

MAS3902: Bayesian Inference

Dr. Lee Fawcett

School of Mathematics, Statistics & Physics

Semester 2, 2022/23

Contact details

My name: Lee Fawcett

Office: Room 2.07 Herschel Building

Phone: 0191 208 7228

Email: `lee.fawcett@ncl.ac.uk`

www: `www.mas.ncl.ac.uk/~nlf8`

Timetable and Administrative arrangements

- **Classes** take place on Tuesdays (3pm, Herschel LT2) and Thursdays (9am, online – fortnightly only; 12pm, Herschel LT1)
- Two of these classes will be **lectures**, and the other session will be a **problems class/drop-in**
 - Lectures will *ordinarily* take place in the in-person slots in the lecture theatres
 - Problems classes/drop-ins will *ordinarily* take place in the online slot
- Due to outreach/recruitment activities, I will not be at work on Thursday 2nd February; Thursday 9th February and Thursday 11th May. Fear not! **Replacement lectures** have been booked:
 - Friday 3rd February (11am, Herschel LT2)
 - Friday 17th February (11am, Herschel LT2)
 - Friday 3rd March (11am, Herschel LT2)
- **Office hours** will be scheduled soon, but just come along and give me a knock!

Timetable and Administrative arrangements

- **Classes** take place on Tuesdays (3pm, Herschel LT2) and Thursdays (9am, online – fortnightly only; 12pm, Herschel LT1)
- Two of these classes will be **lectures**, and the other session will be a **problems class/drop-in**
 - Lectures will *ordinarily* take place in the in-person slots in the lecture theatres
 - Problems classes/drop-ins will *ordinarily* take place in the online slot
- Due to outreach/recruitment activities, I will not be at work on Thursday 2nd February; Thursday 9th February and Thursday 11th May. Fear not! **Replacement lectures** have been booked:
 - Friday 3rd February (11am, Herschel LT2)
 - Friday 17th February (11am, Herschel LT2)
 - Friday 3rd March (11am, Herschel LT2)
- **Office hours** will be scheduled soon, but just come along and give me a knock!

Timetable and Administrative arrangements

- **Classes** take place on Tuesdays (3pm, Herschel LT2) and Thursdays (9am, online – fortnightly only; 12pm, Herschel LT1)
- Two of these classes will be **lectures**, and the other session will be a **problems class/drop-in**
 - Lectures will *ordinarily* take place in the in-person slots in the lecture theatres
 - Problems classes/drop-ins will *ordinarily* take place in the online slot
- Due to outreach/recruitment activities, I will not be at work on Thursday 2nd February; Thursday 9th February and Thursday 11th May. Fear not! **Replacement lectures** have been booked:
 - Friday 3rd February (11am, Herschel LT2)
 - Friday 17th February (11am, Herschel LT2)
 - Friday 3rd March (11am, Herschel LT2)
- **Office hours** will be scheduled soon, but just come along and give me a knock!

Timetable and Administrative arrangements

- **Classes** take place on Tuesdays (3pm, Herschel LT2) and Thursdays (9am, online – fortnightly only; 12pm, Herschel LT1)
- Two of these classes will be **lectures**, and the other session will be a **problems class/drop-in**
 - Lectures will *ordinarily* take place in the in-person slots in the lecture theatres
 - Problems classes/drop-ins will *ordinarily* take place in the online slot
- Due to outreach/recruitment activities, I will not be at work on Thursday 2nd February; Thursday 9th February and Thursday 11th May. Fear not! **Replacement lectures** have been booked:
 - Friday 3rd February (11am, Herschel LT2)
 - Friday 17th February (11am, Herschel LT2)
 - Friday 3rd March (11am, Herschel LT2)
- Office hours will be scheduled soon, but just come along and give me a knock!

