NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2012/2013

MAS2317/3317

Introduction to Bayesian Statistics: Mid–semester test

Specimen – 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.

This test is open–book; you may use two pages of your own notes to help.

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

Geometric distribution: If $X \sim Geo(\theta)$ then it has probability function

$$f(x|\theta) = (1-\theta)^{x-1}\theta, \quad x = 1, 2, \dots, \quad 0 \le \theta \le 1.$$

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

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 $Var(X) = 1/\tau$.

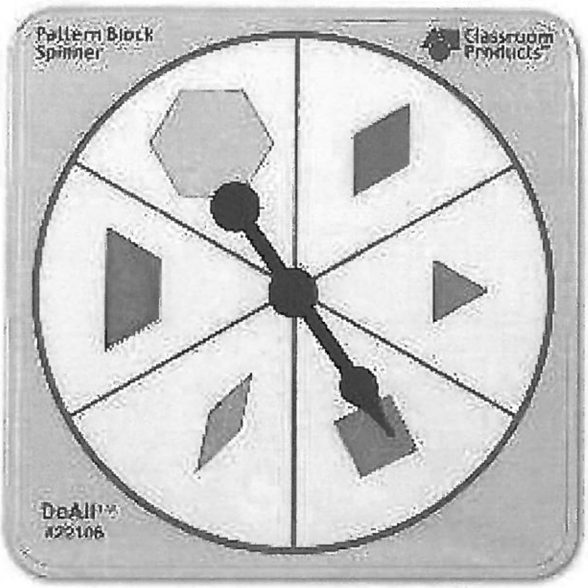
Poisson distribution: If $X \sim Po(\theta)$ then it has probability function

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

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

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

Scenario	Classical? Frequentist? Subjective?
A local bus company is concerned about the number of its bus services running late. To investigate, from its entire list of records it estimates $\theta_1 = \Pr(\text{a bus runs late})$ as the proportion of its services that have finished their route at least 10 minutes late.	Frequentist.
Your friend is running the *Great North Run* this year for the first time. You will give her £20 in sponsorship if she completes the run in under 100 minutes; you therefore try to evaluate $\theta_2 = \Pr(\text{she completes the run in under 100 mins})$.	Subjective.
You are playing a board game with your friends, and need to spin a quadrilateral on the spinner shown below to "get out of jail". You assess that $\theta_3 = \Pr(\text{spin a quadrilateral}) = \frac{2}{3}$.	Classical.



[Total Q1: 3 marks]

2. Three different machines (numbered 1, 2 and 3) were used for producing a large batch of similar manufactured items. Suppose that 20 percent of these items were produced by machine 1, 30 percent by machine 2 and 50 percent by machine 3.

Suppose further that 1 percent of the items produced by machine 1 are defective, that 2 percent of the items produced by machine 2 are defective, and that 3 percent of the items produced by machine 3 are defective.

Finally, suppose that one item is selected at random from the entire batch and it is found to be defective. What are the probabilities that this item was produced by each machine? Express these probabilities in the form $a/23$, for some integer a.

Answer:

Let D = Defective, and M_j = Machine j, $j = 1,2,3$.

By the law of total probability, we have

$$\sum_{j=1}^{3} Pr(D|M_j)\,Pr(M_j) = 0{\cdot}01 \times 0{\cdot}2 + 0{\cdot}02 \times 0{\cdot}3 + 0{\cdot}03 \times 0{\cdot}5$$
$$= 0{\cdot}023.$$

So, by Bayes' Theorem,

$$Pr(M_1|D) = \frac{Pr(D|M_1)\,Pr(M_1)}{0{\cdot}023} = \frac{0{\cdot}01 \times 0{\cdot}2}{0{\cdot}023} = \frac{2}{23}.$$

$$Pr(M_2|D) = \frac{Pr(D|M_2)\,Pr(M_2)}{0{\cdot}023} = \frac{0{\cdot}02 \times 0{\cdot}3}{0{\cdot}023} = \frac{6}{23}.$$

So

$$Pr(M_3|D) = 1 - \left\{\frac{2}{23} + \frac{6}{23}\right\} = \frac{15}{23}.$$

[Total Q2: 6 marks]

3. Suppose we have a random sample from a gamma distribution, i.e. $X_i \sim Ga(\alpha, \lambda)$, $i = 1, 2, \ldots$, where α is known.

