

MAS3317

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

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

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?

[5 marks]

Answer:

Using Bayes' Theorem,

$$\begin{aligned}\Pr(XT|TR) &= \frac{\Pr(TR|XT)\Pr(XT)}{\Pr(TR|XT)\Pr(XT) + \Pr(TR|XT^c)\Pr(XT^c)} \\ &= \frac{0.45 \times \frac{n}{90}}{0.45 \times \frac{n}{90} + 0.1 \times \frac{90-n}{90}} \\ &= \frac{0.45n}{0.45n + 9 - 0.1n} \\ &= \frac{0.45n}{9 + 0.35n}\end{aligned}$$

Very straightforward application of Bayes' Theorem – I imagine most students will manage this question. However, the fact that we have n rather than a number might throw some.

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

[4 marks]

Answer:

We require

$$\begin{aligned}\frac{0.45n}{0.35n + 9} &< 0.5 \\ 0.275n &< 4.5 \\ n &< 16.36,\end{aligned}$$

i.e. at most 16 pigs. Easy, but completely unseen.

[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

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

[5 marks]

Answer:

Let $\lambda \sim Ga(a, b)$. Then

$$\frac{a}{b} = 1.5 \rightarrow a = 1.5b.$$

Substitution into the variance gives

$$\begin{aligned} \frac{1.5b}{b^2} &= \frac{6}{16}, & \text{i.e.} \\ 6b &= 24 \\ b &= 4. \end{aligned}$$

Therefore, $a = 1.5 \times 4 = 6$, and so $\lambda \sim Ga(6, 4)$, giving

$$\pi(\lambda) = \frac{4^6 \lambda^5 e^{-4\lambda}}{\Gamma(6)} = \frac{4096}{120} \lambda^5 e^{-4\lambda} = \frac{512}{15} \lambda^5 e^{-4\lambda}, \quad \lambda > 0,$$

as required.

Though this question is routine, the fact that the question does not say “(i) Find a and b ; (ii) use these values to construct the prior distribution for λ ” will throw some students; I imagine some will not remember that $\Gamma(A) = (A-1)!$, and that some will use $Var(\lambda) = \sqrt{6}/4$.

- (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

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

[6 marks]

Answer:

We have

$$f(\mathbf{x} | \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \times \dots \times \frac{e^{-\lambda} \lambda^{x_{60}}}{x_{60}!}$$

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

$$= \frac{e^{-60\lambda} \lambda^{60\bar{x}}}{\prod_{i=1}^{60} x_i!} \propto k e^{-60\lambda} \lambda^{132},$$

where k is a constant that does not depend on λ . Thus

$$\pi(\lambda|\mathbf{x}) \propto \pi(\lambda) \times f(\mathbf{x}|\lambda) = \lambda^5 e^{-4\lambda} \times \lambda^{132} e^{-60\lambda} = \lambda^{138-1} e^{-64\lambda},$$

and so $\lambda|\mathbf{x} \sim Ga(138, 64)$.

Again – routine, but some might struggle with the fact that they are not explicitly told to find the likelihood, and then use this, and the prior, to get the posterior.

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

[4 marks]

$$E[\lambda|\mathbf{x}] = \frac{138}{64} = \alpha E[\lambda] + (1 - \alpha)\bar{x},$$

where $E[\lambda] = 1.5$ and $\bar{x} = 2.2$. Thus

$$0.7\alpha = 0.04375 \Rightarrow \alpha = 0.0625$$

I expect quite a few students to have forgotten/neglected the Bayes linear rule. It was mentioned in class, with an example, but there have been few examples given in problems sheets.

[Total A2: 15 marks]

A3. Suppose we have data $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ available and we want to make inferences about the parameter ζ in the statistical model $f(\mathbf{x}|\zeta)$. 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 ”.
[2 marks]

Answer:

All the information in the data about ζ is contained in the sufficient statistic T .

This should be OK – bookwork.

- (b) State the Factorisation Theorem.

[3 marks]

Answer:

Under certain regularity conditions,

$$T(\mathbf{X}) \text{ is sufficient for } \zeta \iff f(\mathbf{x}|\zeta) = h(\mathbf{x}) g(t(\mathbf{x}), \zeta)$$

for some functions h and g .

Bookwork.