Timetable and Administrative arrangements

- **Classes** take place on Tuesdays (3pm, Herschel LT2) and Thursdays (9am, online – fortnightly only; 12pm, Herschel LT1)
- Two of these classes will be **lectures**, and the other session will be a **problems class/drop-in**
 - Lectures will *ordinarily* take place in the in-person slots in the lecture theatres
 - Problems classes/drop-ins will *ordinarily* take place in the online slot
- Due to outreach/recruitment activities, I will not be at work on Thursday 2nd February; Thursday 9th February and Thursday 11th May. Fear not! **Replacement lectures** have been booked:
 - Friday 3rd February (11am, Herschel LT2)
 - Friday 17th February (11am, Herschel LT2)
 - Friday 3rd March (11am, Herschel LT2)
- **Office hours** will be scheduled soon, but just come along and give me a knock!

Assessment

Assessment is by:

- 2-hour end-of-semester exam in May/June (**80%**)
- In course assessment (**20%**), comprising four short written assignments:

Date set	Deadline	Marks/weighting
Friday 10th February	4pm, Friday 24th February	20/5%
Friday 24th February	4pm, Friday 10th March	20/5%
Friday 10th March	4pm, Friday 24th March	20/5%
Friday 24th March	4pm, Friday 5th May	20/5%

Assessment

Assessment is by:

- 2-hour end-of-semester exam in May/June (**80%**)
- In course assessment (**20%**), comprising four short written assignments:

Date set	Deadline	Marks/weighting
Friday 10th February	4pm, Friday 24th February	20/5%
Friday 24th February	4pm, Friday 10th March	20/5%
Friday 10th March	4pm, Friday 24th March	20/5%
Friday 24th March	4pm, Friday 5th May	20/5%

Recommended textbooks

- “*Bayes’ Rule: A Tutorial Introduction to Bayesian Analysis*” - James Stone
- “*Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*” - John Krushke
- “*Bayesian Statistics: An Introduction*” - Peter Lee

“Bayes’ rule” is a good introduction to the main concepts in Bayesian statistics but doesn’t cover everything in this course. The other books are broader references which go well beyond the contents of this course.

Other stuff

- Notes (with gaps) will be available for use in lectures – you can fill in the gaps during lectures
- A (very!) summarised version of the notes will be used in lectures as slides
- These notes and slides will be posted on the course Canvas page after each topic is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- Notes (with gaps) will be available for use in lectures – you can fill in the gaps during lectures
- A (very!) summarised version of the notes will be used in lectures as slides
- These notes and slides will be posted on the course Canvas page after each topic is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Other stuff

- Notes (with gaps) will be available for use in lectures – you can fill in the gaps during lectures
- A (very!) summarised version of the notes will be used in lectures as slides
- These notes and slides will be posted on the course Canvas page after each topic is finished, along with any other course material – such as problems sheets, model solutions to assignment questions, supplementary handouts etc.

Chapter 1

Single parameter problems

Chapter 1. Single parameter problems

1.1 Prior and posterior distributions

Data: $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$

Model: pdf/pf $f(\mathbf{x}|\theta)$ depends on a single parameter θ

→ Likelihood function $f(\mathbf{x}|\theta)$

considered as a function of θ for known $\mathbf{x}$

Prior beliefs: pdf/pf $\pi(\theta)$

Combine using Bayes Theorem

Posterior beliefs: pdf/pf $\pi(\theta|\mathbf{x})$

Posterior distribution summarises all our current knowledge about the parameter θ

Bayes Theorem

The posterior probability (density) function for θ is

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

where

$$f(\mathbf{x}) = \begin{cases} \int_{\Theta} \pi(\theta) f(\mathbf{x}|\theta) d\theta & \text{if } \theta \text{ is continuous,} \\ \sum_{\Theta} \pi(\theta) f(\mathbf{x}|\theta) & \text{if } \theta \text{ is discrete.} \end{cases}$$

