NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2012/2013

MAS3317

Introduction to Bayesian Statistics

Time allowed: 1 hour 30 minutes

Candidates should attempt all questions. Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are FOUR questions in Section A and THREE questions in Section B.

Answers to questions should be entered directly on this question paper in the spaces provided. Rough work should be done on the blank sides of the pages. The rough work will not be marked. This question paper must be handed in, attached inside an anonymised cover sheet, at the end of the examination.

Beta distribution: If $X \sim Be(\alpha, \beta)$ then it has density

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1, \quad \alpha > 0, \beta > 0.$$

Also, $E(X) = \alpha/(\alpha+\beta)$, $Var(X) = \alpha\beta/\{(\alpha+\beta)^2(\alpha+\beta+1)\}$ and $Mode(X) = (\alpha-1)/(\alpha+\beta-2)$.

Exponential distribution: If $X \sim Exp(\lambda)$ then it has density

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

Also, $E(X) = 1/\lambda$ and $Var(X) = 1/\lambda^2$.

Gamma distribution: If $X \sim Ga(\alpha, \lambda)$ then it has density

$$f(x|\alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x > 0, \quad \alpha > 0, \lambda > 0.$$

Also, $E(X) = \alpha/\lambda$ and $Var(X) = \alpha/\lambda^2$.

Normal distribution: If $X \sim N(\mu, 1/\tau)$ then it has density

$$f(x|\mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau}{2}(x-\mu)^2\right\}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \tau > 0.$$

Also, $E(X) = \mu$ and $Var(X) = 1/\tau$.

Poisson distribution: If $X \sim Po(\lambda)$ then it has probability function

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \ldots, \quad \lambda > 0.$$

Also, $E(X) = \lambda$ and $Var(X) = \lambda$.

SECTION A

A1. *Econase XT* is an enzyme added to pig feed that has been specially designed to make pigs gain weight more quickly. In a recent trial, it was noticed that some pigs may have developed trotter rot as a result of consuming feed containing the enzyme. The company developing the enzyme report that the probability of a pig developing trotter rot can increase from 0.1 (for a pig given standard feed) to 0.45 (for a pig given feed containing *Econase XT*). Of the 90 pigs in the trial, suppose n were given feed containing *Econase XT*.

Let XT : Pig given the enzyme, and TR : Pig has trotter rot.

(a) A vet examines a pig at random. Given that it has trotter rot, what is the probability it had been given special feed containing the enzyme?

Answer:

(b) How many pigs should be given feed containing the enzyme if we require $\Pr(XT|TR) < 0.5$?

Answer:

[**Total A1: 9 marks**]

A2. As an employee of the American Federal Aviation Authority (FAA), you need to model the arrival rate (λ per minute) of passenger planes arriving at U.S. airports. Experience suggests that a gamma distribution might be suitable for λ.

After a consultation with an air traffic controller at New York's John F. Kennedy airport, you are able to elicit that $E(\lambda) = 1.5$ and $SD(\lambda) = \sqrt{6}/4$.

(a) Show that

$$\pi(\lambda) = \frac{512}{15}\lambda^5 e^{-4\lambda}, \qquad \lambda > 0.$$

Answer:

(b) Assume that the number of passenger planes, X per minute, arriving at John F. Kennedy airport is Poisson distributed with rate λ (and planes arrive independently of each other). In a one hour period, you observe an average of 2.2 planes arriving per minute. Combine this information with the elicited prior distribution in part (a) to show that

$$\lambda|\boldsymbol{x} \sim Ga(138, 64).$$

Answer:

Question A2 continued on next page

[Some space is left here for your solution to the previous question]

(c) Find α, the weight in the Bayes linear rule for the posterior mean.

[**Total A2: 15 marks**]

A3. Suppose we have data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ available and we want to make inferences about the parameter ζ in the statistical model $f(\boldsymbol{x}|\zeta)$. Suppose also that T is a sufficient statistic for ζ.

(a) Give an intuitive explanation of the statement "T is a sufficient statistic for ζ".

Answer:

(b) State the Factorisation Theorem.

