Chapter 2

Bayes' Theorem for Distributions

Bayes' Theorem for Distributions

Suppose we have data $\boldsymbol{x}$ which we model using the probability (density) function $f(\boldsymbol{x}|\theta)$, which depends on a single parameter θ.

Once we have observed the data, $f(\boldsymbol{x}|\theta)$

- is the **likelihood function** for θ
- is a function of θ (for fixed $\boldsymbol{x}$) *rather than*
- a function of $\boldsymbol{x}$ (for fixed θ).

Bayes' Theorem for Distributions

Suppose we have data $\boldsymbol{x}$ which we model using the probability (density) function $f(\boldsymbol{x}|\theta)$, which depends on a single parameter θ.

Once we have observed the data, $f(\boldsymbol{x}|\theta)$

- is the **likelihood function** for θ
- is a function of θ (for fixed $\boldsymbol{x}$) *rather than*
- a function of $\boldsymbol{x}$ (for fixed θ).

Bayes’ Theorem for Distributions

Suppose we have data $\boldsymbol{x}$ which we model using the probability (density) function $f(\boldsymbol{x}|\theta)$, which depends on a single parameter θ.

Once we have observed the data, $f(\boldsymbol{x}|\theta)$

- is the **likelihood function** for θ
- is a function of θ (for fixed $\boldsymbol{x}$) *rather than*
- a function of $\boldsymbol{x}$ (for fixed θ).

Suppose we have data $\boldsymbol{x}$ which we model using the probability (density) function $f(\boldsymbol{x}|\theta)$, which depends on a single parameter θ.

Once we have observed the data, $f(\boldsymbol{x}|\theta)$

- is the **likelihood function** for θ
- is a function of θ (for fixed $\boldsymbol{x}$) *rather than*
- a function of $\boldsymbol{x}$ (for fixed θ).

Bayes' Theorem for Distributions

Also, suppose we have prior beliefs about likely values of θ expressed by a probability (density) function $\pi(\theta)$.

We can combine both pieces of information using the following version of Bayes Theorem.

The resulting distribution for θ is called the **posterior distribution** for θ as it expresses our beliefs about θ *after* seeing the data.

It summarises all our current knowledge about the parameter θ.

Bayes' Theorem for Distributions

The posterior probability (density) function for θ is

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

where

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

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

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

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

Bayes' Theorem for Distributions

The posterior probability (density) function for θ is

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

where

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

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

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

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

Thus, to obtain the posterior distribution, we need:

(1) data, from which we can form the **likelihood** $f(\boldsymbol{x}|\theta)$, and

(2) a suitable distribution, $\pi(\theta)$, that represents our **prior beliefs** about θ.

Bayes' Theorem for Distributions

Thus, to obtain the posterior distribution, we need:

(1) data, from which we can form the **likelihood** $f(\boldsymbol{x}|\theta)$, and
(2) a suitable distribution, $\pi(\theta)$, that represents our **prior beliefs** about θ.

Thus, to obtain the posterior distribution, we need:

(1) data, from which we can form the **likelihood** $f(\boldsymbol{x}|\theta)$, and
(2) a suitable distribution, $\pi(\theta)$, that represents our **prior beliefs** about θ.

Some important continuous distributions

Definition (Continuous Uniform distribution)

The random variable Y follows a Uniform $U(a, b)$ distribution if it has probability density function

$$f(y|a, b) = \frac{1}{b-a}, \qquad a \leq y \leq b.$$

This form of probability density function ensures that all values in the range $[a, b]$ are **equally likely**, hence the name "uniform".

This distribution is sometimes called the **rectangular distribution** because of its shape.

You should remember from MAS1604 that

$$E(Y) = \frac{a+b}{2} \quad \text{and} \quad Var(Y) = \frac{(b-a)^2}{12}.$$

Some important continuous distributions

Definition (Continuous Uniform distribution)

The random variable Y follows a Uniform $U(a, b)$ distribution if it has probability density function

$$f(y|a, b) = \frac{1}{b-a}, \qquad a \leq y \leq b.$$

This form of probability density function ensures that all values in the range $[a, b]$ are **equally likely**, hence the name "uniform".

This distribution is sometimes called the **rectangular distribution** because of its shape.

You should remember from MAS1604 that

$$E(Y) = \frac{a+b}{2} \quad \text{and} \quad Var(Y) = \frac{(b-a)^2}{12}.$$

Some important continuous distributions

Definition (Continuous Uniform distribution)

The random variable Y follows a Uniform $U(a, b)$ distribution if it has probability density function

$$f(y|a, b) = \frac{1}{b-a}, \qquad a \leq y \leq b.$$

This form of probability density function ensures that all values in the range $[a, b]$ are **equally likely**, hence the name "uniform".

This distribution is sometimes called the **rectangular distribution** because of its shape.

You should remember from MAS1604 that

$$E(Y) = \frac{a+b}{2} \quad \text{and} \quad Var(Y) = \frac{(b-a)^2}{12}.$$

Some important continuous distributions

Definition (Continuous Uniform distribution)

The random variable Y follows a Uniform $U(a, b)$ distribution if it has probability density function

$$f(y|a, b) = \frac{1}{b-a}, \qquad a \leq y \leq b.$$

This form of probability density function ensures that all values in the range $[a, b]$ are **equally likely**, hence the name “uniform”.

This distribution is sometimes called the **rectangular distribution** because of its shape.

You should remember from MAS1604 that

$$E(Y) = \frac{a+b}{2} \quad \text{and} \quad Var(Y) = \frac{(b-a)^2}{12}.$$

Some important continuous distributions

Definition (Continuous Uniform distribution)

The random variable Y follows a Uniform $U(a, b)$ distribution if it has probability density function

$$f(y|a, b) = \frac{1}{b-a}, \qquad a \leq y \leq b.$$

This form of probability density function ensures that all values in the range $[a, b]$ are **equally likely**, hence the name “uniform”.

This distribution is sometimes called the **rectangular distribution** because of its shape.

You should remember from MAS1604 that

$$E(Y) = \frac{a+b}{2} \quad \text{and} \quad Var(Y) = \frac{(b-a)^2}{12}.$$

Some important continuous distributions

Definition (Continuous Uniform distribution)

The random variable Y follows a Uniform $U(a, b)$ distribution if it has probability density function

$$f(y|a,b) = \frac{1}{b-a}, \qquad a \leq y \leq b.$$

This form of probability density function ensures that all values in the range $[a, b]$ are **equally likely**, hence the name “uniform”.

This distribution is sometimes called the **rectangular distribution** because of its shape.

You should remember from MAS1604 that

$$E(Y) = \frac{a+b}{2} \quad \text{and} \quad Var(Y) = \frac{(b-a)^2}{12}.$$

Definition (Beta distribution)

The random variable Y follows a Beta $Be(a, b)$ distribution ($a > 0$, $b > 0$) if it has probability density function

$$f(y|a, b) = \frac{y^{a-1}(1-y)^{b-1}}{B(a, b)}, \qquad 0 < y < 1. \tag{2.1}$$

The constant term $B(a, b)$, also known as the **beta function**, ensures that the density integrates to one. Therefore

$$B(a, b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,dy. \tag{2.2}$$

Definition (Beta distribution)

The random variable Y follows a Beta $Be(a,b)$ distribution ($a > 0$, $b > 0$) if it has probability density function

$$f(y|a,b) = \frac{y^{a-1}(1-y)^{b-1}}{B(a,b)}, \qquad 0 < y < 1. \tag{2.1}$$

The constant term $B(a,b)$, also known as the **beta function**, ensures that the density integrates to one. Therefore

$$B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,dy. \tag{2.2}$$

Some important continuous distributions

Definition (Beta distribution)

The random variable Y follows a Beta $Be(a, b)$ distribution ($a > 0$, $b > 0$) if it has probability density function

$$f(y|a,b) = \frac{y^{a-1}(1-y)^{b-1}}{B(a,b)}, \qquad 0 < y < 1. \tag{2.1}$$

The constant term $B(a, b)$, also known as the **beta function**, ensures that the density integrates to one. Therefore

$$B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,dy. \tag{2.2}$$

Some important continuous distributions

It can be shown that the beta function can be expressed in terms of another function, called the **gamma function** $\Gamma(\cdot)$, as

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

where

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,dx. \tag{2.3}$$

It can be shown that the beta function can be expressed in terms of another function, called the **gamma function** $\Gamma(\cdot)$, as

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

where

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\, dx. \tag{2.3}$$

Tables are available for both $B(a, b)$ and $\Gamma(a)$.

However, these functions are very simple to evaluate when a and b are integers since the gamma function is a generalisation of the factorial function.

In particular, when a and b are integers, we have

$$\Gamma(a) = (a-1)! \quad \text{and} \quad B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}.$$

Some important continuous distributions

Tables are available for both $B(a, b)$ and $\Gamma(a)$.

However, these functions are very simple to evaluate when a and b are integers since the gamma function is a generalisation of the factorial function.

In particular, when a and b are integers, we have

$$\Gamma(a) = (a-1)! \quad \text{and} \quad B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}.$$

Some important continuous distributions

It can be shown, using the identity $\Gamma(a) = (a-1)\Gamma(a-1)$, that

$$E(Y) = \frac{a}{a+b}, \qquad \text{and} \qquad Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Also

$$Mode(Y) = \frac{a-1}{a+b-2}, \quad \text{if } a > 1 \text{ and } b > 1.$$

Some important continuous distributions

Definition (Gamma distribution)

The random variable Y follows a Gamma $Ga(a, b)$ distribution ($a > 0$, $b > 0$) if it has probability density function

$$f(y|a,b) = \frac{b^a y^{a-1} e^{-by}}{\Gamma(a)}, \qquad y > 0,$$

where $\Gamma(a)$ is the gamma function defined in Equation (2.3).

It can be shown that

$$E(Y) = \frac{a}{b} \qquad \text{and} \qquad Var(Y) = \frac{a}{b^2}.$$

Also

$$Mode(Y) = \frac{a-1}{b}, \quad \text{if } a \geq 1.$$

Some important continuous distributions

Definition (Gamma distribution)

The random variable Y follows a Gamma $Ga(a, b)$ distribution ($a > 0$, $b > 0$) if it has probability density function

$$f(y|a,b) = \frac{b^a y^{a-1} e^{-by}}{\Gamma(a)}, \qquad y > 0,$$

where $\Gamma(a)$ is the gamma function defined in Equation (2.3).