As $f(\mathbf{x})$ is not a function of θ , Bayes Theorem can be rewritten as

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

i.e. **posterior** $\propto$ **prior** $\times$ **likelihood**

Bayes Theorem

The posterior probability (density) function for θ is

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

where

$$f(\mathbf{x}) = \begin{cases} \int_{\Theta} \pi(\theta) f(\mathbf{x}|\theta) d\theta & \text{if } \theta \text{ is continuous,} \\ \sum_{\Theta} \pi(\theta) f(\mathbf{x}|\theta) & \text{if } \theta \text{ is discrete.} \end{cases}$$

As $f(\mathbf{x})$ is not a function of θ , Bayes Theorem can be rewritten as

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

i.e. **posterior** $\propto$ **prior** $\times$ **likelihood**

Example 1.1

- Data

Year	1998	1999	2000	2001	2002	2003	2004	2005
Cases	2	0	0	0	1	0	2	1

Table: Number of cases of foodborne botulism in England and Wales, 1998–2005

- Assume that cases occur at random at a constant rate θ in time
- This means the data are from a Poisson process and so are a random sample from a Poisson distribution with rate θ
- An expert was consulted to construct a suitable prior distribution for the rate θ

Example 1.1

- (a) What might be an appropriate probability model for the rate θ ? Why is this termed the *prior distribution*?
- (b) Suppose the epidemiologist tells you that she expects around two cases of foodbourne botulism per year in England and Wales. After further discussions with the epidemiologist, you are also able to elicit that $SD(\theta) \approx \sqrt{2}$. Using your answer to part (a) find your full prior distribution for θ , and write down its density.
- (c) Determine the posterior distribution for θ .

Solution

• • • Solution...

Summary

- Model: $X_i|\theta \sim \text{Po}(\theta)$, $i = 1, 2, \dots, 8$ (independent)
- Prior: $\theta \sim \text{Ga}(2, 1)$
- Data: in Table above
- Posterior: $\theta|\mathbf{x} \sim \text{Ga}(8, 9)$

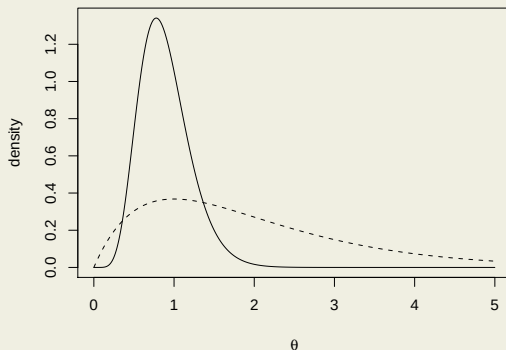


Figure: Prior (dashed) and posterior (solid) densities for θ

	Prior	Likelihood	Posterior
$Mode(\theta)$	1.00	0.75	0.78
$E(\theta)$	2.00	–	0.89
$SD(\theta)$	1.41	–	0.31

Table: Changes in beliefs about θ

- Likelihood mode $<$ prior mode $\rightarrow$ posterior mode moves in direction of likelihood mode $\rightarrow$ posterior mode $<$ prior mode
- Reduction in variability from the prior to the posterior

Example 1.2

We now consider the general case (of Example 1.1). Suppose $X_i|\theta \sim Po(\theta)$, $i = 1, 2, \dots, n$ (independent) and our prior beliefs about θ are summarised by a $Ga(g, h)$ distribution (with g and h known), with density

$$\pi(\theta) = \frac{h^g \theta^{g-1} e^{-h\theta}}{\Gamma(g)}, \quad \theta > 0. \quad (1.1)$$

Determine the posterior distribution for θ .

Solution

...