Answer:

(c) Show that our posterior beliefs about ζ having observed all the data $\boldsymbol{x}$ are the same as if we had only observed the sufficient statistic T. *[Hint: You may use the fact that, if $f(\boldsymbol{x}|\zeta) = h(\boldsymbol{x})g(t, \zeta)$, for some functions h and g, then $g(t, \zeta) \propto f_T(t|\zeta)$]*

Answer:

[**Total A3: 10 marks**]

A4. (a) Consider a statistical model for data which depends on an unknown parameter θ. Describe the differences in interpretation of C_F, a 95% frequentist confidence interval for θ and C_B, a 95% Bayesian confidence intervals for θ.

Answer:

(b) Briefly describe how subjective probabilities might be evaluated using bets.

Answer:

[**Total A4: 6 marks**]

SECTION B

B5. As part of a research project, geologists are interested in rock which degrades at a rate of more than 0.12mm per year. Gravestones provide a useful means of measuring this rate. Let θ be the probability that, in a particular cemetery, a gravestone indicates an annual rate of degradation of more than 0.12mm.

(a) *Briefly* explain why we might use the prior distribution $\pi_A(\theta) = 1$ to reflect prior ignorance about θ.

Answer:

(b) A statistician on the research team uses the *trial roulette method* to elicit a prior distribution for θ. A consultation with one of the geologists results in the screenshot shown overleaf, taken from the *MATCH Uncertainty Elicitation Tool*.

(i) Write down the elicited prior distribution for θ, rounding the parameters of this distribution to the nearest integer. Also write down the corresponding density function $\pi_B(\theta)$.

Answer:

(ii) Use the `Feedback percentiles` to phrase a question to the expert to check the suitability of the fitted distribution.

Answer:

Question B5 continued on next page

Question B5 continued on next page

(c) Of the twenty gravestones in a randomly chosen cemetery, twelve show an annual rate of degradation of more than 0.12mm. Given this information, obtain the posterior distribution for θ assuming

(i) the 'ignorance' prior $\pi_A(\theta)$;

Answer:

(ii) the elicited prior $\pi_B(\theta)$.

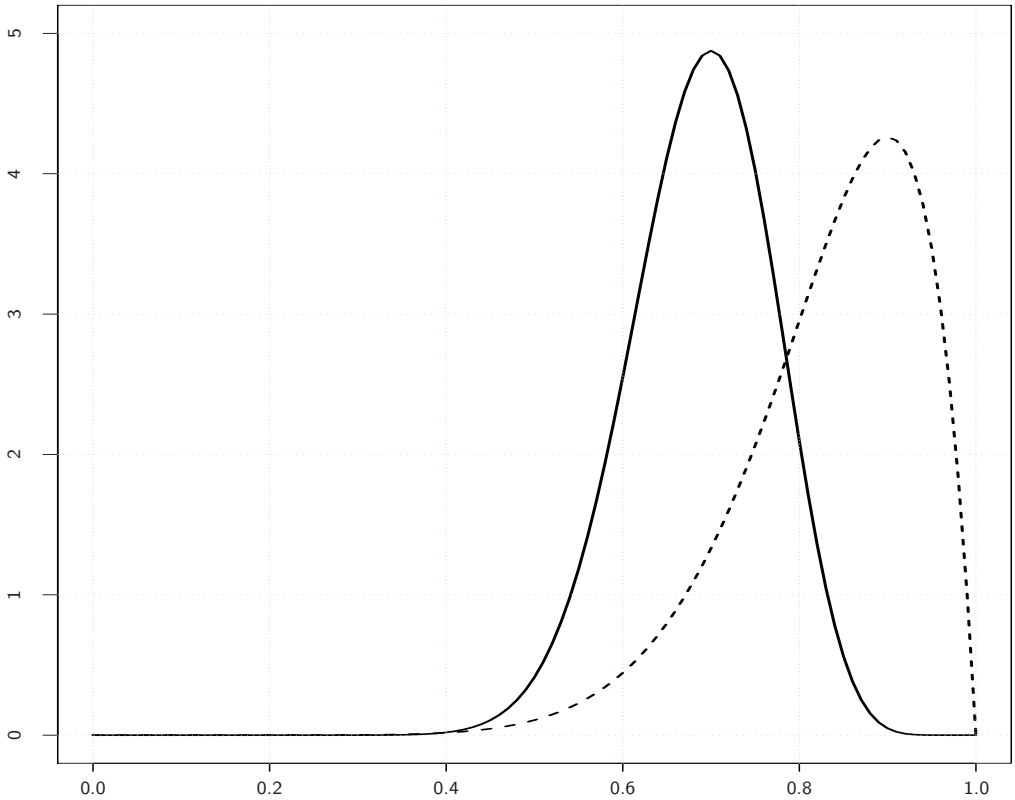
Answer:

(iii) Is the 'ignorance' prior *conjugate*? Explain.

