

MAS3317

NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2013/14

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 $Mode(X) = (\alpha - 1)/(\alpha + \beta - 2)$.

Beta-Binomial distribution: If $X \sim BetaBin(n, \alpha, \beta)$ then it has density

$$f(x|n, \alpha, \beta) = {}^nC_x \frac{B(x + \alpha, \beta + n - x)}{B(\alpha, \beta)}, \quad x = 0, \dots, n, \quad \alpha > 0, \beta > 0.$$

Also, $E(X) = n\alpha/(\alpha + \beta)$, $Var(X) = n\alpha\beta(\alpha + \beta + n)(\alpha + \beta)^2(\alpha + \beta + 1)$.

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

Note: If $X \sim N(0, 1)$ then the following quantiles apply:

x	1.2816	1.6449	1.9600	2.3263	2.5758
$\Pr(X < x)$	0.9	0.95	0.975	0.99	0.995

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

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

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

SECTION A

- A1. A sports scientist measures the rate of oxygen consumption of some randomly chosen athletes immediately after exercise. She finds the following 95% frequentist confidence interval C_F for the population mean rate of oxygen consumption μ :

(1.997, 2.503) litres per minute.

In her report, she makes the following statement:

“My study revealed that the population mean rate of oxygen consumption for athletes lies in the range (1.997, 2.503) with probability 0.95”

- (a) Briefly explain why the sports scientist’s interpretation of C_F is incorrect.

Answer:

In a classical/frequentist analysis, we assume the parameter of interest is a fixed (but unknown) constant. Since it is not a random variable, we either ‘capture’ it or we don’t – giving a probability of 0, or 1 (respectively). Such confidence intervals have a *repeated sampling* interpretation: if we were to compute 100 such intervals, around 95 of these would capture the true population mean.

Bookwork – but I expect this could be overlooked by some.

- (b) After formulating an appropriate prior distribution for μ , a Bayesian statistician updates his beliefs about μ using the data in the sports scientist’s study, giving

$$\mu|\mathbf{x} \sim N(2.39, 0.01).$$

- (i) Find C_B , a 95% Bayesian confidence interval for μ .

Answer:

We have

$$2.39 \pm 1.96 \times \sqrt{0.01} \longrightarrow (2.194, 2.586)$$

Easy – but again, I expect this to be overlooked by some.

- (ii) Briefly explain how C_B differs to C_F in interpretation.

Answer:

Now we *do* have a probability interpretation, as we treat μ as a random variable which, after observing the data, follows the specified distribution.

Bookwork.

[Total A1: 9 marks]

A2. *Sapelo Island* is a barrier island off the southeast coast of the USA, and is on the migratory path of the rare black pelican. The random variable X represents the number of birds that are seen, per day, over the course of a week at an observation spot on the island.

(a) In the first week of April 2014, the following observations were made on X :

2 1 3 0 1 2 3

(i) Why might the Poisson distribution be a suitable model for X ?

Answer:

Poisson – often used to model count data.

(ii) Assuming that $X \sim Po(\theta)$, obtain the likelihood function $f(\mathbf{x}|\theta)$ assuming your observations are independent.

Answer:

Poisson – often used to model count data.

$$f(\mathbf{x}|\theta) = \frac{\theta^2 e^{-\theta}}{2!} \times \cdots \times \frac{\theta^3 e^{-\theta}}{3!} = \frac{\theta^{12} e^{-7\theta}}{144} \quad \theta > 0.$$

Very easy – all students should manage this without difficulty.

(b) After a consultation with an ornithologist (a zoologist with an expertise in birds), we are able to elicit a prior distribution for θ . Combining this prior distribution with the likelihood in part (a) results in the following posterior:

$$\theta|\mathbf{x} \sim Ga(18, 9).$$

Obtain the ornithologist's prior distribution for θ . How do the ornithologist's beliefs compare to our observations?

