Chapter 3

Prior elicitation

In this chapter we will think about how to **construct** a suitable **prior distribution** $\pi(\theta)$ for our parameter of interest θ.

For example:

- Why did we use a $Be(77, 5)$ distribution for θ in the music expert example?
- Why did we use a $Be(2.5, 12)$ distribution for θ in the video game pirate example?
- Why did we assume a $Ga(10, 4000)$ distribution for θ in the earthquake example?

Introduction

In this chapter we will think about how to **construct** a suitable **prior distribution** $\pi(\theta)$ for our parameter of interest θ.

For example:

- Why did we use a $Be(77, 5)$ distribution for θ in the music expert example?
- Why did we use a $Be(2.5, 12)$ distribution for θ in the video game pirate example?
- Why did we assume a $Ga(10, 4000)$ distribution for θ in the earthquake example?

Introduction

In this chapter we will think about how to **construct** a suitable **prior distribution** $\pi(\theta)$ for our parameter of interest θ.

For example:

- Why did we use a $Be(77, 5)$ distribution for θ in the music expert example?
- Why did we use a $Be(2.5, 12)$ distribution for θ in the video game pirate example?
- Why did we assume a $Ga(10, 4000)$ distribution for θ in the earthquake example?

In this chapter we will think about how to **construct** a suitable **prior distribution** $\pi(\theta)$ for our parameter of interest θ.

For example:

- Why did we use a *Be*$(77, 5)$ distribution for θ in the music expert example?
- Why did we use a *Be*$(2.5, 12)$ distribution for θ in the video game pirate example?
- Why did we assume a *Ga*$(10, 4000)$ distribution for θ in the earthquake example?

Prior elicitation – the process by which we attempt to construct the most suitable prior distribution for θ – is a huge area of research in Bayesian Statistics.

The aim in this course is to give a brief (and relatively simple) overview.

We will

- have another case study lecture (before Easter)
- have an interactive music session (using the TurningPoint voting system, after Easter)

to demonstrate the techniques.

We will

- have another case study lecture (before Easter)
- have an interactive music session (using the TURNINGPOINT voting system, after Easter)

to demonstrate the techniques.

We will

- have another case study lecture (before Easter)
- have an interactive music session (using the TURNINGPOINT voting system, after Easter)

to demonstrate the techniques.

We will

- have another case study lecture (before Easter)
- have an interactive music session (using the TURNINGPOINT voting system, after Easter)

to demonstrate the techniques.

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

Introduction

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which "makes sense" and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

Introduction

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which "makes sense" and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

We will consider the cases of:

- **Substantial prior knowledge**
  - turning expert opinion into a probability distribution for θ
  - re–visit the examples about the music expert, the video game pirate and earthquakes in Chapter 2.
- **Vague prior knowledge**
  - No expert available
  - Choose a prior which “makes sense” and keeps the maths simple!
- **Prior ignorance**
  - Assume all values of θ are equally likely

Substantial prior knowledge

We will consider various methods for constructing prior distributions when we have **substantial prior knowledge**:

- Use of **suggested prior summaries**
- The **trial roulette method**
- The **bisection method**

We will make use of an online elicitation tool called `MATCH`, courtesy of Professor Tony O'Hagan and Dr Jeremy Oakley (Sheffield University).

We will consider various methods for constructing prior distributions when we have **substantial prior knowledge**:

- Use of **suggested prior summaries**
- The **trial roulette method**
- The **bisection method**

We will make use of an online elicitation tool called `MATCH`, courtesy of Professor Tony O'Hagan and Dr Jeremy Oakley (Sheffield University).

Substantial prior knowledge

We will consider various methods for constructing prior distributions when we have **substantial prior knowledge**:

- Use of **suggested prior summaries**
- The **trial roulette method**
- The **bisection method**

We will make use of an online elicitation tool called `MATCH`, courtesy of Professor Tony O'Hagan and Dr Jeremy Oakley (Sheffield University).

Substantial prior knowledge

We will consider various methods for constructing prior distributions when we have **substantial prior knowledge**:

- Use of **suggested prior summaries**
- The **trial roulette method**
- The **bisection method**

We will make use of an online elicitation tool called `MATCH`, courtesy of Professor Tony O'Hagan and Dr Jeremy Oakley (Sheffield University).

Substantial prior knowledge

We will consider various methods for constructing prior distributions when we have **substantial prior knowledge**:

- Use of **suggested prior summaries**
- The **trial roulette method**
- The **bisection method**

We will make use of an online elicitation tool called `MATCH`, courtesy of Professor Tony O'Hagan and Dr Jeremy Oakley (Sheffield University).

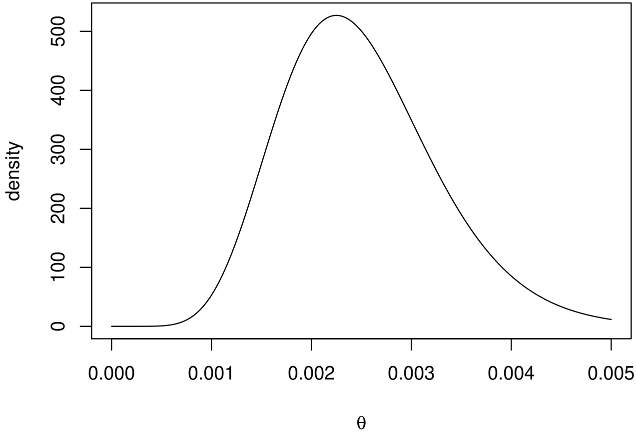
Example 3.1: Using suggested prior summaries

Let us return to **Example 2.4** of Chapter 2.

Recall that we were given some data on the “waiting times”, in days, between 21 earthquakes, and we discussed why an exponenetial distribution $Exp(\theta)$ might be appropriate to model the waiting times.

Further, we were told that an **expert on earthquakes** has prior beliefs about the rate θ, described by a $Ga(10, 4000)$ distribution.

Example 3.1: Using suggested prior summaries



How did we get from the expert's beliefs to a $Ga(10, 4000)$?

Example 3.1: Using suggested prior summaries

Suppose the expert tells us that earthquakes in the region we are interested in usually occur **less than once per year**.

In fact, they occur on average **once every 400 days**.

This gives us a rate of occurrence of about $1/400 = 0.0025$ per day.

Further, he is fairly certain about this and specifies a **very small variance** of 6.25×10^{-7}.

Example 3.1: Using suggested prior summaries

Suppose the expert tells us that earthquakes in the region we are interested in usually occur **less than once per year**.

In fact, they occur on average **once every 400 days**.

This gives us a rate of occurrence of about $1/400 = 0.0025$ per day.

Further, he is fairly certain about this and specifies a **very small variance** of 6.25×10^{-7}.

Example 3.1: Using suggested prior summaries

Suppose the expert tells us that earthquakes in the region we are interested in usually occur **less than once per year**.

In fact, they occur on average **once every 400 days**.

This gives us a rate of occurrence of about 1/400 = 0.0025 per day.

Further, he is fairly certain about this and specifies a **very small variance** of 6.25×10^{-7}.

Example 3.1: Using suggested prior summaries

Suppose the expert tells us that earthquakes in the region we are interested in usually occur **less than once per year**.

In fact, they occur on average **once every 400 days**.

This gives us a rate of occurrence of about $1/400 = 0.0025$ per day.

Further, he is fairly certain about this and specifies a **very small variance** of 6.25×10^{-7}.

Example 3.1: Using suggested prior summaries

A $Ga(a, b)$ distribution seems sensible, since we can't observe a negative daily earthquake rate and the Gamma distribution is specified over positive values only.

Using the information provided by the expert, verify our use of $a = 10$ and $b = 4000$.

Example 3.1: Using suggested prior summaries

A $Ga(a, b)$ distribution seems sensible, since we can't observe a negative daily earthquake rate and the Gamma distribution is specified over positive values only.

Using the information provided by the expert, verify our use of $a = 10$ and $b = 4000$.

Solution to Example 3.1 (1/1)

We know that, if $\theta \sim Ga(a, b)$, then $E(\theta) = a/b$ and $Var(\theta) = a/b^2$. Thus

$$\frac{a}{b} = 0.0025 \implies a = 0.0025b.$$

Substituting into $a/b^2 = 0.000000625$ gives

$$\begin{aligned} \frac{0.0025b}{b^2} &= 0.000000625, \qquad \text{giving} \\ b &= 4000. \end{aligned}$$

Thus $a = 0.0025 \times 4000 = 10$.

Solution to Example 3.1 (1/1)

We know that, if $\theta \sim Ga(a, b)$, then $E(\theta) = a/b$ and $Var(\theta) = a/b^2$. Thus

$$\frac{a}{b} = 0.0025 \implies a = 0.0025b.$$

Substituting into $a/b^2 = 0.000000625$ gives

$$\frac{0.0025b}{b^2} = 0.000000625, \qquad \text{giving}$$

$$b = 4000.$$

Thus $a = 0.0025 \times 4000 = 10$.

Solution to Example 3.1 (1/1)

We know that, if $\theta \sim Ga(a, b)$, then $E(\theta) = a/b$ and $Var(\theta) = a/b^2$. Thus

$$\frac{a}{b} = 0.0025 \implies a = 0.0025b.$$

Substituting into $a/b^2 = 0.000000625$ gives

$$\begin{aligned} \frac{0.0025b}{b^2} &= 0.000000625, \qquad \text{giving} \\ b &= 4000. \end{aligned}$$

Thus $a = 0.0025 \times 4000 = 10$.

Example 3.2: Using suggested prior summaries

Now let us return to **Example 2.2** of Chapter 2.

We considered an experiment to determine how good a music expert is at distinguishing between pages from Haydn and Mozart scores.

When presented with a score from each composer, the expert makes the correct choice with probability θ.

Example 3.2: Using suggested prior summaries

Before conducting the experiment, we were told that the expert is very competent. In fact, we were told that

- θ should have a prior distribution **peaking at around 0.95**
- $\Pr(\theta < 0.8)$ should be **very small**

To achieve this, we assumed that $\theta \sim \textit{Be}(77, 5)$, with density given in Figure 2.4.

Example 3.2: Using suggested prior summaries

Before conducting the experiment, we were told that the expert is very competent. In fact, we were told that

- θ should have a prior distribution **peaking at around 0.95**
- $\Pr(\theta < 0.8)$ should be **very small**

To achieve this, we assumed that $\theta \sim Be(77, 5)$, with density given in Figure 2.4.

Example 3.2: Using suggested prior summaries

Before conducting the experiment, we were told that the expert is very competent. In fact, we were told that

- θ should have a prior distribution **peaking at around 0.95**
- $\Pr(\theta < 0.8)$ should be **very small**

To achieve this, we assumed that $\theta \sim Be(77, 5)$, with density given in Figure 2.4.

Example 3.2: Using suggested prior summaries

Before conducting the experiment, we were told that the expert is very competent. In fact, we were told that

- θ should have a prior distribution **peaking at around 0.95**
- Pr($\theta < 0.8$) should be **very small**

To achieve this, we assumed that $\theta \sim Be(77, 5)$, with density given in Figure 2.4.

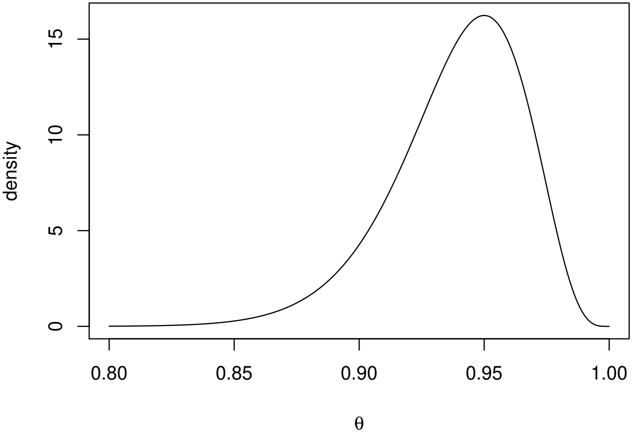
Example 3.2: Using suggested prior summaries

Before conducting the experiment, we were told that the expert is very competent. In fact, we were told that

- θ should have a prior distribution **peaking at around 0.95**
- $\Pr(\theta < 0.8)$ should be **very small**

To achieve this, we assumed that $\theta \sim Be(77, 5)$, with density given in Figure 2.4.

Example 3.2: Using suggested prior summaries



How did we know a $Be(77, 5)$ would work?

Example 3.2: Using suggested prior summaries

We are told that the mode of the distribution should be around 0.95. Using the formulae on **page 23**, we can write

$$
\begin{aligned}
\frac{a-1}{a+b-2} &= 0.95 \\
\Rightarrow a-1 &= 0.95(a+b-2) \\
\Rightarrow a-0.95a &= 0.95b-1.9+1 \\
\Rightarrow 0.05a &= 0.95b-0.9 \\
\Rightarrow a &= 19b-18.
\end{aligned}
$$

We are told that the mode of the distribution should be around 0.95. Using the formulae on **page 23**, we can write

$$\begin{aligned}
\frac{a-1}{a+b-2} &= 0.95 \\
\Rightarrow a-1 &= 0.95(a+b-2) \\
\Rightarrow a-0.95a &= 0.95b-1.9+1 \\
\Rightarrow 0.05a &= 0.95b-0.9 \\
\Rightarrow a &= 19b-18.
\end{aligned}$$

Example 3.2: Using suggested prior summaries

We are told that the mode of the distribution should be around 0.95. Using the formulae on **page 23**, we can write

$$\frac{a-1}{a+b-2} = 0.95$$

$$\begin{aligned} \Rightarrow a-1 &= 0.95(a+b-2) \\ \Rightarrow a-0.95a &= 0.95b-1.9+1 \\ \Rightarrow 0.05a &= 0.95b-0.9 \\ \Rightarrow a &= 19b-18. \end{aligned}$$

Example 3.2: Using suggested prior summaries

We are told that the mode of the distribution should be around 0.95. Using the formulae on **page 23**, we can write

$$\begin{aligned}
\frac{a-1}{a+b-2} &= 0.95 \\
\Rightarrow a-1 &= 0.95(a+b-2) \\
\Rightarrow a-0.95a &= 0.95b-1.9+1 \\
\Rightarrow 0.05a &= 0.95b-0.9 \\
\Rightarrow a &= 19b-18.
\end{aligned}$$

Example 3.2: Using suggested prior summaries

We are told that the mode of the distribution should be around 0.95. Using the formulae on **page 23**, we can write

$$\begin{aligned}
\frac{a-1}{a+b-2} &= 0.95 \\
\Rightarrow a-1 &= 0.95(a+b-2) \\
\Rightarrow a-0.95a &= 0.95b-1.9+1 \\
\Rightarrow 0.05a &= 0.95b-0.9 \\
\Rightarrow a &= 19b-18.
\end{aligned}$$

Example 3.2: Using suggested prior summaries

We are told that the mode of the distribution should be around 0.95. Using the formulae on **page 23**, we can write

$$
\begin{aligned}
\frac{a-1}{a+b-2} &= 0.95 \\
\Rightarrow a-1 &= 0.95(a+b-2) \\
\Rightarrow a-0.95a &= 0.95b-1.9+1 \\
\Rightarrow 0.05a &= 0.95b-0.9 \\
\Rightarrow a &= 19b-18.
\end{aligned}
$$

Example 3.2: Using suggested prior summaries

We are told that the mode of the distribution should be around 0.95. Using the formulae on **page 23**, we can write

$$\begin{aligned}\frac{a-1}{a+b-2} &= 0.95\\ \Rightarrow a-1 &= 0.95(a+b-2)\\ \Rightarrow a-0.95a &= 0.95b-1.9+1\\ \Rightarrow 0.05a &= 0.95b-0.9\\ \Rightarrow a &= 19b-18.\end{aligned}$$

Example 3.2: Using suggested prior summaries

We are also told that $\Pr(\theta < 0.8)$ must be small.

In fact, suppose we are told that $\theta < 0.8$ might occur with probability 0.0001.

This means that if we integrate the probability density function for our beta distribution between 0 and 0.8, we would get 0.0001; from Equation (2.1) on page 23, we can write this as

$$\int_0^{0.8} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta = 0.0001, \qquad \text{i.e.}$$

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18,b)} d\theta = 0.0001.$$

We are also told that $\Pr(\theta < 0.8)$ must be small.

In fact, suppose we are told that $\theta < 0.8$ might occur with probability 0.0001.

This means that if we integrate the probability density function for our beta distribution between 0 and 0.8, we would get 0.0001; from Equation (2.1) on page 23, we can write this as

$$\int_0^{0.8} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta = 0.0001, \qquad \text{i.e.}$$

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18,b)} d\theta = 0.0001.$$

Example 3.2: Using suggested prior summaries

We are also told that $\Pr(\theta < 0.8)$ must be small.

In fact, suppose we are told that $\theta < 0.8$ might occur with probability 0.0001.

This means that if we integrate the probability density function for our beta distribution between 0 and 0.8, we would get 0.0001; from Equation (2.1) on **page 23**, we can write this as

$$\int_0^{0.8} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta = 0.0001, \qquad \text{i.e.}$$

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18,b)} d\theta = 0.0001.$$

Example 3.2: Using suggested prior summaries

We are also told that $\Pr(\theta < 0.8)$ must be small.

In fact, suppose we are told that $\theta < 0.8$ might occur with probability 0.0001.

This means that if we integrate the probability density function for our beta distribution between 0 and 0.8, we would get 0.0001; from Equation (2.1) on **page 23**, we can write this as

$$\int_0^{0.8} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta = 0.0001, \qquad \text{i.e.}$$

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18,b)} d\theta = 0.0001.$$

Example 3.2: Using suggested prior summaries

We are also told that $\Pr(\theta < 0.8)$ must be small.

In fact, suppose we are told that $\theta < 0.8$ might occur with probability 0.0001.

This means that if we integrate the probability density function for our beta distribution between 0 and 0.8, we would get 0.0001; from Equation (2.1) on **page 23**, we can write this as

$$\int_0^{0.8} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta = 0.0001, \qquad \text{i.e.}$$

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18,b)} d\theta = 0.0001.$$

Example 3.2: Using suggested prior summaries

In other words, we set the cumulative distribution function for a $Be(19b - 18, b)$, evaluated at 0.8, equal to 0.0001 and solve for b.

Although this would be tricky to do by hand, we can do this quite easily in R.

Recall that the R command:

- `dbeta(x,a,b)` evaluates the **density** of the $Be(a, b)$ distribution at the point `x`
- `pbeta(x,a,b)` evaluates the **cumulative distribution function** at `x`

Example 3.2: Using suggested prior summaries

In other words, we set the cumulative distribution function for a $Be(19b-18, b)$, evaluated at 0.8, equal to 0.0001 and solve for b.

Although this would be tricky to do by hand, we can do this quite easily in R.

Recall that the R command:

- `dbeta(x,a,b)` evaluates the **density** of the $Be(a, b)$ distribution at the point `x`
- `pbeta(x,a,b)` evaluates the **cumulative distribution function** at `x`

Example 3.2: Using suggested prior summaries

In other words, we set the cumulative distribution function for a $Be(19b - 18, b)$, evaluated at 0.8, equal to 0.0001 and solve for b.

Although this would be tricky to do by hand, we can do this quite easily in R.

Recall that the R command:

- `dbeta(x,a,b)` evaluates the **density** of the $Be(a, b)$ distribution at the point `x`
- `pbeta(x,a,b)` evaluates the **cumulative distribution function** at `x`

Example 3.2: Using suggested prior summaries

In other words, we set the cumulative distribution function for a $Be(19b - 18, b)$, evaluated at 0.8, equal to 0.0001 and solve for b.

Although this would be tricky to do by hand, we can do this quite easily in R.

Recall that the R command:

- `dbeta(x,a,b)` evaluates the **density** of the $Be(a, b)$ distribution at the point `x`
- `pbeta(x,a,b)` evaluates the **cumulative distribution function** at `x`

Example 3.2: Using suggested prior summaries

In other words, we set the cumulative distribution function for a $Be(19b - 18, b)$, evaluated at 0.8, equal to 0.0001 and solve for b.

Although this would be tricky to do by hand, we can do this quite easily in R.

Recall that the R command:

- `dbeta(x,a,b)` evaluates the **density** of the $Be(a, b)$ distribution at the point `x`
- `pbeta(x,a,b)` evaluates the **cumulative distribution function** at `x`

In other words, we set the cumulative distribution function for a $Be(19b-18, b)$, evaluated at 0.8, equal to 0.0001 and solve for b.

