NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2012/2013

MAS2317/3317

Introduction to Bayesian Statistics: Mid–Semester Test

SOLUTIONS

Time allowed: 50 minutes

Candidates should attempt all questions. Marks for each question are indicated.

There are SIX questions on this paper.

Answers to questions should be entered directly on this question paper in the spaces provided. This question paper must be handed in at the end of the test.

Name: Lee Fawcett

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

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 $X \sim Po(\theta)$ then it has probability mass function

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

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

1. Consider the scenarios below. State whether a classical, frequentist or subjective interpretation of probability is being used to estimate θ_1, θ_2 and θ_3.

Scenario	**Classical? Frequentist? Subjective?**
Your friend plays the piano and has her Grade 8 exam in the morning. You and some friends try to evaluate $\theta_1 = \Pr(\text{she passes her exam})$.	Subjective
You work in the outbound sales team of a call centre. From a list of all potential customers, the computer selects one completely at random. $\theta_2 = \Pr(\text{customer is female})$.	Classical
You have an interest in horse racing. You visit three bookmakers in an attempt to determine $\theta_3 = \Pr(\textit{Bayesian Beauty} \text{ wins at the Grand National})$.	Subjective

[Total Q1: 3 marks]

2. Scientists in Venezuela know that, in any given year, there is a 25% chance of an earth tremour within a 50 mile radius of the capital city, Caracas. Such an event can trigger catastrophic landslides: given an earth tremour has occurred, a scientist specifies that there is a probability of 0.9 that this will be immediately followed by a catastrophic landslide. If an earth tremour does *not* occur, the probability of a catastrophic landslide drops to just 0.15.

Given that a catastrophic landslide occurs in Caracas in 2012, what is the probability that this was preceded by an earth tremour?

Answer:

$$\left.\begin{array}{l} ET = \text{Earth tremour}, \quad LS = \text{landslide.} \\ P(ET) = 0.25 \\ P(LS|ET) = 0.9\ ; \quad P(LS|ET^c) = 0.15 \end{array}\right\} \text{Information given.}$$

$$P(ET|LS) = \frac{P(LS|ET) \times P(ET)}{P(LS|ET) \times P(ET) + P(LS|ET^c)\, P(ET^c)}$$

$$= \frac{0.9 \times 0.25}{(0.9 \times 0.25) + (0.15 \times 0.75)}$$

[Total Q2: 4 marks]

$$= \frac{2}{3}$$

3. Suppose we have a random sample from the *Rayleigh distribution*, i.e. $X_i|\theta \sim$ *Rayleigh*(θ), $i = 1, 2, \ldots$ (independent), where the probability density function is given by

$$f(x|\theta) = \frac{2xe^{-x^2/\theta}}{\theta}, \qquad x, \theta > 0.$$

Use the Factorisation Theorem to find a sufficient statistic for θ.

Answer:

$$f(\underline{x}|\theta) = \prod_{i=1}^{n} \frac{2x_i e^{-x_i^2/\theta}}{\theta}, \quad \theta > 0.$$

$$= \prod_{i=1}^{n} 2x_i \times \theta^{-n} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n} x_i^2\right\}$$

$$= h(\underline{x}) \times g\left(\theta, \sum_{i=1}^{n} x_i^2\right),$$

where $h(\underline{x}) = \prod_{i=1}^{n} 2x_i$ and $g(\theta, t) = \theta^{-n} e^{-t/\theta}$,

where $t = \sum_{i=1}^{n} x_i^2$.

Hence, by the F.T., $T = \sum_{i=1}^{n} X_i^2$ is sufficient for θ.

[Total Q3: 5 marks]

4. On any day, the number of loggerhead turtles X_i observed off the northeast coast of Australia is assumed Poisson with rate θ, i.e. $X_i|\theta \sim Po(\theta)$, X_i independent.

(a) On each day of a three day diving trip, a diver sees 1, 3 and 2 loggerhead turtles respectively. Find the likelihood function $f(\boldsymbol{x}|\theta)$.

[4 marks]

Answer:

$$f(\underline{x}|\theta) = \frac{e^{-\theta}.\theta^1}{1!} \times \frac{e^{-\theta}.\theta^3}{3!} \times \frac{e^{-\theta}.\theta^2}{2!}, \quad \theta > 0$$

$$= \frac{\theta^6.e^{-3\theta}}{12}.$$

(b) You interview a marine biologist in an attempt to elicit her beliefs about the rate of occurrence of loggerhead turtles, θ. For this location, she tells you that she would expect to see about 5 turtles per day; thus $E[\theta] = 5$. After further discussions, you agree that $SD(\theta) \approx \sqrt{5/2}$.

(i) Use this information to suggest a suitable prior distribution for θ, specifying fully the parameters in your model.

[5 marks]

Answer:

We often use a gamma distribution for rates.

So use $\theta \sim Ga(a, b)$.

We want $\frac{a}{b} = 5$ and $\frac{a}{b^2} = \frac{5}{2}$.