Answer:

We have

$$\begin{aligned} \text{Posterior} &\propto \text{Prior} \times \text{Likelihood} \\ \theta^{17} e^{-9\theta} &\propto \text{Prior} \times \theta^{12} e^{-7\theta}. \end{aligned}$$

Therefore, we must have $\text{Prior} \propto \theta^5 e^{-2\theta}$, that is

$$\theta \sim Ga(6, 2).$$

[Part of this page has been left blank for your solutions to the last question]

According to the ornithologist's prior beliefs, we would expect to see around 3 black pelicans per day (this is our prior mean).

We observed $\bar{x} = 1.71$, a bit lower than might be expected!

Should be very easy – although I expect some to be caught out because the question isn't the standard "Find the posterior" ...

- (c) Find α , the weight in the Bayes' linear rule for the posterior mean. Does this give more weight to the prior or the data?

Answer:

According to the BLR, we have

$$E[\theta|\mathbf{x}] = \alpha E[\theta] + (1 - \alpha)\bar{x} = \frac{18}{9} = 2.$$

Thus,

$$3\alpha + 1.71 - 1.71\alpha = 2 \longrightarrow \alpha \approx 0.225,$$

giving more weight to the data.

Should be easy, as they've seen this sort of thing before...

- (d) In a new analysis of the same data, a statistician attempts to use $\theta \sim N(3, 1.5)$. Briefly explain the similarities between this prior and the prior you obtained in part (b). What is the main problem with this new prior?

Answer:

Same prior mean and variance – however, θ is a rate and so must be strictly positive; this Normal prior can take negative values.

Will catch some out.

[Total A4: 14 marks]

A3. Suppose we have a random sample from an exponential distribution, that is,
 $X_i|\theta \sim \text{Exp}(\theta)$, $i = 1, \dots, n$.

- (a) Use the Factorisation Theorem to show that $T = \sum_{i=1}^n X_i$ is sufficient for θ , and give an intuitive interpretation of T .

Answer:

We have

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n \theta e^{-\theta x_i}, \quad \theta > 0 \\ &= 1 \times \theta^n \exp \left\{ -\theta \sum_{i=1}^n x_i \right\} \\ &= h(\mathbf{x}) \times g \left(\theta, \sum_{i=1}^n x_i \right), \end{aligned}$$

where $h(\mathbf{x}) = 1$ and $g(\theta, t) = \theta^n e^{-t\theta}$ and $t = \sum_{i=1}^n x_i$.

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

Interpretation: All the information in the data about θ is contained in the sum of the observations.

Should be easy for most – seen in class.

- (b) Suppose we determine that a unit exponential distribution adequately represents our prior beliefs about the parameter θ . It can be shown that, after observing data $\mathbf{x}$, the posterior distribution for θ is

$$\theta|\mathbf{x} \sim \text{Ga} \left(n + 1, \sum_{i=1}^n x_i + 1 \right).$$

Use the fact that $T|\theta \sim \text{Ga}(n, \theta)$ to show that the same posterior is obtained knowing just the sufficient statistic T .

Answer:

We need to use

$$\pi(\theta|t) \propto \pi(\theta) \times f_T(t|\theta).$$

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

Now

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}.$$

Therefore

$$\pi(\theta|t) \propto e^{-\theta} \times \theta^n e^{-\theta t} = \theta^n e^{-\theta(t+1)}.$$

This is proportional to the PDF of a $Ga(n+1, t+1)$, which – on noting that $T = \sum_{i=1}^n X_i$ – is exactly the same as that for $\theta|\mathbf{x}$.

Not difficult, but some will struggle with this – and it is unseen.

[Total A3: 11 marks]

- A4. (a) In the space below, briefly explain the terms *substantial prior knowledge* and *vague prior knowledge*. You may use diagrams to aid your descriptions.

SUBSTANTIAL PRIOR KNOWLEDGE

Answer:

We have *substantial prior information* for θ when the prior distribution *dominates* the posterior distribution, that is $\pi(\theta|\mathbf{x}) \sim \pi(\theta)$. Observing the data does not alter our beliefs about θ very much.