It can be shown that

$$E(Y) = \frac{a}{b} \qquad \text{and} \qquad Var(Y) = \frac{a}{b^2}.$$

Also

$$Mode(Y) = \frac{a-1}{b}, \quad \text{if } a \geq 1.$$

Bayes Theorem for distributions in action

We will now see Bayes' Theorem for distributions "in action".

Recall that we have

$$\pi(\theta|\boldsymbol{x}) \quad \propto \quad \pi(\theta) \times f(\boldsymbol{x}|\theta), \qquad \text{i.e.}$$

$$\textbf{posterior} \quad \propto \quad \textbf{prior} \times \textbf{likelihood}$$

Recall also that, for now, we will assume our prior for θ has been given to us – we will look at how such priors are constructed, or **elicited**, in the next chapter.

Bayes Theorem for distributions in action

We will now see Bayes' Theorem for distributions "in action".

Recall that we have

$$\pi(\theta|\boldsymbol{x}) \quad \propto \quad \pi(\theta) \times f(\boldsymbol{x}|\theta), \qquad \text{i.e.}$$

posterior $\propto$ **prior** $\times$ **likelihood**

Recall also that, for now, we will assume our prior for θ has been given to us – we will look at how such priors are constructed, or **elicited**, in the next chapter.

Bayes Theorem for distributions in action

We will now see Bayes' Theorem for distributions "in action".

Recall that we have

$$\pi(\theta|\boldsymbol{x}) \quad \propto \quad \pi(\theta) \times f(\boldsymbol{x}|\theta), \qquad \text{i.e.}$$

posterior $\propto$ **prior** $\times$ **likelihood**

Recall also that, for now, we will assume our prior for θ has been given to us – we will look at how such priors are constructed, or **elicited**, in the next chapter.

Bayes Theorem for distributions in action

We will now see Bayes' Theorem for distributions "in action".

Recall that we have

$$\pi(\theta|\boldsymbol{x}) \quad \propto \quad \pi(\theta) \times f(\boldsymbol{x}|\theta), \qquad \text{i.e.}$$

$$\textbf{posterior} \quad \propto \quad \textbf{prior} \times \textbf{likelihood}$$

Recall also that, for now, we will assume our prior for θ has been given to us – we will look at how such priors are constructed, or **elicited**, in the next chapter.

Example 2.1

Consider an experiment with a possibly biased coin. Let $\theta = \text{Pr(Head)}$.

Suppose that, before conducting the experiment, we believe that all values of θ are equally likely.

This gives a prior distribution $\theta \sim U(0, 1)$, and so

$$\pi(\theta) = 1, \qquad 0 < \theta < 1. \tag{2.4}$$

Note that with this prior distribution $E(\theta) = 0.5$.

We now toss the coin 5 times and observe 1 head. Determine the posterior distribution for θ given this data.

Example 2.1

Consider an experiment with a possibly biased coin. Let $\theta = \Pr(\text{Head})$.

Suppose that, before conducting the experiment, we believe that all values of θ are equally likely.

This gives a prior distribution $\theta \sim U(0, 1)$, and so

$$\pi(\theta) = 1, \qquad 0 < \theta < 1. \tag{2.4}$$

Note that with this prior distribution $E(\theta) = 0.5$.

We now toss the coin 5 times and observe 1 head. Determine the posterior distribution for θ given this data.

Example 2.1

Consider an experiment with a possibly biased coin. Let $\theta = \text{Pr(Head)}$.

Suppose that, before conducting the experiment, we believe that all values of θ are equally likely.

This gives a prior distribution $\theta \sim U(0, 1)$, and so

$$\pi(\theta) = 1, \qquad 0 < \theta < 1. \tag{2.4}$$

Note that with this prior distribution $E(\theta) = 0.5$.

We now toss the coin 5 times and observe 1 head. Determine the posterior distribution for θ given this data.

Example 2.1

Consider an experiment with a possibly biased coin. Let $\theta = \text{Pr(Head)}$.

Suppose that, before conducting the experiment, we believe that all values of θ are equally likely.

This gives a prior distribution $\theta \sim U(0, 1)$, and so

$$\pi(\theta) = 1, \qquad 0 < \theta < 1. \tag{2.4}$$

Note that with this prior distribution $E(\theta) = 0.5$.

We now toss the coin 5 times and observe 1 head. Determine the posterior distribution for θ given this data.

Example 2.1

Consider an experiment with a possibly biased coin. Let $\theta = \text{Pr(Head)}$.

Suppose that, before conducting the experiment, we believe that all values of θ are equally likely.

This gives a prior distribution $\theta \sim U(0, 1)$, and so

$$\pi(\theta) = 1, \qquad 0 < \theta < 1. \tag{2.4}$$

Note that with this prior distribution $E(\theta) = 0.5$.

We now toss the coin 5 times and observe 1 head. Determine the posterior distribution for θ given this data.

“Full” solution to Example 2.1 (1/4)

The data is an observation on the random variable $X|\theta \sim Bin(5, \theta)$. This gives a likelihood function

$$f(x = 1|\theta) = {}^5C_1\theta^1(1-\theta)^4$$

$$= 5\theta(1-\theta)^4 \qquad (2.5)$$

which favours values of θ near its maximum $\theta = 0.2$ (we observed 1 head out of 5 tosses).

Therefore, we have a conflict of opinions: the prior distribution (2.4) suggests that θ is probably around 0.5 and the data (2.5) suggest that it is around 0.2.

We can use Bayes' Theorem to combine these two sources of information in a coherent way.

“Full” solution to Example 2.1 (1/4)

The data is an observation on the random variable $X|\theta \sim Bin(5, \theta)$. This gives a likelihood function

$$f(x = 1|\theta) = {}^5C_1\theta^1(1-\theta)^4$$

$$= 5\theta(1-\theta)^4 \qquad (2.5)$$

which favours values of θ near its maximum $\theta = 0.2$ (we observed 1 head out of 5 tosses).

Therefore, we have a conflict of opinions: the prior distribution (2.4) suggests that θ is probably around 0.5 and the data (2.5) suggest that it is around 0.2.

We can use Bayes’ Theorem to combine these two sources of information in a coherent way.

"Full" solution to Example 2.1 (1/4)

The data is an observation on the random variable $X|\theta \sim Bin(5, \theta)$. This gives a likelihood function

$$f(x = 1|\theta) = {}^5C_1\theta^1(1-\theta)^4$$

$$= 5\theta(1-\theta)^4 \qquad (2.5)$$

which favours values of θ near its maximum $\theta = 0.2$ (we observed 1 head out of 5 tosses).

Therefore, we have a conflict of opinions: the prior distribution (2.4) suggests that θ is probably around 0.5 and the data (2.5) suggest that it is around 0.2.

We can use Bayes' Theorem to combine these two sources of information in a coherent way.

Recall the “full” version of Bayes’ Theorem for distributions:

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

where

$$f(\boldsymbol{x}) = \int_{\Theta} \pi(\theta)f(\boldsymbol{x}|\theta)d\theta;$$

“Full” solution to Example 2.1 (2/4)

Recall the “full” version of Bayes’ Theorem for distributions:

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

where

$$f(\boldsymbol{x}) = \int_{\Theta} \pi(\theta)f(\boldsymbol{x}|\theta)d\theta;$$

“Full” solution to Example 2.1 (2/4)

Recall the “full” version of Bayes’ Theorem for distributions:

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

where

$$f(\boldsymbol{x}) = \int_{\Theta} \pi(\theta)f(\boldsymbol{x}|\theta)d\theta;$$

"Full" solution to Example 2.1 (3/4)

First the denominator:

$$f(x=1) = \int_{\Theta} \pi(\theta) f(x=1|\theta)\, d\theta \tag{2.6}$$

$$= \int_0^1 1 \times 5\theta(1-\theta)^4\, d\theta$$

$$= \int_0^1 \theta \times 5(1-\theta)^4\, d\theta$$

which, using integration by parts, gives

$$f(x=1) = \left[-(1-\theta)^5\,\theta\right]_0^1 + \int_0^1 (1-\theta)^5\, d\theta$$

$$= 0 + \left[-\frac{(1-\theta)^6}{6}\right]_0^1$$

$$= \frac{1}{6}.$$

“Full” solution to Example 2.1 (3/4)

First the denominator:

$$
\begin{aligned}
f(x=1) &= \int_{\Theta} \pi(\theta) f(x=1|\theta)\, d\theta \qquad (2.6)\\
&= \int_0^1 1 \times 5\theta(1-\theta)^4\, d\theta\\
&= \int_0^1 \theta \times 5(1-\theta)^4\, d\theta
\end{aligned}
$$

which, using integration by parts, gives

$$
\begin{aligned}
f(x=1) &= \left[-(1-\theta)^5\,\theta\right]_0^1 + \int_0^1 (1-\theta)^5\, d\theta\\
&= 0 + \left[-\frac{(1-\theta)^6}{6}\right]_0^1\\
&= \frac{1}{6}.
\end{aligned}
$$

“Full” solution to Example 2.1 (3/4)

First the denominator:

$$f(x=1) = \int_\Theta \pi(\theta) f(x=1|\theta)\,d\theta \qquad (2.6)$$

$$= \int_0^1 1 \times 5\theta(1-\theta)^4\,d\theta$$

$$= \int_0^1 \theta \times 5(1-\theta)^4\,d\theta$$

which, using integration by parts, gives

$$f(x=1) = \left[-(1-\theta)^5\,\theta\right]_0^1 + \int_0^1 (1-\theta)^5\,d\theta$$

$$= 0 + \left[-\frac{(1-\theta)^6}{6}\right]_0^1$$

$$= \frac{1}{6}.$$

“Full” solution to Example 2.1 (4/4)

Therefore, the posterior density is

$$\begin{aligned}
\pi(\theta|x=1) &= \frac{\pi(\theta)f(x=1|\theta)}{f(x=1)} \\
&= \frac{1 \times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\
&= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\
&= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,
\end{aligned}$$

as

$$B(2,5) = \frac{(2-1)!(5-1)!}{(2+5-1)!} = \frac{24}{720} = \frac{1}{30},$$

and so the posterior distribution is $\theta|x=1 \sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta = 0.2$, and mean at $E[\theta|x=1] = 2/7 = 0.286$.

“Full” solution to Example 2.1 (4/4)

Therefore, the posterior density is

$$
\begin{aligned}
\pi(\theta|x=1) &= \frac{\pi(\theta)f(x=1|\theta)}{f(x=1)} \\
&= \frac{1\times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\
&= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\
&= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,
\end{aligned}
$$

as

$$B(2,5)=\frac{(2-1)!(5-1)!}{(2+5-1)!}=\frac{24}{720}=\frac{1}{30},$$

and so the posterior distribution is $\theta|x=1\sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta=0.2$, and mean at $E[\theta|x=1]=2/7=0.286$.