Although this would be tricky to do by hand, we can do this quite easily in R.

Recall that the R command:

- `dbeta(x,a,b)` evaluates the **density** of the $Be(a,b)$ distribution at the point `x`
- `pbeta(x,a,b)` evaluates the **cumulative distribution function** at `x`

First of all, we re–write (3.3) to set it equal to zero:

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18,b)}\,d\theta - 0.0001 = 0. \qquad (3.4)$$

We then write a function in R which computes the left–hand–side of Equation (3.4):

```
f=function(b){
    answer=pbeta(0.8,((19*b)-18),b)-0.0001
    return(answer)}
```

Example 3.2: Using suggested prior summaries

First of all, we re–write (3.3) to set it equal to zero:

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18,b)}\,d\theta - 0.0001 = 0. \tag{3.4}$$

We then write a function in R which computes the left–hand–side of Equation (3.4):

```
f=function(b){
    answer=pbeta(0.8,((19*b)-18),b)-0.0001
    return(answer)}
```

Example 3.2: Using suggested prior summaries

First of all, we re–write (3.3) to set it equal to zero:

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18, b)} d\theta - 0.0001 = 0. \qquad (3.4)$$

We then write a function in R which computes the left–hand–side of Equation (3.4):

```
f=function(b){
    answer=pbeta(0.8,((19*b)-18),b)-0.0001
    return(answer)}
```

First of all, we re–write (3.3) to set it equal to zero:

$$\int_0^{0.8} \frac{\theta^{(19b-18)-1}(1-\theta)^{b-1}}{B(19b-18, b)} d\theta - 0.0001 = 0. \qquad (3.4)$$

We then write a function in R which computes the left–hand–side of Equation (3.4):

```
f=function(b){
     answer=pbeta(0.8,((19*b)-18),b)-0.0001
     return(answer)}
```

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on page 23 that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

Example 3.2: Using suggested prior summaries

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on **page 23** that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

Example 3.2: Using suggested prior summaries

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on **page 23** that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

Example 3.2: Using suggested prior summaries

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on **page 23** that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

Example 3.2: Using suggested prior summaries

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on **page 23** that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

Example 3.2: Using suggested prior summaries

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on **page 23** that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

Example 3.2: Using suggested prior summaries

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on **page 23** that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

The trick now is to use a **numerical procedure** to find the root of `answer` in our R function.

In other words, find the value `b` which equates `answer` to zero.

The R function `uniroot(f, lower=, upper=)`

- uses a numerical search algorithm to find the root of the expression provided by the function `f`
- requires the user to provide a `lower` and `upper` bound to search within

We know from the formulae on **page 23** that $a, b > 1$ when using expression (3.1) for the mode $\rightarrow$ so we search for a root over some specified domain > 1.

Example 3.2: Using suggested prior summaries

For example, we might use `lower=1` and `upper=100`, giving:

```
> uniroot(f,lower=1,upper=100)
$root
[1] 5.06513

$f.root
[1] 6.008134e-09

$iter
[1] 14

$estim.prec
[1] 6.103516e-05
```

Example 3.2: Using suggested prior summaries

For example, we might use `lower=1` and `upper=100`, giving:

```
> uniroot(f,lower=1,upper=100)
$root
[1] 5.06513

$f.root
[1] 6.008134e-09

$iter
[1] 14

$estim.prec
[1] 6.103516e-05
```

Thus, the solution to Equation (3.3) is $b = 5.06513$.

For simplicity, rounding down to $b = 5$ and then substituting into (3.2) gives

$$a = 19 \times 5 - 18 = 77,$$

hence the use of $\theta \sim Be(77, 5)$ in Example 2.2 in Chapter 2.

Thus, the solution to Equation (3.3) is $b = 5.06513$.

For simplicity, rounding down to $b = 5$ and then substituting into (3.2) gives

$$a = 19 \times 5 - 18 = 77,$$

hence the use of $\theta \sim Be(77, 5)$ in Example 2.2 in Chapter 2.

Example 3.2: Using suggested prior summaries

Thus, the solution to Equation (3.3) is $b = 5.06513$.

For simplicity, rounding down to $b = 5$ and then substituting into (3.2) gives

$$a = 19 \times 5 - 18 = 77,$$

hence the use of $\theta \sim Be(77, 5)$ in Example 2.2 in Chapter 2.

Example 3.2: Using suggested prior summaries

Thus, the solution to Equation (3.3) is $b = 5.06513$.

For simplicity, rounding down to $b = 5$ and then substituting into (3.2) gives

$$a = 19 \times 5 - 18 = 77,$$

hence the use of $\theta \sim Be(77, 5)$ in Example 2.2 in Chapter 2.

Example 3.3: Trial roulette method

We now return to **Example 2.3** in Chapter 2.

Recall that Max is a video game pirate, and he is trying to identify the proportion θ of potential customers who might be interested in buying *Call of Duty: Elite* next month.

Why did we use $\theta \sim \mathit{Be}(2.5, 12)$?

Example 3.3: Trial roulette method

For each month over the last two years Max knows the proportion of his customers who have bought similar games; these proportions are shown below in Table 3.1.

0.32	0.25	0.28	0.15	0.33	0.12	0.14	0.18	0.12	0.05	0.25	0.08
0.07	0.16	0.24	0.38	0.18	0.15	0.22	0.05	0.01	0.19	0.08	0.15

Example 3.3: Trial roulette method

For each month over the last two years Max knows the proportion of his customers who have bought similar games; these proportions are shown below in Table 3.1.

0.32	0.25	0.28	0.15	0.33	0.12	0.14	0.18	0.12	0.05	0.25	0.08
0.07	0.16	0.24	0.38	0.18	0.15	0.22	0.05	0.01	0.19	0.08	0.15

Example 3.3: Trial roulette method

The **trial roulette method** of elicitation works in the following way:

- Divide the sample space for θ into m "bins"
- Ask the expert/person "in the know" to distribute n "chips" amongst the bins
- The proportion of chips in a particular bin represents the probability that θ lies in that bin
- Done graphically, we can see the shape of the distribution forming as the expert allocates the chips
- We then find a model that closely matches the distribution of chips

Example 3.3: Trial roulette method

The **trial roulette method** of elicitation works in the following way:

- Divide the sample space for θ into m “bins”
- Ask the expert/person “in the know” to distribute n “chips” amongst the bins
- The proportion of chips in a particular bin represents the probability that θ lies in that bin
- Done graphically, we can see the shape of the distribution forming as the expert allocates the chips
- We then find a model that closely matches the distribution of chips

Example 3.3: Trial roulette method

The **trial roulette method** of elicitation works in the following way:

- Divide the sample space for θ into m "bins"
- Ask the expert/person "in the know" to distribute n "chips" amongst the bins
- The proportion of chips in a particular bin represents the probability that θ lies in that bin
- Done graphically, we can see the shape of the distribution forming as the expert allocates the chips
- We then find a model that closely matches the distribution of chips

Example 3.3: Trial roulette method

The **trial roulette method** of elicitation works in the following way:

- Divide the sample space for θ into m “bins”
- Ask the expert/person “in the know” to distribute n “chips” amongst the bins
- The proportion of chips in a particular bin represents the probability that θ lies in that bin
- Done graphically, we can see the shape of the distribution forming as the expert allocates the chips
- We then find a model that closely matches the distribution of chips

Example 3.3: Trial roulette method

The **trial roulette method** of elicitation works in the following way:

- Divide the sample space for θ into m "bins"
- Ask the expert/person "in the know" to distribute n "chips" amongst the bins
- The proportion of chips in a particular bin represents the probability that θ lies in that bin
- Done graphically, we can see the shape of the distribution forming as the expert allocates the chips
- We then find a model that closely matches the distribution of chips

Example 3.3: Trial roulette method

The **trial roulette method** of elicitation works in the following way:

- Divide the sample space for θ into *m* “bins”
- Ask the expert/person “in the know” to distribute *n* “chips” amongst the bins
- The proportion of chips in a particular bin represents the probability that θ lies in that bin
- Done graphically, we can see the shape of the distribution forming as the expert allocates the chips
- We then find a model that closely matches the distribution of chips

Example 3.3: Trial roulette method

We will implement the trial roulette method using the **MATCH Uncertainty Elicitation Tool**.

This was developed by **Dr. Jeremy Oakley** and **Professor Tony O'Hagan** at Sheffield University.

This can be accessed via any web browser:

http://optics.eee.nottingham.ac.uk/match/uncertainty.php

Example 3.3: Trial roulette method

We will implement the trial roulette method using the **MATCH Uncertainty Elicitation Tool**.

This was developed by **Dr. Jeremy Oakley** and **Professor Tony O'Hagan** at Sheffield University.

This can be accessed via any web browser:

http://optics.eee.nottingham.ac.uk/match/uncertainty.php

Example 3.3: Trial roulette method

We will implement the trial roulette method using the **MATCH Uncertainty Elicitation Tool**.

This was developed by **Dr. Jeremy Oakley** and **Professor Tony O'Hagan** at Sheffield University.

This can be accessed via any web browser:

http://optics.eee.nottingham.ac.uk/match/uncertainty.php

Example 3.3: Trial roulette method

We will implement the trial roulette method using the **MATCH Uncertainty Elicitation Tool**.

This was developed by **Dr. Jeremy Oakley** and **Professor Tony O'Hagan** at Sheffield University.

This can be accessed via any web browser:

http://optics.eee.nottingham.ac.uk/match/uncertainty.php

Feedback on your feedback: The good

- **Lecture materials**
  - Booklet clearly structured
  - Good amount of examples
  - Like the amount of writing we have to do in notes
  - Good cross-referencing
  - Like the chapter summaries

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free

- **Chris**
  - Prefer you to Chris anyway
  - Lee >>> Chris

Feedback on your feedback: The good

- **Lee**
  - Quite engaging (at times)
  - Jordy not an issue for me, Warm Jordy vocals, Soothing Jordy tones
  - Good pace
  - Occasionally enthusiastic
  - Oscar-worthy lectures
  - At last, you're free
- **Chris**
  - Prefer you to Chris anyway
  - Lee $>>>$ Chris

Feedback on your feedback: The bad

- **Lecture materials**
  - Remove the boxes in notes!
  - Mathematical font - difficult to tell difference between X and x
  - Do written stuff on visualiser, NOT SLIDES
  - Less Geography please
  - More time to write down from slides
  - Need more examples. Use NUMBAS perhaps?

    `https://mas-shiny.ncl.ac.uk/2903Questions`
  - Too many formulas. Need formula sheet

Feedback on your feedback: The bad

- **Lecture materials**
  - Remove the boxes in notes!
  - Mathematical font - difficult to tell difference between X and x
  - Do written stuff on visualiser, NOT SLIDES
  - Less Geography please
  - More time to write down from slides
  - Need more examples. Use NUMBAS perhaps?

    https://mas-shiny.ncl.ac.uk/2903Questions

  - Too many formulas. Need formula sheet

Feedback on your feedback: The bad

- **Lecture materials**
  - Remove the boxes in notes!
  - Mathematical font - difficult to tell difference between X and x
  - Do written stuff on visualiser, NOT SLIDES
  - Less Geography please
  - More time to write down from slides
  - Need more examples. Use NUMBAS perhaps?

    `https://mas-shiny.ncl.ac.uk/2903Questions`

  - Too many formulas. Need formula sheet

Feedback on your feedback: The bad

- **Lecture materials**
  - Remove the boxes in notes!
  - Mathematical font - difficult to tell difference between X and x
  - Do written stuff on visualiser, NOT SLIDES
  - Less Geography please
  - More time to write down from slides
  - Need more examples. Use NUMBAS perhaps?

    https://mas-shiny.ncl.ac.uk/2903Questions

  - Too many formulas. Need formula sheet

Feedback on your feedback: The bad

- **Lecture materials**
  - Remove the boxes in notes!
  - Mathematical font - difficult to tell difference between X and x
  - Do written stuff on visualiser, NOT SLIDES
  - Less Geography please
  - More time to write down from slides
  - Need more examples. Use NUMBAS perhaps?

    https://mas-shiny.ncl.ac.uk/2903Questions

  - Too many formulas. Need formula sheet

Feedback on your feedback: The bad

- **Lee**
  - Speak. More. S l o w l y
  - You over-explain the simple things
  - Always late, sort it out mate
  - Owe us 20 minutes from MAS2602
  - Go faster plz
- **Chris**
  - Where's Chris? Like him
  - Any more of Chris please?
  - Still love Chris, my climbing boi

Feedback on your feedback: The bad

- **Lee**
  - Speak. More. S l o w l y
  - You over-explain the simple things
  - Always late, sort it out mate
  - Owe us 20 minutes from MAS2602
  - Go faster plz

- **Chris**
  - Where's Chris? Like him
  - Any more of Chris please?
  - Still love Chris, my climbing boi

Feedback on your feedback: The bad

- **Lee**
  - Speak. More. S l o w l y
  - You over-explain the simple things
  - Always late, sort it out mate
  - Owe us 20 minutes from MAS2602
  - Go faster plz
- **Chris**
  - Where's Chris? Like him
  - Any more of Chris please?
  - Still love Chris, my climbing boi

Feedback on your feedback: The bad

- **Lee**
  - Speak. More. S l o w l y
  - You over-explain the simple things
  - Always late, sort it out mate
  - Owe us 20 minutes from MAS2602
  - Go faster plz
- **Chris**
  - Where's Chris? Like him
  - Any more of Chris please?
  - Still love Chris, my climbing boi

Feedback on your feedback: The bad

- **Lee**
  - Speak. More. S l o w l y
  - You over-explain the simple things
  - Always late, sort it out mate
  - Owe us 20 minutes from MAS2602
  - Go faster plz
- **Chris**
  - Where's Chris? Like him
  - Any more of Chris please?
  - Still love Chris, my climbing boi

Feedback on your feedback: The bad

- **Lee**
  - Speak. More. S l o w l y
  - You over-explain the simple things
  - Always late, sort it out mate
  - Owe us 20 minutes from MAS2602
  - Go faster plz
- **Chris**
  - Where's Chris? Like him
  - Any more of Chris please?
  - Still love Chris, my climbing boi

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we're copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it's in here
- Bayesian sux frequentist 4 lyf
- You're much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we're copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it's in here
- Bayesian sux frequentist 4 lyf
- You're much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we're copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it's in here
- Bayesian sux frequentist 4 lyf
- You're much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we're copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it's in here
- Bayesian sux frequentist 4 lyf
- You're much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we're copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it's in here
- Bayesian sux frequentist 4 lyf
- You're much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we're copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it's in here
- Bayesian sux frequentist 4 lyf
- You're much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we’re copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it’s in here
- Bayesian sux frequentist 4 lyf
- You’re much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Feedback on your feedback: The ugly

- Was really disappointed to learn that Lee was teaching us again
- Please shut up while we're copying down
- Engage us more and give us more breaks
- Less R code, it makes me angry that it's in here
- Bayesian sux frequentist 4 lyf
- You're much better at this, you were rubbish at R
- Too much irrelevant talking
- Not very engaging,... he just talks at us,...

Taking stock

- **Posterior $\propto$ Prior $\times$ Likelihood**
  - Always look for a gamma or a beta distribution first
  - Notation: $\pi(\theta)$ is just notation to represent our PDF for θ; we usually use $f(x)$ to represent the PDF of the data
  - Prior distribution should not have x's in it – the random variable is the parameter!
  - Notation: $E[\theta]$ = prior mean, $E[\theta|\boldsymbol{x}]$ = posterior mean
  - Notation: $f(x|\theta)$ is just our likelihood – form the product over the PDF for each observation, if you have multiple observations (careful with the Binomial!)
  - Intepretation: Compare prior/posterior means and variances; is the posterior closer to the data, or the prior?

- **Posterior $\propto$ Prior $\times$ Likelihood**
  - Always look for a gamma or a beta distribution first
  - Notation: $\pi(\theta)$ is just notation to represent our PDF for θ; we usually use $f(x)$ to represent the PDF of the data
  - Prior distribution should not have x's in it – the random variable is the parameter!
  - Notation: $E[\theta]$ = prior mean, $E[\theta|\boldsymbol{x}]$ = posterior mean
  - Notation: $f(x|\theta)$ is just our likelihood – form the product over the PDF for each observation, if you have multiple observations (careful with the Binomial!)
  - Intepretation: Compare prior/posterior means and variances; is the posterior closer to the data, or the prior?

Taking stock

- **Posterior $\propto$ Prior $\times$ Likelihood**
  - Always look for a gamma or a beta distribution first
  - Notation: $\pi(\theta)$ is just notation to represent our PDF for θ; we usually use $f(x)$ to represent the PDF of the data
  - Prior distribution should not have x's in it – the random variable is the parameter!
  - Notation: $E[\theta]$ = prior mean, $E[\theta|\boldsymbol{x}]$ = posterior mean
  - Notation: $f(x|\theta)$ is just our likelihood – form the product over the PDF for each observation, <u>if</u> you have multiple observations (careful with the Binomial!)
  - Intepretation: Compare prior/posterior means and variances; is the posterior closer to the data, or the prior?

Taking stock

- **Posterior $\propto$ Prior $\times$ Likelihood**
  - Always look for a gamma or a beta distribution first
  - Notation: $\pi(\theta)$ is just notation to represent our PDF for θ; we usually use $f(x)$ to represent the PDF of the data
  - Prior distribution should not have x's in it – the random variable is the parameter!
  - Notation: $E[\theta]$ = prior mean, $E[\theta|\boldsymbol{x}]$ = posterior mean
  - Notation: $f(x|\theta)$ is just our likelihood – form the product over the PDF for each observation, if you have multiple observations (careful with the Binomial!)
  - Intepretation: Compare prior/posterior means and variances; is the posterior closer to the data, or the prior?

Taking stock

- **Posterior $\propto$ Prior $\times$ Likelihood**
  - Always look for a gamma or a beta distribution first
  - Notation: $\pi(\theta)$ is just notation to represent our PDF for θ; we usually use $f(x)$ to represent the PDF of the data
  - Prior distribution should not have x's in it – the random variable is the parameter!
  - Notation: $E[\theta] =$ prior mean, $E[\theta|\boldsymbol{x}] =$ posterior mean
  - Notation: $f(x|\theta)$ is just our likelihood – form the product over the PDF for each observation, if you have multiple observations (careful with the Binomial!)
  - Intepretation: Compare prior/posterior means and variances; is the posterior closer to the data, or the prior?

Taking stock

- **Posterior $\propto$ Prior $\times$ Likelihood**
  - Always look for a gamma or a beta distribution first
  - Notation: $\pi(\theta)$ is just notation to represent our PDF for θ; we usually use $f(x)$ to represent the PDF of the data
  - Prior distribution should not have x's in it – the random variable is the parameter!
  - Notation: $E[\theta]$ = prior mean, $E[\theta|\boldsymbol{x}]$ = posterior mean
  - Notation: $f(x|\theta)$ is just our likelihood – form the product over the PDF for each observation, if you have multiple observations (careful with the Binomial!)
  - Intepretation: Compare prior/posterior means and variances; is the posterior closer to the data, or the prior?

Taking stock

- **Posterior $\propto$ Prior $\times$ Likelihood**
  - Always look for a gamma or a beta distribution first
  - Notation: $\pi(\theta)$ is just notation to represent our PDF for θ; we usually use $f(x)$ to represent the PDF of the data
  - Prior distribution should not have x's in it – the random variable is the parameter!
  - Notation: $E[\theta] =$ prior mean, $E[\theta|\boldsymbol{x}] =$ posterior mean
  - Notation: $f(x|\theta)$ is just our likelihood – form the product over the PDF for each observation, if you have multiple observations (careful with the Binomial!)
  - Intepretation: Compare prior/posterior means and variances; is the posterior closer to the data, or the prior?

Taking stock

- **Constructing priors**
  - For a gamma or a beta prior: two parameters, so two bits of information needed (e.g. mean/variance? mode/probability?)
  - Reference data: linear regression for the prior mean?
  - Historical records: Trial roulette method?
  - Bisection method (today)

- **Sufficiency**
  - Posterior using the likelihood for t is identical to that using the full dataset $\boldsymbol{x}$
  - Much more efficient for "Bayesian updating" – e.g. $T = \sum X_i$

Taking stock

- **Constructing priors**
  - For a gamma or a beta prior: two parameters, so two bits of information needed (e.g. mean/variance? mode/probability?)
  - Reference data: linear regression for the prior mean?
  - Historical records: Trial roulette method?
  - Bisection method (today)
- **Sufficiency**
  - Posterior using the likelihood for t is identical to that using the full dataset $\boldsymbol{x}$
  - Much more efficient for “Bayesian updating" – e.g. $T = \sum X_i$

Example 3.4: Bisection Method

Over the past 15 years there has been considerable scientific interest in the rate of retreat, θ (feet per year), of glaciers in Greenland (as discussed in the recent *Frozen Planet* series shown on the BBC).