(1.2)

(1.3)

Summary

- Model: $X_i|\theta \sim Po(\theta)$, $i = 1, 2, \dots, n$ (independent)
- Prior: $\theta \sim Ga(g, h)$
- Data: observe $\mathbf{x}$
- Posterior: $\theta|\mathbf{x} \sim Ga(g + n\bar{x}, h + n)$
- Taking $g \geq 1$

	Prior (1.1)	Likelihood (1.2)	Posterior (1.3)
$Mode(\theta)$	$(g - 1)/h$	$\bar{x}$	$(g + n\bar{x} - 1)/(h + n)$
$E(\theta)$	g/h	—	$(g + n\bar{x})/(h + n)$
$SD(\theta)$	$\sqrt{g}/h$	—	$\sqrt{g + n\bar{x}}/(h + n)$

Table: Changes in beliefs about θ

Comments

- Posterior mean is greater than the prior mean if and only if the likelihood mode is greater than the prior mean, that is,

$$E(\theta|\mathbf{x}) > E(\theta) \iff \text{Mode}_{\theta}\{f(\mathbf{x}|\theta)\} > E(\theta)$$

- Standard deviation of the posterior distribution is smaller than that of the prior distribution if and only if the sample mean is not too large, that is

$$SD(\theta|\mathbf{x}) < SD(\theta) \iff \text{Mode}_{\theta}\{f(\mathbf{x}|\theta)\} < \left(2 + \frac{n}{h}\right) E(\theta),$$

and that this will be true in large samples.

Example 1.3

Suppose we have a random sample from a normal distribution with unknown mean μ but known precision τ : $X_i|\mu \sim N(\mu, 1/\tau)$, $i = 1, 2, \dots, n$ (independent).

Suppose our prior beliefs about μ can be summarised by a $N(b, 1/d)$ distribution, with probability density function

$$\pi(\mu) = \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\}. \quad (1.4)$$

Determine the posterior distribution for μ .

Hint:

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

where c does not depend on μ .

Solution

...

(1.5)

(1.6)

(1.7)

Summary

- Model: $X_i|\mu \sim N(\mu, 1/\tau)$, $i = 1, 2, \dots, n$ (independent), with τ known
- Prior: $\mu \sim N(b, 1/d)$
- Data: observe $\mathbf{x}$
- Posterior: $\mu|\mathbf{x} \sim N(B, 1/D)$, where

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

	Prior (1.4)	Likelihood (1.5)	Posterior (1.7)
$Mode(\mu)$	b	$\bar{x}$	$(db + n\tau\bar{x})/(d + n\tau)$
$E(\mu)$	b	–	$(db + n\tau\bar{x})/(d + n\tau)$
$Precision(\mu)$	d	–	$d + n\tau$

Table: Changes in beliefs about μ

- Posterior mean is greater than the prior mean if and only if the likelihood mode (sample mean) is greater than the prior mean, that is

$$E(\mu|\mathbf{x}) > E(\mu) \quad \Longleftrightarrow \quad Mode_{\mu}\{f(\mathbf{x}|\mu)\} > E(\mu)$$

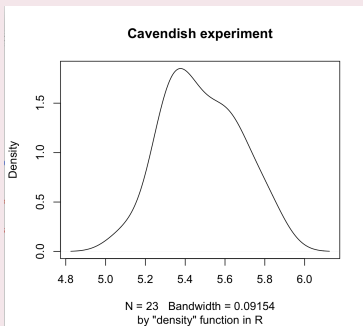
- Standard deviation of the posterior distribution is smaller than that of the prior distribution, that is

$$SD(\mu|\mathbf{x}) < SD(\mu)$$

Example 1.4

The 18th century physicist Henry Cavendish made 23 experimental determinations of the earth's density, and these data (in g/cm^3) are shown below.

5.36	5.29	5.58	5.65	5.57	5.53	5.62	5.29
5.44	5.34	5.79	5.10	5.27	5.39	5.42	5.47
5.63	5.34	5.46	5.30	5.78	5.68	5.85	



Example 1.4

- Cavendish asserts that the error standard deviation of these measurements is 0.2 g/cm^3
- Assume that they are normally distributed with mean equal to the true earth density μ , that is, $X_i|\mu \sim N(\mu, 0.2^2)$, $i = 1, 2, \dots, 23$
- Use a normal prior distribution for μ with mean 5.41 g/cm^3 and standard deviation 0.4 g/cm^3
- Derive the posterior distribution for μ

Solution

...