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

[5 marks]

Answer:

$$\begin{aligned} \pi(\zeta|\mathbf{x}) &\propto \pi(\zeta)f(\mathbf{x}|\zeta) \\ &\propto \pi(\zeta)h(\mathbf{x})g(t, \zeta) \\ &\propto \pi(\zeta)g(t, \zeta), \end{aligned}$$

due to the Factorisation Theorem. Now it can be shown that $g(t, \zeta) \propto f_T(t|\zeta)$, giving

$$\pi(\zeta|\mathbf{x}) \propto \pi(\zeta)f_T(t|\zeta).$$

Applying Bayes' Theorem to the data t gives

$$\pi(\zeta|t) \propto \pi(\zeta)f_T(t|\zeta)$$

and so, since $\pi(\zeta|\mathbf{x}) \propto \pi(\zeta|t)$, and both are probability density functions, we have

$$\pi(\zeta|\mathbf{x}) = \pi(\zeta|t).$$

Bookwork.

[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 .

[3 marks]

Answer:

C_F : $\Pr(\theta \in C_F) = 0$ or 1 , whereas C_B : $\Pr(\theta \in C_B) = 0.95$. In a frequentist confidence interval, the level of ‘confidence’ refers to repeated sampling; that is, about 95% of confidence intervals taken from samples of the same population will contain θ .

Bookwork.

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

[3 marks]

Answer:

Suppose A is the event of interest. Pay $\mathcal{L}p$ for the bet

win $\mathcal{L}1$ if A occurs and $\mathcal{L}0$ if A^c occurs;

randomise between whether the person ‘hosts’ or ‘plays’ the bet.

[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 .

[2 marks]

Answer:

A Uniform $U(0, 1)$ distribution has a density of 1; thus, if $\theta \sim U(0, 1)$ then we are assuming all values of θ , $\theta \in (0, 1)$, are equally likely – hence prior ignorance.

I expect that some students will not recognise this as a $U(0, 1)$ prior.

- (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 $\pi_B(\theta)$.

[2 marks]

Answer:

$$\begin{aligned}\theta &\sim \text{Beta}(10, 2) \\ \pi_B(\theta) &= \frac{\theta^9(1-\theta)}{B(10, 2)}, \quad 0 < \theta < 1.\end{aligned}$$

Should be easy marks for most students.

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

[2 marks]

Answer:

We see that $\Pr(\theta < 0.537) = 0.01$. Thus, we might ask:

“From our fitted model, it is very unlikely (less than a 1/100 chance) that fewer than half of the gravestones will show an annual rate of degradation of more than 0.12mm. Does this seem reasonable?”

Some students will find this difficult – they have struggled with such interpretations in assignments.



- (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 $\pi_A(\theta)$;

[3 marks]

Answer:

$$\pi(\theta|x = 12) = 1 \times {}^{20}C_{12}\theta^{12}(1 - \theta)^8 \propto \theta^{13-1}(1 - \theta)^{9-1},$$

i.e. $\theta|x = 12 \sim Be(13, 9)$.

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

[3 marks]

Answer:

$$\pi(\theta|x = 12) \propto \theta^9(1 - \theta) \times \theta^{12}(1 - \theta)^8 = \theta^{22-1}(1 - \theta)^{10-1},$$

i.e. $\theta|x = 12 \sim Be(22, 10)$.

Both (i) and (ii) are very straightforward, but I know some students will struggle to recognise the binomial likelihood.

- (iii) Is the ‘ignorance’ prior *conjugate*? Explain.

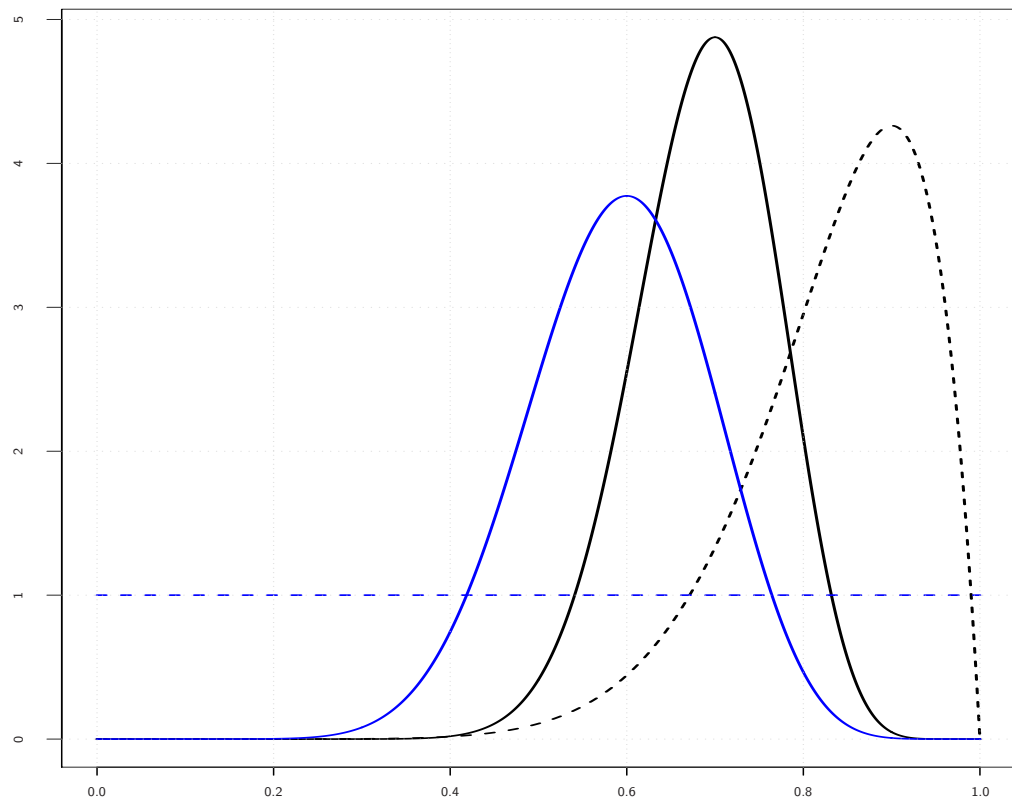
[2 marks]