“Full” solution to Example 2.1 (4/4)

Therefore, the posterior density is

$$\begin{aligned}
\pi(\theta|x=1) &= \frac{\pi(\theta)f(x=1|\theta)}{f(x=1)} \\
&= \frac{1\times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\
&= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\
&= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,
\end{aligned}$$

as

$$B(2,5) = \frac{(2-1)!(5-1)!}{(2+5-1)!} = \frac{24}{720} = \frac{1}{30},$$

and so the posterior distribution is $\theta|x=1 \sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta=0.2$, and mean at $E[\theta|x=1]=2/7=0.286$.

"Full" solution to Example 2.1 (4/4)

Therefore, the posterior density is

$$\begin{aligned}\pi(\theta|x=1) &= \frac{\pi(\theta)f(x=1|\theta)}{f(x=1)} \\ &= \frac{1\times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\ &= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\ &= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,\end{aligned}$$

as

$$B(2,5)=\frac{(2-1)!(5-1)!}{(2+5-1)!}=\frac{24}{720}=\frac{1}{30},$$

and so the posterior distribution is $\theta|x=1 \sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta=0.2$, and mean at $E[\theta|x=1]=2/7=0.286$.

"Full" solution to Example 2.1 (4/4)

Therefore, the posterior density is

$$\begin{aligned}
\pi(\theta|x=1) &= \frac{\pi(\theta) f(x=1|\theta)}{f(x=1)} \\
&= \frac{1 \times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\
&= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\
&= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,
\end{aligned}$$

as

$$B(2,5) = \frac{(2-1)!(5-1)!}{(2+5-1)!} = \frac{24}{720} = \frac{1}{30},$$

and so the posterior distribution is $\theta|x=1 \sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta = 0.2$, and mean at $E[\theta|x=1] = 2/7 = 0.286$.

Therefore, the posterior density is

$$\begin{aligned}\pi(\theta|x=1) &= \frac{\pi(\theta)f(x=1|\theta)}{f(x=1)} \\ &= \frac{1\times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\ &= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\ &= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,\end{aligned}$$

as

$$B(2,5)=\frac{(2-1)!(5-1)!}{(2+5-1)!}=\frac{24}{720}=\frac{1}{30},$$

and so the posterior distribution is $\theta|x=1 \sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta=0.2$, and mean at $E[\theta|x=1]=2/7=0.286$.

“Full” solution to Example 2.1 (4/4)

Therefore, the posterior density is

$$\begin{aligned}
\pi(\theta|x=1) &= \frac{\pi(\theta)f(x=1|\theta)}{f(x=1)} \\
&= \frac{1 \times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\
&= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\
&= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,
\end{aligned}$$

as

$$B(2,5) = \frac{(2-1)!(5-1)!}{(2+5-1)!} = \frac{24}{720} = \frac{1}{30},$$

and so the posterior distribution is $\theta|x=1 \sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta = 0.2$, and mean at $E[\theta|x=1] = 2/7 = 0.286$.

"Full" solution to Example 2.1 (4/4)

Therefore, the posterior density is

$$\begin{aligned}
\pi(\theta|x=1) &= \frac{\pi(\theta)f(x=1|\theta)}{f(x=1)} \\
&= \frac{1 \times 5\theta(1-\theta)^4}{1/6}, \qquad 0<\theta<1 \\
&= 30\,\theta(1-\theta)^4, \qquad 0<\theta<1 \\
&= \frac{\theta^{2-1}(1-\theta)^{5-1}}{B(2,5)}, \qquad 0<\theta<1,
\end{aligned}$$

as

$$B(2,5) = \frac{(2-1)!(5-1)!}{(2+5-1)!} = \frac{24}{720} = \frac{1}{30},$$

and so the posterior distribution is $\theta|x=1 \sim Be(2,5)$ – see Equation 2.1. This distribution has its mode at $\theta = 0.2$, and mean at $E[\theta|x=1] = 2/7 = 0.286$.

Example 2.1

There were two "difficulties" in calculating the posterior distribution:

- Obtaining the $f(x)$ term for the denominator in the "full" version of Bayes' Theorem
- Manipulating the result to realise that we had a $Be(2, 5)$ distribution

Example 2.1

There were two “difficulties” in calculating the posterior distribution:

- Obtaining the $f(x)$ term for the denominator in the “full” version of Bayes’ Theorem
- Manipulating the result to realise that we had a $Be(2,5)$ distribution

Example 2.1

There were two "difficulties" in calculating the posterior distribution:

- Obtaining the $f(x)$ term for the denominator in the "full" version of Bayes' Theorem
- Manipulating the result to realise that we had a $Be(2,5)$ distribution

Example 2.1: "Quick" solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)f(x=1|\theta)\\ &\propto 1\times 5\theta(1-\theta)^4, \qquad 0<\theta<1\\ &= k\theta(1-\theta)^4, \qquad 0<\theta<1.\end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x=1 \sim Be(2,5)$.

Example 2.1: "Quick" solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)f(x=1|\theta)\\ &\propto 1\times 5\theta(1-\theta)^4, \qquad 0<\theta<1\\ &= k\theta(1-\theta)^4, \qquad 0<\theta<1.\end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x=1 \sim Be(2,5)$.

Example 2.1: "Quick" solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(x=1|\theta)\\ &\propto 1 \times 5\theta(1-\theta)^4, \qquad 0<\theta<1\\ &= k\theta(1-\theta)^4, \qquad 0<\theta<1.\end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x=1 \sim Be(2,5)$.

Example 2.1: “Quick” solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(x=1|\theta) \\
&\propto 1 \times 5\theta(1-\theta)^4, \qquad 0 < \theta < 1 \\
&= k\theta(1-\theta)^4, \qquad 0 < \theta < 1.
\end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x = 1 \sim Be(2,5)$.

Example 2.1: "Quick" solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(x=1|\theta) \\ &\propto 1 \times 5\theta(1-\theta)^4, \qquad 0<\theta<1 \\ &= k\theta(1-\theta)^4, \qquad 0<\theta<1. \end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x=1 \sim Be(2,5)$.

Example 2.1: "Quick" solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)f(x=1|\theta) \\ &\propto 1 \times 5\theta(1-\theta)^4, \qquad 0<\theta<1 \\ &= k\theta(1-\theta)^4, \qquad 0<\theta<1. \end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x=1 \sim Be(2,5)$.

Example 2.1: "Quick" solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)f(x=1|\theta)\\ &\propto 1\times 5\theta(1-\theta)^4, \qquad 0<\theta<1\\ &= k\theta(1-\theta)^4, \qquad 0<\theta<1.\end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x=1 \sim Be(2,5)$.

Example 2.1: "Quick" solution

However, in many cases we can recognise the posterior distribution without the need to calculate $f(x)$.

In this example, we can calculate the posterior distribution as

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(x=1|\theta) \\ &\propto 1 \times 5\theta(1-\theta)^4, \qquad 0<\theta<1 \\ &= k\theta(1-\theta)^4, \qquad 0<\theta<1. \end{aligned}$$

You should be able to recognise this as a $Be(2,5)$ distribution – we can re–write the above as

$$\pi(\theta|\boldsymbol{x}) = k\theta^{2-1}(1-\theta)^{5-1},$$

i.e. $\theta|x=1 \sim Be(2,5)$.

It is possible that we have a biased coin.

Recall that if $Y \sim U(a,b)$,

$$E(Y) = \frac{a+b}{2} \qquad \text{and} \qquad Var(Y) = \frac{(b-a)^2}{12}.$$

Also, if $Y \sim Be(a,b)$,

$$E(Y) = \frac{a}{a+b} \qquad \text{and} \qquad Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Thus:

	Prior $U(0,1)$	Observed $\theta = 0.2$	Posterior $Be(2,5)$
Mean	0.5	$\longmapsto$	0.286
St. dev.	0.289	$\longmapsto$	0.160

Example 2.1: Summary

It is possible that we have a biased coin.

Recall that if $Y \sim U(a, b)$,

$$E(Y) = \frac{a+b}{2} \qquad \text{and} \qquad Var(Y) = \frac{(b-a)^2}{12}.$$

Also, if $Y \sim Be(a, b)$,

$$E(Y) = \frac{a}{a+b} \qquad \text{and} \qquad Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Thus:

	Prior $U(0,1)$	Observed $\theta = 0.2$	Posterior $Be(2,5)$
Mean	0.5	$\longmapsto$	0.286
St. dev.	0.289	$\longmapsto$	0.160

Example 2.1: Summary

It is possible that we have a biased coin.

Recall that if $Y \sim U(a, b)$,

$$E(Y) = \frac{a+b}{2} \qquad \text{and} \qquad Var(Y) = \frac{(b-a)^2}{12}.$$

Also, if $Y \sim Be(a, b)$,

$$E(Y) = \frac{a}{a+b} \qquad \text{and} \qquad Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Thus:

	Prior $U(0,1)$	Observed $\theta = 0.2$	Posterior $Be(2,5)$
Mean	0.5	$\longmapsto$	0.286
St. dev.	0.289	$\longmapsto$	0.160

Example 2.1: Summary

It is possible that we have a biased coin.

Recall that if $Y \sim U(a, b)$,

$$E(Y) = \frac{a+b}{2} \qquad \text{and} \qquad Var(Y) = \frac{(b-a)^2}{12}.$$

Also, if $Y \sim Be(a, b)$,

$$E(Y) = \frac{a}{a+b} \qquad \text{and} \qquad Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Thus:

	Prior $U(0,1)$	Observed $\theta = 0.2$	Posterior $Be(2,5)$
Mean	0.5	$\longmapsto$	0.286
St. dev.	0.289	$\longmapsto$	0.160

Example 2.1: Summary

It is possible that we have a biased coin.