Indeed, this has often been used as an indicator of **global warming**.

We are interested in eliciting a suitable prior distribution for θ for the *Zachariae Isstrøm* glacier in Greenland.

Example 3.4: Bisection Method

Records from an expert glaciologist show that glaciers in Greenland have been retreating at a rate of between 0 and 70 feet per year since 1995.

We will use these values as the lower and upper limits for θ, respectively. We now attempt to elicit the **median** and **quartiles** for θ from the glaciologist.

Example 3.4: Bisection Method

Step 1: Eliciting the median

Ask the expert to provide a value m (in the range of permissable values for θ), such that

$$\Pr(\text{minimum} < \theta < m) = \Pr(m < \theta < \text{maximum}) = \frac{1}{2}.$$

- The value m **bisects** the range for θ into two halves of equal probability
- If the expert is "statistically aware", it might be possible to ask them for their **median** for θ
- Otherwise, m might be the value that the expert believes θ is most likely to take.

Example 3.4: Bisection Method

Step 1: Eliciting the median

Ask the expert to provide a value m (in the range of permissable values for θ), such that

$$\Pr(\text{minimum} < \theta < m) = \Pr(m < \theta < \text{maximum}) = \frac{1}{2}.$$

- The value m **bisects** the range for θ into two halves of equal probability
- If the expert is "statistically aware", it might be possible to ask them for their **median** for θ
- Otherwise, m might be the value that the expert believes θ is most likely to take.

Step 1: Eliciting the median

Ask the expert to provide a value m (in the range of permissable values for θ), such that

$$\Pr(\text{minimum} < \theta < m) = \Pr(m < \theta < \text{maximum}) = \frac{1}{2}.$$

- The value m **bisects** the range for θ into two halves of equal probability
- If the expert is "statistically aware", it might be possible to ask them for their **median** for θ
- Otherwise, m might be the value that the expert believes θ is most likely to take.

Example 3.4: Bisection Method

Step 1: Eliciting the median

Ask the expert to provide a value m (in the range of permissable values for θ), such that

$$\Pr(\text{minimum} < \theta < m) = \Pr(m < \theta < \text{maximum}) = \frac{1}{2}.$$

- The value m **bisects** the range for θ into two halves of equal probability
- If the expert is "statistically aware", it might be possible to ask them for their **median** for θ
- Otherwise, m might be the value that the expert believes θ is most likely to take.

Example 3.4: Bisection Method

Step 1: Eliciting the median from the glaciologist

- Glaciers in Greenland have been retreating at a rate of between 0 and 70 feet per year since 1995, depending on how far north the glacier is. Thus, we will say that $\theta \in (0, 70)$.
- The *Zachariae Isstrøm* glacier lies in quite a northerly location, so is not quite so prone to rapid retreat.
- The glaciologist specifies that $m = 24$ might be suitable for bisecting the range for θ – notice how m is closer to the lower bound than the upper.

Example 3.4: Bisection Method

Step 1: Eliciting the median from the glaciologist

- Glaciers in Greenland have been retreating at a rate of between 0 and 70 feet per year since 1995, depending on how far north the glacier is. Thus, we will say that $\theta \in (0, 70)$.

- The *Zachariae Isstrøm* glacier lies in quite a northerly location, so is not quite so prone to rapid retreat.

- The glaciologist specifies that $m = 24$ might be suitable for bisecting the range for θ – notice how m is closer to the lower bound than the upper.

Example 3.4: Bisection Method

Step 1: Eliciting the median from the glaciologist

- Glaciers in Greenland have been retreating at a rate of between 0 and 70 feet per year since 1995, depending on how far north the glacier is. Thus, we will say that $\theta \in (0, 70)$.
- The *Zachariae Isstrøm* glacier lies in quite a northerly location, so is not quite so prone to rapid retreat.
- The glaciologist specifies that $m = 24$ might be suitable for bisecting the range for θ – notice how m is closer to the lower bound than the upper.

Example 3.4: Bisection Method

Step 1: Eliciting the median from the glaciologist

- Glaciers in Greenland have been retreating at a rate of between 0 and 70 feet per year since 1995, depending on how far north the glacier is. Thus, we will say that $\theta \in (0, 70)$.
- The *Zachariae Isstrøm* glacier lies in quite a northerly location, so is not quite so prone to rapid retreat.
- The glaciologist specifies that $m = 24$ might be suitable for bisecting the range for θ – notice how m is closer to the lower bound than the upper.

Step 2: Eliciting the lower quartile

Ask the expert to provide a value ℓ such that

$$\text{Pr}(\text{minimum} < \theta < \ell) = \text{Pr}(\ell < \theta < m),$$

i.e. ℓ **bisects** the lower half of the range for θ.

- This can be more tricky for the expert to do – it's not quite so intuitive a task.
- If the expert struggles, help him/her a bit:
  - Split the lower half into two, and ask them in which part θ is most likely to occur
  - Then ℓ should probably lie in the part which is more likely to occur
- Note that the more certain the expert is, the closer ℓ will be to m

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile

Ask the expert to provide a value ℓ such that

$$\text{Pr}(\text{minimum} < \theta < \ell) = \text{Pr}(\ell < \theta < m),$$

i.e. ℓ **bisects** the lower half of the range for θ.

- This can be more tricky for the expert to do – it's not quite so intuitive a task.
- If the expert struggles, help him/her a bit:
  - Split the lower half into two, and ask them in which part θ is most likely to occur
  - Then ℓ should probably lie in the part which is more likely to occur
- Note that the more certain the expert is, the closer ℓ will be to m

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile

Ask the expert to provide a value ℓ such that

$$\Pr(\text{minimum} < \theta < \ell) = \Pr(\ell < \theta < m),$$

i.e. ℓ **bisects** the lower half of the range for θ.

- This can be more tricky for the expert to do – it's not quite so intuitive a task.
- If the expert struggles, help him/her a bit:
  - Split the lower half into two, and ask them in which part θ is most likely to occur
  - Then ℓ should probably lie in the part which is more likely to occur
- Note that the more certain the expert is, the closer ℓ will be to m

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile

Ask the expert to provide a value ℓ such that

$$\Pr(\text{minimum} < \theta < \ell) = \Pr(\ell < \theta < m),$$

i.e. ℓ **bisects** the lower half of the range for θ.

- This can be more tricky for the expert to do – it's not quite so intuitive a task.
- If the expert struggles, help him/her a bit:
  - Split the lower half into two, and ask them in which part θ is most likely to occur
  - Then ℓ should probably lie in the part which is more likely to occur
- Note that the more certain the expert is, the closer ℓ will be to m

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile

Ask the expert to provide a value ℓ such that

$$\Pr(\text{minimum} < \theta < \ell) = \Pr(\ell < \theta < m),$$

i.e. ℓ **bisects** the lower half of the range for θ.

- This can be more tricky for the expert to do – it's not quite so intuitive a task.
- If the expert struggles, help him/her a bit:
  - Split the lower half into two, and ask them in which part θ is most likely to occur
  - Then ℓ should probably lie in the part which is more likely to occur
- Note that the more certain the expert is, the closer ℓ will be to m

Step 2: Eliciting the lower quartile

Ask the expert to provide a value ℓ such that

$$\text{Pr}(\text{minimum} < \theta < \ell) = \text{Pr}(\ell < \theta < m),$$

i.e. ℓ **bisects** the lower half of the range for θ.

- This can be more tricky for the expert to do – it's not quite so intuitive a task.
- If the expert struggles, help him/her a bit:
  - Split the lower half into two, and ask them in which part θ is most likely to occur
  - Then ℓ should probably lie in the part which is more likely to occur
- Note that the more certain the expert is, the closer ℓ will be to m

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile from the glaciologist

The glaciologist found this task a bit more difficult...

- We ask the expert whether $[0, 12]$ or $[12, 24]$ is more likely
- The expert is fairly sure about $m = 24$, so says $[12, 24]$ is more likely for θ than $[0, 12]$
  - Areas further north than the *Zachariae Isstrøm* glacier have much slower rates of retreat
  - Only the most northerly glaciers have zero retreat
- Focussing on $[12, 24]$, the glaciologist settles on $\ell = 19$.

Step 2: Eliciting the lower quartile from the glaciologist

The glaciologist found this task a bit more difficult...

- We ask the expert whether $[0, 12]$ or $[12, 24]$ is more likely
- The expert is fairly sure about $m = 24$, so says $[12, 24]$ is more likely for θ than $[0, 12]$
  - Areas further north than the *Zachariae Isstrøm* glacier have much slower rates of retreat
  - Only the most northerly glaciers have zero retreat
- Focussing on $[12, 24]$, the glaciologist settles on $\ell = 19$.

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile from the glaciologist

The glaciologist found this task a bit more difficult...

- We ask the expert whether $[0, 12]$ or $[12, 24]$ is more likely
- The expert is fairly sure about $m = 24$, so says $[12, 24]$ is more likely for θ than $[0, 12]$
  - Areas further north than the *Zachariae Isstrøm* glacier have much slower rates of retreat
  - Only the most northerly glaciers have zero retreat
- Focussing on $[12, 24]$, the glaciologist settles on $\ell = 19$.

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile from the glaciologist

The glaciologist found this task a bit more difficult...

- We ask the expert whether $[0, 12]$ or $[12, 24]$ is more likely
- The expert is fairly sure about $m = 24$, so says $[12, 24]$ is more likely for θ than $[0, 12]$
  - Areas further north than the *Zachariae Isstrøm* glacier have much slower rates of retreat
  - Only the most northerly glaciers have zero retreat
- Focussing on $[12, 24]$, the glaciologist settles on $\ell = 19$.

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile from the glaciologist

The glaciologist found this task a bit more difficult...

- We ask the expert whether $[0, 12]$ or $[12, 24]$ is more likely
- The expert is fairly sure about $m = 24$, so says $[12, 24]$ is more likely for θ than $[0, 12]$
  - Areas further north than the *Zachariae Isstrøm* glacier have much slower rates of retreat
  - Only the most northerly glaciers have zero retreat
- Focussing on $[12, 24]$, the glaciologist settles on $\ell = 19$.

Example 3.4: Bisection Method

Step 2: Eliciting the lower quartile from the glaciologist

The glaciologist found this task a bit more difficult...

- We ask the expert whether $[0, 12]$ or $[12, 24]$ is more likely
- The expert is fairly sure about $m = 24$, so says $[12, 24]$ is more likely for θ than $[0, 12]$
  - Areas further north than the *Zachariae Isstrøm* glacier have much slower rates of retreat
  - Only the most northerly glaciers have zero retreat
- Focussing on $[12, 24]$, the glaciologist settles on $\ell = 19$.

Example 3.4: Bisection Method

Step 3: Eliciting the upper quartile

Same sort of process for u as for ℓ.

Step 3: Eliciting the lower quartile from the glaciologist

Using the same process as for ℓ, the glaciologist settles on $u = 30$.

Step 3: Eliciting the upper quartile

Same sort of process for u as for ℓ.

Step 3: Eliciting the lower quartile from the glaciologist

Using the same process as for ℓ, the glaciologist settles on $u = 30$.

Example 3.4: Bisection Method

Step 4: Reflection

Based on the elicited values for ℓ, m and u, the expert should be asked to **reflect**, i.e., does the following seem plausible:

$$\Pr(\min < \theta < \ell) = \Pr(\ell < \theta < m) = \Pr(m < \theta < u) = \Pr(u < \theta < \max)?$$

Step 4: Let the glaciologist reflect

The glaciologist seems fine with this!

Step 4: Reflection

Based on the elicited values for ℓ, m and u, the expert should be asked to **reflect**, i.e., does the following seem plausible:

$$\Pr(\min < \theta < \ell) = \Pr(\ell < \theta < m) = \Pr(m < \theta < u) = \Pr(u < \theta < \max)?$$

Step 4: Let the glaciologist reflect

The glaciologist seems fine with this!

Example 3.4: Bisection Method

Step 5: Fit a parametric distribution to these judgements

We can use the *MATCH* software for this.

Step 5: Fitting a parametric distribution to the glaciologist's judgements

Doing this in *MATCH* gives $\theta \sim Ga(9, 0.36)$.

Step 6: Feedback and refinement

- From the fitted parametric distribution, provide the expert with some summaries: for example, tail probabilities.
- See if these tail probabilities correspond closely to the expert's intuition!
- If not, perhaps ask the expert to refine their choices of ℓ or m or u, or perhaps all three!

Step 6: Feedback and refinement

- From the fitted parametric distribution, provide the expert with some summaries: for example, tail probabilities.
- See if these tail probabilities correspond closely to the expert's intuition!
- If not, perhaps ask the expert to refine their choices of ℓ or m or u, or perhaps all three!

Step 6: Feedback and refinement

- From the fitted parametric distribution, provide the expert with some summaries: for example, tail probabilities.
- See if these tail probabilities correspond closely to the expert's intuition!
- If not, perhaps ask the expert to refine their choices of ℓ or m or u, or perhaps all three!

Example 3.4: Bisection Method

Step 6: Feedback to the glaciologist, and possible refinement

- We show the glaciologist the plot of the $Ga(9, 0.36)$ density. Does this look OK? Yes!
- Now feedback some specific properties:
  - The 1%–ile and 99%–iles are about 10 feet and 48 feet, respectively. This means that

    $$\Pr(\theta < 10) = \Pr(\theta > 48) = 0.01, \text{ or once in a hundred years.}$$

  - Does this seem OK?
  - The glaciologist thinks this is "imaginable"...
- No refinement needed here!

Example 3.4: Bisection Method

Step 6: Feedback to the glaciologist, and possible refinement

- We show the glaciologist the plot of the $Ga(9, 0.36)$ density. Does this look OK? Yes!
- Now feedback some specific properties:
  - The 1%–ile and 99%–iles are about 10 feet and 48 feet, respectively. This means that

    $\Pr(\theta < 10) = \Pr(\theta > 48) = 0.01,$ or once in a hundred years.
  - Does this seem OK?
  - The glaciologist thinks this is “imaginable”...
- No refinement needed here!

Example 3.4: Bisection Method

Step 6: Feedback to the glaciologist, and possible refinement

- We show the glaciologist the plot of the $Ga(9, 0.36)$ density. Does this look OK? Yes!
- Now feedback some specific properties:
  - The 1%–ile and 99%–iles are about 10 feet and 48 feet, respectively. This means that

$$\Pr(\theta < 10) = \Pr(\theta > 48) = 0.01, \text{or once in a hundred years.}$$

  - Does this seem OK?
  - The glaciologist thinks this is "imaginable"...
- No refinement needed here!

Example 3.4: Bisection Method

Step 6: Feedback to the glaciologist, and possible refinement

- We show the glaciologist the plot of the $Ga(9, 0.36)$ density. Does this look OK? Yes!
- Now feedback some specific properties:
  - The 1%–ile and 99%–iles are about 10 feet and 48 feet, respectively. This means that

    $$\Pr(\theta < 10) = \Pr(\theta > 48) = 0.01,$$ or once in a hundred years.
  - Does this seem OK?
  - The glaciologist thinks this is "imaginable"...
- No refinement needed here!

Example 3.4: Bisection Method

Step 6: Feedback to the glaciologist, and possible refinement

- We show the glaciologist the plot of the $Ga(9, 0.36)$ density. Does this look OK? Yes!
- Now feedback some specific properties:
  - The 1%–ile and 99%–iles are about 10 feet and 48 feet, respectively. This means that

    $\Pr(\theta < 10) = \Pr(\theta > 48) = 0.01,$ or once in a hundred years.
  - Does this seem OK?
  - The glaciologist thinks this is “imaginable”...
- No refinement needed here!

Example 3.4: Bisection Method

Step 6: Feedback to the glaciologist, and possible refinement

- We show the glaciologist the plot of the $Ga(9, 0.36)$ density. Does this look OK? Yes!
- Now feedback some specific properties:
  - The 1%–ile and 99%–iles are about 10 feet and 48 feet, respectively. This means that

    $$\Pr(\theta < 10) = \Pr(\theta > 48) = 0.01, \text{or once in a hundred years.}$$

  - Does this seem OK?
  - The glaciologist thinks this is "imaginable"...
- No refinement needed here!

Example 3.5

Let Y be the retreat, in feet, of the *Zachariae Isstrøm* glacier. A *Pareto* distribution with rate θ is often used to model such geophysical activity, with probability density function

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

(a) Obtain the likelihood function for θ given the parameter κ and some observed data $y_1, y_2, \ldots, y_n$ (independent).

(b) Suppose we observe a retreat of 20 feet at the *Zachariae Isstrøm* glacier in 2012. Write down the likelihood function for θ.

(c) Using the elicted prior for the rate of retreat we obtained from the expert glaciologist in Example 3.4, and assuming κ is known to be 12, obtain the posterior distribution $\pi(\theta|y_1 = 20)$.

We have

$$f(\boldsymbol{y}|\theta,\kappa) = \theta\kappa^\theta y_1^{-(\theta+1)} \times \cdots \times \theta\kappa^\theta y_n^{-(\theta+1)}$$

$$= \theta^n \kappa^{n\theta} \prod_{i=1}^{n} y_i^{-(\theta+1)}. \qquad (3.5)$$

Solution to Example 3.5(a) (1/1)

We have

$$f(\boldsymbol{y}|\theta,\kappa) = \theta\kappa^{\theta} y_1^{-(\theta+1)} \times \cdots \times \theta\kappa^{\theta} y_n^{-(\theta+1)}$$

$$= \theta^n \kappa^{n\theta} \prod_{i=1}^{n} y_i^{-(\theta+1)}. \tag{3.5}$$

Solution to Example 3.5(b) (1/1)

We simply substitute $n = 1$ and $y_1 = 20$ into Equation (3.5), giving

$$f(y_1 = 20|\theta, \kappa) = \theta\kappa^{\theta}20^{-(\theta+1)}.$$

Using Bayes' Theorem, and following the examples in Chapter 2, we know that

$$\pi(\theta|y_1 = 20) \propto \pi(\theta) \times f(y_1 = 20|\theta, \kappa).$$

Recall from **Example 3.4** that our elicited prior for θ is $Ga(9, 0.36)$, which has density

$$\pi(\theta) = \frac{0.36^9 \theta^8 e^{-0.36\theta}}{\Gamma(9)}.$$

Solution to Example 3.5(c) (1/3)

Using Bayes' Theorem, and following the examples in Chapter 2, we know that

$$\pi(\theta|y_1 = 20) \propto \pi(\theta) \times f(y_1 = 20|\theta, \kappa).$$

Recall from **Example 3.4** that our elicited prior for θ is $Ga(9, 0.36)$, which has density

$$\pi(\theta) = \frac{0.36^9 \theta^8 e^{-0.36\theta}}{\Gamma(9)}.$$

Solution to Example 3.5(c) (1/3)

Using Bayes' Theorem, and following the examples in Chapter 2, we know that

$$\pi(\theta|y_1 = 20) \propto \pi(\theta) \times f(y_1 = 20|\theta, \kappa).$$

Recall from **Example 3.4** that our elicited prior for θ is $Ga(9, 0.36)$, which has density

$$\pi(\theta) = \frac{0.36^9 \theta^8 e^{-0.36\theta}}{\Gamma(9)}.$$

Solution to Example 3.5(c) (1/3)

Using Bayes' Theorem, and following the examples in Chapter 2, we know that

$$\pi(\theta|y_1 = 20) \propto \pi(\theta) \times f(y_1 = 20|\theta, \kappa).$$

Recall from **Example 3.4** that our elicited prior for θ is $Ga(9, 0.36)$, which has density

$$\pi(\theta) = \frac{0.36^9 \theta^8 e^{-0.36\theta}}{\Gamma(9)}.$$

Solution to Example 3.5(c) (1/3)

Using Bayes' Theorem, and following the examples in Chapter 2, we know that

$$\pi(\theta|y_1 = 20) \propto \pi(\theta) \times f(y_1 = 20|\theta, \kappa).$$

Recall from **Example 3.4** that our elicited prior for θ is $Ga(9, 0.36)$, which has density

$$\pi(\theta) = \frac{0.36^9 \theta^8 e^{-0.36\theta}}{\Gamma(9)}.$$

Solution to Example 3.5(c) (2/3)

Combining this with the likelihood above (and using $\kappa = 12$) gives

$$\pi(\theta|y_1 = 20) = \frac{0.36^9\theta^8 e^{-0.36\theta}}{\Gamma(9)} \times \theta 12^\theta 20^{-(\theta+1)}$$

$$\propto \theta^9 e^{-0.36\theta} 12^\theta 20^{-(\theta+1)}$$

$$\propto \theta^9 e^{-0.36\theta} 12^\theta 20^{-\theta}. \qquad (3.6)$$

Solution to Example 3.5(c) (2/3)

Combining this with the likelihood above (and using $\kappa = 12$) gives

$$\pi(\theta|y_1 = 20) = \frac{0.36^9\theta^8 e^{-0.36\theta}}{\Gamma(9)} \times \theta 12^\theta 20^{-(\theta+1)}$$

$$\propto \theta^9 e^{-0.36\theta} 12^\theta 20^{-(\theta+1)}$$

$$\propto \theta^9 e^{-0.36\theta} 12^\theta 20^{-\theta}. \tag{3.6}$$

Solution to Example 3.5(c) (3/3)

Now consider the term $12^{\theta}20^{-\theta}$. Taking logs, we get

$$\theta \ln 12 - \theta \ln 20 = (\ln 12 - \ln 20)\theta;$$

exponentiating to 're–balance', you should see that

$$12^{\theta}20^{-\theta} = e^{(\ln 12 - \ln 20)\theta}.$$

Substituting back into (3.6) gives

$$\begin{aligned} \pi(\theta|y_1 = 20) &\propto \theta^9 e^{-0.36\theta} e^{(\ln 12 - \ln 20)\theta} \qquad \text{i.e.} \\ &\propto \theta^9 e^{-0.36\theta + (\ln 12 - \ln 20)\theta} \\ &\propto \theta^9 e^{-0.87\theta}, \end{aligned}$$

i.e. $\theta|y_1 = 20 \sim Ga(10, 0.87)$.

