Queen Mary
University of London

MTH709U / MTHM042: Bayesian Statistics

Duration: 3 hours

Date and time: 10 May 2016, 2:30-5:30

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, J Griffin

Question 1 (5 marks). Explain what is meant by a non-informative prior. Give the formula for finding a Jeffreys' prior. Give one advantage and one disadvantage of using a non-informative prior.

Question 2 (17 marks). (a) Show that a normal distribution with known mean 0 and unknown precision θ is an exponential family. [**3**]

(b) Identify the natural parameter. [**1**]

(c) If a random sample of n observations $x_1, x_2, \ldots, x_n$ from such a normal distribution is collected write down a minimal sufficient statistic for θ. [**1**]

(d) Show that a Gamma distribution $Ga(a/2, b/2)$ is a conjugate prior for θ by finding the density of the posterior distribution of θ. [**4**]

(e) Find the density of the predictive distribution of another independent observation y. [**4**]

(f) Identify this distribution and its parameters. [**4**]

Question 3 (20 marks). (a) A random sample of observations $x_1, \ldots, x_n$ is assumed to have uniform distribution on the interval $(0, \theta)$. The prior distribution for θ is taken as Pareto ($Par(\alpha, \theta_0)$) with probability density function

$$p(\theta) = \frac{\alpha\theta_0^\alpha}{\theta^{\alpha+1}} \qquad \theta \geq \theta_0.$$

Show that the posterior distribution of θ is $Par(\alpha + n, S)$ where S is to be determined. [**4**]

(b) Show that the Bayes estimate of θ under the loss function

$$L(t, \theta) = \frac{(t - \theta)^2}{\theta},$$

where t is any estimate of θ, is given by $\{E(\theta^{-1} \mid \underline{x})\}^{-1}$. [**3**]

(c) Find the Bayes estimate of θ for the posterior in (a) using the loss function in (b). [**3**]

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

$$p(x_1, \ldots, x_n) = \frac{\alpha{\theta_0}^\alpha}{(\alpha + n)S^{\alpha+n}}.$$

[**3**]

(e) If $\alpha = 3$, $\theta_0 = 2$, $n = 5$, and the 5 observations were 0.6, 1.7, 0.5, 1.9, 0.2, test the null hypothesis that $\theta = 2$ against an alternative that θ has a $Par(\alpha, \theta_0)$ prior. [**5**]

(f) What would your conclusion have been if the largest observation was 2.1 instead of 1.9? [**2**]

Question 4 (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 $A_1, A_2, C_1, C_2, C_3, \mu$ are known. Using the results for such a three stage model and the matrix lemma given in the note at the end of this question, show that when $C_3^{-1} \to 0$ the posterior distribution of $\underline{\theta}_1$ is $N(D_0\underline{d}_0, D_0)$ where

$$\begin{aligned} D_0^{-1} &= A_1^T C_1^{-1} A_1 + C_2^{-1} - C_2^{-1} A_2 (A_2^T C_2^{-1} A_2)^{-1} A_2^T C_2^{-1} \\ \underline{d}_0 &= A_1^T C_1^{-1} \underline{y} \end{aligned}$$

[**5**]

(b) Write the following problem in terms of a three stage linear model.

$$y_i \sim N(\beta x_i + \gamma z_i, 1) \qquad i = 1, \ldots, n$$

$$\beta \sim N(\mu, 2) \qquad \gamma \sim N(\mu, 2) \qquad \beta \text{ and } \gamma \text{ independent}$$

and μ has a prior with zero precision. [**4**]

If 5 observations of y and the associated regressors were as follows

y	x	z
3	−2	−2
5	−1	0
8	0	2
11	1	3
15	2	5

find the joint posterior distribution of β and γ. [**6**]

Hence calculate a 95% highest posterior density interval for $\beta - \gamma$. [**4**]

NOTE For a three stage linear model given in the question the posterior distribution of θ_1 is $N(D\underline{d}, D)$ where

$$\begin{aligned} D^{-1} &= A_1^T C_1^{-1} A_1 + (C_2 + A_2 C_3 A_2^T)^{-1}, \\ \underline{d} &= A_1^T C_1^{-1} \underline{y} + (C_2 + A_2 C_3 A_2^T)^{-1} A_2 \underline{\mu}. \end{aligned}$$

Matrix lemma: For any matrices A_1, C_1, C_2 of appropriate dimensions for which the inverses stated in the result exist we have

$$C_1^{-1} - C_1^{-1} A_1 (A_1^T C_1^{-1} A_1 + C_2^{-1})^{-1} A_1^T C_1^{-1} = (C_1 + A_1 C_2 A_1^T)^{-1}$$

Question 5 (23 marks). (a) Suppose the posterior density of a one-dimensional parameter θ, $p(\theta|\underline{y})$, is unimodal and roughly symmetric. By considering a Taylor series expansion of $\log p(\theta|\underline{y})$ about the posterior mode $\hat{\theta}$ show that $p(\theta|\underline{y})$ can be approximated by a normal distribution with mean $\hat{\theta}$ and a variance which you should determine. [**6**]

(b) A biased coin is tossed n times and results in y heads. Let θ be the chance the coin lands heads. If the prior for θ is a beta distribution

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \qquad 0 < \theta < 1,$$

find the posterior of θ and its mean and variance. [**3**]

(c) Use result (a) to determine a normal approximation to the posterior in (b). [**6**]

(d) If $n = 20$, $y = 7$, $a = 2$, $b = 1$ find a 99% highest posterior density interval for θ based on this normal approximation. [**4**]

(e) Calculate the highest posterior density interval for θ using Table 29 in the New Cambridge Statistical Tables. [**2**]

(f) Comment on the differences between the two intervals. [**2**]

Question 6 (16 marks). Suppose we have observations x_{ij} which have a Poisson distribution with mean θ_j for $j = 1, 2, \ldots, p$ and $i = 1, 2, \ldots, n_j$. The θ_j are independent and distributed with Gamma distributions $Ga(\alpha_j, \beta)$. The parameters α_j are known but β is unknown and assigned a Gamma distribution $Ga(\gamma, \delta)$.

(a) Derive the necessary conditional distributions for Gibbs sampling and explain how this would proceed. [**9**]

(b) How would you estimate the median of θ_1? [**3**]

(c) Explain how the algorithm would have to change if instead of the θ_j being independent it is known that $\theta_1 < \theta_2 < \cdots < \theta_p$. [**4**]

End of Paper—An appendix of 2 pages 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

Distribution	Density	Range of 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 $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp[\frac{-(x-\mu)^2}{2\sigma^2}]$	$-\infty < \mu < \infty$ $\sigma > 0,\ -\infty < x < \infty$	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)	$\frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}$	$\beta > 0,\ \alpha > 0,\ x > 0$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$

Distribution	Density	Range of Variates	Mean	Variance
Beta (a, b)	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$	$a > 0,\ b > 0,\ 0 < x < 1$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$
$t_\nu(m, g)$	$\frac{g^{1/2}\Gamma((\nu+1)/2)}{\sqrt{(\nu\pi)}\Gamma(\nu/2)}$	$-\infty < x < \infty$	location m	precision g,
	$\times \left[1 + \frac{g}{\nu}(x-m)^2\right]^{-(\nu+1)/2}$	dof ν		
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.