Recall that if $Y \sim U(a, b)$,

$$E(Y) = \frac{a+b}{2} \qquad \text{and} \qquad Var(Y) = \frac{(b-a)^2}{12}.$$

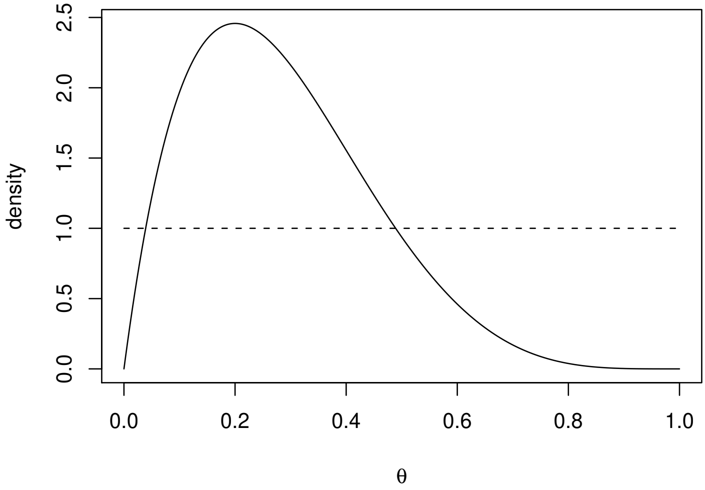
Also, if $Y \sim Be(a, b)$,

$$E(Y) = \frac{a}{a+b} \qquad \text{and} \qquad Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Thus:

	Prior $U(0,1)$	Observed $\theta = 0.2$	Posterior $Be(2,5)$
Mean	0.5	$\longmapsto$	0.286
St. dev.	0.289	$\longmapsto$	0.160

Example 2.1: Summary



Example 2.2

Consider an experiment to determine how good a music expert is at distinguishing between pages from Haydn and Mozart scores. Let $\theta = \Pr(\text{correct choice})$.

Suppose that, before conducting the experiment, we have been told that the expert is very competent.

In fact, it is suggested that we should have a prior distribution which has a mode around $\theta = 0.95$ and for which $\Pr(\theta < 0.8)$ is very small.

We choose $\theta \sim Be(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

$$\begin{aligned} \pi(\theta) &= \frac{\theta^{76}(1-\theta)^4}{B(77,5)}, \qquad 0 < \theta < 1 \\ &= 128107980\theta^{76}(1-\theta)^4, \qquad 0 < \theta < 1. \qquad (2.7) \end{aligned}$$

Example 2.2

Consider an experiment to determine how good a music expert is at distinguishing between pages from Haydn and Mozart scores. Let $\theta = \text{Pr(correct choice)}$.

Suppose that, before conducting the experiment, we have been told that the expert is very competent.

In fact, it is suggested that we should have a prior distribution which has a mode around $\theta = 0.95$ and for which $\Pr(\theta < 0.8)$ is very small.

We choose $\theta \sim Be(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

$$\begin{aligned}\pi(\theta) &= \frac{\theta^{76}(1-\theta)^4}{B(77,5)}, \qquad 0 < \theta < 1\\ &= 128107980\theta^{76}(1-\theta)^4, \qquad 0 < \theta < 1. \qquad (2.7)\end{aligned}$$

Example 2.2

Consider an experiment to determine how good a music expert is at distinguishing between pages from Haydn and Mozart scores. Let $\theta = \text{Pr(correct choice)}$.

Suppose that, before conducting the experiment, we have been told that the expert is very competent.

In fact, it is suggested that we should have a prior distribution which has a mode around $\theta = 0.95$ and for which $\Pr(\theta < 0.8)$ is very small.

We choose $\theta \sim Be(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

$$\begin{aligned}\pi(\theta) &= \frac{\theta^{76}(1-\theta)^4}{B(77,5)}, \qquad 0 < \theta < 1 \\ &= 128107980\theta^{76}(1-\theta)^4, \qquad 0 < \theta < 1. \qquad (2.7)\end{aligned}$$

Example 2.2

Consider an experiment to determine how good a music expert is at distinguishing between pages from Haydn and Mozart scores. Let $\theta = \text{Pr(correct choice)}$.

Suppose that, before conducting the experiment, we have been told that the expert is very competent.

In fact, it is suggested that we should have a prior distribution which has a mode around $\theta = 0.95$ and for which $\Pr(\theta < 0.8)$ is very small.

We choose $\theta \sim Be(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

$$\begin{aligned}\pi(\theta) &= \frac{\theta^{76}(1-\theta)^4}{B(77,5)}, \qquad 0 < \theta < 1 \\ &= 128107980\theta^{76}(1-\theta)^4, \qquad 0 < \theta < 1. \qquad (2.7)\end{aligned}$$

Example 2.2

Consider an experiment to determine how good a music expert is at distinguishing between pages from Haydn and Mozart scores. Let $\theta = \text{Pr}(\text{correct choice})$.

Suppose that, before conducting the experiment, we have been told that the expert is very competent.

In fact, it is suggested that we should have a prior distribution which has a mode around $\theta = 0.95$ and for which $\text{Pr}(\theta < 0.8)$ is very small.

We choose $\theta \sim Be(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

$$\pi(\theta) = \frac{\theta^{76}(1-\theta)^4}{B(77,5)}, \qquad 0 < \theta < 1$$

$$= 128107980\theta^{76}(1-\theta)^4, \qquad 0 < \theta < 1. \qquad (2.7)$$

Example 2.2

Consider an experiment to determine how good a music expert is at distinguishing between pages from Haydn and Mozart scores. Let $\theta = \Pr(\text{correct choice})$.

Suppose that, before conducting the experiment, we have been told that the expert is very competent.

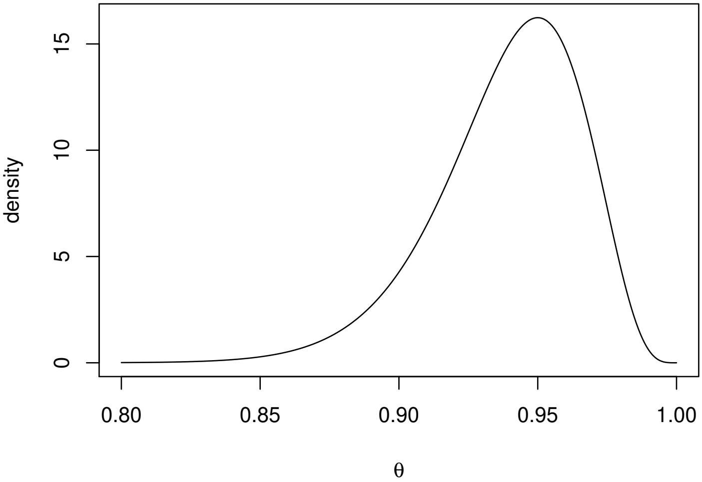
In fact, it is suggested that we should have a prior distribution which has a mode around $\theta = 0.95$ and for which $\Pr(\theta < 0.8)$ is very small.

We choose $\theta \sim Be(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

$$\begin{aligned} \pi(\theta) &= \frac{\theta^{76}(1-\theta)^4}{B(77,5)}, \qquad 0 < \theta < 1 \\ &= 128107980\theta^{76}(1-\theta)^4, \qquad 0 < \theta < 1. \end{aligned} \tag{2.7}$$

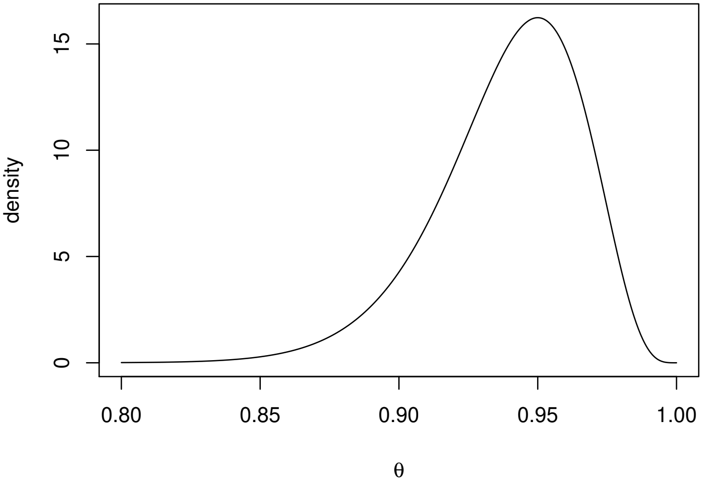
Example 2.2

A graph of this prior density is given in Figure 2.4.



Example 2.2

A graph of this prior density is given in Figure 2.4.



Example 2.2

In the experiment, the music expert makes the correct choice 9 out of 10 times.

Determine the posterior distribution for θ given this information.

Example 2.2

In the experiment, the music expert makes the correct choice 9 out of 10 times.

Determine the posterior distribution for θ given this information.

Solution to Example 2.2 (1/2)

We have an observation on the random variable $X|\theta \sim Bin(10, \theta)$.

This gives a likelihood function of

$$f(x = 9|\theta) = {}^{10}C_9 \theta^9 (1-\theta)^1$$

$$= 10\theta^9(1-\theta) \qquad (2.8)$$

which favours values of θ near its maximum $\theta = 0.9$.

We combine these two sources of information using Bayes' Theorem.

Solution to Example 2.2 (1/2)

We have an observation on the random variable $X|\theta \sim Bin(10, \theta)$.

This gives a likelihood function of

$$f(x = 9|\theta) = {}^{10}C_9\theta^9(1-\theta)^1$$

$$= 10\theta^9(1-\theta) \qquad (2.8)$$

which favours values of θ near its maximum $\theta = 0.9$.

We combine these two sources of information using Bayes' Theorem.

Solution to Example 2.2 (1/2)

We have an observation on the random variable $X|\theta \sim Bin(10, \theta)$.

This gives a likelihood function of

$$f(x = 9|\theta) = {}^{10}C_9 \theta^9 (1-\theta)^1$$

$$= 10\theta^9(1-\theta) \qquad (2.8)$$

which favours values of θ near its maximum $\theta = 0.9$.

We combine these two sources of information using Bayes' Theorem.

Solution to Example 2.2 (1/2)

We have an observation on the random variable $X|\theta \sim Bin(10, \theta)$.

This gives a likelihood function of

$$f(x = 9|\theta) = {}^{10}C_9\theta^9(1-\theta)^1$$

$$= 10\theta^9(1-\theta) \qquad (2.8)$$

which favours values of θ near its maximum $\theta = 0.9$.

We combine these two sources of information using Bayes' Theorem.

Solution to Example 2.2 (2/2)

The posterior density function is

$$\pi(\theta|x=9) \propto \pi(\theta) f(x=9|\theta)$$

$$\propto \frac{\theta^{76}(1-\theta)^4}{B(77,5)} \times 10\,\theta^9(1-\theta), \qquad 0<\theta<1$$

$$= 128107980\theta^{76}(1-\theta)^4 \times 10\,\theta^9(1-\theta), \qquad 0<\theta<1$$

$$= k\theta^{85}(1-\theta)^5, \qquad 0<\theta<1. \tag{2.9}$$

We can recognise this density function as one from the Beta family. In fact, the posterior distribution is $\theta|x=9 \sim Be(86,6)$.