Solution to Example 3.5(c) (3/3)

Now consider the term $12^{\theta}20^{-\theta}$. Taking logs, we get

$$\theta \ln 12 - \theta \ln 20 = (\ln 12 - \ln 20)\theta;$$

exponentiating to 're–balance', you should see that

$$12^{\theta}20^{-\theta} = e^{(\ln 12 - \ln 20)\theta}.$$

Substituting back into (3.6) gives

$$\begin{aligned} \pi(\theta|y_1 = 20) &\propto \theta^9 e^{-0.36\theta} e^{(\ln 12 - \ln 20)\theta} \qquad \text{i.e.} \\ &\propto \theta^9 e^{-0.36\theta + (\ln 12 - \ln 20)\theta} \\ &\propto \theta^9 e^{-0.87\theta}, \end{aligned}$$

i.e. $\theta|y_1 = 20 \sim Ga(10, 0.87)$.

Solution to Example 3.5(c) (3/3)

Now consider the term $12^\theta 20^{-\theta}$. Taking logs, we get

$$\theta \ln 12 - \theta \ln 20 = (\ln 12 - \ln 20)\theta;$$

exponentiating to 're–balance', you should see that

$$12^\theta 20^{-\theta} = e^{(\ln 12 - \ln 20)\theta}.$$

Substituting back into (3.6) gives

$$\begin{aligned} \pi(\theta|y_1 = 20) &\propto \theta^9 e^{-0.36\theta} e^{(\ln 12 - \ln 20)\theta} \qquad \text{i.e.} \\ &\propto \theta^9 e^{-0.36\theta + (\ln 12 - \ln 20)\theta} \\ &\propto \theta^9 e^{-0.87\theta}, \end{aligned}$$

i.e. $\theta|y_1 = 20 \sim Ga(10, 0.87)$.

Solution to Example 3.5(c) (3/3)

Now consider the term $12^{\theta}20^{-\theta}$. Taking logs, we get

$$\theta \ln 12 - \theta \ln 20 = (\ln 12 - \ln 20)\theta;$$

exponentiating to 're–balance', you should see that

$$12^{\theta}20^{-\theta} = e^{(\ln 12 - \ln 20)\theta}.$$

Substituting back into (3.6) gives

$$\begin{aligned} \pi(\theta | y_1 = 20) &\propto \theta^9 e^{-0.36\theta} e^{(\ln 12 - \ln 20)\theta} \qquad \text{i.e.} \\ &\propto \theta^9 e^{-0.36\theta + (\ln 12 - \ln 20)\theta} \\ &\propto \theta^9 e^{-0.87\theta}, \end{aligned}$$

i.e. $\theta | y_1 = 20 \sim Ga(10, 0.87)$.

Solution to Example 3.5(c) (3/3)

Now consider the term $12^{\theta}20^{-\theta}$. Taking logs, we get

$$\theta \ln 12 - \theta \ln 20 = (\ln 12 - \ln 20)\theta;$$

exponentiating to 're–balance', you should see that

$$12^{\theta}20^{-\theta} = e^{(\ln 12 - \ln 20)\theta}.$$

Substituting back into (3.6) gives

$$\begin{aligned} \pi(\theta|y_1 = 20) &\propto \theta^9 e^{-0.36\theta} e^{(\ln 12 - \ln 20)\theta} \qquad \text{i.e.} \\ &\propto \theta^9 e^{-0.36\theta + (\ln 12 - \ln 20)\theta} \\ &\propto \theta^9 e^{-0.87\theta}, \end{aligned}$$

i.e. $\theta|y_1 = 20 \sim Ga(10, 0.87)$.

Solution to Example 3.5(c) (3/3)

Now consider the term $12^{\theta}20^{-\theta}$. Taking logs, we get

$$\theta \ln 12 - \theta \ln 20 = (\ln 12 - \ln 20)\theta;$$

exponentiating to 're–balance', you should see that

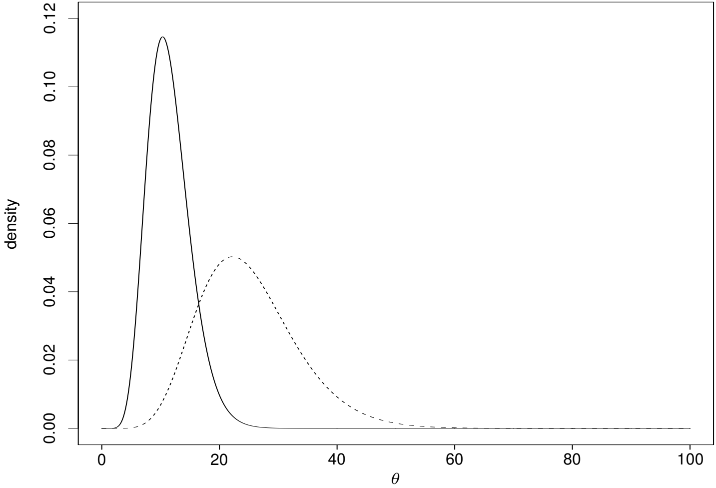
$$12^{\theta}20^{-\theta} = e^{(\ln 12 - \ln 20)\theta}.$$

Substituting back into (3.6) gives

$$\begin{aligned}\pi(\theta|y_1 = 20) &\propto \theta^9 e^{-0.36\theta} e^{(\ln 12 - \ln 20)\theta} \qquad \text{i.e.} \\ &\propto \theta^9 e^{-0.36\theta + (\ln 12 - \ln 20)\theta} \\ &\propto \theta^9 e^{-0.87\theta},\end{aligned}$$

i.e. $\theta|y_1 = 20 \sim Ga(10, 0.87)$.

Example 3.5



Substantial prior information

Definition (Substantial prior information)

We have **substantial prior information** for θ when the prior distribution *dominates* the posterior distribution, that is $\pi(\theta|\boldsymbol{x}) \sim \pi(\theta)$.

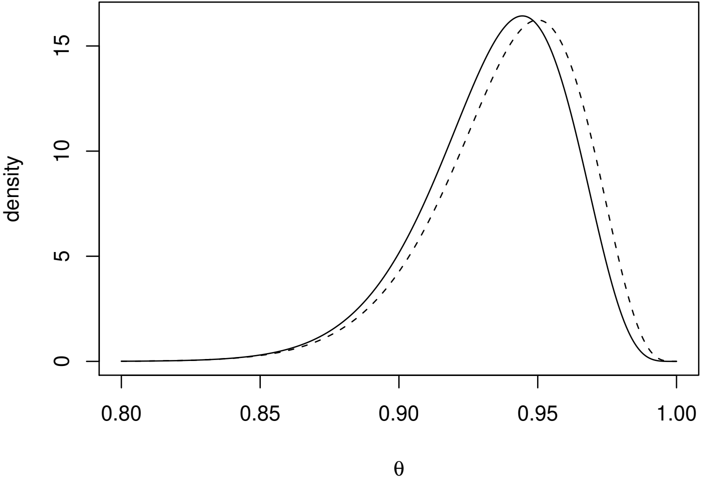
Substantial prior information

An example of substantial prior knowledge was given in **Example 2.2** where a music expert was trying to distinguish between pages from Mozart and Haydn scores.

Figure 3.9 shows the prior and posterior distributions for θ, the probability that the expert makes the correct choice.

Notice the similarity between the prior and posterior distributions. Observing the data has not altered our beliefs about θ very much.

Substantial prior information



Substantial prior information

When we have substantial prior information there can be some difficulties:

1. the **intractability of the mathematics** in deriving the posterior distribution — though with modern computing facilities this is less of a problem,
2. the **practical formulation of the prior distribution** — coherently specifying prior beliefs in the form of a probability distribution is far from straightforward although, as we have seen, this can be attempted using computer software.

Substantial prior information

When we have substantial prior information there can be some difficulties:

1. the **intractability of the mathematics** in deriving the posterior distribution — though with modern computing facilities this is less of a problem,
2. the **practical formulation of the prior distribution** — coherently specifying prior beliefs in the form of a probability distribution is far from straightforward although, as we have seen, this can be attempted using computer software.

Parameter Constraints

[We will come back to this soon... For now, turn to page 73!]

Vague Prior Knowledge/Prior Ignorance

If we have very little or no prior information about the model parameter θ, we must still choose a prior distribution in order to operate Bayes Theorem.

Obviously, it would be sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance.

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

If we have very little or no prior information about the model parameter θ, we must still choose a prior distribution in order to operate Bayes Theorem.

Obviously, it would be sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance.

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

Vague Prior Knowledge/Prior Ignorance

If we have very little or no prior information about the model parameter θ, we must still choose a prior distribution in order to operate Bayes Theorem.

Obviously, it would be sensible to choose a prior distribution which is not concentrated about any particular value, that is, one with a very large variance.

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

Vague Prior Knowledge/Prior Ignorance

An example of vague prior knowledge was given in **Example 2.1** where a possibly biased coin was assessed.

Figure 3.13 shows the prior and posterior distributions for $\theta = \text{Pr(Head)}$.

Notice that the prior and posterior distributions look very different.

In fact, in this example, the posterior distribution is simply a scaled version of the likelihood function – likelihood functions are not usually proper probability (density) functions and so scaling is required to ensure that it integrates to one.

Most of our beliefs about θ have come from observing the data.

Vague Prior Knowledge/Prior Ignorance

An example of vague prior knowledge was given in **Example 2.1** where a possibly biased coin was assessed.

Figure 3.13 shows the prior and posterior distributions for $\theta = \text{Pr(Head)}$.

Notice that the prior and posterior distributions look very different.

In fact, in this example, the posterior distribution is simply a scaled version of the likelihood function – likelihood functions are not usually proper probability (density) functions and so scaling is required to ensure that it integrates to one.

Most of our beliefs about θ have come from observing the data.

Vague Prior Knowledge/Prior Ignorance

An example of vague prior knowledge was given in **Example 2.1** where a possibly biased coin was assessed.

Figure 3.13 shows the prior and posterior distributions for $\theta = \text{Pr(Head)}$.

Notice that the prior and posterior distributions look very different.

In fact, in this example, the posterior distribution is simply a scaled version of the likelihood function – likelihood functions are not usually proper probability (density) functions and so scaling is required to ensure that it integrates to one.

Most of our beliefs about θ have come from observing the data.

Vague Prior Knowledge/Prior Ignorance

An example of vague prior knowledge was given in **Example 2.1** where a possibly biased coin was assessed.

Figure 3.13 shows the prior and posterior distributions for $\theta = \text{Pr(Head)}$.

Notice that the prior and posterior distributions look very different.

In fact, in this example, the posterior distribution is simply a scaled version of the likelihood function – likelihood functions are not usually proper probability (density) functions and so scaling is required to ensure that it integrates to one.

Most of our beliefs about θ have come from observing the data.

Vague Prior Knowledge/Prior Ignorance

An example of vague prior knowledge was given in **Example 2.1** where a possibly biased coin was assessed.

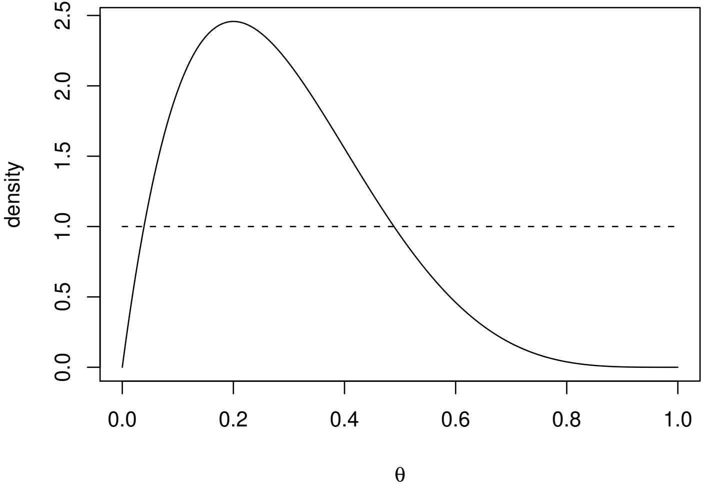
Figure 3.13 shows the prior and posterior distributions for $\theta = \text{Pr(Head)}$.

Notice that the prior and posterior distributions look very different.

In fact, in this example, the posterior distribution is simply a scaled version of the likelihood function – likelihood functions are not usually proper probability (density) functions and so scaling is required to ensure that it integrates to one.

Most of our beliefs about θ have come from observing the data.

Vague Prior Knowledge/Prior Ignorance



Vague prior knowledge

We represent **vague prior knowledge** by using a prior distribution which is conjugate to the model for $\boldsymbol{x}$ and which has "infinite" variance.

Example 3.9

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.

Conjugate prior: Normal distribution. From **Example 2.6**, if $\mu \sim N(b, 1/d)$ then $\mu|\boldsymbol{x} \sim N(B, 1/D)$ where

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

If we now make our prior knowledge vague about μ by letting the prior variance tend to infinity ($d \to 0$), we obtain

$$B \to \bar{x} \qquad \text{and} \qquad D \to n\tau.$$

giving $\mu|\boldsymbol{x} \sim N(\bar{x}, 1/(n\tau))$ posterior distribution. Notice that the posterior mean is the sample mean (the likelihood mode) and that the posterior variance $1/D \to 0$ as $n \to \infty$.

Solution to Example 3.9 (1/1)

Conjugate prior: Normal distribution. From **Example 2.6**, if $\mu \sim N(b, 1/d)$ then $\mu|\boldsymbol{x} \sim N(B, 1/D)$ where

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

If we now make our prior knowledge vague about μ by letting the prior variance tend to infinity ($d \to 0$), we obtain

$$B \to \bar{x} \qquad \text{and} \qquad D \to n\tau.$$

giving $\mu|\boldsymbol{x} \sim N(\bar{x}, 1/(n\tau))$ posterior distribution. Notice that the posterior mean is the sample mean (the likelihood mode) and that the posterior variance $1/D \to 0$ as $n \to \infty$.

Solution to Example 3.9 (1/1)

Conjugate prior: Normal distribution. From **Example 2.6**, if $\mu \sim N(b, 1/d)$ then $\mu|\boldsymbol{x} \sim N(B, 1/D)$ where

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

If we now make our prior knowledge vague about μ by letting the prior variance tend to infinity ($d \to 0$), we obtain

$$B \to \bar{x} \qquad \text{and} \qquad D \to n\tau.$$

giving $\mu|\boldsymbol{x} \sim N(\bar{x}, 1/(n\tau))$ posterior distribution. Notice that the posterior mean is the sample mean (the likelihood mode) and that the posterior variance $1/D \to 0$ as $n \to \infty$.

Solution to Example 3.9 (1/1)

Conjugate prior: Normal distribution. From **Example 2.6**, if $\mu \sim N(b, 1/d)$ then $\mu|\boldsymbol{x} \sim N(B, 1/D)$ where

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

If we now make our prior knowledge vague about μ by letting the prior variance tend to infinity ($d \to 0$), we obtain

$$B \to \bar{x} \qquad \text{and} \qquad D \to n\tau.$$

giving $\mu|\boldsymbol{x} \sim N(\bar{x}, 1/(n\tau))$ posterior distribution. Notice that the posterior mean is the sample mean (the likelihood mode) and that the posterior variance $1/D \to 0$ as $n \to \infty$.

Solution to Example 3.9 (1/1)

Conjugate prior: Normal distribution. From **Example 2.6**, if $\mu \sim N(b, 1/d)$ then $\mu|\boldsymbol{x} \sim N(B, 1/D)$ where

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

If we now make our prior knowledge vague about μ by letting the prior variance tend to infinity ($d \to 0$), we obtain

$$B \to \bar{x} \qquad \text{and} \qquad D \to n\tau.$$

giving $\mu|\boldsymbol{x} \sim N(\bar{x}, 1/(n\tau))$ posterior distribution. Notice that the posterior mean is the sample mean (the likelihood mode) and that the posterior variance $1/D \to 0$ as $n \to \infty$.

Solution to Example 3.9 (1/1)

Conjugate prior: Normal distribution. From **Example 2.6**, if $\mu \sim N(b, 1/d)$ then $\mu|\boldsymbol{x} \sim N(B, 1/D)$ where

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

If we now make our prior knowledge vague about μ by letting the prior variance tend to infinity ($d \to 0$), we obtain

$$B \to \bar{x} \qquad \text{and} \qquad D \to n\tau.$$

giving $\mu|\boldsymbol{x} \sim N(\bar{x}, 1/(n\tau))$ posterior distribution. Notice that the posterior mean is the sample mean (the likelihood mode) and that the posterior variance $1/D \to 0$ as $n \to \infty$.

Example 3.10

Suppose we have a random sample from an exponential distribution, that is, $X_i|\theta \sim Exp(\theta)$, $i = 1, 2, \ldots, n$ (independent).

Determine the posterior distribution assuming a vague prior for θ.

Solution to Example 3.10 (1/1)

Conjugate prior: Gamma distribution. Recall that a $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$.

Rearranging these formulae we obtain

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Clearly $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m).

We have seen how taking a $Ga(g, h)$ prior distribution results in a $Ga(g + n, h + n\bar{x})$ posterior distribution (**Example 2.5**).

Therefore, taking a vague prior distribution will give a $Ga(n, n\bar{x})$ posterior distribution.

Note that the posterior mean is $1/\bar{x}$ (the likelihood mode) and that the posterior variance $1/(n\bar{x}^2) \to 0$ and $n \to \infty$.

Solution to Example 3.10 (1/1)

Conjugate prior: Gamma distribution. Recall that a $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$.

Rearranging these formulae we obtain

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Clearly $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m).

We have seen how taking a $Ga(g, h)$ prior distribution results in a $Ga(g + n, h + n\bar{x})$ posterior distribution (**Example 2.5**).

Therefore, taking a vague prior distribution will give a $Ga(n, n\bar{x})$ posterior distribution.

Note that the posterior mean is $1/\bar{x}$ (the likelihood mode) and that the posterior variance $1/(n\bar{x}^2) \to 0$ and $n \to \infty$.

Solution to Example 3.10 (1/1)

Conjugate prior: Gamma distribution. Recall that a $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$.

Rearranging these formulae we obtain

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Clearly $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m).

We have seen how taking a $Ga(g, h)$ prior distribution results in a $Ga(g + n, h + n\bar{x})$ posterior distribution (**Example 2.5**).

Therefore, taking a vague prior distribution will give a $Ga(n, n\bar{x})$ posterior distribution.

Note that the posterior mean is $1/\bar{x}$ (the likelihood mode) and that the posterior variance $1/(n\bar{x}^2) \to 0$ and $n \to \infty$.