VAGUE PRIOR KNOWLEDGE

Answer:

If we have *vague prior information*, most of the information about θ will be passed through to the posterior distribution via the data, and so we have $\pi(\theta|\mathbf{x}) \sim f(\mathbf{x}|\theta)$.

Should be OK – bookwork.

- (b) A *Jeffreys prior* is often used when we are *prior ignorant*. Give one advantage, and one disadvantage, of using such a prior.

Answer:

- Local uniformity property satisfied – does not change much in the region over which the likelihood is significant
- Can be cumbersome to use in high dimensions

Bookwork.

[Total A4: 6 marks]

SECTION B

B5. You work as a statistician for the Canadian spreadbetting company *ZenTrader*. You are interested in θ , the probability that the *Montreal Canadiens* ice hockey team win their next match against the *Boston Bruins*.

- (a) A discussion with a fan of the *Montreal Canadiens* results in the screenshot overleaf, taken from the *MATCH Uncertainty Elicitation Tool*. Briefly explain the benefits of the elicitation procedure used over the *bisection* (*quartile*) method, and write down the prior distribution for θ as suggested by *MATCH* (round your parameters to the nearest integer).

Answer:

Trial roulette method:

- Expert can see the distribution “take shape” dynamically
- No need for the expert to think about quartiles/other summaries, as in the bisection method

We have $\theta \sim Be(3, 13)$.

Should be easy for most – bookwork.

- (b) The *Montreal Canadiens* have won two of their last six matches against the *Boston Bruins*. Given this information, find the posterior distribution for θ using the elicited prior from part (a).

Answer:

Using the elicited prior, we have

$$\pi(\theta|x=2) \propto \theta^2(1-\theta)^{12} \times \theta^2(1-\theta)^4 = \theta^4(1-\theta)^{16},$$

and so we have $\theta|x=2 \sim Be(5, 17)$.



- (c) In the absence of any expert prior information, a *Haldane prior* for θ , with density

$$\pi(\theta) \propto \frac{1}{\theta(1-\theta)},$$

can be used.

- (i) Find the posterior distribution for θ using the Haldane prior.

Answer:

This prior results in:

$$\pi(\theta|x=2) \propto \theta^{-1}(1-\theta)^{-1} \times \theta^2(1-\theta)^4 = \theta^{2-1}(1-\theta)^{4-1},$$

i.e. $\theta|x=2 \sim Be(2, 4)$.

- (ii) Why might a Haldane prior be attractive in the absence of substantial prior information?

Answer:

The resulting posterior has the same mean as the data (1/3). Thus, using such a prior allows the data to “speak”.

- (d) Compare and contrast the posterior distributions for θ obtained in parts (b) and (c). You should use numerical summaries, as well as sketches of the posteriors using the graph paper overleaf, in your answer.

[Hint: the posterior resulting from the elicited prior has been drawn for you.]

Answer:

We can summarise these two posteriors numerically:

	Posterior using	
	Haldane prior	Elicited prior
Mean	0.333	0.227
Mode	0.25	0.2
SD	0.178	0.095