Solution to Example 2.2 (2/2)

The posterior density function is

$$\pi(\theta|x=9) \propto \pi(\theta)f(x=9|\theta)$$

$$\propto \frac{\theta^{76}(1-\theta)^4}{B(77,5)} \times 10\,\theta^9(1-\theta), \qquad 0<\theta<1$$

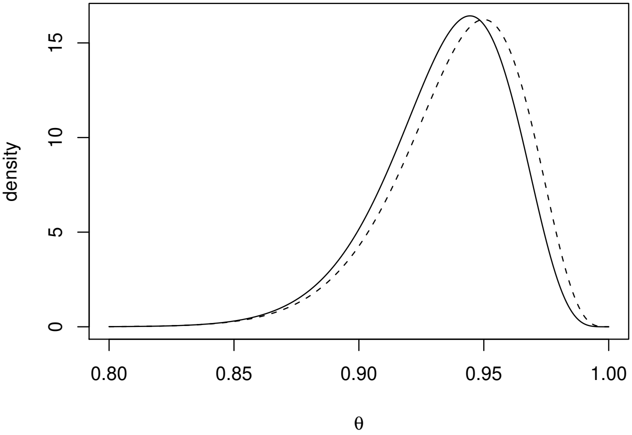
$$= 128107980\theta^{76}(1-\theta)^4 \times 10\,\theta^9(1-\theta), \qquad 0<\theta<1$$

$$= k\theta^{85}(1-\theta)^5, \qquad 0<\theta<1. \qquad (2.9)$$

We can recognise this density function as one from the Beta family. In fact, the posterior distribution is $\theta|x=9 \sim Be(86,6)$.

Example 2.2: Summary

The changes in our beliefs about θ are described by the prior and posterior distributions shown in Figure 2.5 and summarised in Table 2.1.



Example 2.2: Summary

	Prior (2.7)	Likelihood (2.8)	Posterior (2.9)
Mode(θ)	0.950	0.900	0.944
$E(\theta)$	0.939	–	0.935
$SD(\theta)$	0.0263	–	0.0256

Notice that, having observed only a 90% success rate in the experiment, the posterior mode and mean are smaller than their prior values.

Also, the experiment has largely confirmed our ideas about θ, with the uncertainty about θ being only very slightly reduced.

Example 2.2: Summary

	Prior (2.7)	Likelihood (2.8)	Posterior (2.9)
$Mode(\theta)$	0.950	0.900	0.944
$E(\theta)$	0.939	–	0.935
$SD(\theta)$	0.0263	–	0.0256

Notice that, having observed only a 90% success rate in the experiment, the posterior mode and mean are smaller than their prior values.

Also, the experiment has largely confirmed our ideas about θ, with the uncertainty about θ being only very slightly reduced.

Example 2.2: Summary

	Prior (2.7)	Likelihood (2.8)	Posterior (2.9)
$Mode(\theta)$	0.950	0.900	0.944
$E(\theta)$	0.939	–	0.935
$SD(\theta)$	0.0263	–	0.0256

Notice that, having observed only a 90% success rate in the experiment, the posterior mode and mean are smaller than their prior values.

Also, the experiment has largely confirmed our ideas about θ, with the uncertainty about θ being only very slightly reduced.

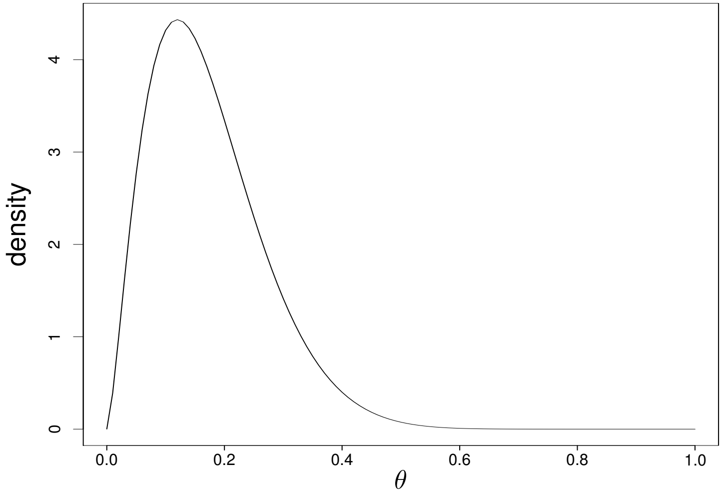
Example 2.3

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

Based on the proportion of customers who have bought similarly violent games from him in the past, he assumes that $\theta \sim Be(2.5, 12)$ (see Example 3.3, Chapter 3).

Example 2.3

A plot of this prior density is shown in Figure 2.6.



Example 2.3

Max asks five potential customers if they would buy *Call of Duty: Elite* from him, and four say they would.

Using this information, what is Max's posterior distribution for θ?

Solution to Example 2.3 (1/2)

We have been told that the prior for θ is a *Be*$(2.5, 12)$ distribution – this has density given by

$$\pi(\theta) = \frac{\theta^{2.5-1}(1-\theta)^{12-1}}{B(2.5, 12)}$$

$$= 435.1867\theta^{1.5}(1-\theta)^{11}. \qquad (2.10)$$

We have an observation on the random variable $X|\theta \sim Bin(5, \theta)$. This gives a likelihood function of

$$f(x=4|\theta) = {}^5C_4\theta^4(1-\theta)^1$$

$$= 5\theta^4(1-\theta), \qquad (2.11)$$

which favours values of θ near its maximum 0.8.

Solution to Example 2.3 (1/2)

We have been told that the prior for θ is a $Be(2.5, 12)$ distribution – this has density given by

$$\pi(\theta) = \frac{\theta^{2.5-1}(1-\theta)^{12-1}}{B(2.5, 12)}$$

$$= 435.1867\theta^{1.5}(1-\theta)^{11}. \qquad (2.10)$$

We have an observation on the random variable $X|\theta \sim Bin(5, \theta)$. This gives a likelihood function of

$$f(x=4|\theta) = {}^5C_4\theta^4(1-\theta)^1$$

$$= 5\theta^4(1-\theta), \qquad (2.11)$$

which favours values of θ near its maximum 0.8.

Solution to Example 2.3 (1/2)

We have been told that the prior for θ is a $Be(2.5, 12)$ distribution – this has density given by

$$\pi(\theta) = \frac{\theta^{2.5-1}(1-\theta)^{12-1}}{B(2.5, 12)}$$

$$= 435.1867\theta^{1.5}(1-\theta)^{11}. \quad (2.10)$$

We have an observation on the random variable $X|\theta \sim Bin(5, \theta)$. This gives a likelihood function of

$$f(x=4|\theta) = {}^5C_4\theta^4(1-\theta)^1$$

$$= 5\theta^4(1-\theta), \quad (2.11)$$

which favours values of θ near its maximum 0.8.

Solution to Example 2.3 (2/2)

We combine our prior information (2.10) with the data (2.11) – to obtain our posterior distribution – using Bayes' Theorem.

The posterior density function is

$$\pi(\theta|x=4) \propto \pi(\theta)f(x=4|\theta)$$

$$\propto 435.1867\theta^{1.5}(1-\theta)^{11} \times 5\theta^4(1-\theta), \qquad 0<\theta<1,$$

$$= k\theta^{5.5}(1-\theta)^{12}, \qquad 0<\theta<1. \qquad (2.12)$$

You should recognise this density function as one from the beta family. In fact, we have a $Be(6.5, 13)$, i.e. $\theta|x=4 \sim Be(6.5, 13)$.

Solution to Example 2.3 (2/2)

We combine our prior information (2.10) with the data (2.11) – to obtain our posterior distribution – using Bayes' Theorem.

The posterior density function is

$\pi(\theta|x=4) \propto \pi(\theta)f(x=4|\theta)$

$$\propto 435.1867\theta^{1.5}(1-\theta)^{11} \times 5\theta^4(1-\theta), \qquad 0<\theta<1,$$

$$= k\theta^{5.5}(1-\theta)^{12}, \qquad 0<\theta<1. \qquad (2.12)$$

You should recognise this density function as one from the beta family. In fact, we have a $Be(6.5, 13)$, i.e. $\theta|x=4 \sim Be(6.5, 13)$.

Solution to Example 2.3 (2/2)

We combine our prior information (2.10) with the data (2.11) – to obtain our posterior distribution – using Bayes' Theorem.

The posterior density function is

$$\begin{aligned}\pi(\theta|x=4) &\propto \pi(\theta)f(x=4|\theta) \\ &\propto 435.1867\theta^{1.5}(1-\theta)^{11} \times 5\theta^{4}(1-\theta), \qquad 0<\theta<1, \\ &= k\theta^{5.5}(1-\theta)^{12}, \qquad 0<\theta<1. \qquad (2.12)\end{aligned}$$

You should recognise this density function as one from the beta family. In fact, we have a $Be(6.5, 13)$, i.e. $\theta|x=4 \sim Be(6.5, 13)$.

Solution to Example 2.3 (2/2)

We combine our prior information (2.10) with the data (2.11) – to obtain our posterior distribution – using Bayes' Theorem.

The posterior density function is

$$\begin{aligned}\pi(\theta|x=4) &\propto \pi(\theta)f(x=4|\theta)\\ &\propto 435.1867\theta^{1.5}(1-\theta)^{11}\times 5\theta^{4}(1-\theta), \qquad 0<\theta<1,\\ &= k\theta^{5.5}(1-\theta)^{12}, \qquad 0<\theta<1. \qquad (2.12)\end{aligned}$$

You should recognise this density function as one from the beta family. In fact, we have a $Be(6.5, 13)$, i.e. $\theta|x=4 \sim Be(6.5, 13)$.

Solution to Example 2.3 (2/2)

We combine our prior information (2.10) with the data (2.11) – to obtain our posterior distribution – using Bayes' Theorem.