Solution to Example 3.10 (1/1)

Conjugate prior: Gamma distribution. Recall that a $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$.

Rearranging these formulae we obtain

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Clearly $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m).

We have seen how taking a $Ga(g, h)$ prior distribution results in a $Ga(g + n, h + n\bar{x})$ posterior distribution (**Example 2.5**).

Therefore, taking a vague prior distribution will give a $Ga(n, n\bar{x})$ posterior distribution.

Note that the posterior mean is $1/\bar{x}$ (the likelihood mode) and that the posterior variance $1/(n\bar{x}^2) \to 0$ and $n \to \infty$.

Solution to Example 3.10 (1/1)

Conjugate prior: Gamma distribution. Recall that a $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$.

Rearranging these formulae we obtain

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Clearly $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m).

We have seen how taking a $Ga(g, h)$ prior distribution results in a $Ga(g + n, h + n\bar{x})$ posterior distribution (**Example 2.5**).

Therefore, taking a vague prior distribution will give a $Ga(n, n\bar{x})$ posterior distribution.

Note that the posterior mean is $1/\bar{x}$ (the likelihood mode) and that the posterior variance $1/(n\bar{x}^2) \to 0$ and $n \to \infty$.

Solution to Example 3.10 (1/1)

Conjugate prior: Gamma distribution. Recall that a $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$.

Rearranging these formulae we obtain

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Clearly $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m).

We have seen how taking a $Ga(g, h)$ prior distribution results in a $Ga(g + n, h + n\bar{x})$ posterior distribution (**Example 2.5**).

Therefore, taking a vague prior distribution will give a $Ga(n, n\bar{x})$ posterior distribution.

Note that the posterior mean is $1/\bar{x}$ (the likelihood mode) and that the posterior variance $1/(n\bar{x}^2) \to 0$ and $n \to \infty$.

Solution to Example 3.10 (1/1)

Conjugate prior: Gamma distribution. Recall that a $Ga(g, h)$ distribution has mean $m = g/h$ and variance $v = g/h^2$.

Rearranging these formulae we obtain

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Clearly $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m).

We have seen how taking a $Ga(g, h)$ prior distribution results in a $Ga(g + n, h + n\bar{x})$ posterior distribution (**Example 2.5**).

Therefore, taking a vague prior distribution will give a $Ga(n, n\bar{x})$ posterior distribution.

Note that the posterior mean is $1/\bar{x}$ (the likelihood mode) and that the posterior variance $1/(n\bar{x}^2) \to 0$ and $n \to \infty$.

Prior ignorance

We could represent ignorance by the concept "all values of θ are equally likely".

If θ were **discrete** with m possible values then we could assign each value the same probability $1/m$.

However, if θ is **continuous**, we need some limiting argument (from the discrete case).

Prior ignorance

We could represent ignorance by the concept "all values of θ are equally likely".

If θ were **discrete** with m possible values then we could assign each value the same probability $1/m$.

However, if θ is **continuous**, we need some limiting argument (from the discrete case).

Prior ignorance

We could represent ignorance by the concept "all values of θ are equally likely".

If θ were **discrete** with m possible values then we could assign each value the same probability $1/m$.

However, if θ is **continuous**, we need some limiting argument (from the discrete case).

Prior ignorance

Suppose that θ can take values between a and b, where $-\infty < a < b < \infty$.

Letting all (permitted) values of θ be equally likely results in taking a uniform $U(a, b)$ distribution as our prior distribution for θ.

However, if the parameter space is not finite then we cannot do this: there is no such thing as a $U(-\infty, \infty)$ distribution.

Prior ignorance

Suppose that θ can take values between a and b, where $-\infty < a < b < \infty$.

Letting all (permitted) values of θ be equally likely results in taking a uniform $U(a, b)$ distribution as our prior distribution for θ.

However, if the parameter space is not finite then we cannot do this: there is no such thing as a $U(-\infty, \infty)$ distribution.

Prior ignorance

Suppose that θ can take values between *a* and *b*, where $-\infty < a < b < \infty$.

Letting all (permitted) values of θ be equally likely results in taking a uniform $U(a, b)$ distribution as our prior distribution for θ.

However, if the parameter space is not finite then we cannot do this: there is no such thing as a $U(-\infty, \infty)$ distribution.

Prior ignorance

Suppose that θ can take values between a and b, where $-\infty < a < b < \infty$.

Letting all (permitted) values of θ be equally likely results in taking a uniform $U(a, b)$ distribution as our prior distribution for θ.

However, if the parameter space is not finite then we cannot do this: there is no such thing as a $U(-\infty, \infty)$ distribution.

Prior ignorance

Convention suggests that we should use the "**improper**" uniform prior distribution

$$\pi(\theta) = \text{constant.}$$

This distribution is improper because

$$\int_{-\infty}^{\infty} \pi(\theta)\, d\theta$$

is not a convergent integral, let alone equal to one.

Prior ignorance

Convention suggests that we should use the "**improper**" uniform prior distribution

$$\pi(\theta) = \text{constant.}$$

This distribution is improper because

$$\int_{-\infty}^{\infty} \pi(\theta)\, d\theta$$

is not a convergent integral, let alone equal to one.

Prior ignorance

Convention suggests that we should use the "**improper**" uniform prior distribution

$$\pi(\theta) = \text{constant}.$$

This distribution is improper because

$$\int_{-\infty}^{\infty} \pi(\theta)\, d\theta$$

is not a convergent integral, let alone equal to one.

Prior ignorance

We have a similar problem if θ takes positive values — we cannot use a $U(0, \infty)$ prior distribution.

Now if $\theta \in (0, \infty)$ then $\phi = \log \theta \in (-\infty, \infty)$, and so we could use an "improper" uniform prior for ϕ: $\pi(\phi) = constant$.

In turn, this induces a distribution on θ. Recall the result from **Distribution Theory**:

Fact (Distribution of a transformation)

Suppose that X is a random variable with probability density function $f_X(x)$. If g is a bijective (1–1) function then the random variable $Y = g(X)$ has probability density function

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|. \qquad (3.7)$$

Prior ignorance

We have a similar problem if θ takes positive values — we cannot use a $U(0, \infty)$ prior distribution.

Now if $\theta \in (0, \infty)$ then $\phi = \log \theta \in (-\infty, \infty)$, and so we could use an "improper" uniform prior for ϕ: $\pi(\phi) = \textit{constant}$.

In turn, this induces a distribution on θ. Recall the result from **Distribution Theory**:

Fact (Distribution of a transformation)

Suppose that X is a random variable with probability density function $f_X(x)$. If g is a bijective (1–1) function then the random variable $Y = g(X)$ has probability density function

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|. \tag{3.7}$$

Prior ignorance

We have a similar problem if θ takes positive values — we cannot use a $U(0, \infty)$ prior distribution.

Now if $\theta \in (0, \infty)$ then $\phi = \log \theta \in (-\infty, \infty)$, and so we could use an "improper" uniform prior for ϕ: $\pi(\phi) =$ *constant*.

In turn, this induces a distribution on θ. Recall the result from **Distribution Theory**:

Fact (Distribution of a transformation)

Suppose that X is a random variable with probability density function $f_X(x)$. If g is a bijective (1–1) function then the random variable $Y = g(X)$ has probability density function

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|. \tag{3.7}$$

Prior ignorance

Applying this result to $\theta = e^{\phi}$ gives

$$\begin{aligned}
\pi_\theta(\theta) &= \pi_\phi(\log\theta)\left|\frac{d}{d\theta}\,\log\theta\right|, \qquad \theta > 0 \\
&= \textit{constant} \times \left|\frac{1}{\theta}\right|, \qquad \theta > 0 \\
&\propto \frac{1}{\theta}, \qquad \theta > 0.
\end{aligned}$$

This too is an improper distribution.

Prior ignorance

There is a drawback of using uniform or improper priors to represent prior ignorance: if we are "ignorant" about θ then we are also "ignorant" about any function of θ, for example, about $\phi_1 = \theta^3$, $\phi_2 = e^\theta$, $\phi_3 = 1/\theta, \ldots$.

Is it **possible** to choose a distribution where we are ignorant about all these functions of θ?

If not, on which function of θ should we place the uniform/improper prior distribution?

Prior ignorance

There is a drawback of using uniform or improper priors to represent prior ignorance: if we are "ignorant" about θ then we are also "ignorant" about any function of θ, for example, about $\phi_1 = \theta^3$, $\phi_2 = e^\theta$, $\phi_3 = 1/\theta, \ldots$.

Is it **possible** to choose a distribution where we are ignorant about all these functions of θ?

If not, on which function of θ should we place the uniform/improper prior distribution?

Prior ignorance

There is a drawback of using uniform or improper priors to represent prior ignorance: if we are "ignorant" about θ then we are also "ignorant" about any function of θ, for example, about $\phi_1 = \theta^3$, $\phi_2 = e^\theta$, $\phi_3 = 1/\theta$, ...

Is it **possible** to choose a distribution where we are ignorant about all these functions of θ?

If not, on which function of θ should we place the uniform/improper prior distribution?

Prior ignorance

- There is no distribution which can represent ignorance for all functions of θ
- Assigning an ignorance prior to ϕ means we do not have an ignorance prior for e^{ϕ}

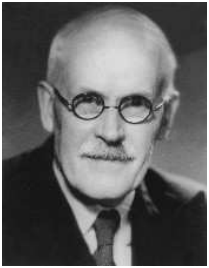
A solution to problems of this type was suggested by **Sir Harold Jeffreys**.

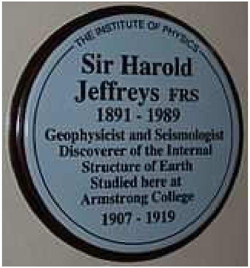
Prior ignorance

- There is no distribution which can represent ignorance for all functions of θ
- Assigning an ignorance prior to ϕ means we do not have an ignorance prior for e^{ϕ}

A solution to problems of this type was suggested by **Sir Harold Jeffreys**.

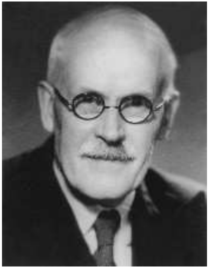
Sir Harold Jeffreys, FRS

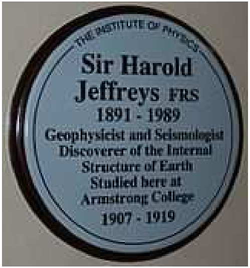




- April 1891 – March 1989
- Mathematician, **Statistician**, Geophysicist, Astronomer
- Studied at **Armstrong College**, **Durham**, now Newcastle University (see plaque)
- Seminal book – **Theory of Probability** – revived the Bayesian view of probability
- Married a Physicist – Bertha Swirles – together they wrote **Methods of Mathematical Physics**

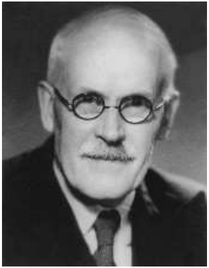
Sir Harold Jeffreys, FRS

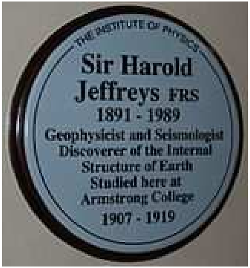




- April 1891 – March 1989
- Mathematician, **Statistician**, Geophysicist, Astronomer
- Studied at **Armstrong College**, **Durham**, now Newcastle University (see plaque)
- Seminal book – **Theory of Probability** – revived the Bayesian view of probability
- Married a Physicist – Bertha Swirles – together they wrote **Methods of Mathematical Physics**

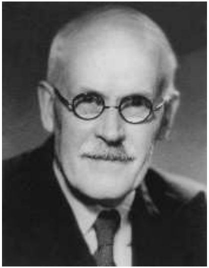
Sir Harold Jeffreys, FRS

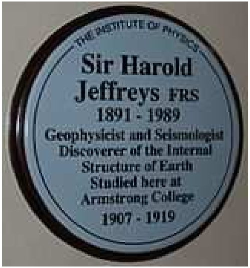




- April 1891 – March 1989
- Mathematician, **Statistician**, Geophysicist, Astronomer
- Studied at **Armstrong College**, **Durham**, now Newcastle University (see plaque)
- Seminal book – **Theory of Probability** – revived the Bayesian view of probability
- Married a Physicist – Bertha Swirles – together they wrote **Methods of Mathematical Physics**

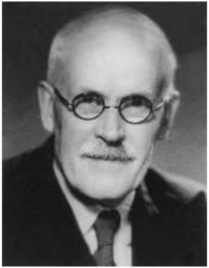
Sir Harold Jeffreys, FRS

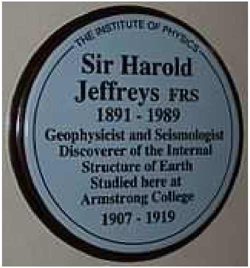




- April 1891 – March 1989
- Mathematician, **Statistician**, Geophysicist, Astronomer
- Studied at **Armstrong College**, **Durham**, now Newcastle University (see plaque)
- Seminal book – **Theory of Probability** – revived the Bayesian view of probability
- Married a Physicist – Bertha Swirles – together they wrote **Methods of Mathematical Physics**

Sir Harold Jeffreys, FRS





- April 1891 – March 1989
- Mathematician, **Statistician**, Geophysicist, Astronomer
- Studied at **Armstrong College**, **Durham**, now Newcastle University (see plaque)
- Seminal book – **Theory of Probability** – revived the Bayesian view of probability
- Married a Physicist – Bertha Swirles – together they wrote **Methods of Mathematical Physics**

Jeffrey's prior

Jeffreys' suggestion was specified in terms of **Fisher's Information**

$$I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right]. \tag{3.8}$$

He recommended that we represent prior ignorance by the prior distribution

$$\pi(\theta) \propto \sqrt{I(\theta)}. \tag{3.9}$$

Such a prior distribution is known as a **Jeffreys prior distribution**.

Jeffrey's prior

Jeffreys' suggestion was specified in terms of **Fisher's Information**

$$I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right]. \tag{3.8}$$

He recommended that we represent prior ignorance by the prior distribution

$$\pi(\theta) \propto \sqrt{I(\theta)}. \tag{3.9}$$

Such a prior distribution is known as a **Jeffreys prior distribution**.

Jeffreys' suggestion was specified in terms of **Fisher's Information**

$$I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right]. \tag{3.8}$$

He recommended that we represent prior ignorance by the prior distribution

$$\pi(\theta) \propto \sqrt{I(\theta)}. \tag{3.9}$$

Such a prior distribution is known as a **Jeffreys prior distribution**.

Advantages of using a Jeffrey's prior

- satisfies the **local uniformity property**: it does not change much in the region over which the likelihood is significant $\Rightarrow$ represents ignornace
- It is **invariant** with respect to one–to–one transformations

Disadvantages

- Often **improper**, and can lead to improper posteriors
- Can be cumbersome to use in high dimensions

Jeffrey's prior

Advantages of using a Jeffrey's prior

- satisfies the **local uniformity property**: it does not change much in the region over which the likelihood is significant $\Rightarrow$ represents ignornace
- It is **invariant** with respect to one–to–one transformations

Disadvantages

- Often **improper**, and can lead to improper posteriors
- Can be cumbersome to use in high dimensions

Advantages of using a Jeffrey's prior

- satisfies the **local uniformity property**: it does not change much in the region over which the likelihood is significant $\Rightarrow$ represents ignornace
- It is **invariant** with respect to one–to–one transformations

Disadvantages

- Often **improper**, and can lead to improper posteriors
- Can be cumbersome to use in high dimensions

Advantages of using a Jeffrey's prior

- satisfies the **local uniformity property**: it does not change much in the region over which the likelihood is significant $\Rightarrow$ represents ignornace
- It is **invariant** with respect to one–to–one transformations

Disadvantages

- Often **improper**, and can lead to improper posteriors
- Can be cumbersome to use in high dimensions

Advantages of using a Jeffrey's prior

- satisfies the **local uniformity property**: it does not change much in the region over which the likelihood is significant $\Rightarrow$ represents ignornace
- It is **invariant** with respect to one–to–one transformations

Disadvantages

- Often **improper**, and can lead to improper posteriors
- Can be cumbersome to use in high dimensions

Advantages of using a Jeffrey's prior

- satisfies the **local uniformity property**: it does not change much in the region over which the likelihood is significant $\Rightarrow$ represents ignornace
- It is **invariant** with respect to one–to–one transformations

Disadvantages

- Often **improper**, and can lead to improper posteriors
- Can be cumbersome to use in high dimensions

Example 3.11

Suppose we have a random sample from a distribution with probability density function

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

Determine the Jeffreys prior for this model.

The likelihood function is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \frac{2x_i\, e^{-x_i^2/\theta}}{\theta}$$

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

Solution to Example 3.11 (1/5)

The likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{2x_i\, e^{-x_i^2/\theta}}{\theta} \\
&= \frac{2^n}{\theta^n} \left(\prod_{i=1}^{n} x_i \right) \exp\left\{ -\frac{1}{\theta} \sum_{i=1}^{n} x_i^2 \right\}.
\end{aligned}
$$

Solution to Example 3.11 (2/5)

Therefore

$$
\begin{aligned}
\log f(\boldsymbol{x}|\theta) &= n\log 2 - n\log\theta + \sum_{i=1}^{n}\log x_i - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2 \\
\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2 \\
\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) &= \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n} x_i^2
\end{aligned}
$$

Therefore

$$\log f(\boldsymbol{x}|\theta) = n\log 2 - n\log\theta + \sum_{i=1}^{n}\log x_i - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n} x_i^2$$

Solution to Example 3.11 (2/5)

Therefore

$$\begin{aligned}
\log f(\boldsymbol{x}|\theta) &= n\log 2 - n\log\theta + \sum_{i=1}^{n}\log x_i - \frac{1}{\theta}\sum_{i=1}^{n}x_i^2 \\
\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}x_i^2 \\
\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) &= \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n}x_i^2
\end{aligned}$$

Solution to Example 3.11 (3/5)

$$
\begin{aligned}
\Rightarrow \quad I(\theta) &= E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&= -\frac{n}{\theta^2} + \frac{2}{\theta^3} E_{X|\theta}\left[\sum_{i=1}^{n} X_i^2\right] \\
&= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\left(E_{X|\theta}\left[X_1^2\right] + \ldots + E_{X|\theta}\left[X_n^2\right]\right) \\
&= -\frac{n}{\theta^2} + \frac{2n}{\theta^3} E_{X|\theta}\left[X^2\right]
\end{aligned}
$$

since the X_i are identically distributed.

Solution to Example 3.11 (3/5)

$$
\begin{aligned}
\Rightarrow \quad I(\theta) &= E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&= -\frac{n}{\theta^2} + \frac{2}{\theta^3} E_{\boldsymbol{X}|\theta}\left[\sum_{i=1}^{n} X_i^2\right] \\
&= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\left(E_{X|\theta}\left[X_1^2\right] + \ldots + E_{X|\theta}\left[X_n^2\right]\right) \\
&= -\frac{n}{\theta^2} + \frac{2n}{\theta^3} E_{X|\theta}\left[X^2\right]
\end{aligned}
$$

since the X_i are identically distributed.

Solution to Example 3.11 (3/5)

$$\begin{aligned}
\Rightarrow \quad I(\theta) &= E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&= -\frac{n}{\theta^2} + \frac{2}{\theta^3} E_{\boldsymbol{X}|\theta}\left[\sum_{i=1}^{n} X_i^2\right] \\
&= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\left(E_{X|\theta}\left[X_1^2\right] + \ldots + E_{X|\theta}\left[X_n^2\right]\right) \\
&= -\frac{n}{\theta^2} + \frac{2n}{\theta^3} E_{X|\theta}\left[X^2\right]
\end{aligned}$$

since the X_i are identically distributed.

$$
\begin{aligned}
\Rightarrow \quad I(\theta) &= E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&= -\frac{n}{\theta^2}+\frac{2}{\theta^3}E_{\boldsymbol{X}|\theta}\left[\sum_{i=1}^{n}X_i^2\right] \\
&= -\frac{n}{\theta^2}+\frac{2}{\theta^3}\left(E_{X|\theta}\left[X_1^2\right]+\ldots+E_{X|\theta}\left[X_n^2\right]\right) \\
&= -\frac{n}{\theta^2}+\frac{2n}{\theta^3}E_{X|\theta}\left[X^2\right]
\end{aligned}
$$

since the X_i are identically distributed.