At the posterior mode, we have

$$\pi(\theta = 0.25|x = 2) = \frac{0.25 \times 0.75^3}{B(2, 4)} = 2.1094$$

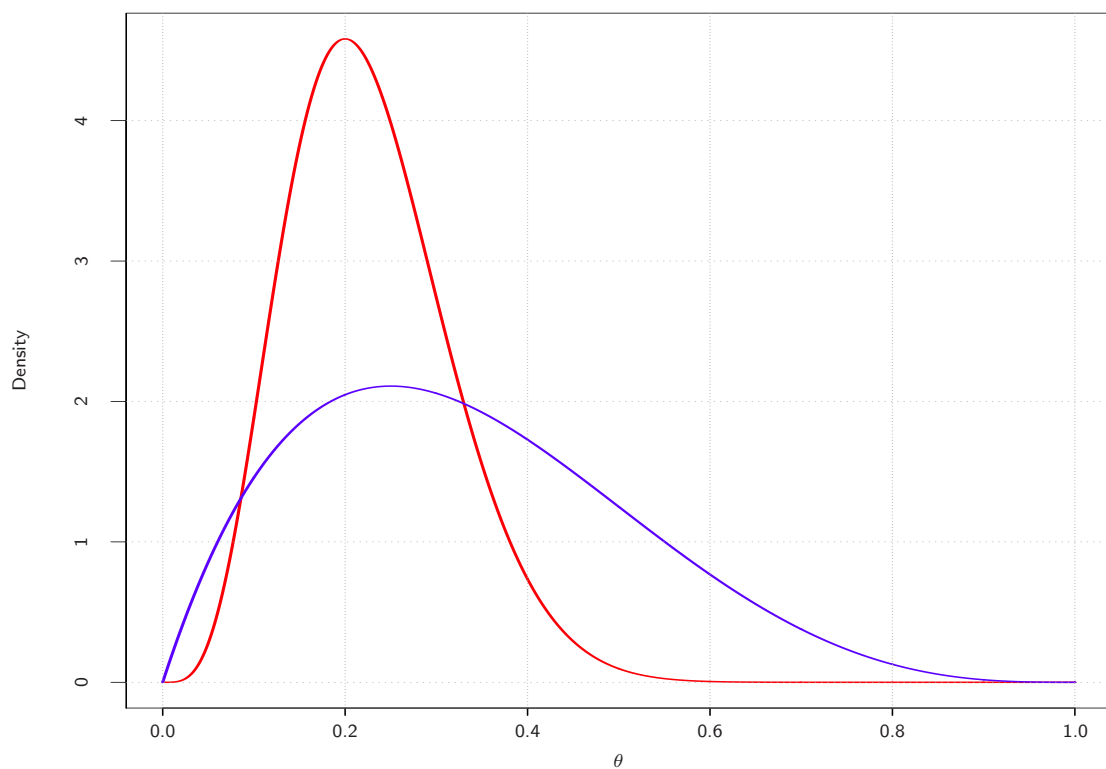
when using the Haldane prior, and

$$\pi(\theta = 0.2|x = 2) = \frac{0.2^4 0.8^{16}}{B(5, 17)} = 4.582$$

when using the elicited prior.

Comments

- The posterior resulting from the ignorance prior peaks at the likelihood mode – that is, $2/6 = 1/3$
- Using the elicited prior gives something a bit more “informative” - that is, the prior beliefs of the fan have been taken into account to give a higher posterior mean/mode
- The posterior resulting from the elicited prior is much less variable – showing that we are more certain about our beliefs about θ after having observed the data
- Either way, it still seems unlikely that the *Canadiens* will beat the *Bruins*!



This question isn't difficult, but we haven't focussed on plotting/sketching the distributions by hand and so I imagine some might struggle. Some will also struggle with the comments. Students will not need to give as much detail to get full marks: simple summaries of average and spread, a reasonable plot and some sensible comments will suffice.

(d) Look at the following R code:

```
> g=function(a,b){
+   f=function(x){
+     l=x[1]
+     u=x[2]
+     (pbeta(u,a,b)-pbeta(l,a,b)-0.95)^2
+     +0.0001*(dbeta(u,a,b)-dbeta(l,a,b))^2
+   }
+   initial.l=qbeta(0.025,a,b)
+   initial.u=qbeta(0.975,a,b)
+   A=optim(c(initial.l,initial.u),f)
+   return(A$par)

> g(2,4)
[1] 0.02602514 0.67014432
> g(5,17)
[1] 0.06921317 0.39948449
```

Very briefly, explain what the function `g` does, and interpret the output for `g(2,4)` and `g(5,17)`. What does this tell us about the likely outcome of the next hockey match between the *Canadiens* and the *Bruins*?

