Queen Mary
University of London

M. Sci. Examination by course unit 2015

MTH709U: Bayesian Statistics

Duration: 3 hours

Date and time: 15 May 2015, 14.30-17.30

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 awarded are shown next to the questions.

Calculators ARE permitted in this examination. The unauthorised use of material stored in pre-programmable memory constitutes an examination offence. Please state on your answer book the name and type of machine used.
Statistical functions provided by the calculator may be used provided that you state clearly where you have used them.
The New Cambridge Statistical Tables 2nd Edition are provided.
A table of common distributions is provided as an appendix.

Complete all rough workings 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 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 shall 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.

Examiner(s): L I Pettit, H Maruri-Aguilar

Question 1 (5 marks). Explain what is meant by a conjugate prior. Give an example of a conjugate prior and the corresponding distribution for the data. Give one advantage and one disadvantage of using a conjugate prior.

Question 2 (24 marks). (a) Show that an exponential distribution with mean θ^{-1} is an exponential family. [4]

(b) Identify the natural parameter. [1]

(c) If a random sample of n exponentially distributed observations is collected write down a minimal sufficient statistic for θ. [1]

(d) Six batteries are put on test for 200 hours. Four fail after 96, 130, 160 and 180 hours. The other two batteries are still working after 200 hours. Assuming the lifetime of these batteries has an exponential distribution with mean θ^{-1} find the likelihood function. [3]

(e) The prior distribution of θ is given by a gamma distribution with density

$$p(\theta) = \frac{\beta^{\alpha}\theta^{\alpha-1}\exp(-\beta\theta)}{\Gamma(\alpha)}, \quad \theta \geq 0.$$

It is thought the lifetime of the batteries have mean 180 and the probability it lies between 110 and 470 is approximately 0.95. Show that a prior with $\alpha = 10$ and $\beta = 1800$ is appropriate. [5]

(f) Find the posterior distribution of θ and its mean and variance. [4]

(g) Find the density of the predictive distribution of another independent battery and find the probability it has a lifetime more than 200 hours. [6]

Question 3 (17 marks). Two scientists are interested in the decay rate of a rare isotope. They do a series of 20 independent experiments giving observations $X_1, X_2, \ldots, X_{20}$ with each X_i having a Poisson distribution with mean θ. They find $\sum x_i = 30$.

(a) The first scientist has little prior knowledge about θ and decides to use the Jeffreys prior. Show that this is proportional to $\theta^{-1/2}$ and find the posterior distribution. [5]

(b) The second scientist has more experience and uses a prior which is Gamma(4, 2). Find the posterior distribution for the second distribution. [2]

(c) Using Table 30 in the New Cambridge Statistical Tables find for the second scientist a 95% highest posterior density interval for θ. [3]

(d) If the scientist didn't have access to the Cambridge Tables an alternative way of finding an interval estimate would be to approximate the posterior distribution by a normal distribution with mean the posterior mode and variance the reciprocal of the observed information evaluated at the posterior mode. Find the parameters of this normal distribution and hence an approximate 95% highest posterior density interval. [4]

(e) Compare the intervals you have found in (c) and (d) and comment. [3]

Question 4 (16 marks). In a mass production process items are made to a nominal weight of 1.0kg. All underweight items are rejected but the remaining items may be slightly overweight. It is believed that the excess weight, X in grammes, has a continuous uniform distribution on $(0, \theta)$ but the value of θ is unknown. The prior density for θ is

$$p(\theta) = \begin{cases} 0 & \theta \le 0 \\ C/100 & 0 < \theta \le 10 \\ C\theta^{-2} & 10 < \theta \end{cases}$$

(a) Find the value of C [4]

(b) We observe 10 items and their excess weights, in grammes, are as follows.

3.9 2.3 4.9 1.6 3.7 0.4 1.4 3.4 4.3 0.8

Assume that these are independent observations given θ.

(i) Find the likelihood. [4]

(ii) Find a function $h(\theta)$ such that the posterior density of θ is

$$p(\theta|\underline{x}) = Dh(\theta)$$

where D is a constant. [4]

(iii) Evaluate the constant D. [4]

Question 5 (19 marks). (a) A three stage linear model is given by

$$\begin{aligned} \underline{y} \mid \underline{\theta}_1 &\sim N(A_1\underline{\theta}_1, C_1) \\ \underline{\theta}_1 \mid \underline{\theta}_2 &\sim N(A_2\underline{\theta}_2, C_2) \\ \underline{\theta}_2 &\sim N(\underline{\mu}, C_3) \end{aligned}$$

where $A_1, A_2, C_1, C_2, C_3, \underline{\mu}$ are known. Using the results for the two stage model given below, show that the posterior distribution of $\underline{\theta}_1$ is $N(D\underline{d}, D)$ where D^{-1} and $\underline{d}$ are to be determined. [6]

(b) Suppose six independent observations are available with means as follows

$$E[y_1] = \theta_1, \quad E[y_2] = \theta_2, \quad E[y_3] = \theta_3, \quad E[y_4] = \theta_1 - \theta_2,$$

$$E[y_5] = \theta_2 - \theta_3, \quad E[y_6] = \theta_3 - \theta_1.$$

The observations are assumed normally distributed with variance 2. The prior beliefs about the θ_i's are such that $E[\theta_i] = 0, Var[\theta_i] = 3$ for $i = 1, 2, 3$ and they are independent and normally distributed.

Write this problem as a two stage linear model and hence find the posterior means of θ_1, θ_2 and θ_3. [8]

Find a 95% HPD interval for $\theta_1 + \theta_2 + \theta_3$. [5]

NOTE 1 For the two-stage linear model

$$\begin{aligned} \underline{y}|\underline{\theta}_1 &\sim N(A_1\underline{\theta}_1, C_1) \\ \underline{\theta}_1|\underline{\mu} &\sim N(\underline{\mu}, C_2) \end{aligned}$$

where A_1, C_1, C_2 and $\underline{\mu}$ are known.