Solution to Example 3.11 (4/5)

Now

$$E_{X|\theta}\left[X^2\right] = \int_0^\infty x^2 \times \left(\frac{2x\,e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x} dy.$$

Substituting into the integral above gives

$$\begin{aligned}\int_0^\infty y\,2xe^{-y}\frac{\theta}{2x}dy &= \theta\int_0^\infty ye^{-y}dy \\ &= \theta \times 1 = \theta,\end{aligned}$$

since the remaining integral is the mean of a unit exponential.

Solution to Example 3.11 (4/5)

Now

$$E_{X|\theta}\left[X^2\right] \;=\; \int_0^\infty x^2 \times \left(\frac{2x\, e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x}dy.$$

Substituting into the integral above gives

$$\begin{aligned}\int_0^\infty y\, 2xe^{-y}\frac{\theta}{2x}dy \;&=\; \theta\int_0^\infty ye^{-y}dy \\ &=\; \theta \times 1 = \theta,\end{aligned}$$

since the remaining integral is the mean of a unit exponential.

Now

$$E_{X|\theta}\left[X^2\right] = \int_0^\infty x^2 \times \left(\frac{2x\, e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x} dy.$$

Substituting into the integral above gives

$$\begin{aligned}\int_0^\infty y\, 2xe^{-y}\frac{\theta}{2x} dy &= \theta \int_0^\infty ye^{-y} dy \\ &= \theta \times 1 = \theta,\end{aligned}$$

since the remaining integral is the mean of a unit exponential.

Now

$$E_{X|\theta}\left[X^2\right] \;=\; \int_0^\infty x^2 \times \left(\frac{2x\, e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x} dy.$$

Substituting into the integral above gives

$$\int_0^\infty y\, 2xe^{-y} \frac{\theta}{2x} dy \;=\; \theta \int_0^\infty ye^{-y} dy$$

$$= \;\theta \times 1 = \theta,$$

since the remaining integral is the mean of a unit exponential.

Now

$$E_{X|\theta}\left[X^2\right] = \int_0^\infty x^2 \times \left(\frac{2x\,e^{-x^2/\theta}}{\theta}\right)\,dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x}dy.$$

Substituting into the integral above gives

$$\int_0^\infty y\,2xe^{-y}\frac{\theta}{2x}dy = \theta\int_0^\infty ye^{-y}dy$$

$$= \theta \times 1 = \theta,$$

since the remaining integral is the mean of a unit exponential.

Solution to Example 3.11 (4/5)

Now

$$E_{X|\theta}\left[X^2\right] = \int_0^\infty x^2 \times \left(\frac{2x\,e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x} dy.$$

Substituting into the integral above gives

$$\int_0^\infty y\, 2xe^{-y}\frac{\theta}{2x}dy = \theta\int_0^\infty ye^{-y}dy$$

$$= \theta \times 1 = \theta,$$

since the remaining integral is the mean of a unit exponential.

Now

$$E_{X|\theta}\left[X^2\right] = \int_0^\infty x^2 \times \left(\frac{2x\, e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x}dy.$$

Substituting into the integral above gives

$$\int_0^\infty y\, 2xe^{-y}\frac{\theta}{2x}dy = \theta \int_0^\infty ye^{-y}dy$$

$$= \theta \times 1 = \theta,$$

since the remaining integral is the mean of a unit exponential.

Solution to Example 3.11 (4/5)

Now

$$E_{X|\theta}\left[X^2\right] = \int_0^\infty x^2 \times \left(\frac{2x\, e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x} dy.$$

Substituting into the integral above gives

$$\begin{aligned}\int_0^\infty y\, 2xe^{-y} \frac{\theta}{2x} dy &= \theta \int_0^\infty ye^{-y} dy \\ &= \theta \times 1 = \theta,\end{aligned}$$

since the remaining integral is the mean of a unit exponential.

Solution to Example 3.11 (4/5)

Now

$$E_{X|\theta}\left[X^2\right] \;=\; \int_0^\infty x^2 \times \left(\frac{2x\, e^{-x^2/\theta}}{\theta}\right) dx.$$

If we let $y = \dfrac{x^2}{\theta}$, then

$$\frac{dy}{dx} = \frac{2x}{\theta} \Rightarrow dx = \frac{\theta}{2x} dy.$$

Substituting into the integral above gives

$$\begin{aligned}\int_0^\infty y\, 2xe^{-y}\frac{\theta}{2x}dy \;&=\; \theta\int_0^\infty ye^{-y}dy \\ &=\; \theta \times 1 = \theta,\end{aligned}$$

since the remaining integral is the mean of a unit exponential.

Solution to Example 3.11 (5/5)

Therefore

$$I(\theta) = -\frac{n}{\theta^2} + \left(\frac{2n}{\theta^3} \times \theta\right) = \frac{n}{\theta^2}.$$

Hence, the Jeffreys prior for this model is

$$\begin{aligned} \pi(\theta) &\propto I(\theta)^{1/2} \\ &\propto \frac{\sqrt{n}}{\theta}, \qquad \theta > 0 \\ &\propto \frac{1}{\theta}, \qquad \theta > 0. \end{aligned}$$

Therefore

$$I(\theta) = -\frac{n}{\theta^2} + \left(\frac{2n}{\theta^3} \times \theta\right) = \frac{n}{\theta^2}.$$

Hence, the Jeffreys prior for this model is

$$\begin{aligned} \pi(\theta) &\propto I(\theta)^{1/2} \\ &\propto \frac{\sqrt{n}}{\theta}, \qquad \theta > 0 \\ &\propto \frac{1}{\theta}, \qquad \theta > 0. \end{aligned}$$

Solution to Example 3.11 (5/5)

Therefore

$$I(\theta) = -\frac{n}{\theta^2} + \left(\frac{2n}{\theta^3} \times \theta\right) = \frac{n}{\theta^2}.$$

Hence, the Jeffreys prior for this model is

$$\begin{aligned} \pi(\theta) &\propto I(\theta)^{1/2} \\ &\propto \frac{\sqrt{n}}{\theta}, \qquad \theta > 0 \\ &\propto \frac{1}{\theta}, \qquad \theta > 0. \end{aligned}$$

Solution to Example 3.11 (5/5)

Therefore

$$I(\theta) = -\frac{n}{\theta^2} + \left(\frac{2n}{\theta^3} \times \theta\right) = \frac{n}{\theta^2}.$$

Hence, the Jeffreys prior for this model is

$$\begin{aligned} \pi(\theta) &\propto I(\theta)^{1/2} \\ &\propto \frac{\sqrt{n}}{\theta}, \qquad \theta > 0 \\ &\propto \frac{1}{\theta}, \qquad \theta > 0. \end{aligned}$$

Example 3.11

Notice that this distribution is **improper** since $\int_0^\infty d\theta/\theta$ is a divergent integral, and so we cannot find a constant which ensures that the density function integrates to one.

Example 3.12

Suppose we have a random sample from an exponential distribution, that is, $X_i|\theta \sim Exp(\theta)$, $i = 1, 2, \ldots, n$ (independent).

Determine the Jeffreys prior for this model.

Solution to Example 3.12 (1/2)

Recall that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = \theta^n e^{-n\bar{x}\theta},$$

and therefore

$$\log f(\boldsymbol{x}|\theta) = n\log\theta - n\bar{x}\theta$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = \frac{n}{\theta} - n\bar{x}$$

$$\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta^2}$$

$$\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] = \frac{n}{\theta^2}.$$

Solution to Example 3.12 (1/2)

Recall that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = \theta^n e^{-n\bar{x}\theta},$$

and therefore

$$\log f(\boldsymbol{x}|\theta) = n \log \theta - n\bar{x}\theta$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta} \log f(\boldsymbol{x}|\theta) = \frac{n}{\theta} - n\bar{x}$$

$$\Rightarrow \quad \frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta^2}$$

$$\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{X}|\theta)\right] = \frac{n}{\theta^2}.$$

Solution to Example 3.12 (1/2)

Recall that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = \theta^n e^{-n\bar{x}\theta},$$

and therefore

$$\log f(\boldsymbol{x}|\theta) = n\log\theta - n\bar{x}\theta$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = \frac{n}{\theta} - n\bar{x}$$

$$\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta^2}$$

$$\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] = \frac{n}{\theta^2}.$$

Solution to Example 3.12 (2/2)

Hence, the Jeffreys prior for this model is

$$\begin{aligned}\pi(\theta) &\propto I(\theta)^{1/2} \\ &\propto \frac{\sqrt{n}}{\theta}, \qquad \theta > 0 \\ &\propto \frac{1}{\theta}, \qquad \theta > 0.\end{aligned}$$

Solution to Example 3.12 (2/2)

Hence, the Jeffreys prior for this model is

$$\begin{aligned}
\pi(\theta) &\propto I(\theta)^{1/2} \\
&\propto \frac{\sqrt{n}}{\theta}, \qquad \theta > 0 \\
&\propto \frac{1}{\theta}, \qquad \theta > 0.
\end{aligned}$$

Example 3.12

Notice that this distribution is **improper** since $\int_0^\infty d\theta/\theta$ is a divergent integral, and so we cannot find a constant which ensures that the density function integrates to one.

Notice also that this density is, in fact, a limiting form of a $Ga(g, h)$ density (ignoring the integration constant) since

$$\frac{h^g\,\theta^{g-1}e^{-h\theta}}{\Gamma(g)} \propto \theta^{g-1}e^{-h\theta} \to \frac{1}{\theta}, \qquad \text{as } g \to 0,\ h \to 0.$$

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Example 3.12

Notice that this distribution is **improper** since $\int_0^\infty d\theta/\theta$ is a divergent integral, and so we cannot find a constant which ensures that the density function integrates to one.

Notice also that this density is, in fact, a limiting form of a *Ga*(g, h) density (ignoring the integration constant) since

$$\frac{h^g\,\theta^{g-1}e^{-h\theta}}{\Gamma(g)} \propto \theta^{g-1}e^{-h\theta} \to \frac{1}{\theta}, \qquad \text{as } g \to 0,\ h \to 0.$$

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Example 3.12

Notice that this distribution is **improper** since $\int_0^\infty d\theta/\theta$ is a divergent integral, and so we cannot find a constant which ensures that the density function integrates to one.

Notice also that this density is, in fact, a limiting form of a $Ga(g,h)$ density (ignoring the integration constant) since

$$\frac{h^g\,\theta^{g-1}e^{-h\theta}}{\Gamma(g)} \propto \theta^{g-1}e^{-h\theta} \to \frac{1}{\theta}, \qquad \text{as } g\to 0,\ h\to 0.$$

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Example 3.12

Notice that this distribution is **improper** since $\int_0^\infty d\theta/\theta$ is a divergent integral, and so we cannot find a constant which ensures that the density function integrates to one.

Notice also that this density is, in fact, a limiting form of a $Ga(g,h)$ density (ignoring the integration constant) since

$$\frac{h^g\,\theta^{g-1}e^{-h\theta}}{\Gamma(g)} \propto \theta^{g-1}e^{-h\theta} \to \frac{1}{\theta}, \qquad \text{as } g \to 0,\ h \to 0.$$

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Example 3.13

Suppose we have a random sample from a $N(\mu, 1/\tau)$ distribution (with τ known).

Determine the Jeffreys prior for this model.

Solution to Example 3.13 (1/2)

Recall from Example 2.6 that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},$$

and therefore

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

$$\Rightarrow \quad \frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu) = -\frac{\tau}{2}\times\sum_{i=1}^{n}-2(x_i-\mu)$$

$$= \tau\sum_{i=1}^{n}(x_i-\mu)$$

$$= \tau(n\bar{x}-n\mu)$$

$$= n\tau(\bar{x}-\mu)$$

Solution to Example 3.13 (1/2)

Recall from Example 2.6 that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},$$

and therefore

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

$$\Rightarrow \quad \frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu) = -\frac{\tau}{2}\times\sum_{i=1}^{n}-2(x_i-\mu)$$

$$= \tau\sum_{i=1}^{n}(x_i-\mu)$$

$$= \tau(n\bar{x}-n\mu)$$

$$= n\tau(\bar{x}-\mu)$$

Solution to Example 3.13 (1/2)

Recall from Example 2.6 that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},$$

and therefore

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

$$\Rightarrow \quad \frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu) = -\frac{\tau}{2}\times\sum_{i=1}^{n}-2(x_i-\mu)$$

$$= \tau\sum_{i=1}^{n}(x_i-\mu)$$

$$= \tau(n\bar{x} - n\mu)$$

$$= n\tau(\bar{x} - \mu)$$

Solution to Example 3.13 (1/2)

Recall from Example 2.6 that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},$$

and therefore

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

$$\Rightarrow \quad \frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu) = -\frac{\tau}{2}\times\sum_{i=1}^{n}-2(x_i-\mu)$$

$$= \tau\sum_{i=1}^{n}(x_i-\mu)$$

$$= \tau(n\bar{x}-n\mu)$$

$$= n\tau(\bar{x}-\mu)$$

Solution to Example 3.13 (1/2)

Recall from Example 2.6 that

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right\},$$

and therefore

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

$$\Rightarrow \quad \frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu) = -\frac{\tau}{2} \times \sum_{i=1}^{n} -2(x_i - \mu)$$

$$= \tau\sum_{i=1}^{n}(x_i - \mu)$$

$$= \tau(n\bar{x} - n\mu)$$

$$= n\tau(\bar{x} - \mu)$$

Solution to Example 3.13 (2/2)

Also

$$\Rightarrow \quad \frac{\partial^2}{\partial \mu^2} \log f(\boldsymbol{x}|\mu) = -n\tau$$

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

Hence, the Jeffreys prior for this model is

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

Solution to Example 3.13 (2/2)

Also

$$\Rightarrow \quad \frac{\partial^2}{\partial \mu^2} \log f(\boldsymbol{x}|\mu) = -n\tau$$

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

Hence, the Jeffreys prior for this model is

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

Solution to Example 3.13 (2/2)

Also

$$\Rightarrow \quad \frac{\partial^2}{\partial \mu^2} \log f(\boldsymbol{x}|\mu) = -n\tau$$

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

Hence, the Jeffreys prior for this model is

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

Solution to Example 3.13 (2/2)

Also

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

Hence, the Jeffreys prior for this model is

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

Example 3.13

Notice that 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.

Also it is a limiting form of a $N(b, 1/d)$ density (ignoring the integration constant) since

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

as $d \to 0$.

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Example 3.13

Notice that 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.

Also it is a limiting form of a $N(b, 1/d)$ density (ignoring the integration constant) since

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

as $d \to 0$.

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Example 3.13

Notice that 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.

Also it is a limiting form of a $N(b, 1/d)$ density (ignoring the integration constant) since

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

as $d \to 0$.

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Example 3.13

Notice that 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.

Also it is a limiting form of a $N(b, 1/d)$ density (ignoring the integration constant) since

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

as $d \to 0$.

Therefore, we obtain the same posterior distribution whether we adopt the Jeffreys prior or vague prior knowledge.

Announcements

- **Assignment 1**
  - Currently being marked, almost done
  - Will try to release marks/return scripts this week/early next week

- **Assignment 2**
  - Questions to complete: **14**, **40**, **41** and also **30**
  - Due in by 4pm, Friday 3rd May
  - Associated practical: currently scheduled at 1pm, Friday 3rd May (not helpful!), so trying to reschedule (perhaps next week), check emails!

- **Mid-semester test**
  - Work looking good so far
  - Hoping to get the work back to you very soon

Announcements

- **Assignment 1**
  - Currently being marked, almost done
  - Will try to release marks/return scripts this week/early next week
- **Assignment 2**
  - Questions to complete: **14**, **40**, **41** and also **30**
  - Due in by 4pm, Friday 3rd May
  - Associated practical: currently scheduled at 1pm, Friday 3rd May (not helpful!), so trying to reschedule (perhaps next week), check emails!
- **Mid-semester test**
  - Work looking good so far
  - Hoping to get the work back to you very soon

Announcements

- **Assignment 1**
  - Currently being marked, almost done
  - Will try to release marks/return scripts this week/early next week
- **Assignment 2**
  - Questions to complete: **14**, **40**, **41** and also **30**
  - Due in by 4pm, Friday 3rd May
  - Associated practical: currently scheduled at 1pm, Friday 3rd May (not helpful!), so trying to reschedule (perhaps next week), check emails!
- **Mid-semester test**
  - Work looking good so far
  - Hoping to get the work back to you very soon

Announcements

- **Assignment 1**
  - Currently being marked, almost done
  - Will try to release marks/return scripts this week/early next week
- **Assignment 2**
  - Questions to complete: **14**, **40**, **41** and also **30**
  - Due in by 4pm, Friday 3rd May
  - Associated practical: currently scheduled at 1pm, Friday 3rd May (not helpful!), so trying to reschedule (perhaps next week), check emails!
- **Mid-semester test**
  - Work looking good so far
  - Hoping to get the work back to you very soon

Announcements

- **Assignment 1**
  - Currently being marked, almost done
  - Will try to release marks/return scripts this week/early next week
- **Assignment 2**
  - Questions to complete: **14**, **40**, **41** and also **30**
  - Due in by 4pm, Friday 3rd May
  - Associated practical: currently scheduled at 1pm, Friday 3rd May (not helpful!), so trying to reschedule (perhaps next week), check emails!
- **Mid-semester test**
  - Work looking good so far
  - Hoping to get the work back to you very soon

Professor Richard Boys

- Sadly died on Tuesday 5th March
- Preface of your lecture notes:

  "Since 1980, the number of academic staff in Mathematics & Statistics at Newcastle publishing advanced research using Bayesian methods has increased dramatically. In the 1980s, there was only one Bayesian at Newcastle. Now there are at least 12."

- "...one Bayesian at Newcastle": Professor Boys

Professor Richard Boys

- Sadly died on Tuesday 5th March
- Preface of your lecture notes:

''Since 1980, the number of academic staff in Mathematics & Statistics at Newcastle publishing advanced research using Bayesian methods has increased dramatically. In the 1980s, there was only one Bayesian at Newcastle. Now there are at least 12."

- "...one Bayesian at Newcastle": Professor Boys

Professor Richard Boys

- Sadly died on Tuesday 5th March
- Preface of your lecture notes:

  ''Since 1980, the number of academic staff in Mathematics & Statistics at Newcastle publishing advanced research using Bayesian methods has increased dramatically. In the 1980s, there was only one Bayesian at Newcastle. Now there are at least 12."

- "...one Bayesian at Newcastle": Professor Boys

Professor Richard Boys

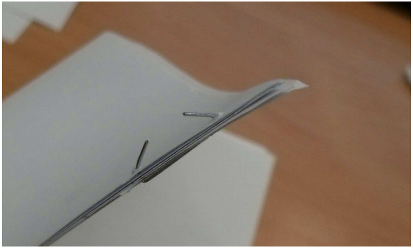
- Richard was a big personality in the department, an extremely talented Statistician and a good friend and colleague to many staff
- His funeral will be taking place this coming Thursday, so our scheduled session at 2pm is **cancelled**

Professor Richard Boys

- Richard was a big personality in the department, an extremely talented Statistician and a good friend and colleague to many staff
- His funeral will be taking place this coming Thursday, so our scheduled session at 2pm is **cancelled**

Dangerous stapling

Several assignments were poorly stapled



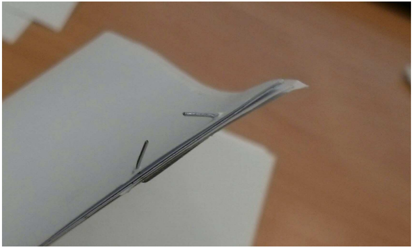
Staples like this are dangerous and can cause pages to go missing. Many of your assignments had to be re-stapled.

It is your responsibility to make sure your work is held together securely and safely, by pushing the stapler down firmly.

Assignments with unsafe staples will not be marked!

Dangerous stapling

Several assignments were poorly stapled



Staples like this are dangerous and can cause pages to go missing. Many of your assignments had to be re-stapled.
It is <u>your responsibility</u> to make sure your work is held together securely and safely, by pushing the stapler down <u>firmly</u>.
Assignments with unsafe staples will not be marked!

Asymptotic posterior distribution

There are many limiting results in Statistics.

The one you will probably remember is the **Central Limit Theorem**.

This concerns the distribution of $\bar{X}_n$, the mean of n independent and identically distributed random variables (each with known mean μ and known variance σ^2), as the sample size $n \to \infty$.

It is easy to show that $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$, and so if we define

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma},$$

then we know that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

The following theorem gives a similar result for the posterior distribution.