Thus $a = 5b$, and substituting into the variance gives

$$\frac{5b}{b^2} = \frac{5}{2} \Rightarrow \frac{5}{b} = 2.5 \Rightarrow b = 2.$$

If $a = 5b$, then $a = 5 \times 2 = 10$, giving

$$\theta \sim Ga(10, 2).$$

Question 4 continued on next page

(ii) Show that the posterior distribution for the rate of occurrence of loggerhead turtles is

$$\theta|\boldsymbol{x} \sim Ga(16, 5).$$

[3 marks]

Answer:

$$\pi(\theta|\underline{x}) \propto \pi(\theta) \times f(\underline{x}|\theta)$$

$$\propto \theta^{9}e^{-2\theta} \times \theta^{6}e^{-3\theta}$$

$$\propto \theta^{16-1}e^{-5\theta}, \quad \theta > 0.$$

so $\theta|\underline{x} \sim Ga(16,5)$, as required.

(iii) The posterior mean can be written according to the *Bayes linear rule*

$$E(\theta|\boldsymbol{x}) = \alpha E(\theta) + (1-\alpha)\bar{x},$$

where $\bar{x}$ is the sample mean. Find α, and comment on whether this gives more weight to the prior or the ~~posterior~~. data.

[4 marks]

Answer:

We have $E(\theta|\underline{x}) = \frac{16}{5}$; $E(\theta) = \frac{10}{2}$; $\bar{x} = \frac{6}{3} = 2$.

So
$$5\alpha + 2(1-\alpha) = \frac{16}{5}$$

$$5\alpha + 2 - 2\alpha = \frac{16}{5}$$

$$3\alpha = 1.2$$

$$\alpha = 0.4$$

$\Rightarrow$ more weight to the data

Question 4 continued on next page

(c) Suppose you did not have time to interview the marine biologist. What would be your posterior distribution assuming *vague prior knowledge*? How does this affect your posterior variability?

[7 marks]

Answer:

Generally, we have

$$f(\underline{x}|\theta) \propto \theta^{\sum x_i} e^{-n\theta}, \quad \underline{\text{and}} \quad \pi(\theta) \propto \theta^{a-1} e^{-b\theta},$$

giving

$$\pi(\theta|\underline{x}) \propto \theta^{(a+\sum x_i)-1} e^{-(n+b)\theta},$$

$\underline{\text{so}}$

$$\theta|\underline{x} \sim Ga(a+\sum x_i, n+b).$$

For vague prior knowledge, we need $a \to 0$ and $b \to 0$ (see Example 3.7 in lecture notes), giving

$$\theta|\underline{x} \sim Ga(\sum x_i, n).$$

$$\sim Ga(6, 3).$$

[Total Q4: 23 marks]

$$Var(\theta|\underline{x}) = \frac{6}{9} = 0.667,$$

compared to

$Var(\theta|\underline{x}) = \frac{16}{25} = 0.64$, as we previously had.

5. The parameter θ represents the rate of the gamma or Pareto distributions, or the binomial success probability. Complete the table below, choosing **one** word from the following list for each empty cell in the table:

Gamma Normal Pareto

Beta Uniform Exponential

[Note: There is one mark for each empty cell. You must enter ***one*** *word only into each cell. Although more than one word might be appropriate, no marks will be given if you enter multiple answers.]*

Model: $X_i \sim$	**Prior for θ**	**Posterior for θ**
Exponential	Exponential	Gamma
Pareto	Gamma	Gamma
Binomial	Uniform	Beta

[Total Q5: 4 marks]

6. Hurricane–strength wind speeds (X miles per hour) for locations in the Gulf of Mexico, are modelled using a distribution with probability density function

$$f(x|\sigma) = \frac{x}{\sigma^2}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \quad x > 0, \quad \sigma > 0.$$

(a) During a hurricane, you collect wind speeds at n randomly chosen locations in the Gulf of Mexico, giving $x_1, x_2, \ldots, x_n$. Find the likelihood function for σ, assuming your observations are independent.

[3 marks]

Answer:

$$f(\underline{x}|\sigma) = \prod_{i=1}^{n} \frac{x_i}{\sigma^2} \exp\left\{-\frac{x_i^2}{2\sigma^2}\right\}$$

$$= \prod_{i=1}^{n} x_i \,(\sigma^{-2n})\, \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\}$$

(b) A meteorologist suggests using the *square root inverted gamma* distribution as a prior for σ, that is

$$\sigma \sim SqIG(a, b),$$

where

$$\pi(\sigma) \propto \sigma^{-(2b+1)}\exp\left\{-\frac{a}{2\sigma^2}\right\}, \qquad a, b > 0.$$

Find $\pi(\sigma|\boldsymbol{x})$, and explain why the prior suggested by the meteorologist is *conjugate*.

[8 marks]

Answer:

$$\pi(\sigma|\underline{x}) \propto \sigma^{-(2b+1)} \exp\left\{-\frac{a}{2\sigma^2}\right\} \times (\sigma^{-2n}) \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\}$$

$$= \sigma^{-(2b+1)-2n} \cdot \exp\left\{-\frac{1}{2\sigma^2}\left(a + \sum_{i=1}^{n} x_i^2\right)\right\}.$$

[The next page has been left blank for your solution to this question]

[This page has been left blank for your solution to the last question]

$$\pi(\sigma|\underline{x}) \propto \sigma^{-(2[b+n]+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(a+\sum_{i=1}^{n} x_i^2\right)\right\}$$

This is another SqIG distribution – in fact,

$$\sigma|\underline{x} \sim SqIG\left(a+\sum_{i=1}^{n} x_i^2,\ b+n\right).$$

Thus, the prior is conjugate as the prior and posterior are from the same family.

[Total Q6: 11 marks]

THE END