Use the Factorisation Theorem to show that $S = \sum_{i=1}^{n} X_i$ is sufficient for λ.

Answer:

We have

$$f(\underline{x}|\lambda) = \prod_{i=1}^{n} \frac{\lambda^\alpha x_i^{\alpha-1} e^{-\lambda x_i}}{\Gamma(\alpha)}$$

$$= \prod_{i=1}^{n} x_i^{\alpha-1} \times \prod_{i=1}^{n} \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x_i}$$

$$= \prod_{i=1}^{n} x_i^{\alpha-1} \times (\Gamma(\alpha))^{-n} \lambda^{\alpha n} e^{-\lambda \sum x_i}$$

$$= h(\underline{x}) \times g\left(\sum x_i, \lambda\right),$$

where $h(\underline{x}) = \prod_{i=1}^{n} x_i^{\alpha-1}$ and $g(s, \lambda) = (\Gamma(\alpha))^{-n} \lambda^{\alpha n} e^{-\lambda s}$.

Therefore, by the Factorisation Theorem,

$S = \sum_{i=1}^{n} X_i$ is sufficient for λ.

[Total Q3: 5 marks]

4. *Wassaw Island* is a barrier island off the southeast coast of the USA, and is on the migratory path of the rare black pelican as it heads north (in April) to its breeding ground. As part of a study of these birds, we are interested in X, the number of birds we count, once the first black pelican is observed. We assume a geometric distribution for X with success probability p, that is, $X \sim Geo(p)$.

   (a) At four (independent) observation stations on the island we count a total of $\boldsymbol{x} = (8, 6, 9, 4)$ birds once the first black pelican is observed. Use this information to find the likelihood function $f(\boldsymbol{x}|p)$.

   [2 marks]

   Answer:

$$f(\underline{x}|p) = (1-p)^7 \cdot p * (1-p)^5 p \times (1-p)^8 \cdot p \times (1-p)^3 \cdot p$$

$$= (1-p)^{23} \cdot p^4, \quad 0 \leq p \leq 1.$$

   (b) Suppose, prior to any data collection, we have no knowledge about the chances of observing a black pelican on the island, that is, we are *prior ignorant* about p. In the space below, suggest a suitable prior distribution for p, and write down it's density (π_A). Also write down your prior mean, and briefly explain why this prior is appropriate for p.

   [6 marks]

   Answer:

$$p \sim U(0,1)$$

$$\pi_A(p) = \frac{1}{1-0} = 1, \quad 0 \leq p \leq 1.$$

$$E[p] = \frac{1+0}{2} = \frac{1}{2}$$

   Explanation:

   Appropriate because all values of p are equally likely - reflects our lack of prior knowledge.

(c) We would now like to elicit a more appropriate prior for p, say π_{B}.

(i) After consulting with an ornithologist (a zoologist with an expertise in birds), we find that the most likely value for p is around 0.2. Assuming that $p \sim Beta(a, b)$, show that

$$a = \frac{1}{4}(b + 3).$$

[3 marks]

Answer:

$$\text{Mode}(p) = \frac{a-1}{a+b-2} = 0.2$$

$$\Rightarrow a-1 = 0.2a + 0.2b - 0.4$$

$$\Rightarrow 0.8a = 0.2b + 0.6$$

$$\Rightarrow 8a = 2b + 6$$

$$a = \tfrac{2}{8}b + \tfrac{6}{8}$$

$$= \tfrac{1}{4}(b+3), \text{ as required.}$$

(ii) The ornithologist also tells us that it is unlikely that p will be greater than 0.6. In fact, we elicit from her that $\Pr(p > 0.6) = 0.01$. Using this information, complete the following statement by filling in the blanks, making sure the integrand is a function of b and p only.