Answer:

The function `g` finds a 95% Bayesian confidence interval (in fact, a highest density interval) for a probability θ when the posterior is a beta distribution with parameters `a` and `b`. It does this using the `optim` numerical solver, attempting to minimise

$$f(l, u) = \{f(l) - f(u) - 0.95\}^2 + k\{f(u) - f(l)\}^2,$$

where l and u are the lower and upper bounds of the confidence interval respectively. For the posterior using the Haldane prior, we have (0.026, 0.670); for the posterior which has used the elicited prior, we have (0.069, 0.399). We are much more certain that the *Canadiens* will lose their next match having incorporated the fan's beliefs!

Not difficult, but completely unseen.

[Total B5: 20 marks]

B6. 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 $\bar{x}$. It can be shown that the likelihood function for μ is

$$f(\mathbf{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2} [s^2 + (\bar{x} - \mu)^2]\right\},$$

where

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- (a) Suppose our prior beliefs about μ can be summarised by a Normal $N(b, 1/d)$ distribution. Show that the posterior distribution is $\mu|\mathbf{x} \sim N(B, 1/D)$, where

$$B = \frac{db + n\tau\bar{x}}{D} \quad \text{and} \quad D = d + n\tau.$$

Answer:

Applying Bayes Theorem, the posterior density function is

$$\begin{aligned} \pi(\mu|\mathbf{x}) &\propto \pi(\mu)f(\mathbf{x}|\mu) \\ &\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \times \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2} [s^2 + (\bar{x} - \mu)^2]\right\} \\ &= k_1 \exp\left\{-\frac{d}{2}(\mu - b)^2 - \frac{n\tau}{2} [s^2 + (\bar{x} - \mu)^2]\right\} \\ &= k_1 \exp\left\{-\frac{1}{2} [d(\mu - b)^2 + n\tau(\bar{x} - \mu)^2]\right\} \end{aligned}$$

where k_1 is a constant that does not depend on μ .

We can also simplify the exponent:

$$\begin{aligned} d(\mu - b)^2 + n\tau(\bar{x} - \mu)^2 &= (d + n\tau)\mu^2 - 2(db + n\tau\bar{x})\mu + db^2 + n\tau\bar{x}^2 \\ &= (d + n\tau) \left\{ \mu - \left(\frac{db + n\tau\bar{x}}{d + n\tau} \right) \right\}^2 + c \end{aligned}$$

where c does not depend on μ .

Let

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} \quad \text{and} \quad D = d + n\tau.$$

Then

$$\begin{aligned} \pi(\mu|\mathbf{x}) &= k_1 \exp\left\{-\frac{D}{2}(\mu - B)^2 - \frac{c}{2}\right\} \\ &= k \exp\left\{-\frac{D}{2}(\mu - B)^2\right\}, \end{aligned}$$

where k is a constant that does not depend on μ . Therefore, the posterior distribution takes the form $k \exp\{-D(\mu - B)^2/2\}$, $-\infty < \mu < \infty$ and so must be a normal distribution: we have $\mu|\mathbf{x} \sim N(B, 1/D)$.

Bookwork.

[Part of this page has been left blank for your solution to the last question]

- (b) A sports scientist measures the rate of oxygen consumption of 10 randomly chosen athletes immediately after exercise. The sample mean is $\bar{x} = 2.25$ litres per minute. She assumes a Normal distribution for these measurements with a standard deviation of $1/\sqrt{6}$ litres per minute.

A prior study suggests that $\mu \sim N(2.6, 0.025)$.

- (i) Using this information, confirm the posterior distribution used in question **A1(b)**, that is

$$\mu|\mathbf{x} \sim N(2.39, 0.01).$$

Answer:

We have $\tau = 6$, $n = 10$ and $\bar{x} = 2.25$. We are also told that $b = 2.6$ and $d = 1/0.025$, giving

$$D = d + n\tau = \frac{1}{0.025} + 10 \times 6 = 100,$$

and

$$B = \frac{\frac{2.6}{0.025} + 10 \times 6 \times 2.25}{100} = 2.39,$$

giving $\mu|\mathbf{x} \sim N(2.39, 1/100)$, as required.