Asymptotic posterior distribution

There are many limiting results in Statistics.

The one you will probably remember is the **Central Limit Theorem**.

This concerns the distribution of $\bar{X}_n$, the mean of n independent and identically distributed random variables (each with known mean μ and known variance σ^2), as the sample size $n \to \infty$.

It is easy to show that $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$, and so if we define

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma},$$

then we know that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

The following theorem gives a similar result for the posterior distribution.

Asymptotic posterior distribution

There are many limiting results in Statistics.

The one you will probably remember is the **Central Limit Theorem**.

This concerns the distribution of $\bar{X}_n$, the mean of n independent and identically distributed random variables (each with known mean μ and known variance σ^2), as the sample size $n \to \infty$.

It is easy to show that $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$, and so if we define

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma},$$

then we know that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

The following theorem gives a similar result for the posterior distribution.

Asymptotic posterior distribution

There are many limiting results in Statistics.

The one you will probably remember is the **Central Limit Theorem**.

This concerns the distribution of $\bar{X}_n$, the mean of n independent and identically distributed random variables (each with known mean μ and known variance σ^2), as the sample size $n \to \infty$.

It is easy to show that $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$, and so if we define

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma},$$

then we know that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

The following theorem gives a similar result for the posterior distribution.

Asymptotic posterior distribution

There are many limiting results in Statistics.

The one you will probably remember is the **Central Limit Theorem**.

This concerns the distribution of $\bar{X}_n$, the mean of n independent and identically distributed random variables (each with known mean μ and known variance σ^2), as the sample size $n \to \infty$.

It is easy to show that $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$, and so if we define

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma},$$

then we know that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

The following theorem gives a similar result for the posterior distribution.

Asymptotic posterior distribution

Theorem (Asymptotic posterior distribution)

Suppose we have a statistical model $f(\boldsymbol{x}|\theta)$ *for data* $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, *together with a prior distribution* $\pi(\theta)$ *for* θ. *Then*

$$\sqrt{J(\hat{\theta})}\,(\theta - \hat{\theta})|\boldsymbol{x} \xrightarrow{\mathcal{D}} N(0, 1) \qquad \text{as } n \to \infty,$$

where $\hat{\theta}$ *is the likelihood mode and* $J(\theta)$ *is the observed information*

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

Outline proof (1/7)

Using Bayes Theorem, the posterior distribution for θ is

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)\, f(\boldsymbol{x}|\theta).$$

Let $\psi = \sqrt{n}(\theta - \hat{\theta})$ and

$$\ell_n(\theta) = \frac{1}{n} \log f(\boldsymbol{x}|\theta)$$

be the average log–likelihood per observation, in which case,

$$f(\boldsymbol{x}|\theta) = e^{n\ell_n(\theta)}.$$

Outline proof (1/7)

Using Bayes Theorem, the posterior distribution for θ is

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)\, f(\boldsymbol{x}|\theta).$$

Let $\psi = \sqrt{n}(\theta - \hat{\theta})$ and

$$\ell_n(\theta) = \frac{1}{n}\log f(\boldsymbol{x}|\theta)$$

be the average log–likelihood per observation, in which case,

$$f(\boldsymbol{x}|\theta) = e^{n\ell_n(\theta)}.$$

Outline proof (2/7)

Recall Equation (3.7), which tells us about the distribution of a random variable $Y = g(X)$:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|.$$

We want to know the distribution of $\psi = g(\theta)$, where

$$g(\theta) = \sqrt{n}(\theta - \hat{\theta}).$$

Now

$$g^{-1}(\psi) = \hat{\theta} + \frac{\psi}{\sqrt{n}} \qquad \text{and} \qquad \frac{d}{d\psi}g^{-1}(\psi) = \frac{1}{\sqrt{n}},$$

giving

$$\pi_\psi(\psi) = \pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\middle|\, \boldsymbol{x}\right) \times \frac{1}{\sqrt{n}}.$$

Outline proof (2/7)

Recall Equation (3.7), which tells us about the distribution of a random variable $Y = g(X)$:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}\,g^{-1}(y)\right|.$$

We want to know the distribution of $\psi = g(\theta)$, where

$$g(\theta) = \sqrt{n}(\theta - \hat{\theta}).$$

Now

$$g^{-1}(\psi) = \hat{\theta} + \frac{\psi}{\sqrt{n}} \qquad \text{and} \qquad \frac{d}{d\psi}g^{-1}(\psi) = \frac{1}{\sqrt{n}},$$

giving

$$\pi_\psi(\psi) = \pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\middle|\,\boldsymbol{x}\right) \times \frac{1}{\sqrt{n}}.$$

Outline proof (2/7)

Recall Equation (3.7), which tells us about the distribution of a random variable $Y = g(X)$:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}\,g^{-1}(y)\right|.$$

We want to know the distribution of $\psi = g(\theta)$, where

$$g(\theta) = \sqrt{n}(\theta - \hat{\theta}).$$

Now

$$g^{-1}(\psi) = \hat{\theta} + \frac{\psi}{\sqrt{n}} \qquad \text{and} \qquad \frac{d}{d\psi}g^{-1}(\psi) = \frac{1}{\sqrt{n}},$$

giving

$$\pi_\psi(\psi) = \pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\middle|\,\boldsymbol{x}\right) \times \frac{1}{\sqrt{n}}.$$

Outline proof (2/7)

Recall Equation (3.7), which tells us about the distribution of a random variable $Y = g(X)$:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}\,g^{-1}(y)\right|.$$

We want to know the distribution of $\psi = g(\theta)$, where

$$g(\theta) = \sqrt{n}(\theta - \hat{\theta}).$$

Now

$$g^{-1}(\psi) = \hat{\theta} + \frac{\psi}{\sqrt{n}} \qquad \text{and} \qquad \frac{d}{d\psi}g^{-1}(\psi) = \frac{1}{\sqrt{n}},$$

giving

$$\pi_\psi(\psi) = \pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\middle|\,\boldsymbol{x}\right) \times \frac{1}{\sqrt{n}}.$$

Recall Equation (3.7), which tells us about the distribution of a random variable $Y = g(X)$:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}\,g^{-1}(y)\right|.$$

We want to know the distribution of $\psi = g(\theta)$, where

$$g(\theta) = \sqrt{n}(\theta - \hat{\theta}).$$

Now

$$g^{-1}(\psi) = \hat{\theta} + \frac{\psi}{\sqrt{n}} \qquad \text{and} \qquad \frac{d}{d\psi}g^{-1}(\psi) = \frac{1}{\sqrt{n}},$$

giving

$$\pi_\psi(\psi) = \pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\middle|\,\boldsymbol{x}\right) \times \frac{1}{\sqrt{n}}.$$

Outline proof (3/7)

Thus

$$
\begin{aligned}
\pi_\psi(\psi) &= \pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}} \middle| \boldsymbol{x}\right) \times \frac{1}{\sqrt{n}} \\
&\propto \pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\right) \exp\left\{n\ell_n\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\right)\right\}.
\end{aligned}
$$

Taking Taylor series expansions about $\psi = 0$ gives

$$
\begin{aligned}
\pi_\theta\left(\hat{\theta} + \frac{\psi}{\sqrt{n}}\right) &= \pi_\theta(\hat{\theta}) + \pi'_\theta(\hat{\theta})\frac{\psi}{\sqrt{n}} + \frac{1}{2!}\pi''_\theta(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2 \\
&\quad + \frac{1}{3!}\pi'''_\theta(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^3 + \ldots \\
&\approx \pi_\theta(\hat{\theta}).
\end{aligned}
$$

Outline proof (3/7)

Thus

$$\begin{aligned}
\pi_\psi(\psi) &= \pi_\theta\left(\hat\theta + \frac{\psi}{\sqrt{n}}\middle|\, \boldsymbol{x}\right) \times \frac{1}{\sqrt{n}} \\
&\propto \pi_\theta\left(\hat\theta + \frac{\psi}{\sqrt{n}}\right) \exp\left\{n\ell_n\left(\hat\theta + \frac{\psi}{\sqrt{n}}\right)\right\}.
\end{aligned}$$

Taking Taylor series expansions about $\psi = 0$ gives

$$\begin{aligned}
\pi_\theta\left(\hat\theta + \frac{\psi}{\sqrt{n}}\right) &= \pi_\theta(\hat\theta) + \pi_\theta'(\hat\theta)\frac{\psi}{\sqrt{n}} + \frac{1}{2!}\pi_\theta''(\hat\theta)\left[\frac{\psi}{\sqrt{n}}\right]^2 \\
&\quad + \frac{1}{3!}\pi_\theta'''(\hat\theta)\left[\frac{\psi}{\sqrt{n}}\right]^3 + \dots \\
&\approx \pi_\theta(\hat\theta).
\end{aligned}$$

Outline proof (3/7)

Thus

$$\begin{aligned}\pi_\psi(\psi) &= \pi_\theta\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\middle|\,\boldsymbol{x}\right)\times\frac{1}{\sqrt{n}}\\ &\propto \pi_\theta\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\exp\left\{n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\right\}.\end{aligned}$$

Taking Taylor series expansions about $\psi = 0$ gives

$$\begin{aligned}\pi_\theta\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right) &= \pi_\theta(\hat{\theta})+\pi'_\theta(\hat{\theta})\frac{\psi}{\sqrt{n}}+\frac{1}{2!}\pi''_\theta(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2\\ &\qquad+\frac{1}{3!}\pi'''_\theta(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^3+\ldots\\ &\approx \pi_\theta(\hat{\theta}).\end{aligned}$$

Outline proof (4/7)

Similarly,

$$
\begin{aligned}
n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right) &= n\left\{\ell_n(\hat{\theta})+\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2\right. \\
&\qquad \left. +\frac{1}{3!}\ell_n'''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^3+\ldots\right\} \\
&\approx n\ell_n(\hat{\theta})+n\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+n\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2 \\
&\approx n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2.
\end{aligned}
$$

Similarly,

$$n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right) = n\left\{\ell_n(\hat{\theta})+\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2 + \frac{1}{3!}\ell_n'''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^3+\ldots\right\}$$

$$\approx n\ell_n(\hat{\theta})+n\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+n\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2$$

$$\approx n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2.$$

Outline proof (4/7)

Similarly,

$$
\begin{aligned}
n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right) &= n\left\{\ell_n(\hat{\theta})+\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2\right. \\
&\qquad\left.+\frac{1}{3!}\ell_n'''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^3+\dots\right\} \\
&\approx n\ell_n(\hat{\theta})+n\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+n\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2 \\
&\approx n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2.
\end{aligned}
$$

Outline proof (4/7)

Similarly,

$$\begin{aligned} n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right) &= n\left\{\ell_n(\hat{\theta})+\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2\right. \\ &\qquad \left. +\frac{1}{3!}\ell_n'''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^3+\dots\right\} \\ &\approx n\ell_n(\hat{\theta})+n\ell_n'(\hat{\theta})\frac{\psi}{\sqrt{n}}+n\frac{1}{2!}\ell_n''(\hat{\theta})\left[\frac{\psi}{\sqrt{n}}\right]^2 \\ &\approx n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2. \end{aligned}$$

Outline proof (5/7)

This gives

$$\begin{aligned}
\pi_\psi(\psi|\boldsymbol{x}) &\propto \pi_\theta\left(\hat\theta + \frac{\psi}{\sqrt{n}}\right)\exp\left\{n\ell_n\left(\hat\theta + \frac{\psi}{\sqrt{n}}\right)\right\} \\
&\approx \pi_\theta(\hat\theta)\exp\left\{n\ell_n(\hat\theta) + \frac{1}{2}\ell_n''(\hat\theta)\psi^2\right\} \\
&\propto \exp\left\{\frac{1}{2}\ell_n''(\hat\theta)\psi^2\right\} \\
&\propto \exp\left\{-\frac{\left[-\ell_n''(\hat\theta)\right]}{2}(\psi-0)^2\right\},
\end{aligned}$$

showing that

$$\psi|\boldsymbol{x} \sim N\left(0, [-\ell_n''(\hat\theta)]^{-1}\right) \qquad \text{as } n\to\infty.$$

Outline proof (5/7)

This gives

$$\pi_\psi(\psi|\boldsymbol{x}) \propto \pi_\theta\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\exp\left\{n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\right\}$$

$$\approx \pi_\theta(\hat{\theta})\exp\left\{n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\}$$

$$\propto \exp\left\{\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\}$$

$$\propto \exp\left\{-\frac{\left[-\ell_n''(\hat{\theta})\right]}{2}(\psi-0)^2\right\},$$

showing that

$$\psi|\boldsymbol{x} \sim N\left(0,[-\ell_n''(\hat{\theta})]^{-1}\right) \qquad \text{as } n\to\infty.$$

Outline proof (5/7)

This gives

$$\begin{aligned}\pi_\psi(\psi|\boldsymbol{x}) &\propto \pi_\theta\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\exp\left\{n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\right\}\\ &\approx \pi_\theta(\hat{\theta})\exp\left\{n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\}\\ &\propto \exp\left\{\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\}\\ &\propto \exp\left\{-\frac{\left[-\ell_n''(\hat{\theta})\right]}{2}(\psi-0)^2\right\},\end{aligned}$$

showing that

$$\psi|\boldsymbol{x} \sim N\left(0,[-\ell_n''(\hat{\theta})]^{-1}\right) \qquad \text{as } n\to\infty.$$

Outline proof (5/7)

This gives

$$
\begin{aligned}
\pi_\psi(\psi|\boldsymbol{x}) &\propto \pi_\theta\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\exp\left\{n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\right\} \\
&\approx \pi_\theta(\hat{\theta})\exp\left\{n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\} \\
&\propto \exp\left\{\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\} \\
&\propto \exp\left\{-\frac{\left[-\ell_n''(\hat{\theta})\right]}{2}(\psi-0)^2\right\},
\end{aligned}
$$

showing that

$$\psi|\boldsymbol{x} \sim N\left(0,[-\ell_n''(\hat{\theta})]^{-1}\right) \qquad \text{as } n\to\infty.$$

Outline proof (5/7)

This gives

$$\begin{aligned}
\pi_\psi(\psi|\boldsymbol{x}) &\propto \pi_\theta\left(\hat\theta + \frac{\psi}{\sqrt{n}}\right)\exp\left\{n\ell_n\left(\hat\theta+\frac{\psi}{\sqrt{n}}\right)\right\} \\
&\approx \pi_\theta(\hat\theta)\exp\left\{n\ell_n(\hat\theta) + \frac{1}{2}\ell_n''(\hat\theta)\psi^2\right\} \\
&\propto \exp\left\{\frac{1}{2}\ell_n''(\hat\theta)\psi^2\right\} \\
&\propto \exp\left\{-\frac{\left[-\ell_n''(\hat\theta)\right]}{2}(\psi-0)^2\right\},
\end{aligned}$$

showing that

$$\psi|\boldsymbol{x} \sim N\left(0, [-\ell_n''(\hat\theta)]^{-1}\right) \qquad \text{as } n\to\infty.$$

Outline proof (5/7)

This gives

$$\begin{aligned}
\pi_\psi(\psi|\boldsymbol{x}) &\propto \pi_\theta\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\exp\left\{n\ell_n\left(\hat{\theta}+\frac{\psi}{\sqrt{n}}\right)\right\} \\
&\approx \pi_\theta(\hat{\theta})\exp\left\{n\ell_n(\hat{\theta})+\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\} \\
&\propto \exp\left\{\frac{1}{2}\ell_n''(\hat{\theta})\psi^2\right\} \\
&\propto \exp\left\{-\frac{\left[-\ell_n''(\hat{\theta})\right]}{2}(\psi-0)^2\right\},
\end{aligned}$$

showing that

$$\psi|\boldsymbol{x} \sim N\left(0,[-\ell_n''(\hat{\theta})]^{-1}\right) \qquad \text{as } n\to\infty.$$

Outline proof (6/7)

The result on the last slide gives

$$Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\}=\left[-\ell_n''(\hat{\theta})\right]^{-1}.$$

Multiplying the term inside {} by $\sqrt{-\ell_n''(\hat{\theta})}$ gives

$$\begin{aligned}
Var\left\{\sqrt{-\ell_n''(\hat{\theta})}\times\sqrt{n}(\theta-\hat{\theta})\right\} &= -\ell_n''(\hat{\theta})\times Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\}, \text{ i.e.} \\
Var\left\{\sqrt{-n\ell_n''(\hat{\theta})}(\theta-\hat{\theta})\right\} &= -\ell_n''(\hat{\theta})\times\left[-\ell_n''(\hat{\theta})\right]^{-1}, \quad \text{i.e.} \\
Var\left\{\sqrt{J(\hat{\theta})}(\theta-\hat{\theta})\right\} &= 1.
\end{aligned}$$

Outline proof (6/7)

The result on the last slide gives

$$Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\} = \left[-\ell_n''(\hat{\theta})\right]^{-1}.$$

Multiplying the term inside {} by $\sqrt{-\ell_n''(\hat{\theta})}$ gives

$$Var\left\{\sqrt{-\ell_n''(\hat{\theta})}\times\sqrt{n}(\theta-\hat{\theta})\right\} = -\ell_n''(\hat{\theta})\times Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\}, \text{ i.e.}$$

$$Var\left\{\sqrt{-n\ell_n''(\hat{\theta})}(\theta-\hat{\theta})\right\} = -\ell_n''(\hat{\theta})\times\left[-\ell_n''(\hat{\theta})\right]^{-1}, \quad \text{i.e.}$$

$$Var\left\{\sqrt{J(\hat{\theta})}(\theta-\hat{\theta})\right\} = 1.$$

Outline proof (6/7)

The result on the last slide gives

$$Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\} = \left[-\ell_n''(\hat{\theta})\right]^{-1}.$$

Multiplying the term inside $\{\}$ by $\sqrt{-\ell_n''(\hat{\theta})}$ gives

$$Var\left\{\sqrt{-\ell_n''(\hat{\theta})}\times\sqrt{n}(\theta-\hat{\theta})\right\} = -\ell_n''(\hat{\theta})\times Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\}, \text{ i.e.}$$

$$Var\left\{\sqrt{-n\ell_n''(\hat{\theta})}(\theta-\hat{\theta})\right\} = -\ell_n''(\hat{\theta})\times\left[-\ell_n''(\hat{\theta})\right]^{-1}, \qquad \text{i.e.}$$

$$Var\left\{\sqrt{J(\hat{\theta})}(\theta-\hat{\theta})\right\} = 1.$$

Outline proof (6/7)

The result on the last slide gives

$$Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\}=\left[-\ell_n''(\hat{\theta})\right]^{-1}.$$

Multiplying the term inside $\{\}$ by $\sqrt{-\ell_n''(\hat{\theta})}$ gives

$$\begin{aligned}
Var\left\{\sqrt{-\ell_n''(\hat{\theta})}\times\sqrt{n}(\theta-\hat{\theta})\right\} &= -\ell_n''(\hat{\theta})\times Var\left\{\sqrt{n}(\theta-\hat{\theta})\right\}, \text{ i.e.}\\
Var\left\{\sqrt{-n\ell_n''(\hat{\theta})}(\theta-\hat{\theta})\right\} &= -\ell_n''(\hat{\theta})\times\left[-\ell_n''(\hat{\theta})\right]^{-1}, \qquad \text{i.e.}\\
Var\left\{\sqrt{J(\hat{\theta})}(\theta-\hat{\theta})\right\} &= 1.
\end{aligned}$$

Thus, we have the equivalent result

$$\sqrt{J(\hat{\theta})}\;(\theta-\hat{\theta})\Big|\boldsymbol{x} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n\to\infty.$$

Dividing by $\sqrt{J(\hat{\theta})}$ and adding $\hat{\theta}$ also gives

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

MEMORISE!

Outline proof (7/7)

Thus, we have the equivalent result

$$\sqrt{J(\hat{\theta})}\;\left(\theta - \hat{\theta}\right)\Big|\,\boldsymbol{x} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

Dividing by $\sqrt{J(\hat{\theta})}$ and adding $\hat{\theta}$ also gives

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

MEMORISE!

Outline proof (7/7)

Thus, we have the equivalent result

$$\sqrt{J(\hat{\theta})}\ \left(\theta - \hat{\theta}\right)\Big|\,\boldsymbol{x} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

Dividing by $\sqrt{J(\hat{\theta})}$ and adding $\hat{\theta}$ also gives

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

MEMORISE!