Answer:

Question B5 continued on next page

(d) The graph below shows the elicited prior $\pi_B(\theta)$ as provided by *MATCH* (dashed line), as well as the corresponding posterior (as obtained in part (c)(ii) of this question – solid line). On the same graph, sketch the 'ignorance' prior $\pi_A(\theta)$ and the resulting posterior (as obtained in part (c)(i) of this question), and label the axes of the graph.



[Total B5: 18 marks]

B6. In hydrology, the *Pareto* distribution, with probability density function

$$f(x|\kappa,\theta) = \theta\kappa^{\theta}x^{-(\theta+1)}, \qquad \theta,\kappa > 0 \text{ and } x > \kappa,$$

is often applied to extreme events, such as annual maximum one day rainfall accumulations. For Eskdale, a location in the Lake District, suppose we are told that $\kappa = 30$ mm is a plausible lower bound for annual maximum daily rainfall totals.

(a) Obtain the likelihood function for θ, given that we observe a sample of annual maximum rainfall accumulations $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ at this location (and we assume these observations are independent).

[**2 marks**]

Answer:

(b) Suppose our prior beliefs about θ can be described by an exponential $Exp(g)$ distribution. Show that the posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim Ga\left(n+1, g - n\log 30 + \sum_{i=1}^{n} \log x_i\right).$$

[**5 marks**]

Answer:

Question B6 continued on next page

[Some space is left here for your solution to the previous question]

(c) Describe the term *vague prior knowledge* and explain why, in this example, this is represented as $g \to 0$.

[**3 marks**]

Answer:

(d) In 2010, 2011 and 2012, annual maximum daily rainfall accumulations at Eskdale were 61mm, 55mm and 60mm (respectively). Assuming vague prior knowledge,

(i) find the posterior distribution for θ, rounding all parameters in this posterior to the nearest integer for convenience;

[**2 marks**]

Answer:

(ii) show that, for a 95% Bayesian confidence interval (a, b) for θ,

$$\int_a^b \theta^3 e^{-2\theta} d\theta = \frac{57}{160};$$

[**4 marks**]

Answer:

(iii) write down an additional relationship between a and b that must hold for the Bayesian confidence interval to be one of *highest density*;

[**3 marks**]

Answer:

Question B6 continued on next page

(iv) use the following output from R to write down the 95% highest density interval for $\theta|\boldsymbol{x}$, and briefly explain the content of the function g.

[5 marks]

```
>g=function(x)
   {
    a=x[1]
    b=x[2]
    (pgamma(b,2,4)-pgamma(a,2,4)-0.95)^2
      +0.0001*(dgamma(b,2,4)-dgamma(a,2,4))^2
   }

>initiala=qgamma(0.025,2,4)
>initialb=qgamma(0.975,2,4)

>optim(c(initiala,initialb),g)
$par
[1] 0.01054675 1.19126680

$value
[1] 5.287579e-11
```

[Total B6: 24 marks]

B7. A random sample of size n is taken from a Normal $N(\mu, 1/\tau)$ distribution (where the precision τ is known) giving a sample mean $\bar{x}$.

(a) Obtain the log–likelihood function for μ as a function of μ, τ and the data $\boldsymbol{x}$.

Answer:

(b) Suppose we are *ignorant* about μ prior to any data collection. Derive the Jeffreys prior for μ.

Answer:

Question B7 continued on next page

[Some space is left here for your solution to the previous question]

(c) Explain why your Jeffreys prior in part (b) is *improper*.

Answer:

(d) What is the main drawback of using uniform or improper priors to represent prior ignorance?

Answer:

Question B7 continued on next page

(e) In this example, why would using a Jeffreys prior result in the same posterior as if we had used a prior based on vague prior knowledge?

Answer:

(f) Explain why, no matter what our degree of prior belief about μ, for large n

$$Var(\mu|\boldsymbol{x}) \approx \frac{1}{n\tau}.$$

Answer:

[**Total B7: 18 marks**]

THE END