Summary

- Small increase in mean from prior to posterior
- Large decrease in uncertainty from prior to posterior

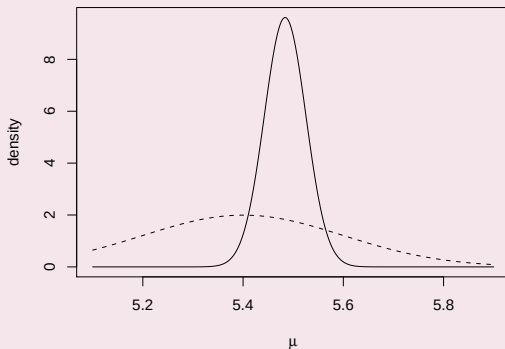


Figure: Prior (dashed) and posterior (solid) densities for the earth's density

Summary

- The actual mean density of the earth is 5.515 g/cm^3 (Wikipedia)
- We can determine the (posterior) probability that the mean density is within 0.1 of this value as follows...
- The posterior distribution is $\mu|x \sim N(5.484, 0.0415^2)$ and so

$$Pr(5.415 < \mu < 5.615|x) = 0.9510,$$

calculated using the R command

```
pnorm(5.615,5.484,0.0415)-pnorm(5.415,5.484,0.0415)
```

Summary

- The actual mean density of the earth is 5.515 g/cm^3 (Wikipedia)
- We can determine the (posterior) probability that the mean density is within 0.1 of this value as follows...
- The posterior distribution is $\mu|x \sim N(5.484, 0.0415^2)$ and so

$$Pr(5.415 < \mu < 5.615|x) = 0.9510,$$

calculated using the R command

```
pnorm(5.615,5.484,0.0415)-pnorm(5.415,5.484,0.0415)
```

Summary

- The actual mean density of the earth is 5.515 g/cm^3 (Wikipedia)
- We can determine the (posterior) probability that the mean density is within 0.1 of this value as follows...
- The posterior distribution is $\mu|\mathbf{x} \sim N(5.484, 0.0415^2)$ and so

$$Pr(5.415 < \mu < 5.615|\mathbf{x}) = 0.9510,$$

calculated using the R command

```
pnorm(5.615,5.484,0.0415)-pnorm(5.415,5.484,0.0415)
```

Summary

- Without the data, the only basis for determining the earth's density is via the prior distribution
- Here the prior distribution is $\mu \sim N(5.4, 0.4^2)$ and so the (prior) probability that the mean density is within 0.1 of the (now known) true value is

$$Pr(5.415 < \mu < 5.615) = 0.1896,$$

calculated using the R command

```
pnorm(5.615,5.4,0.4)-pnorm(5.415,5.4,0.4)
```

Summary

- Without the data, the only basis for determining the earth's density is via the prior distribution
- Here the prior distribution is $\mu \sim N(5.4, 0.4^2)$ and so the (prior) probability that the mean density is within 0.1 of the (now known) true value is

$$Pr(5.415 < \mu < 5.615) = 0.1896,$$

calculated using the R command

```
pnorm(5.615,5.4,0.4)-pnorm(5.415,5.4,0.4)
```

1.2 Different levels of prior knowledge

1.2.1 Substantial Prior Knowledge

- We have *substantial prior information* for θ when the prior distribution *dominates* the likelihood function, that is $\pi(\theta|\mathbf{x}) \sim \pi(\theta)$
- Difficulties:
 - 1 Intractability of mathematics in deriving the posterior
 - 2 Practical formulation of the prior distribution – coherently specifying prior beliefs in the form of a probability distribution is far from straightforward, let alone reconciling differences between experts!