- (ii) Find the posterior probability that the mean rate of oxygen consumption is greater than 2.518 litres per minute.

We want

$$P(\mu|\mathbf{x} > 2.518) = P\left(Z > \frac{2.518 - 2.39}{\sqrt{0.01}}\right) = 1 - P(Z < 1.28).$$

From tables at the front of the exam paper, this gives

$$1 - 0.9 = 0.1.$$

- (c) Obtain the posterior distribution for μ under the assumption of *vague prior knowledge*, and explain why the resulting 95% Bayesian confidence interval would be the same (numerically) as the sports scientist's frequentist confidence interval in question A1.

Answer:

If we have vague prior knowledge, then we want the prior variance $\rightarrow \infty$. Thus, $d \rightarrow 0$, and so we have

$$B = \bar{x} \quad \text{and} \quad D = n\tau,$$

giving $\mu|\mathbf{x} \sim N(\bar{x}, 1/n\tau)$. In our example, this gives

$$\mu|\mathbf{x} \sim N(2.25, 1/60),$$

This gives

$$2.25 \pm \frac{1.96}{\sqrt{60}} \rightarrow (1.997, 2.503).$$

The central limit theorem would give (numerically) the same confidence interval in a frequentist analysis.

This will cause some problems, I would imagine.

(d) Given that

$$\frac{\partial}{\partial \mu} \log f(\mathbf{x}|\mu) = n\tau(\bar{x} - \mu),$$

show that, no matter what prior distribution is used for μ , the posterior is the same as that you obtained in part (c) *when n is large*.

Answer:

This requires the result for the asymptotic posterior distribution: for large n ,

$$\theta|\mathbf{x} \sim N\left(\hat{\theta}, J(\hat{\theta})^{-1}\right),$$

where $\hat{\mu}$ is the maximum likelihood estimator for μ , and

$$J(\theta) = -\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta).$$

We have

$$J(\mu) = -\frac{\partial^2}{\partial \mu^2} \log f(\mathbf{x}|\mu) = n\tau.$$

Also, maximising the log-likelihood, gives $\hat{\mu} = \bar{x}$, and so

$$J(\hat{\mu})^{-1} = 1/n\tau.$$

Thus,

$$\mu|\mathbf{x} \sim N(\hat{x}, 1/n\tau) \longrightarrow N(2.25, 1/60);$$

exactly as before.

This will cause problems for many.

[Total B6: 20 marks]

- B7.** Suppose that X is the number of tropical storms occurring over the Atlantic Ocean in July 2014, and let θ be the daily rate of occurrence of such storms. A meteorologist assumes that $X|\theta \sim Po(31\theta)$, with the following prior distribution for θ :

Level of activity	θ	$\pi(\theta)$
Busy	0.25	0.2
Normal	0.1	0.8

- (a) Suppose that six tropical storms are observed over the Atlantic Ocean in July 2014. Determine the posterior distribution for θ .

Answer:

The likelihood term is

$$f(X = 6|\theta) = \frac{(31\theta)^6 e^{-31\theta}}{6!} = \begin{cases} 0.12963 & \text{if } \theta = 0.25 \\ 0.05553 & \text{if } \theta = 0.1 \end{cases}$$

Then

$$\begin{aligned} \pi(\theta = 0.25|X = 6) &= \frac{\Pr(X = 6|\theta = 0.25)\Pr(\theta = 0.25)}{\Pr(X = 6|\theta = 0.25)\Pr(\theta = 0.25) + \Pr(X = 6|\theta = 0.1)\Pr(\theta = 0.1)} \\ &= \frac{0.12963 \times 0.2}{0.12963 \times 0.2 + 0.05553 \times 0.8} \\ &= 0.3685288 \end{aligned}$$

Thus,

$$\Pr(\theta = 0.1|X = 6) = 1 - 0.3685288 = 0.6314712.$$

- (b) Suppose now that we want to predict the number of tropical storms Y over the Atlantic Ocean during August 2014. Determine the predictive probability function for Y , $f(y|x=6)$, clearly stating any assumptions that you make.