The marginal distribution of $\underline{y}$ is $N(A_1\underline{\mu}, C_1 + A_1 C_2 A_1^T)$.

The posterior distribution of $\underline{\theta}_1$ is $N(B\underline{b}, B)$ where

$$\begin{aligned} B^{-1} &= A_1^T C_1^{-1} A_1 + C_2^{-1}, \\ \underline{b} &= A_1^T C_1^{-1} \underline{y} + C_2^{-1} \underline{\mu}. \end{aligned}$$

NOTE 2 You may use the result

$$(aI_n + bJ_n)^{-1} = \frac{1}{a} I_n - \frac{b}{\{(a + nb)a\}} J_n$$

for $a > 0$, $b \neq -\frac{a}{n}$, where I_n is an $n \times n$ identity matrix and J_n is an $n \times n$ matrix of ones.

Question 6 (19 marks). (a) It is desired to estimate the posterior means of three parameters, θ, δ and ϕ. The full conditional distributions,

$$p(\theta|\delta, \phi, \text{data}), \quad p(\delta|\phi, \theta, \text{data}), \quad p(\phi|\theta, \delta, \text{data})$$

are known. Explain carefully how the conditional distributions can be used to construct a Gibbs sampler to simulate a sample from the joint posterior of the parameters. Explain how the posterior mean and median of θ can be estimated from this sample. [8]

(b) How is the basic algorithm modified if the parameter δ is constrained to lie in the interval (d_1, d_2)? [3]

(c) The brightness (J) of comets may be modelled by a Pareto distribution with density

$$p(J \mid \alpha) = \frac{\alpha J_x^\alpha}{J^{\alpha+1}}, \qquad J > J_x$$

where J_x is the known lower limit in brightness for the sample considered and α is a brightness index.

(i) If a sample of the brightness J_i of n comets is available the maximum likelihood estimate of α is given by

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log_e(J_i/J_x)}.$$

If the prior distribution for α is taken as Gamma(a, b) show the posterior distribution is Gamma$(a + n, b + (n/\hat{\alpha}))$. [3]

(ii) Three groups of comets of sizes n_1, n_2 and n_3 have brightness parameters α_1, α_2 and α_3. The prior distribution for each α_i is taken as Gamma(a, b). For physical reasons the brightness parameters must be ordered $\alpha_1 < \alpha_2 < \alpha_3$. Explain how the posterior means of the parameters can be estimated by adapting the basic Gibbs sampling algorithm. [5]

End of Paper—An appendix of 1 page follows.

Bayesian Statistics – Common Distributions

Discrete Distributions

Distribution	Density	Range of Variates	Mean	Variance
Uniform	$\frac{1}{N}$	$N = 1, 2, \ldots$ $x = 1, 2, \ldots, N$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$
Bernoulli	$p^x(1-p)^{1-x}$	$0 \leq p \leq 1,\ x = 0, 1$	p	$p(1-p)$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	$0 \leq p \leq 1,\ n = 1, 2, \ldots$ $x = 0, 1, \ldots n$	np	$np(1-p)$
Poisson	$\frac{\exp(-\lambda)\lambda^x}{x!}$	$\lambda > 0,\ x = 0, 1, 2, \ldots$	λ	λ
Geometric	$p(1-p)^x$	$0 < p \leq 1,\ x = 0, 1, 2, \ldots$	$\frac{(1-p)}{p}$	$\frac{(1-p)}{p^2}$
Negative Binomial	$\binom{r+x-1}{x}p^r(1-p)^x$	$0 < p \leq 1,\ r > 0$ $x = 0, 1, 2, \ldots$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$

Continuous Distributions

Uniform	$\frac{1}{b-a}$	$-\infty < a < b < \infty$ $a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp[\frac{-(x-\mu)^2}{2\sigma^2}]$	$-\infty < \mu < \infty$ $\sigma > 0,\ -\infty < x < \infty$	μ	σ^2
Normal $No(\mu, h)$	$\frac{\sqrt{h}}{\sqrt{2\pi}}\exp[\frac{-h(x-\mu)^2}{2}]$	$-\infty < \mu < \infty$ $h > 0,\ -\infty < x < \infty$	μ	h^{-1}
Exponential	$\lambda\exp(-\lambda x)$	$\lambda > 0,\ x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma (α, β)	$\frac{\beta^\alpha x^{\alpha-1}\exp(-\beta x)}{\Gamma(\alpha)}$	$\beta > 0,\ \alpha > 0,\ x > 0$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Beta (a, b)	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$	$a > 0,\ b > 0,\ 0 < x < 1$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$
F_n^m	$\frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{\frac{m}{2}}$	$m, n = 1, 2, \ldots$	$\frac{n}{n-2}$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$
	$\times\frac{x^{(m-2)/2}}{[1+(m/n)x]^{(m+n)/2}}$	$x \geq 0$	for $n > 2$	for $n > 4$
χ_k^2	$\frac{1}{\Gamma(k/2)2^{k/2}}x^{k/2-1}\exp(-\frac{x}{2})$	$k = 1, 2, \ldots,\ x > 0$	k	$2k$
Pareto	$\frac{\alpha\beta^\alpha}{x^{\alpha+1}}$	$\alpha > 0, \beta > 0, x > \beta$	$\frac{\beta\alpha}{(\alpha-1)}$	$\frac{\beta^2\alpha}{(\alpha-1)^2(\alpha-2)}$

End of Appendix.