Outline proof (7/7)

Thus, we have the equivalent result

$$\sqrt{J(\hat{\theta})}\ \ (\theta - \hat{\theta})\Big|\,\boldsymbol{x} \xrightarrow{\mathcal{D}} N(0,1) \qquad \text{as } n \to \infty.$$

Dividing by $\sqrt{J(\hat{\theta})}$ and adding $\hat{\theta}$ also gives

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

MEMORISE!

Example 3.14

Suppose we have a random sample from a distribution with probability density function

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

Determine the asymptotic posterior distribution for θ. Note that from Example 3.11 we have

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2,$$

$$J(\theta) = -\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n} x_i^2 = \frac{n}{\theta^3}\left(-\theta + \frac{2}{n}\sum_{i=1}^{n} x_i^2\right).$$

Solution to Example 3.14 (1/2)

The asymptotic posterior distribution is given by

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

First, let's find $\hat{\theta}$. Now

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2;$$

Setting equal to zero and solving for $\theta = \hat{\theta}$ gives

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \overline{x^2}.$$

Solution to Example 3.14 (1/2)

The asymptotic posterior distribution is given by

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

First, let's find $\hat{\theta}$. Now

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2;$$

Setting equal to zero and solving for $\theta = \hat{\theta}$ gives

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \overline{x^2}.$$

Solution to Example 3.14 (1/2)

The asymptotic posterior distribution is given by

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

First, let's find $\hat{\theta}$. Now

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2;$$

Setting equal to zero and solving for $\theta = \hat{\theta}$ gives

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \overline{x^2}.$$

Solution to Example 3.14 (2/2)

Also,

$$\begin{aligned} J(\hat{\theta}) &= \frac{n}{\hat{\theta}^3}\left(-\hat{\theta} + \frac{2}{n}\sum_{i=1}^{n} x_i^2\right) \\ &= \frac{n}{\left(\overline{x^2}\right)^3}\left(-\overline{x^2} + 2\overline{x^2}\right) \\ &= \frac{n}{\left(\overline{x^2}\right)^3}\overline{x^2} = \frac{n}{\left(\overline{x^2}\right)^2}. \end{aligned}$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\overline{x^2}, \frac{1}{n}\left(\overline{x^2}\right)^2\right).$$

Also,

$$
\begin{aligned}
J(\hat{\theta}) &= \frac{n}{\hat{\theta}^3}\left(-\hat{\theta} + \frac{2}{n}\sum_{i=1}^{n} x_i^2\right) \\
&= \frac{n}{\left(\overline{x^2}\right)^3}\left(-\overline{x^2} + 2\overline{x^2}\right) \\
&= \frac{n}{\left(\overline{x^2}\right)^3}\overline{x^2} = \frac{n}{\left(\overline{x^2}\right)^2}.
\end{aligned}
$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\overline{x^2}, \frac{1}{n}\left(\overline{x^2}\right)^2\right).$$

Solution to Example 3.14 (2/2)

Also,

$$
\begin{aligned}
J(\hat{\theta}) &= \frac{n}{\hat{\theta}^3}\left(-\hat{\theta} + \frac{2}{n}\sum_{i=1}^{n} x_i^2\right) \\
&= \frac{n}{\left(\overline{x^2}\right)^3}\left(-\overline{x^2} + 2\overline{x^2}\right) \\
&= \frac{n}{\left(\overline{x^2}\right)^3}\overline{x^2} = \frac{n}{\left(\overline{x^2}\right)^2}.
\end{aligned}
$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\overline{x^2}, \frac{1}{n}\left(\overline{x^2}\right)^2\right).$$

Solution to Example 3.14 (2/2)

Also,

$$
\begin{aligned}
J(\hat{\theta}) &= \frac{n}{\hat{\theta}^3}\left(-\hat{\theta} + \frac{2}{n}\sum_{i=1}^{n} x_i^2\right) \\
&= \frac{n}{\left(\overline{x^2}\right)^3}\left(-\overline{x^2} + 2\overline{x^2}\right) \\
&= \frac{n}{\left(\overline{x^2}\right)^3}\overline{x^2} = \frac{n}{\left(\overline{x^2}\right)^2}.
\end{aligned}
$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$
\theta | \boldsymbol{x} \sim N\left(\overline{x^2}, \frac{1}{n}\left(\overline{x^2}\right)^2\right).
$$

Solution to Example 3.14 (2/2)

Also,

$$\begin{aligned} J(\hat{\theta}) &= \frac{n}{\hat{\theta}^3}\left(-\hat{\theta} + \frac{2}{n}\sum_{i=1}^{n} x_i^2\right) \\ &= \frac{n}{\left(\overline{x^2}\right)^3}\left(-\overline{x^2} + 2\overline{x^2}\right) \\ &= \frac{n}{\left(\overline{x^2}\right)^3}\overline{x^2} = \frac{n}{\left(\overline{x^2}\right)^2}. \end{aligned}$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\overline{x^2}, \frac{1}{n}\left(\overline{x^2}\right)^2\right).$$

Solution to Example 3.14 (2/2)

Also,

$$\begin{aligned} J(\hat{\theta}) &= \frac{n}{\hat{\theta}^3}\left(-\hat{\theta} + \frac{2}{n}\sum_{i=1}^{n} x_i^2\right) \\ &= \frac{n}{\left(\overline{x^2}\right)^3}\left(-\overline{x^2} + 2\overline{x^2}\right) \\ &= \frac{n}{\left(\overline{x^2}\right)^3}\overline{x^2} = \frac{n}{\left(\overline{x^2}\right)^2}. \end{aligned}$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\overline{x^2}, \frac{1}{n}\left(\overline{x^2}\right)^2\right).$$

Example 3.15

Suppose we have a random sample from an exponential distribution, that is, $X_i|\theta \sim Exp(\theta)$, $i = 1, 2, \ldots, n$ (independent).

Determine the asymptotic posterior distribution for θ.

Note that from Example 3.12 we have

$$\begin{aligned}\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) &= \frac{n}{\theta} - n\bar{x},\\ J(\theta) = -\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) &= \frac{n}{\theta^2}.\end{aligned}$$

Solution to Example 3.15 (1/1)

We have

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = 0 \quad \Longrightarrow \quad \hat{\theta} = \frac{1}{\bar{x}}$$

$$\Longrightarrow \quad J(\hat{\theta}) = \frac{n}{\left(\frac{1}{\bar{x}}\right)^2} = n\bar{x}^2$$

$$\Longrightarrow \quad J(\hat{\theta})^{-1} = \frac{1}{n\bar{x}^2}.$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\frac{1}{\bar{x}}, \frac{1}{n\bar{x}^2}\right).$$

Solution to Example 3.15 (1/1)

We have

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = 0 \quad\Longrightarrow\quad \hat{\theta} = \frac{1}{\bar{x}}$$

$$\Longrightarrow\quad J(\hat{\theta}) = \frac{n}{\left(\frac{1}{\bar{x}}\right)^2} = n\bar{x}^2$$

$$\Longrightarrow\quad J(\hat{\theta})^{-1} = \frac{1}{n\bar{x}^2}.$$

Therefore, for large n, the (approximate) posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\frac{1}{\bar{x}},\, \frac{1}{n\bar{x}^2}\right).$$

Example 3.15

Recall that, assuming a vague prior distribution, the posterior distribution is a $Ga(n, n\bar{x})$ distribution, with mean $1/\bar{x}$ and variance $1/(n\bar{x}^2)$.

The Central Limit Theorem tells us that, for large n, the gamma distribution tends to a normal distribution, matched, of course, for mean and variance.

Therefore, we have shown that, for large n, the asymptotic posterior distribution is the same as the posterior distribution under vague prior knowledge. Not a surprising result!

Example 3.16

Suppose we have a random sample from a $N(\mu, 1/\tau)$ distribution (with τ known). Determine the asymptotic posterior distribution for μ. Note that from Example 3.13 we have

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

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

Solution to Example 3.16 (1/1)

We have

$$\frac{\partial}{\partial\mu}\log f(\boldsymbol{x}|\mu)=0 \qquad \Longrightarrow \qquad \hat{\mu}=\bar{x}$$

$$\Longrightarrow \qquad J(\hat{\mu})=n\tau$$

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

Therefore, for large n, the (approximate) posterior distribution for μ is

$$\mu|\boldsymbol{x}\sim N\left(\bar{x},\,\frac{1}{n\tau}\right).$$

Example 3.16

Again, we have shown that the asymptotic posterior distribution is the same as the posterior distribution under vague prior knowledge.

Using a random sample from a $Bin(k, \theta)$ (with k known), determine the posterior distribution for θ assuming

(i) vague prior knowledge;

(ii) the Jeffreys prior distribution;

(iii) a very large sample.

Solution

The **conjugate prior distribution** is a *Beta*(g, h) distribution. Using this prior distribution, the posterior density is

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)\, f(\boldsymbol{x}|\theta) \\
&\propto \theta^{g-1}(1-\theta)^{h-1} \times \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{k-x_i}, \qquad 0<\theta<1 \\
&\propto \theta^{g+n\bar{x}-1}(1-\theta)^{h+nk-n\bar{x}-1}, \qquad 0<\theta<1
\end{aligned}$$

$$i.e.\ \theta|\boldsymbol{x} \sim Beta(G = g+n\bar{x}, H = h+nk-n\bar{x}).$$

Solution

The **conjugate prior distribution** is a *Beta*(g, h) distribution. Using this prior distribution, the posterior density is

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)\, f(\boldsymbol{x}|\theta) \\ &\propto \theta^{g-1}(1-\theta)^{h-1} \times \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{k-x_i}, \qquad 0<\theta<1 \\ &\propto \theta^{g+n\bar{x}-1}(1-\theta)^{h+nk-n\bar{x}-1}, \qquad 0<\theta<1\end{aligned}$$

i.e. $\theta|\boldsymbol{x} \sim Beta(G = g + n\bar{x}, H = h + nk - n\bar{x})$.

Solution

The **conjugate prior distribution** is a $Beta(g, h)$ distribution. Using this prior distribution, the posterior density is

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)\, f(\boldsymbol{x}|\theta)\\ &\propto \theta^{g-1}(1-\theta)^{h-1} \times \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{k-x_i}, \qquad 0<\theta<1\\ &\propto \theta^{g+n\bar{x}-1}(1-\theta)^{h+nk-n\bar{x}-1}, \qquad 0<\theta<1\end{aligned}$$

$$i.e.\ \theta|\boldsymbol{x} \sim Beta(G = g + n\bar{x}, H = h + nk - n\bar{x}).$$

Solution

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range (0,1), there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at $\theta = 0$ and $\theta = 1$ – this distribution is obtained in the limit $g \to 0$ and $h \to 0$.

Thus we will take this limit to represent vague prior information.Hence the posterior distribution under vague prior information is

$$\theta|\boldsymbol{x} \sim Beta(n\bar{x}, nk - n\bar{x}).$$

Solution

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range (0,1), there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at $\theta = 0$ and $\theta = 1$ – this distribution is obtained in the limit $g \to 0$ and $h \to 0$.

Thus we will take this limit to represent vague prior information.Hence the posterior distribution under vague prior information is

$$\theta | \boldsymbol{x} \sim Beta(n\bar{x}, nk - n\bar{x}).$$

Solution

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range (0,1), there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at $\theta = 0$ and $\theta = 1$ – this distribution is obtained in the limit $g \to 0$ and $h \to 0$.

Thus we will take this limit to represent vague prior information.Hence the posterior distribution under vague prior information is

$$\theta|\boldsymbol{x} \sim Beta(n\bar{x}, nk - n\bar{x}).$$

Solution

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range (0,1), there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at $\theta = 0$ and $\theta = 1$ – this distribution is obtained in the limit $g \to 0$ and $h \to 0$.

Thus we will take this limit to represent vague prior information. Hence the posterior distribution under vague prior information is

$$\theta|\boldsymbol{x} \sim Beta(n\bar{x}, nk - n\bar{x}).$$

Solution

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range (0,1), there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at $\theta = 0$ and $\theta = 1$ – this distribution is obtained in the limit $g \to 0$ and $h \to 0$.

Thus we will take this limit to represent vague prior information.Hence the posterior distribution under vague prior information is

$$\theta|\boldsymbol{x} \sim Beta(n\bar{x}, nk - n\bar{x}).$$

Solution

The **Jeffreys prior distribution** is

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

Now

$$\begin{aligned} f(\boldsymbol{x}|\theta) &\propto \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{k-x_i} \\ &\propto \theta^{n\bar{x}}(1-\theta)^{kn-n\bar{x}}. \end{aligned}$$

Solution

The **Jeffreys prior distribution** is

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

Now

$$\begin{aligned} f(\boldsymbol{x}|\theta) &\propto \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{k-x_i} \\ &\propto \theta^{n\bar{x}}(1-\theta)^{kn-n\bar{x}}. \end{aligned}$$

Solution

Therefore

$$\begin{aligned}
\log f(\boldsymbol{x}|\theta) &= constant + n\bar{x}\log\theta + n(k-\bar{x})\log(1-\theta) \\
&\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = \frac{n\bar{x}}{\theta} - \frac{n(k-\bar{x})}{1-\theta} \\
&\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n\bar{x}}{\theta^2} - \frac{n(k-\bar{x})}{(1-\theta)^2} \\
&\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&\qquad = \frac{nE_{\boldsymbol{X}|\theta}(\bar{X})}{\theta^2} + \frac{n[k - E_{\boldsymbol{X}|\theta}(\bar{X})]}{(1-\theta)^2}.
\end{aligned}$$

Therefore

$$
\begin{aligned}
\log f(\boldsymbol{x}|\theta) &= constant + n\bar{x}\log\theta + n(k-\bar{x})\log(1-\theta) \\
&\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = \frac{n\bar{x}}{\theta} - \frac{n(k-\bar{x})}{1-\theta} \\
&\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n\bar{x}}{\theta^2} - \frac{n(k-\bar{x})}{(1-\theta)^2} \\
&\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&\qquad = \frac{nE_{\boldsymbol{X}|\theta}(\bar{X})}{\theta^2} + \frac{n[k-E_{\boldsymbol{X}|\theta}(\bar{X})]}{(1-\theta)^2}.
\end{aligned}
$$

Solution

Therefore

$$\begin{aligned}
\log f(\boldsymbol{x}|\theta) \quad &= \quad \textit{constant} + n\bar{x}\log\theta + n(k-\bar{x})\log(1-\theta) \\
&\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = \frac{n\bar{x}}{\theta} - \frac{n(k-\bar{x})}{1-\theta} \\
&\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n\bar{x}}{\theta^2} - \frac{n(k-\bar{x})}{(1-\theta)^2} \\
&\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&\qquad\quad = \frac{nE_{\boldsymbol{X}|\theta}(\bar{X})}{\theta^2} + \frac{n[k-E_{\boldsymbol{X}|\theta}(\bar{X})]}{(1-\theta)^2}.
\end{aligned}$$

Solution

Therefore

$$\begin{aligned}
\log f(\boldsymbol{x}|\theta) &= constant + n\bar{x}\log\theta + n(k-\bar{x})\log(1-\theta) \\
&\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = \frac{n\bar{x}}{\theta} - \frac{n(k-\bar{x})}{1-\theta} \\
&\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n\bar{x}}{\theta^2} - \frac{n(k-\bar{x})}{(1-\theta)^2} \\
&\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&\qquad = \frac{nE_{\boldsymbol{X}|\theta}(\bar{X})}{\theta^2} + \frac{n[k-E_{\boldsymbol{X}|\theta}(\bar{X})]}{(1-\theta)^2}.
\end{aligned}$$

Solution

Therefore

$$
\begin{aligned}
\log f(\boldsymbol{x}|\theta) &= \mathit{constant} + n\bar{x}\log\theta + n(k-\bar{x})\log(1-\theta) \\
&\Rightarrow \quad \frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = \frac{n\bar{x}}{\theta} - \frac{n(k-\bar{x})}{1-\theta} \\
&\Rightarrow \quad \frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = -\frac{n\bar{x}}{\theta^2} - \frac{n(k-\bar{x})}{(1-\theta)^2} \\
&\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right] \\
&\qquad\quad = \frac{nE_{\boldsymbol{X}|\theta}(\bar{X})}{\theta^2} + \frac{n[k-E_{\boldsymbol{X}|\theta}(\bar{X})]}{(1-\theta)^2}.
\end{aligned}
$$

Solution

Now $E_{\boldsymbol{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$. Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned}\pi(\theta) &\propto \sqrt{I(\theta)} \\ &\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \qquad 0 < \theta < 1 \\ &\propto \theta^{-1/2}(1-\theta)^{-1/2}, \qquad 0 < \theta < 1.\end{aligned}$$

Solution

Now $E_{\boldsymbol{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$. Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned} \pi(\theta) &\propto \sqrt{I(\theta)} \\ &\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \qquad 0 < \theta < 1 \\ &\propto \theta^{-1/2}(1-\theta)^{-1/2}, \qquad 0 < \theta < 1. \end{aligned}$$

Solution

Now $E_{\boldsymbol{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$. Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned}
\pi(\theta) &\propto \sqrt{I(\theta)} \\
&\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \qquad 0 < \theta < 1 \\
&\propto \theta^{-1/2}(1-\theta)^{-1/2}, \qquad 0 < \theta < 1.
\end{aligned}$$

Solution

Now $E_{\boldsymbol{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$. Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned}
\pi(\theta) &\propto \sqrt{I(\theta)} \\
&\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \qquad 0 < \theta < 1 \\
&\propto \theta^{-1/2}(1-\theta)^{-1/2}, \qquad 0 < \theta < 1.
\end{aligned}$$

Solution

Now $E_{\boldsymbol{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$. Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned}
\pi(\theta) &\propto \sqrt{I(\theta)} \\
&\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \qquad 0 < \theta < 1 \\
&\propto \theta^{-1/2}(1-\theta)^{-1/2}, \qquad 0 < \theta < 1.
\end{aligned}$$

Solution

Now $E_{\boldsymbol{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$. Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned}
\pi(\theta) &\propto \sqrt{I(\theta)} \\
&\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \qquad 0 < \theta < 1 \\
&\propto \theta^{-1/2}(1-\theta)^{-1/2}, \qquad 0 < \theta < 1.
\end{aligned}$$

Solution

This is a *Beta*$(1/2, 1/2)$ prior distribution and so the resulting posterior distribution is

$$\theta|\boldsymbol{x} \sim Beta(1/2 + n\bar{x}, 1/2 + nk - n\bar{x}).$$

Solution

The **asymptotic posterior distribution** (as $n \to \infty$) is

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

where

$$J(\theta) = -\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) = \frac{n\bar{x}}{\theta^2} + \frac{n(k-\bar{x})}{(1-\theta)^2}.$$

Solution

Now

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) = 0 \quad\Longrightarrow\quad \frac{n\bar{x}}{\hat{\theta}} - \frac{n(k-\bar{x})}{1-\hat{\theta}} = 0$$

$$\Longrightarrow\quad \hat{\theta} = \frac{\bar{x}}{k}$$

$$\Longrightarrow\quad J(\hat{\theta}) = \frac{nk^3}{\bar{x}(k-\bar{x})}$$

$$\Longrightarrow\quad J(\hat{\theta})^{-1} = \frac{\bar{x}(k-\bar{x})}{nk^3}.$$

Solution

Now

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta)=0 \quad\Longrightarrow\quad \frac{n\bar{x}}{\hat{\theta}}-\frac{n(k-\bar{x})}{1-\hat{\theta}}=0$$

$$\Longrightarrow\quad \hat{\theta}=\frac{\bar{x}}{k}$$

$$\Longrightarrow\quad J(\hat{\theta})=\frac{nk^3}{\bar{x}(k-\bar{x})}$$

$$\Longrightarrow\quad J(\hat{\theta})^{-1}=\frac{\bar{x}(k-\bar{x})}{nk^3}.$$

Solution

Now

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta)=0 \implies \frac{n\bar{x}}{\hat{\theta}}-\frac{n(k-\bar{x})}{1-\hat{\theta}}=0$$

$$\implies \hat{\theta}=\frac{\bar{x}}{k}$$

$$\implies J(\hat{\theta})=\frac{nk^3}{\bar{x}(k-\bar{x})}$$

$$\implies J(\hat{\theta})^{-1}=\frac{\bar{x}(k-\bar{x})}{nk^3}.$$

Solution

Therefore, for large n, the posterior distribution for θ is

$$\theta|\boldsymbol{x} \sim N\left(\frac{\bar{x}}{k}, \frac{\bar{x}(k-\bar{x})}{nk^3}\right) \qquad \text{approximately.}$$