Answer:

Assuming tropical storms occur at the same rate in August as they do in July, we have

$$\begin{aligned} f(y|x=6) &= \sum_{\theta=0.25,0.1} f(y|\theta)\pi(\theta|x=6) \\ &= \left(\frac{7.75^y e^{-7.75}}{y!} \times 0.3685288 \right) + \left(\frac{3.1^y e^{-3.1}}{y!} \times 0.6314712 \right) \end{aligned}$$

- (c) Complete the table below showing the posterior predictive distribution for Y , and use this table to find an approximate 90% prediction interval for Y .

Answer:

y	0	1	2	3	4	5	6
$f(y x=6)$	0.0286	0.0894	0.1415	0.1536	0.1333	0.1049	0.0828
y	7	8	9	10	11	12	13
$f(y x=6)$	0.0684	0.0573	0.0462	0.0348	0.0243	0.0156	0.0093

Using the values giving rise to highest density, we see that (1,10) gives close to a 90% coverage.

- (d) Let ψ be the probability that a tropical storm develops into a hurricane. After a discussion with the meteorologist, we are able to elicit the following prior distribution for ψ :

$$\psi \sim Be(2, 23), \quad 0 < \psi < 1.$$

One of the last five tropical storms over the Atlantic Ocean developed into a hurricane. For the next five tropical storms observed in the future, show that the predictive probability distribution for the number that will develop into a hurricane is Beta-Binomial $BetaBin(5, 3, 27)$.

Answer:

The posterior is

$$\pi(\psi|x=1) \propto \psi(1-\psi)^{22}\psi(1-\psi)^4 = \psi^2(1-\psi)^{26},$$

and so $\psi|x=1 \sim Be(3, 27)$.

The predictive probability function for Y is, for $Y = 0, 1, \dots$,

$$\begin{aligned} f(y|x=1) &= \int_{\psi} f(y|\psi)\pi(\psi|x=1)d\psi \\ &= \int_0^1 {}^5C_y \psi^y (1-\psi)^{5-y} \frac{\psi^2(1-\psi)^{26}}{B(3, 27)} \\ &= \frac{{}^5C_y}{B(3, 27)} \int_0^1 \psi^{y+2} (1-\psi)^{31-y} \\ &= {}^5C_y \frac{B(y+3, 32-y)}{B(3, 27)}, \end{aligned}$$

that is, $Y|x=1 \sim BetaBin(5, 3, 27)$, as required.

- (e) Use Bayes' Theorem to show that the predictive probability distribution for future data y , given past data $\mathbf{x}$, can be written as

$$f(y|\mathbf{x}) = \frac{f(y|\psi)\pi(\psi|\mathbf{x})}{\pi(\psi|\mathbf{x}, y)},$$

assuming $\mathbf{x}$ and y are independent.

Using Bayes' Theorem, the posterior density for θ given $\mathbf{x}$ and y is

$$\begin{aligned}\pi(\theta|\mathbf{x}, y) &= \frac{\pi(\theta)f(\mathbf{x}, y|\theta)}{f(\mathbf{x}, y)} \\ &= \frac{\pi(\theta)f(\mathbf{x}|\theta)f(y|\theta)}{f(\mathbf{x})f(y|\mathbf{x})} && \text{since } \mathbf{X} \text{ and } Y \text{ are independent given } \theta \\ &= \frac{\pi(\theta|\mathbf{x})f(y|\theta)}{f(y|\mathbf{x})}.\end{aligned}$$

Rearranging, we obtain

$$f(y|\mathbf{x}) = \frac{f(y|\theta)\pi(\theta|\mathbf{x})}{\pi(\theta|\mathbf{x}, y)}.$$

Many students will find *all* of this question a challenge

[Total B7: 20 marks]

THE END