Answer:

“Yes”: If $\theta \sim U(0, 1)$ then $\theta \sim Be(1, 1)$; the resulting posterior is also in the Beta family, meaning we have a conjugate prior. Will also accept “No”: The prior is Uniform but the posterior is not. Full marks for the *explanation*, rather than “Yes/No”.

- (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.

[4 marks]



Using the ignorance prior, we have

$$\theta|x = 12 \sim Be(13, 9),$$

which has mode at $12/20 = 0.6$. Also,

$$Var(\theta|x = 12) = \frac{13 \times 9}{22^2 \times 23} = 0.011,$$

which is almost twice that of the posterior obtained from the elicited prior, giving a more dispersed distribution.

Unseen – easy, but they’ve not done this before (other than in R!) and so this will throw lots of students.

[Total B5: 18 marks]

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

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

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 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ at this location (and we assume these observations are independent).

[2 marks]

Answer:

$$f(\mathbf{x}|\theta) = \theta^n 30^{n\theta} \prod_{i=1}^n x_i^{-(\theta+1)}.$$

Unseen distribution, but should be easy.

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

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

[5 marks]

Answer:

We have

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \pi(\theta) \times f(\mathbf{x}|\theta) \\ &= ge^{-g\theta} \times \theta^n 30^{n\theta} \prod_{i=1}^n x_i^{-(\theta+1)} \\ &\propto \theta^n e^{-g\theta} \times 30^{n\theta} \prod_{i=1}^n x_i^{-(\theta+1)}. \end{aligned}$$

Now take $30^{n\theta} \prod_{i=1}^n x_i^{-(\theta+1)}$. Taking logs gives

$$n\theta \log 30 - \theta \sum_{i=1}^n \log x_i - \sum_{i=1}^n \log x_i.$$

Exponentiating, we get

$$\begin{aligned} 30^{n\theta} \prod_{i=1}^n x_i^{-(\theta+1)} &= \exp\left\{n\theta \log 30 - \theta \sum_{i=1}^n \log x_i - \sum_{i=1}^n \log x_i\right\} \\ &= \exp\left\{n\theta \log 30 - \theta \sum_{i=1}^n \log x_i\right\} \exp\left\{-\sum_{i=1}^n \log x_i\right\}. \end{aligned}$$

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

Substitution into the expression for $\pi(\theta|\mathbf{x})$ gives

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \theta^n \exp\{-g\theta\} \times \exp\left\{n\theta\log 30 - \theta \sum_{i=1}^n \log x_i\right\} \\ &\propto \theta^{(n+1)-1} \exp\left\{-\left(g - n\log 30 + \sum_{i=1}^n \log x_i\right)\theta\right\}.\end{aligned}$$

Thus,

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

as required.

Will require confidence, as unseen distribution! But should be OK.

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

[3 marks]

Answer:

Vague prior knowledge – when we have very little prior information regarding the parameter of interest θ , and so $\pi(\theta|\mathbf{x}) \sim f(\mathbf{x}|\theta)$. In order to operate Bayes' Theorem, we still need a prior; for convenience, we choose a conjugate prior, with variance as large as possible. We have $\theta \sim \text{Exp}(g)$, and so

$$\text{Var}(\theta) = \frac{1}{g^2}.$$

Thus, as $g \rightarrow 0$, $\text{Var}(\theta) \rightarrow \infty$.

Description: Bookwork; application to $\text{Exp}(g)$ unseen, but easy.

