Queen Mary
University of London

Main Examination period 2021 – 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.

You are expected to spend about **2 hours** to complete the assessment, plus the time taken to scan and upload your work. 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 [25 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:

$$p(y_i \mid \eta) = \kappa \eta y_i^{\kappa-1} e^{-\eta y_i^\kappa},\ y_i \geq 0.$$

The unknown parameter is η, with κ assumed to be known, and $\eta, \kappa > 0$.

(a) Write down the likelihood for η given y. Find an expression for the maximum likelihood estimate (MLE) $\hat{\eta}$. **[6]**

(b) A Gamma(α,β) distribution is chosen as the prior distribution for η. Show that the posterior distribution is also a gamma distribution with parameters that you should determine. **[6]**

(c) We would like to choose the gamma prior distribution parameters such that the prior mean is $(B+10)/100$, where B is the second-to-last digit of your ID number, and the prior coefficient of variation (standard deviation divided by the mean) is 0.5. Find the values of α and β that are needed. **[4]**

(d) The data are $y = (2, 7, 5, 3, C+1)$, where C is the last digit of your ID number, with $n = 5$. Set $\kappa = 2$.

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

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

  (iii) What are the posterior mean and standard deviation for η? **[3]**

Question 2 [16 marks].
Suppose that the data $y = (y_1, \dots, y_n)$ are a sample from a normal distribution with unknown mean θ and known standard deviation $\sigma = 2$. Our prior distribution $p(\theta)$ is normal with mean 0 and standard deviation σ_0.

(a) For an uninformative prior, do we need a large or small value for σ_0? **[2]**

(b) We want the prior probability $P(|\theta| \leq A+10)$ to be 0.95, where A is the third-to-last digit of your ID number. What value for σ_0 should we choose? **[4]**

(c) A colleague prefers a Cauchy distribution as a prior. What is a possible reason for this preference? **[2]**

Let the sample mean $\bar{y}$ be $B+1$, where B is the second-to-last digit of your ID number, and the sample size be $n = 20$. Use the prior distribution found in part (b).

(d) What is the posterior distribution for θ, $p(\theta \mid y)$? Based on this posterior distribution, find a point estimate for θ. **[5]**

(e) Suppose that we want to find the posterior probability $P(\theta \leq 0 \mid y)$. Write an expression for this probability in terms of Φ, the cumulative distribution function for the standard normal distribution. **[3]**

Question 3 [20 marks].
The data are $y = (y_1, \ldots, y_n)$, a sample from a negative binomial distribution with parameters q and r, where r is assumed to be known. A Beta(α,β) prior distribution is assigned to q. Apart from part (c), the answers do not need any numerical calculations.
In the following R code, the data y is denoted by `y` in the code, `r` is the known parameter, and `alpha` and `beta` are the prior parameters. The posterior distribution for q is Beta(a,b).

```
r = 2
alpha = 4
beta = 4
a = r*length(y) + alpha
b = sum(y) + beta
qbeta(0.5, shape1=a, shape2=b)
qbeta(c(0.025, 0.975), shape1=a, shape2=b)
```

(a) In statistical terms, what will the second-to-last line of code output? **[2]**

(b) In statistical terms, what will the last line of code output? **[3]**

(c) Let B and C be the second-to-last and last digits of your ID number, respectively. Take the sample size $n = B + 10$, and $\sum_{i=1}^{n} y_i = C + 20$. What are the posterior mean and standard deviation for q? **[6]**

The R code below follows on from the code above.

```
q_sim = rbeta(2000, shape1=a, shape2=b)
x = rnbinom(length(q_sim), size=r, prob=q_sim)
mean(x<3)
```

(d) When this code has run, what will `q_sim` contain? What will `x` contain? **[6]**

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

Question 4 [23 marks].
The observed data is $y = (y_1, \ldots, y_n)$, a sample from a geometric distribution with parameter q. The prior distribution for q is uniform on the interval $[0,1]$. Suppose that $y_1 = \cdots = y_n = 0$. Take $n = 20 + A$, where A is the third-to-last digit of your ID number.

(a) What is the posterior probability density function for q? **[4]**

(b) Find an expression for the quantile function for this posterior distribution, and hence find the posterior median for q. **[4]**

(c) Let x be a new data-point generated by the same geometric distribution with parameter q. Find $P(x = 0 \mid y)$, the posterior predictive probability that x is 0. **[6]**

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 geometric 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 $1/2$ that each model is the true model. Find the posterior probability that M_2 is the true model. **[3]**

Question 5 [16 marks].
We have observed data

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

Each y_{ij} is the number of late trains observed out of N_{ij} journeys, for train operating company i on route j, where $j = 1,\dots,m_i$ are the routes operated by company i.
A hierarchical model is used to model the data. We assume that

$$y_{ij} \sim \text{ Binomial}(N_{ij}, q_i).$$

q_i is the probability of being late for operating company i, which varies between companies according to a beta distribution

$$q_i \sim \text{ Beta}(\alpha, \beta), \ i = 1,\dots,n.$$

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 $p(q_1,\dots,q_n,\alpha,\beta \mid y)$.

(a) Explain how to obtain a sample from the marginal posterior distribution $p(\alpha,\beta \mid y)$ using the joint posterior sample. **[2]**

(b) Given a sample from $p(\alpha,\beta \mid y)$, explain how to estimate the following:

(i) The posterior median of α.

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

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

[6]

(c) Explain how to generate a sample from the posterior predictive distribution of the number of late trains out of K journeys for a route not in our dataset, in each of the following two cases:

(i) if the operating company for the route is in our dataset;

(ii) if the operating company is not in our dataset.

[8]

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	$\binom{n}{x} q^x (1-q)^{n-x}$	$0 \le q \le 1$ $x = 0, 1, \ldots, n$	nq	$nq(1-q)$
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	$\lambda > 0$ $x = 0, 1, 2, \ldots$	λ	λ
Geometric	$q(1-q)^x$	$0 < q \le 1$ $x = 0, 1, 2, \ldots$	$\frac{(1-q)}{q}$	$\frac{(1-q)}{q^2}$
Negative binomial	$\binom{r+x-1}{x} q^r (1-q)^x$	$0 < q \le 1$, $r > 0$ $x = 0, 1, 2, \ldots$	$\frac{r(1-q)}{q}$	$\frac{r(1-q)}{q^2}$

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.