Queen Mary
University of London

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

$$\mu_1 = \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma^2}\right) \Bigg/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right), \; \sigma_1^2 = 1 \Bigg/ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right),$$ and $\bar{y}$ 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? 2 **[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]**

 Continue to next page

Solution

Q1) The likelihood function, $p(y|\psi)$, for ψ given $y=(y_1,\dots,y_n)$ is the joint density of y, which by independence is

$$p(y|\psi) = p(y_1,\dots,y_n|\psi) = \prod_{i=1}^{n} p(y_i|\psi)$$

$$= \prod_{i=1}^{n} 2\psi y_i \exp(-\psi y_i^2)$$

(2 marks)

$$= 2^n \psi^n \left(\prod_{i=1}^{n} y_i\right) \exp\left(-\psi \sum_{i=1}^{n} y_i^2\right).$$

The log likelihood is

$$\ell(\psi) = \log p(y|\psi)$$

$$= n\log(2) + n\log(\psi) + \sum_{i=1}^{n} \log(y_i) - \psi \sum_{i=1}^{n} y_i^2$$

$$\psi > 0.$$

To find the MLE, we take the derivative of $\ell(\psi)$ with respect to ψ to find

$$\frac{d}{d\psi}\ell(\psi) = \frac{n}{\psi} - \sum_{i=1}^{n} y_i^2$$

4 marks

The equation $\frac{d}{d\psi} \ell(\psi) = 0$ yields

$$\frac{n}{\psi} - \sum_{i=1}^{n} y_i^2 = 0 \Rightarrow \hat{\psi} = \frac{n}{\sum_{i=1}^{n} y_i^2}$$

$\hat{\psi}$ is a global maximum since

$$\frac{d^2}{d\psi^2} \ell(\psi) = \frac{-n}{\psi^2} < 0 \quad \forall\ \psi > 0.$$

So $\hat{\psi} = n / \sum_{i=1}^{n} y_i^2$ is the MLE for ψ.

b) $\psi \sim \text{Gamma}(a, b)$ with pdf

$$p(\psi) = \frac{b^a}{\Gamma(a)} \psi^{a-1} \exp(-b\psi)$$

The posterior, $p(\psi|y)$, is

$$p(\psi|y) \propto p(\psi) \times p(y|\psi)$$

$$\propto \psi^{a-1} \exp(-b\psi)$$

$$\psi^{n} \times \exp\left(-\psi \sum_{i=1}^{n} y_i^2\right)$$

$$= \psi^{a+n-1} \exp\left(-\psi \left(b + \sum_{i=1}^{n} y_i^2\right)\right)$$

So, the posterior pdf is proportional to a Gamma density with posterior parameters $a+n$ and $b + \sum_{i=1}^{n} y_i^2$

so $p(\psi|y) \sim \text{Gamma}(a+n, \theta+S)$

$$S = \sum_{i=1}^{n} y_i^2$$

(c) The posterior mean, $\hat{\psi}_B$,

is $$\hat{\psi}_B = \frac{a+n}{\theta+S} = \frac{a}{\theta+S} + \frac{n}{\theta+S}$$

$$= \frac{a}{\theta+S} \cdot \frac{\theta}{\theta} + \frac{n}{\theta+S} \cdot \frac{S}{S}$$

$$= \frac{\theta}{\theta+S} \underbrace{\frac{a}{\theta}}_{\text{prior mean}} + \frac{S}{\theta+S} \underbrace{\frac{n}{S}}_{\text{MLE}}$$

$$= w \frac{a}{\theta} + (1-w) \frac{n}{S}, \text{ where}$$

$$w = \frac{\theta}{\theta+S}, \quad 0 \leq w \leq 1$$

So the posterior mean is in between the prior mean and the MLE.

(d) First, $y = (2, 6, 5, 4, C+1)$, $n=5$

$$S = \sum^{n} y_i^2 = 2^2 + 6^2 + 5^2 + 4^2 + (C+1)^2$$
$$= 81 + (C+1)^2$$

So the MLE $\hat{\psi}$ is

$$\hat{\psi} = \frac{n}{\sum_{i=1}^{n} y_i^2} = \frac{5}{81 + (C+1)^2}.$$

(ii) The posterior parameters of the Gamma posterior density are

$$\alpha + n = 2 + 5 = 7$$

$$\beta + \sum_{i=1}^{n} y_i^2 = 2 + 81 + (C+1)^2$$
$$= 83 + (C+1)^2$$

A point estimate for ψ is the posterior mean

$$\hat{\psi}_B = \frac{\alpha + n}{\beta + \sum^{n} y_i^2} = \frac{7}{83 + (C+1)^2}$$

Q2

(a) As the prior distribution becomes less informative (large σ_0) the posterior mean approaches the MLE $\bar{y}$. On the other hand, as the prior distribution becomes more informative (small σ_0), the posterior mean approaches prior mean μ_0.