- (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:

Assuming vague prior knowledge, $g = 0$. Thus

$$\begin{aligned}\theta|\mathbf{x} &\sim Ga(4, -3\log 30 + [\log 61 + \log 55 + \log 60]), \quad \text{i.e.} \\ &\sim Ga(4, 2).\end{aligned}$$

Substitution: should be easy.

(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:

We want

$$\int_a^b \frac{2^4 \theta^3 e^{-2\theta}}{\Gamma(4)} d\theta = \frac{95}{100}$$

$$\frac{16}{6} \int_a^b \theta^3 e^{-2\theta} d\theta = \frac{95}{100}$$

giving

$$\int_a^b \theta^3 e^{-2\theta} d\theta = \frac{95}{100} \times \frac{3}{8} = \frac{57}{160},$$

as required.

Completely unseen: will cause trouble.

(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:

We also require that $\theta|\mathbf{x}$ has the same probability density at a and b , giving:

$$a^3 e^{-2a} = b^3 e^{-2b}.$$

Many will find this a challenge.

- (iv) use the following output from R to write down the 95% highest density interval for $\theta|\mathbf{x}$, 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
```

The 95% HDI for $\theta|\mathbf{x}$ is (0.0105, 1.1913). The `optim` function has been used to minimise

$$g(a, b) = \{F(b) - F(a) - 0.95\}^2 + k \{f(b) - f(a)\}^2,$$

which corresponds to the conditions from parts (ii) and (iii). Here, $k > 0$ is a tuning parameter that tries to ensure that both terms are zeroed.

Students will find this hard. We have done an example similar in class, but they struggled with it.

[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 $\mathbf{x}$.
[2 marks]

Answer:

$$\begin{aligned} f(\mathbf{x}|\mu) &= \prod_{i=1}^n \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau}{2}(x_i - \mu)^2\right\} \\ &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2\right\}, \end{aligned}$$

so

$$\log f(\mathbf{x}|\mu) = \frac{n}{2} \log(\tau) - \frac{n}{2} \log(2\pi) - \frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2.$$

Easy for most/all.

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

[6 marks]

Answer:

Now

$$\begin{aligned} \Rightarrow \quad \frac{\partial}{\partial \mu} \log f(\mathbf{x}|\mu) &= -\frac{\tau}{2} \times \sum_{i=1}^n -2(x_i - \mu) \\ &= \tau \sum_{i=1}^n (x_i - \mu) \\ &= \tau(n\bar{x} - n\mu) \\ &= n\tau(\bar{x} - \mu) \end{aligned}$$

Also

$$\begin{aligned} \Rightarrow \quad \frac{\partial^2}{\partial \mu^2} \log f(\mathbf{x}|\mu) &= -n\tau \\ \Rightarrow \quad I(\mu) &= E_{\mathbf{X}|\mu} \left[-\frac{\partial^2}{\partial \mu^2} \log f(\mathbf{X}|\mu) \right] = n\tau. \end{aligned}$$

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

Hence, the Jeffreys prior for this model is

$$\begin{aligned}\pi(\mu) &\propto I(\mu)^{1/2} \\ &\propto \sqrt{n\tau}, & -\infty < \mu < \infty \\ &= \text{constant}, & -\infty < \mu < \infty.\end{aligned}$$

Bookwork.

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

[2 marks]

Answer:

This distribution is improper since $\int_{-\infty}^{\infty} d\mu$ is a divergent integral, and so we cannot find a constant which ensures that the density function integrates to one.

Will cause trouble for some.

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

[2 marks]

Answer:

If we are ignorant about μ , then we are ignorant about any function of μ ; assigning an ignorance prior to μ means we do not have an ignorance prior for, e.g., e^μ .

Will cause trouble for some.

- (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?

[3 marks]

Answer:

Suppose we now have vague prior knowledge; we would then use a conjugate prior for μ with large variance, that is

$$\pi(\mu) = \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \propto \exp\left\{-\frac{d}{2}(\mu - b)^2\right\}$$

with prior mean and precision b and d . As $Var(\mu) \rightarrow \infty$, $d \rightarrow 0$ and the above expression tends to 1, which would give the same posterior for μ as if we had used the Jeffreys prior (constant).

Will cause trouble for most.

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

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

[3 marks]

Answer:

The asymptotic posterior distribution is Normal, with variance $J^{-1}(\hat{\mu})$, where

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

From part (b), we thus have

$$Var(\mu|\mathbf{x}) \approx (n\tau)^{-1}, \quad \text{for large } n.$$

Bookwork – but will catch some out because it does not say “asymptotic posterior”.

[Total B7: 18 marks]

THE END