The posterior density function is

$$\pi(\theta|x=4) \propto \pi(\theta)f(x=4|\theta)$$

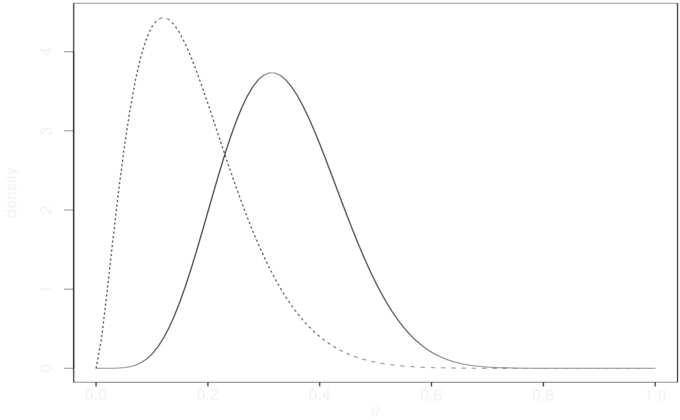
$$\propto 435.1867\theta^{1.5}(1-\theta)^{11} \times 5\theta^4(1-\theta), \qquad 0<\theta<1,$$

$$= k\theta^{5.5}(1-\theta)^{12}, \qquad 0<\theta<1. \tag{2.12}$$

You should recognise this density function as one from the beta family. In fact, we have a $Be(6.5, 13)$, i.e. $\theta|x=4 \sim Be(6.5, 13)$.

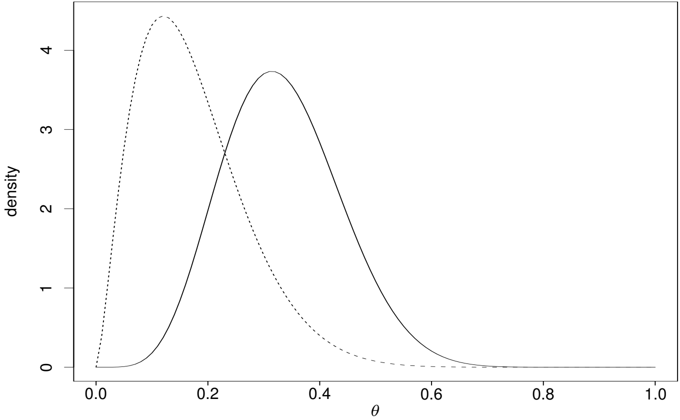
Example 2.3: Summary

The changes in our beliefs about θ are described by the prior and posterior distributions shown in Figure 2.7 and summarised in Table 2.2.



Example 2.3: Summary

The changes in our beliefs about θ are described by the prior and posterior distributions shown in Figure 2.7 and summarised in Table 2.2.



Example 2.3: Summary

	Prior (2.10)	Likelihood (2.11)	Posterior (2.12)
$Mode(\theta)$	0.12	0.8	0.314
$E(\theta)$	0.172	–	0.333
$SD(\theta)$	0.096	–	0.104

Notice how the posterior has been "pulled" from the prior towards the observed value: the mode has moved up from 0.12 to 0.314, and the mean has moved up from 0.172 to 0.333.

Having just one observation in the likelihood, we see that there is hardly any change in the standard deviation from prior to posterior: we would expect to see a decrease in standard deviation with the addition of more data values.

Example 2.3: Summary

	Prior (2.10)	Likelihood (2.11)	Posterior (2.12)
$Mode(\theta)$	0.12	0.8	0.314
$E(\theta)$	0.172	–	0.333
$SD(\theta)$	0.096	–	0.104

Notice how the posterior has been "pulled" from the prior towards the observed value: the mode has moved up from 0.12 to 0.314, and the mean has moved up from 0.172 to 0.333.

Having just one observation in the likelihood, we see that there is hardly any change in the standard deviation from prior to posterior: we would expect to see a decrease in standard deviation with the addition of more data values.

Example 2.3: Summary

	Prior (2.10)	Likelihood (2.11)	Posterior (2.12)
$Mode(\theta)$	0.12	0.8	0.314
$E(\theta)$	0.172	–	0.333
$SD(\theta)$	0.096	–	0.104

Notice how the posterior has been "pulled" from the prior towards the observed value: the mode has moved up from 0.12 to 0.314, and the mean has moved up from 0.172 to 0.333.

Having just one observation in the likelihood, we see that there is hardly any change in the standard deviation from prior to posterior: we would expect to see a decrease in standard deviation with the addition of more data values.

Announcements

- Assignment drop-in this Thursday, 2pm (questions 4, 9, 10, 11, 12). Good place for hints, tips etc.
- First computer practical this Friday, 1pm, Herschel PC cluster
- ReCap problems...?
- Next week: 2 lectures and a problems class (case study 1)

Example 2.4

Table 2.3 shows some data on the times between serious earthquakes.

An earthquake is included if its magnitude is at least 7.5 on the Richter scale or if over 1000 people were killed.

Recording starts on 16 December 1902 (4500 killed in Turkistan).

The table includes data on 21 earthquakes, that is, 20 “waiting times” between earthquakes.

840	157	145	44	33	121	150	280	434	736
584	887	263	1901	695	294	562	721	76	710

Example 2.4

It is believed that earthquakes happen in a random haphazard kind of way and that times between earthquakes can be described by an exponential distribution.

Data over a much longer period suggest that this exponential assumption is plausible.

Therefore, we will assume that these data are a random sample from an exponential distribution with rate θ (and mean $1/\theta$). The parameter θ describes the rate at which earthquakes occur.

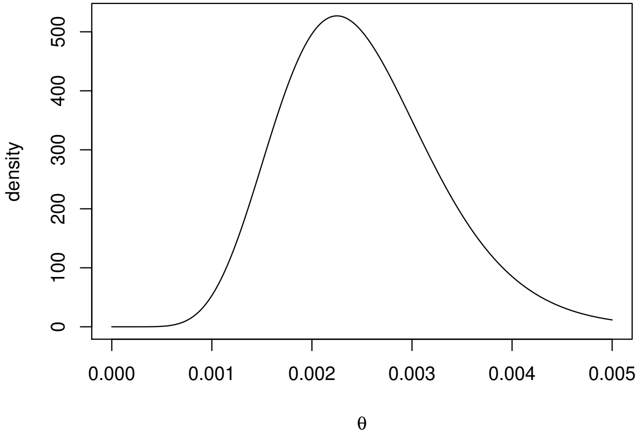
Example 2.4

An expert on earthquakes has prior beliefs about the rate of earthquakes, θ, described by a $Ga(10, 4000)$ distribution (see Example 3.1, Chapter 3), which has density density

$$\pi(\theta) = \frac{4000^{10}\,\theta^{9} e^{-4000\theta}}{\Gamma(10)}, \quad \theta > 0, \qquad (2.13)$$

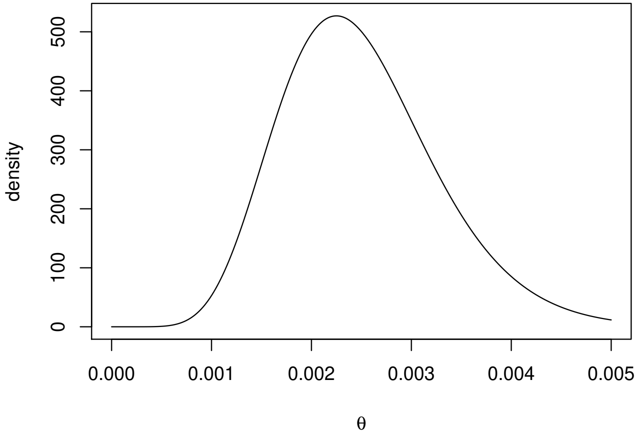
and mean $E(\theta) = 0.0025$.

A plot of this prior distribution can be found in Figure 2.8. As you might expect, the expert believes that only very small values of θ are likely, though larger values are not ruled out!



Example 2.4

A plot of this prior distribution can be found in Figure 2.8. As you might expect, the expert believes that only very small values of θ are likely, though larger values are not ruled out!



Solution to Example 2.4 (1/2)

The data are observations on $X_i|\theta \sim Exp(\theta)$, $i = 1, 2, \ldots, 20$ (independent).

Therefore, the likelihood function for θ is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{20} \theta e^{-\theta x_i}, \qquad \theta > 0$$

$$= \theta^{20} \exp\left(-\theta \sum_{i=1}^{20} x_i\right), \qquad \theta > 0$$

$$= \theta^{20} e^{-9633\theta}, \qquad \theta > 0. \qquad (2.14)$$

The data are observations on $X_i|\theta \sim Exp(\theta)$, $i = 1, 2, \ldots, 20$ (independent).

Therefore, the likelihood function for θ is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{20} \theta e^{-\theta x_i}, \qquad \theta > 0$$

$$= \theta^{20} \exp\left(-\theta \sum_{i=1}^{20} x_i\right), \qquad \theta > 0$$

$$= \theta^{20} e^{-9633\theta}, \qquad \theta > 0. \qquad (2.14)$$

Solution to Example 2.4 (1/2)

The data are observations on $X_i|\theta \sim Exp(\theta)$, $i = 1, 2, \ldots, 20$ (independent).

Therefore, the likelihood function for θ is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{20} \theta e^{-\theta x_i}, \qquad \theta > 0$$

$$= \theta^{20} \exp\left(-\theta \sum_{i=1}^{20} x_i\right), \qquad \theta > 0$$

$$= \theta^{20} e^{-9633\theta}, \qquad \theta > 0. \tag{2.14}$$

Solution to Example 2.4 (2/2)

We now apply Bayes Theorem to combine the expert opinion with the observed data. The posterior density function is

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

$$\propto \frac{4000^{10}\,\theta^{9} e^{-4000\theta}}{\Gamma(10)} \times \theta^{20} e^{-9633\theta}, \qquad \theta > 0$$

$$= k\,\theta^{30-1} e^{-13633\theta}, \qquad \theta > 0. \qquad (2.15)$$

The only continuous distribution which takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ is the $Ga(g, h)$ distribution.

Therefore, the posterior distribution must be $\theta|\boldsymbol{x} \sim Ga(30, 13633)$.

Solution to Example 2.4 (2/2)

We now apply Bayes Theorem to combine the expert opinion with the observed data. The posterior density function is

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

$$\propto \frac{4000^{10}\,\theta^{9} e^{-4000\theta}}{\Gamma(10)} \times \theta^{20} e^{-9633\theta}, \qquad \theta > 0$$

$$= k\,\theta^{30-1} e^{-13633\theta}, \qquad \theta > 0. \tag{2.15}$$

The only continuous distribution which takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ is the $Ga(g, h)$ distribution.

Therefore, the posterior distribution must be $\theta|\boldsymbol{x} \sim Ga(30, 13633)$.

