data is that the die was thrown twice, and it showed a six exactly once.

(a) Write down the likelihood for the observed data. What is the maximum likelihood estimate for k? **[4]**

(b) Derive the normalized posterior distribution for k. What is the posterior mean for k? **[6]**

(c) Find the posterior predictive probability that if the die is thrown again, it will not show a six. **[4]**

Question 2 [24 marks].
Suppose that we have data $y = (y_1, \dots, y_n)$. Each data-point y_i is assumed to be generated by a distribution with the following probability density function:

$$p(y_i \mid \theta) = \frac{\theta^2}{y_i^3} \exp\left(-\frac{\theta}{y_i}\right), \ y_i \geq 0.$$

The unknown parameter is θ, with $\theta > 0$.

(a) Write down the likelihood for θ given y. Find an expression for the maximum likelihood estimate (MLE) $\hat{\theta}$. **[5]**

(b) A Gamma(α,β) distribution is chosen as the prior distribution for θ. Derive the resulting posterior distribution for θ given y. **[6]**

(c) We would like to choose the gamma prior distribution parameters so that $\alpha = 1$, and

$$P(\theta > 50 + B) = 0.05,$$

where B is the second-to-last digit of your ID number. Find the value of β that is needed. **[5]**

(d) The data are $y = (4, 4, 8, 8, 4, C+3)$, where C is the last digit of your ID number, with $n = 6$.

(i) What is the MLE $\hat{\theta}$? **[3]**

(ii) Using the prior distribution from part (c), what are the parameters of the posterior distribution for θ? **[3]**

(iii) What are the posterior mean and standard deviation for θ? **[2]**

Question 3 [20 marks].
We want to estimate a single unknown parameter θ in a certain model. Assume that in R we have defined a function log_post to calculate the log of the unnormalized posterior density as a function of θ. This function and the data y being analysed are not shown in the code extract below. The posterior density is $p(\theta \mid y)$. Consider the following R code:

```
nm = 10000
theta = vector(length=nm)
s = 0.4
theta0 = 2
log_post0 = log_post(theta0)
for(i in 1:nm){
    theta1 = theta0 + s*rnorm(1)
    log_post1 = log_post(theta1)
    if(log(runif(1)) < log_post1-log_post0){
        theta0 = theta1
        log_post0 = log_post1
    }
    theta[i] = theta0
}
quantile(theta, probs=c(0.5, 0.025, 0.975))
```

An explanation in words is all that is needed for this question.

(a) What is the name of the algorithm that the code is carrying out? **[2]**

(b) Explain what the command `theta1 = theta0 + s*rnorm(1)` is doing in the context of the algorithm. **[4]**

(c) When the code has run, what will the vector `theta` contain? **[3]**

(d) In statistical terms, what will the last line of code output? **[5]**

Suppose that the data y was a sample from an exponential distribution with parameter θ. The code below follows from the preceding code.

```
v = rexp(length(theta), rate=theta)
mean(v>5 & v<10)
```

(e) When this code has run, what will `v` contain? **[3]**

(f) What will the last line of code output (in statistical terms)? **[3]**

Question 4 [26 marks].
The observed data is $y = (y_1, \ldots, y_n)$, a sample from a negative binomial distribution with parameters q and r, where r is assumed to be known.
The prior distribution for q is $Beta(\alpha,\beta)$. Suppose that $y_1 = \cdots = y_n = 0$. Take $n = 10 + A$, where A is the third-to-last digit of your ID number; $\alpha = 5 + B$,where B is the second-to-last digit of your ID number; $r = 3$; and $\beta = 1$.

(a) What is the posterior probability density function for q? **[5]**

(b) Find an expression for the quantile function for this posterior distribution, and hence find a 95% credible interval for q. **[6]**

(c) Let x be a new data-point generated by the same negative binomial distribution with parameters q and r. Find $P(x = 0 \mid y)$, the posterior predictive probability that x is 0. **[5]**

Suppose now that we want to compare two models. Model M_1 is the model and prior distribution described above. Model M_2 assumes that the data follow a negative binomial distribution with q known to be $q_0 = 0.9$.

(d) Find the Bayes factor B_{12} for comparing the two models. **[6]**

(e) We assign prior probabilities of $1/3$ that M_1 is the true model, and $2/3$ that M_2 is the true model. Find the posterior probability that M_1 is the true model. **[4]**

Question 5 [16 marks].
We have observed data

$$y = \{y_{ij} : i = 1, \ldots, n, j = 1, \ldots, m_i\}.$$

Each y_{ij} is the number of times a certain type of machine needs to be repaired during length of time T_{ij}, where $j = 1, \ldots, m_i$ are the machines in factory i, for $i = 1, \ldots, n$, with $n \geq 2$.
A hierarchical model is used to model the data. We assume that

$$y_{ij} \sim \text{Poisson}(T_{ij}\mu_i).$$

μ_i is the repair rate for factory i, which varies between factories according to a gamma distribution

$$\mu_i \sim \text{Gamma}(\alpha, \beta),\ i = 1, \ldots, n.$$

The parameters α and β are given prior distributions, $p(\alpha)$ and $p(\beta)$.
Suppose that we have generated a sample of size M from the joint posterior distribution $p(\mu_1, \ldots, \mu_n, \alpha, \beta \mid y)$.

(a) Explain how to estimate the following using the joint posterior sample:

(i) The posterior mean of α.

(ii) The posterior median of $\nu = \dfrac{\alpha}{\beta}$.

(iii) A 95% equal tail credible interval for ν.

[6]

(b) Explain how to generate a sample from the posterior predictive distribution of the number of repairs during time U for a machine not in our dataset, in each of the following two cases:

(i) If the factory containing this machine is in our dataset. **[4]**

(ii) If the factory is not in our dataset. Also explain how to estimate the posterior predictive probability that such a machine will not need any repairs during time U. **[6]**

End of Paper – An appendix of 1 page follows.

Appendix: common distributions

For each distribution, x is the random quantity and the other symbols are parameters.

Discrete distributions

Distribution	Probability mass function	Range of parameters and variates	Mean	Variance
Binomial	$\binom{n}{x} q^x (1-q)^{n-x}$	$0 \leq q \leq 1$ $x = 0, 1, \ldots, n$	nq	$nq(1-q)$
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	$\lambda > 0$ $x = 0, 1, 2, \ldots$	λ	λ
Geometric	$q(1-q)^x$	$0 < q \leq 1$ $x = 0, 1, 2, \ldots$	$\frac{(1-q)}{q}$	$\frac{(1-q)}{q^2}$
Negative binomial	$\binom{r+x-1}{x} q^r (1-q)^x$	$0 < q \leq 1$, $r > 0$ $x = 0, 1, 2, \ldots$	$\frac{r(1-q)}{q}$	$\frac{r(1-q)}{q^2}$

Continuous distributions

Distribution	Probability density function	Range of parameters and variates	Mean	Variance
Uniform	$\frac{1}{b-a}$	$-\infty < a < b < \infty$ $a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$-\infty < \mu < \infty, \sigma > 0$ $-\infty < x < \infty$	μ	σ^2

The 95th and 97.5th percentiles of the standard $N(0,1)$ distribution are 1.64 and 1.96, respectively.

Distribution	Probability density function	Range of parameters and variates	Mean	Variance
Exponential	$\lambda e^{-\lambda x}$	$\lambda > 0$ $x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$	$\alpha > 0, \beta > 0$ $x > 0$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\alpha > 0, \beta > 0$ $0 < x < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

End of Appendix.