1.2.2 Limited prior information

Pragmatic approach:

- Choose a distribution which makes the Bayes updating from prior to posterior mathematically straightforward
- Use what prior information is available to determine the parameters of this distribution

Previous examples:

- ➊ Poisson random sample, Gamma prior distribution $\longrightarrow$ Gamma posterior distribution
- ➋ Normal random sample (known variance), Normal prior distribution $\longrightarrow$ Normal posterior distribution

In these examples, the prior distribution and the posterior distribution come from the same family

1.2.2 Limited prior information

Pragmatic approach:

- Choose a distribution which makes the Bayes updating from prior to posterior mathematically straightforward
- Use what prior information is available to determine the parameters of this distribution

Previous examples:

- 1 Poisson random sample, Gamma prior distribution $\longrightarrow$ Gamma posterior distribution
- 2 Normal random sample (known variance), Normal prior distribution $\longrightarrow$ Normal posterior distribution

In these examples, the prior distribution and the posterior distribution come from the same family

Definition 1.1 (Conjugate priors)

Suppose that data $\mathbf{x}$ are to be observed with distribution $f(\mathbf{x}|\theta)$. A family $\mathfrak{F}$ of prior distributions for θ is said to be *conjugate* to $f(\mathbf{x}|\theta)$ if for every prior distribution $\pi(\theta) \in \mathfrak{F}$, the posterior distribution $\pi(\theta|\mathbf{x})$ is also in $\mathfrak{F}$.

Comment

The conjugate family depends crucially on the model chosen for the data $\mathbf{x}$.

For example, the only family conjugate to the model “random sample from a Poisson distribution” is the Gamma family. Here, the likelihood is $f(\mathbf{x}|\theta) \propto \theta^{n\bar{x}} e^{-n\theta}$, $\theta > 0$. Therefore we need a family with density $f(\theta|\mathbf{a})$ and parameters $\mathbf{a}$ such that

$$f(\theta|\mathbf{A}) \propto f(\theta|\mathbf{a}) \times \theta^{n\bar{x}} e^{-n\theta}, \quad \theta > 0$$

$$\implies f(\theta|\mathbf{a}) \propto \theta^{a_1} e^{-a_2\theta}, \quad \theta > 0$$

that is, the Gamma family of distributions

1.2.3 Vague prior knowledge/prior ignorance

- Have very little or no prior information about θ
- Still must choose a prior distribution
- Sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance
- Most of the information about θ will be passed through to the posterior distribution via the data, and so we have $\pi(\theta|\mathbf{x}) \sim f(\mathbf{x}|\theta)$
- Vague prior: Represent vague prior knowledge by using a prior distribution which is conjugate to the model for $\mathbf{x}$ and which has as large a variance as possible
- Prior ignorance: Uniform prior? Jeffreys prior?

$$I(\theta) = E_{\mathbf{X}|\theta} \left[-\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{X}|\theta) \right], \quad (1.8)$$

and then:

$$\pi(\theta) \propto \sqrt{I(\theta)}. \quad (1.9)$$

1.2.3 Vague prior knowledge/prior ignorance

- Have very little or no prior information about θ
- Still must choose a prior distribution
- Sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance
- Most of the information about θ will be passed through to the posterior distribution via the data, and so we have $\pi(\theta|\mathbf{x}) \sim f(\mathbf{x}|\theta)$
- Vague prior: Represent vague prior knowledge by using a prior distribution which is conjugate to the model for $\mathbf{x}$ and which has as large a variance as possible
- Prior ignorance: Uniform prior? Jeffreys prior?