Solution to Example 2.4 (2/2)

We now apply Bayes Theorem to combine the expert opinion with the observed data. The posterior density function is

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

$$\propto \frac{4000^{10}\,\theta^{9} e^{-4000\theta}}{\Gamma(10)} \times \theta^{20} e^{-9633\theta}, \qquad \theta > 0$$

$$= k\,\theta^{30-1} e^{-13633\theta}, \qquad \theta > 0. \tag{2.15}$$

The only continuous distribution which takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ is the $Ga(g, h)$ distribution.

Therefore, the posterior distribution must be $\theta|\boldsymbol{x} \sim Ga(30, 13633)$.

Solution to Example 2.4 (2/2)

We now apply Bayes Theorem to combine the expert opinion with the observed data. The posterior density function is

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

$$\propto \frac{4000^{10}\,\theta^{9} e^{-4000\theta}}{\Gamma(10)} \times \theta^{20} e^{-9633\theta}, \qquad \theta > 0$$

$$= k\,\theta^{30-1} e^{-13633\theta}, \qquad \theta > 0. \tag{2.15}$$

The only continuous distribution which takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ is the $Ga(g,h)$ distribution.

Therefore, the posterior distribution must be $\theta|\boldsymbol{x} \sim Ga(30, 13633)$.

Example 2.4: Summary

The data have updated our beliefs about θ from a $Ga(10, 4000)$ distribution to a $Ga(30, 13633)$ distribution.

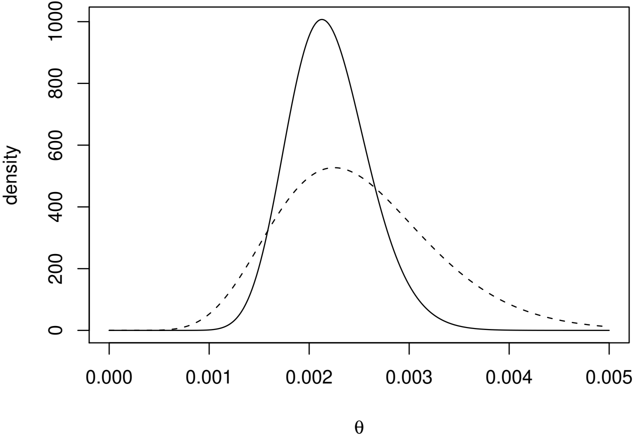
Plots of these distributions are given in Figure 2.9, and Table 2.4 gives a summary of the main changes induced by incorporating the data.

Example 2.4: Summary

The data have updated our beliefs about θ from a $Ga(10, 4000)$ distribution to a $Ga(30, 13633)$ distribution.

Plots of these distributions are given in Figure 2.9, and Table 2.4 gives a summary of the main changes induced by incorporating the data.

Example 2.4: Summary



Example 2.4: Summary

	Prior (2.13)	Likelihood (2.14)	Posterior (2.15)
Mode(θ)	0.00225	0.00208	0.00213
$E(\theta)$	0.00250	–	0.00220
$SD(\theta)$	0.00079	–	0.00040

Notice that, as the mode of the likelihood function is close to that of the prior distribution, the information in the data is consistent with that in the prior distribution.

This results in a reduction in variability from the prior to the posterior distributions.

The similarity between the prior beliefs and the data has reduced the uncertainty we have about the likely earthquake rate θ.

Example 2.4: Summary

	Prior (2.13)	Likelihood (2.14)	Posterior (2.15)
Mode(θ)	0.00225	0.00208	0.00213
$E(\theta)$	0.00250	–	0.00220
$SD(\theta)$	0.00079	–	0.00040

Notice that, as the mode of the likelihood function is close to that of the prior distribution, the information in the data is consistent with that in the prior distribution.

This results in a reduction in variability from the prior to the posterior distributions.

The similarity between the prior beliefs and the data has reduced the uncertainty we have about the likely earthquake rate θ.

Example 2.4: Summary

	Prior (2.13)	Likelihood (2.14)	Posterior (2.15)
Mode(θ)	0.00225	0.00208	0.00213
$E(\theta)$	0.00250	–	0.00220
$SD(\theta)$	0.00079	–	0.00040

Notice that, as the mode of the likelihood function is close to that of the prior distribution, the information in the data is consistent with that in the prior distribution.

This results in a reduction in variability from the prior to the posterior distributions.

The similarity between the prior beliefs and the data has reduced the uncertainty we have about the likely earthquake rate θ.

Example 2.4: Summary

	Prior (2.13)	Likelihood (2.14)	Posterior (2.15)
Mode(θ)	0.00225	0.00208	0.00213
E(θ)	0.00250	–	0.00220
SD(θ)	0.00079	–	0.00040

Notice that, as the mode of the likelihood function is close to that of the prior distribution, the information in the data is consistent with that in the prior distribution.

This results in a reduction in variability from the prior to the posterior distributions.

The similarity between the prior beliefs and the data has reduced the uncertainty we have about the likely earthquake rate θ.

Example 2.5

We now consider the general case of the problem discussed in Example 2.4.

Suppose $X_i|\theta \sim Exp(\theta)$, $i = 1, 2, \ldots, n$ (independent) and our prior beliefs about θ are summarised by a $Ga(g, h)$ distribution (with g and h known), with density

$$\pi(\theta) = \frac{h^g\, \theta^{g-1}\, e^{-h\theta}}{\Gamma(g)}, \quad \theta > 0. \tag{2.16}$$

Determine the posterior distribution for θ.

Solution to Example 2.5 (1/2)

The likelihood function for θ is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i}, \qquad \theta > 0$$

$$= \theta^n e^{-\theta x_1 - \theta x_2 - \ldots - \theta x_n}$$

$$= \theta^n e^{-\theta \sum x}$$

$$= \theta^n e^{-n\bar{x}\theta}, \qquad \theta > 0. \qquad (2.17)$$

Solution to Example 2.5 (2/2)

We now apply Bayes Theorem. The posterior density function is

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

$$\propto \frac{h^g\, \theta^{g-1}\, e^{-h\theta}}{\Gamma(g)} \times \theta^n e^{-n\bar{x}\theta}, \qquad \theta > 0.$$

$$\pi(\theta|\boldsymbol{x}) = k\theta^{g+n-1} e^{-(h+n\bar{x})\theta}, \qquad \theta > 0, \qquad (2.18)$$

where k is a constant that does not depend on θ.

Therefore, the posterior distribution takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ and so must be a gamma distribution.

Thus we have $\theta|\boldsymbol{x} \sim Ga(g+n, h+n\bar{x})$.

Solution to Example 2.5 (2/2)

We now apply Bayes Theorem. The posterior density function is

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

$$\propto \frac{h^g\,\theta^{g-1}\,e^{-h\theta}}{\Gamma(g)} \times \theta^n e^{-n\bar{x}\theta}, \qquad \theta > 0.$$

$$\pi(\theta|\boldsymbol{x}) = k\theta^{g+n-1} e^{-(h+n\bar{x})\theta}, \qquad \theta > 0, \qquad (2.18)$$

where k is a constant that does not depend on θ.

Therefore, the posterior distribution takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ and so must be a gamma distribution.

Thus we have $\theta|\boldsymbol{x} \sim Ga(g+n, h+n\bar{x})$.

Solution to Example 2.5 (2/2)

We now apply Bayes Theorem. The posterior density function is

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

$$\propto \frac{h^g\,\theta^{g-1}e^{-h\theta}}{\Gamma(g)} \times \theta^n e^{-n\bar{x}\theta}, \qquad \theta > 0.$$

$$\pi(\theta|\boldsymbol{x}) = k\theta^{g+n-1}e^{-(h+n\bar{x})\theta}, \qquad \theta > 0, \qquad (2.18)$$

where k is a constant that does not depend on θ.

Therefore, the posterior distribution takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ and so must be a gamma distribution.

Thus we have $\theta|\boldsymbol{x} \sim Ga(g+n, h+n\bar{x})$.

Solution to Example 2.5 (2/2)

We now apply Bayes Theorem. The posterior density function is

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

$$\propto \frac{h^g\,\theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \theta^n e^{-n\bar{x}\theta}, \qquad \theta > 0.$$

$$\pi(\theta|\boldsymbol{x}) = k\theta^{g+n-1} e^{-(h+n\bar{x})\theta}, \qquad \theta > 0, \tag{2.18}$$

where k is a constant that does not depend on θ.

Therefore, the posterior distribution takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ and so must be a gamma distribution.

Thus we have $\theta|\boldsymbol{x} \sim Ga(g+n, h+n\bar{x})$.

Solution to Example 2.5 (2/2)

We now apply Bayes Theorem. The posterior density function is

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

$$\propto \frac{h^g\, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \theta^n e^{-n\bar{x}\theta}, \qquad \theta > 0.$$

$$\pi(\theta|\boldsymbol{x}) = k\theta^{g+n-1} e^{-(h+n\bar{x})\theta}, \qquad \theta > 0, \tag{2.18}$$

where k is a constant that does not depend on θ.

Therefore, the posterior distribution takes the form $k\theta^{g-1}e^{-h\theta}$, $\theta > 0$ and so must be a gamma distribution.

Thus we have $\theta|\boldsymbol{x} \sim Ga(g+n, h+n\bar{x})$.

Example 2.5: Summary

If

- $X_i \sim Exp(\theta)$ and
- $\theta \sim Ga(g, h)$, then
- $\theta|\boldsymbol{x} \sim Ga(g + n, h + n\bar{x})$

The changes in our beliefs about θ are summarised in Table 2.5, taking $g \geq 1$.

	Prior (2.16)	Likelihood (2.17)	Posterior (2.18)
$Mode(\theta)$	$(g-1)/h$	$1/\bar{x}$	$(g+n-1)/(h+n\bar{x})$
$E(\theta)$	g/h	–	$(g+n)/(h+n\bar{x})$
$SD(\theta)$	$\sqrt{g}/h$	–	$\sqrt{g+n}/(h+n\bar{x})$

Example 2.5: Summary

If

- $X_i \sim Exp(\theta)$ and
- $\theta \sim Ga(g, h)$, then
- $\theta|\boldsymbol{x} \sim Ga(g+n, h+n\bar{x})$

The changes in our beliefs about θ are summarised in Table 2.5, taking $g \geq 1$.

	Prior (2.16)	Likelihood (2.17)	Posterior (2.18)
$Mode(\theta)$	$(g-1)/h$	$1/\bar{x}$	$(g+n-1)/(h+n\bar{x})$
$E(\theta)$	g/h	–	$(g+n)/(h+n\bar{x})$
$SD(\theta)$	$\sqrt{g}/h$	–	$\sqrt{g+n}/(h+n\bar{x})$

Example 2.5: Summary

If

- $X_i \sim Exp(\theta)$ and
- $\theta \sim Ga(g, h)$, then
- $\theta|\boldsymbol{x} \sim Ga(g + n, h + n\bar{x})$

The changes in our beliefs about θ are summarised in Table 2.5, taking $g \geq 1$.