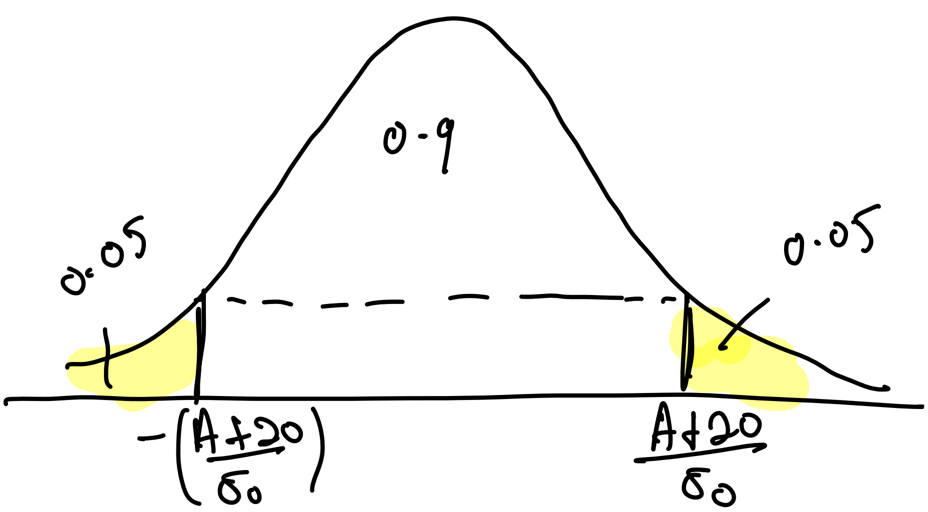
(b) $\mu \sim N(0, \sigma_0^2)$. We want to find $\sigma_0 > 0$ such that

$$P(|\mu| \le A+20) = 0.9$$

$$\Leftrightarrow \quad P(-(A+20) \le \mu \le A+20) = 0.9$$

$$\Leftrightarrow \quad P\left(-\frac{(A+20)}{\sigma_0} \le \frac{\mu}{\sigma_0} \le \frac{A+20}{\sigma_0}\right) = 0.9 \qquad \frac{\mu}{\sigma_0} \sim N(0,1)$$

Z



so $P\left(\frac{\nu}{\sigma_0} \le -\left(\frac{A+20}{\sigma_0}\right)\right) = 0.05$

$= P\left(\frac{\nu}{\sigma_0} \geq \frac{A+20}{\sigma_0}\right) = 0.05$

Thus, $P\left(\frac{\nu}{\sigma_0} \le \frac{A+20}{\sigma_0}\right) = 0.95$

So $\Phi\left(\frac{A+20}{\sigma_0}\right) = 0.95$, Φ is cdf of $N(0,1)$

$\Rightarrow \frac{A+20}{\sigma_0} = \Phi^{-1}(0.95) = 1.64$

$\Rightarrow A+20 = (1.64)\sigma_0$

$$\Rightarrow \boxed{\hat{\sigma}_0 = \frac{A+20}{1.64}}$$

(c) $\bar{y} = B+1$

$n = 40$

$\sigma_0 = \dfrac{A+20}{1.64}$

$\mu_0 = 0$

Use the formula to find μ_1 and σ_1^2

For the normal, the posterior median is equal to the posterior mean μ_1.

(d) Let x be a new data point from $N(\mu, \sigma^2)$.
By the law of iterated expectation, the predictive mean of x is

$$\mathbb{E}(\mathbb{E}(x \mid \cancel{y}, \mu)) = \mathbb{E}(\mathbb{E}(x \mid \mu))$$

since $x \sim \underbrace{N(\mu, \sigma^2)}$ then $\underline{\mathbb{E}(x \mid \mu)} = \underline{\mu}$ ($\mu \sim N(\mu_0, \sigma_0^2)$)

so $$\mathbb{E}(\mathbb{E}(x \mid \mu)) = \underline{\mathbb{E}(\mu)} = \underbrace{\mu_0}_{\text{the mean of } x}.$$

By the law of total variance,

$$\mathbb{E}(\mathrm{var}(x \mid \cancel{y}, \mu)) + \mathrm{Var}(\mathbb{E}(x \mid \cancel{y}, \mu))$$

- $\mathbb{E}(\mathrm{var}(x \mid \cancel{y}, \mu)) = \mathbb{E}(\mathrm{var}(x \mid \mu))$

But $x \sim N(\mu, \sigma^2)$ so $\mathrm{var}(x \mid \mu) = \underline{\sigma^2}$ fixed

so $$\mathbb{E}(\mathrm{var}(x \mid \mu)) = \mathbb{E}(\sigma^2) = \sigma^2$$

- $\mathrm{Var}(\mathbb{E}(x \mid \cancel{y}, \mu)) = \mathrm{Var}(\mathbb{E}(x \mid \mu)) = \underbrace{\mathrm{Var}(\mu)}_{\text{prior variance}} = \sigma_0^2$

The predictive variance of x is

$$\sigma^2 + \sigma_0^2$$

Question 3 [26 marks].
The dataset $y = (y_1, \ldots, y_n)$ 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 $y = (y_1, \ldots, y_n)$, 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 $n = 10 + A$, 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]**

Q3 solution

$y = (y_1, \ldots, y_n) \sim \text{Poisson}(\lambda)$

$\lambda \sim \text{Gamma}(a, b)$

(a) It displays the posterior median and a 95% equal-tail credible interval for λ.

(b) It computes $P(\lambda \leq a \mid y)$ where

$\lambda \sim \text{Gamma}\left(3 + \sum_{i=1}^{n} y_i,\ 3+n\right)$

(c) $n = B+15$, $\sum_{i=1}^{n} y_i = C+30$

The posterior mean is

$$\hat{\lambda}_B = \frac{3 + \sum y_i}{3+n} = \frac{33+C}{18+B}$$

The posterior standard deviation is

$$\sqrt{\frac{33+C}{(18+B)^2}}$$

(d) v will contain an iid sample of size 5000 from the posterior for λ, $p(\lambda \mid y)$.

(f) Bayesian updating reduces to modifying the parameters of the prior distribution.

(g) If we assume some other non-conjugate prior distribution, then the posterior for λ might not be a well-known distribution e.g gamma. In this case, we could use a MCMC method to generate a sample from $p(\lambda|y)$, and use this sample to summarise the posterior.

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

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

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 $p(\mu_C, \sigma_S, \sigma_C, \mu_1, \ldots, \mu_n \mid y)$.

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

Continuous distributions

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

End of Appendix.