Queen Mary
University of London

M. Sci. Examination by course unit 2014

MTH709U Bayesian Statistics

Duration: 3 hours

Date and time: 14 May 2014, 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 unauthorized 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 are provided.
A table of common distributions is provided.

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

Important note: the Academic Regulations state that possession of unauthorized material at any time by a student who is under examination conditions is an assessment offence and can lead to expulsion from QMUL.

Please check now to ensure you do not have any notes, mobile phones or unauthorised electronic devices on your person. If you have any, then please raise your hand and give them to an invigilator immediately. Please be aware that 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. Disruption caused by mobile phones is also an examination offence.

Exam papers must not be removed from the examination room.

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

Question 1 (5 marks) Write down three advantages and two disadvantages of the Bayesian paradigm compared to classical (frequentist) statistics. [5]

Question 2 (15 marks) (a) Show that a Poisson distribution with mean θ is an exponential family. [4]

(b) Identify the natural parameter. [1]

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

(d) It is believed that the number of accidents in a new factory will follow a Poisson distribution with mean θ per month. The prior distribution of θ is given by a gamma distribution with density

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

(i) If $\alpha = 12, \beta = 4$ and there are 27 accidents in the first eight months, derive the posterior distribution of θ and find its mean and variance. [6]

(ii) Show the posterior mean can be written as a weighted average of the prior mean and the sample mean. [3]

Question 3 (17 marks) A chemist wishes to learn about the melting point, $\theta°C$, of a newly developed compound, under specific laboratory conditions. The melting point can be measured using an instrument which is known to have measurement errors which are normally distributed, with mean zero and standard deviation $0.3°$. Errors on separate observations are independent.

(a) Before carrying out any measurements the chemist's judgement of the most likely value of θ is 10. She also assesses her probability that θ is below 5 to be 0.1. Assuming her beliefs can be adequately represented by a normal distribution, obtain the parameters of her prior. With this prior what is the chemist's prior probability that $\theta < 0$? [5]

(b) The chemist makes four observations with values (in $°C$) 6.1, 4.9, 5.5, 5.8. Calculate her posterior distribution for θ based on these observations. What is the chemist's posterior probability that $\theta < 0$? [7]

(c) If instead of the prior distribution obtained above the chemist had used a non-informative uniform prior for θ, what would her posterior mean and variance have been? Briefly compare this posterior distribution with the one you found in (b) and comment. [5]

Question 4 (25 marks) A production line in a factory manufactures a large number of electrical components. The proportion manufactured that are usable is given by θ, which is unknown, the remainder are faulty.

(a) A random sample of n components is to be tested. The number of usable components, X, can be taken to have a binomial distribution with parameters n and θ. If $X = X_1 + \cdots + X_n$ where X_i is 1 if the ith component is usable and 0 otherwise, explain what assumptions are necessary to justify the binomial model. [2]

(b) A manager, who has just started at the factory, has prior beliefs based on similar production lines. He believes with probability 0.95 that the true proportion of usable components lies between 93% and 98%. Explain why a Beta distribution with parameters 250 and 12 (Be(250,12)) might be a suitable representation of his beliefs about θ. [6]

(c) A random sample of size $n = 24$ is found to contain 23 usable components and a single faulty one. Calculate the manager's posterior distribution about θ after this sample has been collected. If he has to make an estimate of θ under a quadratic loss function, what would be his optimal estimate and expected loss? [7]

(d) An alternative approach to sampling involves testing a batch of m components together giving an observation Z defined by

$$Z = \begin{cases} 1 & \text{if all components are usable} \\ 0 & \text{otherwise.} \end{cases}$$

Explain why the likelihood function for θ based on such an observation takes the form

$$p(z|\theta) = \begin{cases} \theta^m & \text{if } z = 1 \\ 1 - \theta^m & \text{if } z = 0, \end{cases}$$

for $0 \leq \theta \leq 1$. [4]

(e) Suppose a beta prior distribution (Be(a, b)) for θ with parameters a and b is assumed, followed by a single observation of Z based on a batch of m components. Calculate the posterior distribution for θ in terms of a, b and m if $z = 0$. [6]

Question 5 (13 marks) (a) A random sample of failure times $t_1, \ldots, t_n$ is observed for n machine components. Each t_i is assumed to have an exponential distribution with mean λ^{-1} and the prior distribution for λ is taken as Gamma ($Ga(\alpha, \beta)$) with parameters α and β. Find the posterior distribution of λ. [3]

(b) Show that for the situation described in (a), the prior predictive density (or marginal likelihood) is given by

$$p(t_1, \ldots, t_n) = \frac{\beta^\alpha \Gamma(\alpha + n)}{\Gamma(\alpha)(\beta + S_n)^{\alpha+n}}$$

where $S_n = \sum_{i=1}^n t_i$. [5]

(c) Explain how you could test the null hypothesis that $\lambda = \alpha/\beta$ against an alternative that λ has a $Ga(\alpha, \beta)$ prior. [5]

Question 6 (25 marks) (a) Consider a two-stage linear model

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

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

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

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

[5]

(b) Observations are taken from a model

$$y_{ij} \sim No(\alpha_i, \xi) \qquad \text{for } i = 1, 2 \text{ and } j = 1, \ldots, n_i.$$

Note that ξ is the precision. The priors for α_1 and α_2 are respectively $No(4, 2\xi)$ and $No(6, 3\xi)$ and they are assumed independent. The prior for ξ is $Ga(a/2, b/2)$.

(i) Find the posterior distributions of α_1 and α_2 given ξ, $p(\alpha_1|\underline{y}, \xi)$ and $p(\alpha_2|\underline{y}, \xi)$. [7]

(ii) Find the posterior distribution of ξ given α_1 and α_2, $p(\xi|\underline{y}, \alpha_1, \alpha_2)$. [5]

(iii) Explain how samples from the unconditional posterior distributions of α_1, α_2 and ξ can be simulated using Gibbs sampling. [5]

(iv) Given these samples how would you estimate the posterior median of $\alpha_1 - \alpha_2$? [3]

End of Paper