Data provided: Tables of Normal Quantiles. Tables of distribution functions

MAS364

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Autumn Semester 2014–15**

Bayesian Statistics **2 hours**

Candidates may bring to the examination a calculator which conforms to University regulations. Marks will be awarded for your best ***three*** *answers. Total marks 84.*

Standard results from the lecture notes may be used without derivation, but must be clearly stated.

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 Consider the following hierarchical model

$$x_i \sim \mathrm{N}(x_i \,|\, \mu_i, 1/\lambda)\ , \quad \text{independent for } i = 1, \ldots, n$$
$$\pi\,(\mu_i) = \mathrm{N}(\mu_i \,|\, \theta, 1/\gamma)\ , \quad \text{independent for } i = 1, \ldots, n$$
$$\pi\,(\lambda) = \mathrm{Ga}(\lambda \,|\, a, b)\ , \quad \pi\,(\theta) = \mathrm{N}(\theta \,|\, m, 1/p) \quad \text{and} \quad \pi\,(\gamma) = \mathrm{Ga}(\gamma \,|\, c, d)\ ,$$

with $\{m, p, a, b, c, d\}$ known constants.

(i) Prove that the full conditionals for

(a) μ_i are $\mathrm{N}\big(\mu_i \,|\, m_i^\star, 1/p^\star\big)$ and give explicit formulae for $m_i^\star$ and $p^\star$. ***(6 marks)***

(b) θ is $\mathrm{N}(\theta \,|\, q^\star, 1/v^\star)$ and give explicit formulae for $q^\star$ and $v^\star$. ***(6 marks)***

(c) λ is $\mathrm{Ga}(\lambda \,|\, a^\star, b^\star)$ and give explicit formulae for $a^\star$ and $b^\star$. ***(4 marks)***

(d) γ is $\mathrm{Ga}(\gamma \,|\, c^\star, d^\star)$ and give explicit formulae for $c^\star$ and $d^\star$. ***(4 marks)***

(ii) Write pseudo-code of a Gibbs sampler for exploring the posterior of the parameters for this model. ***(8 marks)***

2 A physicist studying the expansion of the universe has two sets of measurements covering the same section of the Milky Way. The difference between these measurements is related to redshift and, if current theory is correct, expected to be very close to zero.

(i) Let d_i be the i-th observed difference and assume they are conditionally independent with $d_i \sim \mathrm{N}(d_i \mid \mu, 1/\lambda)$ and $i = 1, \ldots, n$. Write down the joint posterior distribution of (μ, λ) using the conjugate prior,

$$\pi(\mu, \lambda) = \mathrm{N}\left(\mu \,\middle|\, m, \frac{1}{p\lambda}\right) \mathrm{Ga}(\lambda \mid a, b) ,$$

and give explicit expressions for its parameters. ***(10 marks)***

(ii) Given the data, the scientist may report a discrepancy (call this decision d_1) or and agreement (decision d_2) with the current theory. If a real discrepancy is reported there is a good chance it would be published in a top journal; if the discrepancy is not real, his career would suffer a major drawback. After some consideration, he believes that the following loss function reflects well his preferences:

$$\mathrm{L}(d_1, \mu) = \begin{cases} 100 & |\mu| \leq 0.5 \\ 0 & |\mu| > 0.5 \end{cases}, \qquad \mathrm{L}(d_2, \mu) = \begin{cases} 0 & |\mu| \leq 0.5 \\ 30 & |\mu| > 0.5 \end{cases}.$$

A set of $n = 150$ measures is taken and the following statistics are recorded:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = -0.3 , \quad s_d^2 = \frac{1}{n}\sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2 = 10 .$$

After elicitation, the scientist's prior parameters are $m = 0$, $p = 0.5$, $a = 2$ and $b = 0.6$. Should the scientist report a discrepancy? ***(18 marks)***

3 Let x_i be the number of complaints filed to a consumer agency in a given day, and let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be a random sample obtained from the agency's records. Assuming that $\mathrm{Po}(x_i \mid \lambda)$ is a suitable model:

(i) (a) Prove that the posterior from the non-informative (improper) prior $\pi(\lambda) \propto \lambda^{-1}$ is a Gamma distribution and write down the posterior parameters explicitly. ***(5 marks)***

(b) Prove that $\pi(\lambda) = \mathrm{Ga}(\lambda \mid a, b)$ is a conjugate prior and write down the posterior parameters explicitly. ***(5 marks)***

(ii) It is further assumed that the distribution of the x_i arise from a process where the distribution of the waiting time to the next complaint, t, is exponential with the same parameter λ.

(a) Using your results in (i) (a), prove that the predictive distribution of t is

$$f(t \mid \boldsymbol{x}) = n^s s(n+t)^{-(s+1)}\,, \quad \text{where} \quad s = \sum_{i=1}^{n} x_i$$

(6 marks)

(b) The agency's manager claims that it is almost certain that the next complaint will be filed before midday; *i.e.* $t \leq 1/2$. Does the sample $\boldsymbol{x} = \{0, 2, 1, 4, 3, 4, 3, 0, 2, 1\}$ provide evidence to support this statement? ***(6 marks)***

(iii) (a) Using your results in (i)(b), prove that the predictive distribution of y, the number of complaints filed in next day is

$$f(y \mid \boldsymbol{x}) = \frac{b^{\star a^{\star}}}{y!} \frac{\Gamma(a^{\star} + y)}{(a^{\star} + 1)^{(b^{\star} + y)}}$$

with $\{a^{\star}, b^{\star}\}$ the parameters of the posterior distribution. ***(6 marks)***

4 A laboratory supervisor has been asked to validate and calibrate the thermostat of a furnace. The calibration process can be posed as

$$y_i = \beta x_i + \varepsilon_i$$

where y_i is the i-th measurement obtained with the thermometer, x_i is the corresponding reading from the furnace thermostat, $\beta i n \mathbb{R}$ is the calibration coefficient and ε_i is an independent measurement error, assumed to follow a $\mathrm{N}(\varepsilon_i \,|\, 0, 1/\lambda)$ distribution.

30 measurements are taken under homogenous conditions and record the temperature reading from the furnace and a high quality thermometer. The data have been pre-processed so that $\sum x_i = \sum y_i = 0$; the recorded statistics are

$$s_{xx} = \sum_{i=1}^{n} x_i^2 = 1, \quad s_{yy} = \sum_{i=1}^{n} y_i^2 = 11 \quad \text{and} \quad s_{xy} = \sum_{i=1}^{n} x_i y_i = 3 \, .$$

The supervisor has asked for your help and after trying to elicit her prior opinions she has decided to use the non-informative (improper) prior $\pi\,(\beta, \lambda) \propto \lambda^{-1}$ for the analysis.

(i) According to regulations, the thermostat is said to be within normal limits if $|\beta| < 4$. Use a Gaussian approximation to the marginal posterior to calculate a HPD interval of probability 0.95 to advise the supervisor on whether the thermostat is within normal limits or not. ***(15 marks)***

(ii) The calibration is said to be admissible if the precision of the measurement error is larger than 15. Using a normal approximation to the marginal posterior of λ, provide a credible interval of approximate probability 0.95 and decide if the calibration is admissible. ***(13 marks)***

End of Question Paper