Queen Mary
University of London

**Main Examination period 2020 – January – Semester A**

# MTH6102: Bayesian Statistical Methods

**Duration: 2 hours**

**Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.**

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

**Only non-programmable calculators that have been approved from the college list of non-programmable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.**

Complete all rough work in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any unauthorised notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

**Exam papers must not be removed from the examination room.**

**Examiners: J. Griffin, L. Pettit**

**Question 1 [12 marks].**
A box contains $m = 5$ balls, of which $r$ are red and the rest black. The unknown quantity is $r$. Our prior distribution is that each value $r = 0, 1, \ldots, m$ has equal probability. We are told that twice, a ball was taken out and immediately replaced, and both times the ball was red.

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

(b) Derive the normalized posterior distribution for $r$. What is the posterior mean for $r$? **[5]**

(c) Find the posterior predictive probability that if another ball is taken from the box, it is black. **[3]**

**Question 2 [34 marks].**
A biased coin with probability $q$ of landing heads is repeatedly tossed until the first head is seen. The number of tails $X$ before the first head is modelled as a geometric distribution with probability mass function $P(X = x) = q(1-q)^x$. The experiment was repeated $n$ times and $x_1, x_2, \ldots, x_n$ tails were observed.

(a) Write down the likelihood for $q$. Show that the maximum likelihood estimate for $q$ is

$$\hat{q} = \frac{n}{n+S}, \text{ where } S = \sum_{i=1}^{n} x_i.$$ **[6]**

(b) Find the Fisher information and hence the asymptotic variance for $\hat{q}$. **[5]**

(c) A Beta($\alpha_0$,$\beta_0$) distribution is chosen as the prior distribution for $q$. Show that the posterior distribution is Beta($\alpha_1$,$\beta_1$), where you should determine $\alpha_1$ and $\beta_1$. **[6]**

(d) We have $n = 5$ and observed data $x_1, \ldots, x_n = 4, 2, 5, 6, 3$.

(i) What is the maximum likelihood estimate $\hat{q}$? **[3]**

(ii) Find an approximate 95% confidence interval for $q$. **[4]**

(iii) Before seeing the data, our probability distribution for $q$ has mean 0.4 and standard deviation 0.2. Find values of $\alpha_0$ and $\beta_0$ corresponding to this belief. What is then the posterior distribution for $q$? What is the posterior mean? **[8]**

(iv) Comment on the posterior mean compared to the maximum likelihood estimate and the prior mean for this example. No further calculations or formulae are needed here. **[2]**

**Question 3 [26 marks].**

We want to estimate a single unknown parameter $\theta$ 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 $\theta$. 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:

```
nb = 1000
nm = 10000
theta = vector(length=nm)
s = 0.4
theta0 = 2
log_post0 = log_post(theta0)
for(i in 1:(nb+nm)){
    theta1 = rnorm(1, mean=theta0, sd=s)
    log_post1 = log_post(theta1)
    if(log(runif(1)) < log_post1-log_post0){
        theta0 = theta1
        log_post0 = log_post1
    }
    if(i>nb) theta[i-nb] = theta0
}
stheta = sort(theta)
stheta[nm/2]
stheta[nm*0.025]
stheta[nm*0.975]
```

Except where stated, 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? **[3]**

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

(c) Explain what the command `if(log(runif(1)) < log_post1-log_post0)` is doing in the context of the algorithm. In your answer, include a formula involving $p(\theta \mid y)$ that the code is implementing. **[5]**

(d) What are the effects on the behaviour of the algorithm of making the variable called `s` smaller? What are the effects of making it larger? **[4]**

(e) What is the purpose of the variable called `nb`? **[2]**

(f) When the code has run, what will the vector `theta` contain? **[2]**

(g) In statistical terms, what will the command `stheta[nm/2]` output? **[2]**

(h) In statistical terms, what will the last two lines of code output? **[4]**

**Question 4 [17 marks].**
The observed data $y = \{y_{ij} : i = 1,\ldots,n,\ j = 1,\ldots,m_i\}$ are the recorded counts of a disease in district $j$ within county $i$. The population of each district is $N_{ij}$. The following hierarchical model is considered reasonable

$$\begin{aligned} y_{ij} &\sim \text{Poisson}(\lambda_i N_{ij}),\ j = 1,\ldots,m_i \\ \lambda_i &\sim \text{Gamma}(\alpha,\beta),\ i = 1,\ldots,n. \end{aligned}$$

$\alpha$ and $\beta$ are unknown parameters which are given a prior distribution $p(\alpha,\beta)$.
Suppose that we have generated a sample of size $M$ from the joint posterior distribution $p(\alpha,\beta,\lambda_1,\ldots,\lambda_n \mid y)$.

(a) How would we obtain a sample from the marginal posterior distribution $p(\alpha,\beta \mid y)$ using the joint posterior sample? How would we estimate the posterior mean for $\alpha/\beta$? [**5**]

(b) Explain how to generate a sample from the posterior predictive distribution of the disease count for a district not in our dataset with population $P$, in each of the following two cases: if the county containing the district is in our dataset; or if the county is not in our dataset. In the latter case, how would we estimate the posterior predictive probability that the disease count in this district will be zero? [**8**]

(c) Give two reasons why in general we might want to use a hierarchical model instead of a single-level model. [**4**]

**Question 5 [11 marks].**
Two models $M_1$ and $M_2$ are under consideration, with corresponding parameters $\theta$ and $\psi$. $\theta$ is a single parameter with unbounded range. For the prior distribution $p(\theta \mid M_1)$, we assign a normal distribution $N(0,\sigma^2)$ with an extremely large value of $\sigma$ so that the prior is practically flat over the range supported by the likelihood. We also assign a prior distribution $p(\psi \mid M_2)$. The observed data is $y$.

(a) State the formula for the Bayes factor $B_{12}$ for comparing the models, in which large values of $B_{12}$ favour model $M_1$. [**5**]

(b) For inference conditional upon model $M_1$, what is the effect on the posterior mean for $\theta$ if we replace $\sigma$ with $1000\sigma$ in $p(\theta \mid M_1)$? [**3**]

(c) What is the effect on $B_{12}$ if we replace $\sigma$ with $1000\sigma$ in $p(\theta \mid M_1)$? [**3**]

---

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

### Continuous distributions

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

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 |
|---|---|---|---|---|
| Normal $No(\mu, \tau)$ | $\dfrac{\sqrt{\tau}}{\sqrt{2\pi}} \exp\left(-\dfrac{\tau(x-\mu)^2}{2}\right)$ | $-\infty < \mu < \infty, \tau > 0$<br>$-\infty < x < \infty$ | $\mu$ | $\tau^{-1}$<br>(precision $\tau$) |
| Exponential | $\lambda e^{-\lambda x}$ | $\lambda > 0$<br>$x > 0$ | $\dfrac{1}{\lambda}$ | $\dfrac{1}{\lambda^2}$ |
| Gamma | $\dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$ | $\alpha > 0, \beta > 0$<br>$x > 0$ | $\dfrac{\alpha}{\beta}$ | $\dfrac{\alpha}{\beta^2}$ |
| Beta | $\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ | $\alpha > 0, \beta > 0$<br>$0 < x < 1$ | $\dfrac{\alpha}{\alpha+\beta}$ | $\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ |

---

**End of Appendix.**