Main Examination period 2023 – January – Semester A

MTH6102: Bayesian Statistical Methods

Duration: 2 hours

The exam is intended to be completed within **2 hours**. However, you will have a period of **4 hours** to complete the exam and submit your solutions.

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

All work should be **handwritten** and should **include your student number**. Only one attempt is allowed – **once you have submitted your work, it is final**.

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.

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;

Examiners: J. Griffin, D. Stark

Question 1 [24 marks].
Suppose that we have data $y = (y_1, \ldots, y_n)$. Each data-point is assumed to be generated by a distribution with the following probability density function:

$$p(y_i \mid \psi) = 2\psi y_i \exp\left(-\psi y_i^2\right),\ y_i \geq 0,\ i = 1, \ldots, n.$$

The unknown parameter is ψ, with $\psi > 0$.

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

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

(c) Show that the posterior mean for ψ is always in between the prior mean and the MLE for this example. **[5]**

(d) The data are $y = (2, 6, 5, 4, C + 1)$, where C is the last digit of your ID number, with $n = 5$. The prior distribution is Gamma$(2, 2)$.

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

  (ii) What is the posterior distribution for ψ? Based on this posterior distribution, calculate a point estimate for ψ. **[4]**

Question 2 [19 marks].
The data $y = (y_1, \ldots, y_n)$ is a sample from a normal distribution with unknown mean μ and known standard deviation $\sigma = 2$. The prior distribution for μ is normal $N\left(\mu_0, \sigma_0^2\right)$. The posterior distribution is $\mu \mid y \sim N\left(\mu_1, \sigma_1^2\right)$, where

$$\mu_1 = \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma^2}\right) \Big/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right),\ \sigma_1^2 = 1 \Big/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right), \text{ and } \bar{y} \text{ is the sample mean.}$$

(a) As the prior distribution becomes less informative, what value does the posterior mean for μ approach? As the prior distribution becomes more informative, what value does the posterior mean for μ approach? **[4]**

(b) Suppose that we take $\mu_0 = 0$, and we want the prior probability $P(|\mu| \leq A + 20)$ to be 0.9, where A is the third-to-last digit of your ID number. What value for σ_0 should we choose? **[4]**

Let the sample mean be $B + 1$, where B is the second-to-last digit of your ID number, and the sample size be $n = 40$. Use the prior distribution found in part (b).

(c) What is the posterior distribution for μ, $p(\mu \mid y)$? What is the posterior median for μ? **[4]**

(d) Let x be a future data-point from the same $N\left(\mu, \sigma^2\right)$ distribution. Find the posterior predictive mean and variance of x. **[7]**

Question 3 [26 marks].
The dataset $y = (y_1, \dots, y_n)$ is a sample from a Poisson distribution with parameter λ. A Gamma(α,β) prior distribution is assigned to λ. 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, and `alpha` and `beta` are the prior parameters.

```
alpha = 3
beta = 3
a = sum(y) + alpha
b = length(y) + beta
pgamma(2, shape=a, rate=b)
qgamma(c(0.5, 0.025, 0.975), shape=a, rate=b)
```

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

(b) What will the line which starts with `pgamma` output? **[2]**

(c) Let B and C be the second-to-last and last digits of your ID number, respectively. Take the sample size $n = B + 15$, and $\sum_{i=1}^{n} y_i = C + 30$. What are the posterior mean and standard deviation for λ? **[5]**

The R code below follows on from the code above.

```
v = rgamma(5000, shape=a, rate=b)
w = rpois(length(v), lambda=v)
mean(w==0)
```

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

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

(f) State one advantage of using a prior distribution which is conjugate to the likelihood. **[2]**

(g) Suppose that we assumed some other prior distribution instead of a gamma distribution. What method could we use to make inferences based on the resulting posterior distribution for λ? **[3]**

Question 4 [16 marks].
The observed data is $y = (y_1, \dots, y_n)$, 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 $n = 10 + A$, where A is the third-to-last digit of your ID number.

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

Suppose now that we want to compare two models. Model M_1 assumes that the data follow a geometric distribution with q known to be $q_0 = 0.8$. Model M_2 is the model and prior distribution described above.

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

(c) We assign prior probabilities of $1/2$ that each model is the true model. Find the posterior probability that M_1 is the true model. **[3]**

(d) State a drawback of using Bayes factors and posterior probabilities to compare models. **[2]**

Question 5 [15 marks].
The observed data $y = \{y_{ij},\ i = 1,\dots,n,\ j = 1,\dots,m_i\}$ are the average results in an exam for school j within county i. The following hierarchical model is considered reasonable:

$$y_{ij} \sim \text{Normal}\left(\mu_i, \sigma_S^2\right),\ j = 1,\dots,m_i$$
$$\mu_i \sim \text{Normal}\left(\mu_C, \sigma_C^2\right),\ i = 1,\dots,n.$$

where μ_C, σ_S and σ_C are unknown parameters which are each assigned a prior distribution. Suppose that we have generated a sample of size M from the joint posterior distribution $p(\mu_C, \sigma_S, \sigma_C, \mu_1, \dots, \mu_n \mid y)$.

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

(i) the posterior mean for μ_C;

(ii) a 95% credible interval for σ_S / σ_C;

(iii) the posterior probability that $\mu_1 < \mu_2$. [**7**]

(b) Explain how to generate a sample from the posterior predictive distribution of the result for a school not in our dataset, in each of the following two cases:

(i) if the county containing the school is in our dataset;

(ii) or if the county 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 $N\left(\mu, \sigma^2\right)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$-\infty < \mu < \infty, \sigma > 0$ $-\infty < x < \infty$	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.