$$I(\theta) = E_{\mathbf{X}|\theta} \left[-\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{X}|\theta) \right], \quad (1.8)$$

and then:

$$\pi(\theta) \propto \sqrt{I(\theta)}. \quad (1.9)$$

1.2.3 Vague prior knowledge/prior ignorance

- Have very little or no prior information about θ
- Still must choose a prior distribution
- Sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance
- Most of the information about θ will be passed through to the posterior distribution via the data, and so we have $\pi(\theta|\mathbf{x}) \sim f(\mathbf{x}|\theta)$
- Vague prior: Represent vague prior knowledge by using a prior distribution which is conjugate to the model for $\mathbf{x}$ and which has as large a variance as possible
- Prior ignorance: Uniform prior? Jeffreys prior?

$$I(\theta) = E_{\mathbf{X}|\theta} \left[-\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{X}|\theta) \right], \quad (1.8)$$

and then:

$$\pi(\theta) \propto \sqrt{I(\theta)}. \quad (1.9)$$

1.2.3 Vague prior knowledge/prior ignorance

- Have very little or no prior information about θ
- Still must choose a prior distribution
- Sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance
- Most of the information about θ will be passed through to the posterior distribution via the data, and so we have $\pi(\theta|\mathbf{x}) \sim f(\mathbf{x}|\theta)$
- Vague prior: Represent vague prior knowledge by using a prior distribution which is conjugate to the model for $\mathbf{x}$ and which has as large a variance as possible
- Prior ignorance: Uniform prior? Jeffreys prior?

$$I(\theta) = E_{\mathbf{X}|\theta} \left[-\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{X}|\theta) \right], \quad (1.8)$$

and then:

$$\pi(\theta) \propto \sqrt{I(\theta)}. \quad (1.9)$$

1.2.3 Vague prior knowledge/prior ignorance

- Have very little or no prior information about θ
- Still must choose a prior distribution
- Sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance
- Most of the information about θ will be passed through to the posterior distribution via the data, and so we have $\pi(\theta|\mathbf{x}) \sim f(\mathbf{x}|\theta)$
- Vague prior: Represent vague prior knowledge by using a prior distribution which is conjugate to the model for $\mathbf{x}$ and which has as large a variance as possible
- Prior ignorance: Uniform prior? Jeffreys prior?

$$I(\theta) = E_{\mathbf{X}|\theta} \left[-\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{X}|\theta) \right], \quad (1.8)$$

and then:

$$\pi(\theta) \propto \sqrt{I(\theta)}. \quad (1.9)$$

Example 1.5

In constructing the prior in Example 1.1, did we assume (i) substantial prior knowledge; (ii) limited prior knowledge; or (iii) prior ignorance?

Solution

...

Example 1.6

Suppose the parameter θ represents a probability. What might be the simplest prior to use here, if we are prior ignorant?

Solution

...

Example 1.7

Suppose we have a random sample from a $N(\mu, 1/\tau)$ distribution (with τ known). Determine the posterior distribution assuming a vague prior for μ .

Solution

...

Example 1.8

Suppose we have a random sample from an Poisson distribution, that is, $X_i|\theta \sim Po(\theta)$, $i = 1, 2, \dots, n$ (independent). Determine the posterior distribution assuming a vague prior for θ .

Solution

- The conjugate prior distribution is a Gamma distribution
- The $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$
- Rearranging gives $g = m^2/v$ and $h = m/v$
- Clearly $g \rightarrow 0$ and $h \rightarrow 0$ as $v \rightarrow \infty$ (for fixed m)
- We have seen that, for this model, using a $Ga(g, h)$ prior distribution results in a $Ga(g + n\bar{x}, h + n)$ posterior distribution
- Therefore, taking a vague prior distribution will give a $Ga(n\bar{x}, n)$ posterior distribution

Note that the posterior mean is $\bar{x}$ (the likelihood mode) and that the posterior variance $\bar{x}/n \rightarrow 0$ and $n \rightarrow \infty$.

Example 1.9

Suppose we have a random sample from an exponential distribution, that is, $X_i|\theta \sim \text{Exp}(\theta)$, $i = 1, 2, \dots, n$ (independent). Determine the Jeffreys prior for this model.

Solution

...