[4 marks]

Answer:

$$\int_0^{\underline{0.6}} \frac{p^{\frac{1}{4}(b+3)-1}(1-p)^{b-1}}{B(\underline{\tfrac{1}{4}(b+3)}, \underline{b})} dp - \underline{0.99} = 0.$$

Question 4 continued on next page

(iii) We can use R to solve the equation in part (ii) for b. Doing so gives $b = 8$ (to the nearest integer). Show that the full density for p is given by

$$\pi_B(p) = 360p^2(1-p)^7, \qquad 0 < p < 1,$$

where a has also been rounded to the nearest integer.

[4 marks]

Answer:

If $b=8$, $a = \frac{11}{4} = 2.75 = 3$ (to nearest integer).

so $\pi_B(p) = \dfrac{p^2(1-p)^7}{B(3,8)}$; $B(3,8) = \dfrac{2!\,7!}{10!} = \dfrac{1}{360}$.

so $\pi_B(p) = \dfrac{p^2(1-p)^7}{1/360} = 360p^2(1-p)^7$, $0<p<1$.

(d) Obtain the posterior distribution for p, using (i) the ignorance prior for p (π_A), and (ii) the elicited prior for p (π_B), and compare the two posteriors by completing the table below.

[6 marks]

Answer:

(i) $\pi_A(p|\underline{x}) \propto \pi_A(p) \times f(\underline{x}|p) = 1 \times p^4(1-p)^{23}$

$\Rightarrow p|\underline{x} \sim Beta(5,24)$.

(ii) $\pi_B(p|\underline{x}) \propto \pi_B(p) \times f(\underline{x}|p) = 360p^2(1-p)^7 \times p^4(1-p)^{23}$

$\Rightarrow p|\underline{x} \sim Beta(7,31)$ $\propto p^6(1-p)^{30}$.

	Posterior found using	
	Ignorance prior	Elicited prior
Posterior Mean	0.172	0·184
Posterior St. Dev.	0.069	0·062

[Total Q4: 25 marks]

5. (a) Complete the table below, using each of the following words exactly **once**:

Normal Uniform Exponential Pareto Beta Poisson

The parameter θ represents the mean of the Normal, the Pareto rate, the Poisson rate or the binomial success probability.

	Model: $X_i \sim$	**Prior for θ**	**Posterior for θ**
1	Poisson	Exponential	Gamma
2	Binomial	Uniform	Beta
3	Pareto	Gamma	Gamma
4	Normal	Normal	Normal

(b) In which of the situations above is the prior not *conjugate*?

Answer:

Could say situations 1 and 2, where the priors and posteriors are different.

However, could also say none – since both the exponential and uniform distributions are special cases of the gamma and beta (respectively).

[Total Q5: 7 marks]

6. A random sample of size n is to be taken from a $N(\theta, 3^2)$ distribution. The prior for θ is $N(b, 1/d)$.

If $n = 144$, show that no matter how large the prior variance, the standard deviation of the posterior distribution is less than 1/4.

Answer:

Recall (Example 2.6) that the posterior standard deviation is given by

$$\frac{1}{\sqrt{d+n\tau}} = \frac{1}{\sqrt{d+144\times\left(\frac{1}{9}\right)}}$$

$$= \frac{1}{\sqrt{d+16}}$$

The larger the prior variance $\left(\frac{1}{d}\right)$, the smaller the prior precision (d). The smallest this could be is $d=0$, giving

$$\frac{1}{\sqrt{0+16}} = \frac{1}{4}$$, as required!

[Total Q6: 4 marks]

THE END