	Prior (2.16)	Likelihood (2.17)	Posterior (2.18)
$Mode(\theta)$	$(g-1)/h$	$1/\bar{x}$	$(g+n-1)/(h+n\bar{x})$
$E(\theta)$	g/h	–	$(g+n)/(h+n\bar{x})$
$SD(\theta)$	$\sqrt{g}/h$	–	$\sqrt{g+n}/(h+n\bar{x})$

Example 2.5: Summary

If

- $X_i \sim Exp(\theta)$ and
- $\theta \sim Ga(g, h)$, then
- $\theta|\boldsymbol{x} \sim Ga(g+n, h+n\bar{x})$

The changes in our beliefs about θ are summarised in Table 2.5, taking $g \geq 1$.

	Prior (2.16)	Likelihood (2.17)	Posterior (2.18)
$Mode(\theta)$	$(g-1)/h$	$1/\bar{x}$	$(g+n-1)/(h+n\bar{x})$
$E(\theta)$	g/h	–	$(g+n)/(h+n\bar{x})$
$SD(\theta)$	$\sqrt{g}/h$	–	$\sqrt{g+n}/(h+n\bar{x})$

Model	**Prior**	**Posterior**
Binomial(k, θ)	$\theta \sim U(0, 1)$	Beta
	$\theta \sim Beta(a, b)$	Beta
Exponential(θ)	$\theta \sim Ga(g, h)$	Gamma
	$\theta \sim exp(g)$	??
Normal(μ, σ^2)	$\mu \sim N(b, 1/d)$	Normal

Chapter 2 so far...

Model	Prior	Posterior
Binomial(k, θ)	$\theta \sim U(0,1)$	Beta
	$\theta \sim Beta(a,b)$	Beta
Exponential(θ)	$\theta \sim Ga(g,h)$	Gamma
	$\theta \sim exp(g)$	??
Normal(μ, σ^2)	$\mu \sim N(b, 1/d)$	Normal

Model	Prior	Posterior
Binomial(k, θ)	$\theta \sim U(0, 1)$	Beta
	$\theta \sim Beta(a, b)$	Beta
Exponential(θ)	$\theta \sim Ga(g, h)$	Gamma
	$\theta \sim exp(g)$	??
Normal(μ, σ^2)	$\mu \sim N(b, 1/d)$	Normal

Chapter 2 so far...

Model	Prior	Posterior
Binomial(k, θ)	$\theta \sim U(0, 1)$	Beta
	$\theta \sim Beta(a, b)$	Beta
Exponential(θ)	$\theta \sim Ga(g, h)$	Gamma
	$\theta \sim exp(g)$	??
Normal$\left(\mu, \sigma^2\right)$	$\mu \sim N(b, 1/d)$	Normal

Conjugacy

Definition (Conjugacy)

Suppose that data $\boldsymbol{x}$ are to be observed with distribution $f(\boldsymbol{x}|\theta)$.

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

Example 2.6

Suppose we have a random sample from a Normal distribution.

In Bayesian statistics, when dealing with the Normal distribution, the mathematics is more straightforward if we work with the **precision** ($= 1/$variance) of the distribution rather than the variance itself.

So we will assume that this population has unknown mean μ but known precision τ: $X_i|\mu \sim N(\mu, 1/\tau)$, $i = 1, 2, \ldots, n$ (independent), where τ is known.

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

$$\pi(\mu) = \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \qquad (2.19)$$

Determine the posterior distribution for μ.

Example 2.6

Suppose we have a random sample from a Normal distribution.

In Bayesian statistics, when dealing with the Normal distribution, the mathematics is more straightforward if we work with the **precision** $(= 1/\text{variance})$ of the distribution rather than the variance itself.

So we will assume that this population has unknown mean μ but known precision τ: $X_i|\mu \sim N(\mu, 1/\tau)$, $i = 1, 2, \ldots, n$ (independent), where τ is known.

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

$$\pi(\mu) = \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \qquad (2.19)$$

Determine the posterior distribution for μ.

Example 2.6

Suppose we have a random sample from a Normal distribution.

In Bayesian statistics, when dealing with the Normal distribution, the mathematics is more straightforward if we work with the **precision** ($= 1/$variance) of the distribution rather than the variance itself.

So we will assume that this population has unknown mean μ but known precision τ: $X_i|\mu \sim N(\mu, 1/\tau)$, $i = 1, 2, \ldots, n$ (independent), where τ is known.

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

$$\pi(\mu) = \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \qquad (2.19)$$

Determine the posterior distribution for μ.

Example 2.6

Suppose we have a random sample from a Normal distribution.

In Bayesian statistics, when dealing with the Normal distribution, the mathematics is more straightforward if we work with the **precision** $(= 1/\text{variance})$ of the distribution rather than the variance itself.

So we will assume that this population has unknown mean μ but known precision τ: $X_i|\mu \sim N(\mu, 1/\tau)$, $i = 1, 2, \ldots, n$ (independent), where τ is known.

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

$$\pi(\mu) = \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \tag{2.19}$$

Determine the posterior distribution for μ.

Solution to Example 2.6 (1/5)

The likelihood function for μ is

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

Solution to Example 2.6 (1/5)

The likelihood function for μ is

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

Solution to Example 2.6 (1/5)

The likelihood function for μ is

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

Solution to Example 2.6 (1/5)

The likelihood function for μ is

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

Solution to Example 2.6 (2/5)

Giving

$$
\begin{aligned}
f(\boldsymbol{x}|\mu) &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right.\right. \\
&\qquad \left.\left. + \, 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) + \sum_{i=1}^{n}(\bar{x}-\mu)^2\right]\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right]\right\},
\end{aligned}
$$

Let

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

Then

$$
f(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2 + (\bar{x}-\mu)^2\right]\right\}. \qquad (2.20)
$$

Solution to Example 2.6 (2/5)

Giving

$$\begin{aligned} f(\boldsymbol{x}|\mu) &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 \right.\right. \\ &\qquad \left.\left. + \, 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) + \sum_{i=1}^{n}(\bar{x}-\mu)^2\right]\right\} \\ &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right]\right\}, \end{aligned}$$

Let

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

Then

$$f(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2 + (\bar{x}-\mu)^2\right]\right\}. \qquad (2.20)$$

Solution to Example 2.6 (2/5)

Giving

$$
\begin{aligned}
f(\boldsymbol{x}|\mu) &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right.\right. \\
&\qquad \left.\left. + 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) + \sum_{i=1}^{n}(\bar{x}-\mu)^2\right]\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right]\right\},
\end{aligned}
$$

Let

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

Then

$$
f(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2 + (\bar{x}-\mu)^2\right]\right\}. \qquad (2.20)
$$

Solution to Example 2.6 (2/5)

Giving

$$
\begin{aligned}
f(\boldsymbol{x}|\mu) &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right.\right. \\
&\qquad \left.\left. + 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x}) + \sum_{i=1}^{n}(\bar{x}-\mu)^2\right]\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right]\right\},
\end{aligned}
$$

Let

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

Then

$$
f(\boldsymbol{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2+(\bar{x}-\mu)^2\right]\right\}. \tag{2.20}
$$

Solution to Example 2.6 (3/5)

Applying Bayes Theorem, the posterior density function is

$$\pi(\mu|\boldsymbol{x}) \propto \pi(\mu) f(\boldsymbol{x}|\mu)$$

$$\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu-b)^2\right\} \times \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2+(\bar{x}-\mu)^2\right]\right\}$$

$$= k_1 \exp\left\{-\frac{d}{2}(\mu-b)^2 - \frac{n\tau}{2}\left[s^2+(\bar{x}-\mu)^2\right]\right\}$$

$$= k_1 \exp\left\{-\frac{1}{2}\left[d(\mu-b)^2 + n\tau(\bar{x}-\mu)^2\right]\right\}$$

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

Solution to Example 2.6 (3/5)

Applying Bayes Theorem, the posterior density function is

$$\pi(\mu|\boldsymbol{x}) \propto \pi(\mu) f(\boldsymbol{x}|\mu)$$

$$\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \times \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2 + (\bar{x} - \mu)^2\right]\right\}$$

$$= k_1 \exp\left\{-\frac{d}{2}(\mu - b)^2 - \frac{n\tau}{2}\left[s^2 + (\bar{x} - \mu)^2\right]\right\}$$

$$= k_1 \exp\left\{-\frac{1}{2}\left[d(\mu - b)^2 + n\tau(\bar{x} - \mu)^2\right]\right\}$$

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

Solution to Example 2.6 (3/5)

Applying Bayes Theorem, the posterior density function is

$$\pi(\mu|\boldsymbol{x}) \propto \pi(\mu)f(\boldsymbol{x}|\mu)$$

$$\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu-b)^2\right\} \times \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2+(\bar{x}-\mu)^2\right]\right\}$$

$$= k_1 \exp\left\{-\frac{d}{2}(\mu-b)^2 - \frac{n\tau}{2}\left[s^2+(\bar{x}-\mu)^2\right]\right\}$$

$$= k_1 \exp\left\{-\frac{1}{2}\left[d(\mu-b)^2 + n\tau(\bar{x}-\mu)^2\right]\right\}$$

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

Solution to Example 2.6 (4/5)

We can also simplify the exponent. We have

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

where c does not depend on μ.

Solution to Example 2.6 (4/5)

We can also simplify the exponent. We have

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

where c does not depend on μ.

Solution to Example 2.6 (4/5)

We can also simplify the exponent. We have

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

where c does not depend on μ.

Solution to Example 2.6 (4/5)

We can also simplify the exponent. We have

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

where c does not depend on μ.

Solution to Example 2.6 (4/5)

We can also simplify the exponent. We have

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

where c does not depend on μ.

Solution to Example 2.6 (4/5)

We can also simplify the exponent. We have

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

where c does not depend on μ.

Solution to Example 2.6 (5/5)

Let

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} \qquad \text{and} \qquad D = d + n\tau. \qquad (2.21)$$

Then

$$\pi(\mu|\boldsymbol{x}) = k_1 \exp\left\{-\frac{D}{2}(\mu - B)^2 - \frac{c}{2}\right\}$$

$$= k \exp\left\{-\frac{D}{2}(\mu - B)^2\right\}, \qquad (2.22)$$

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

Therefore, the posterior distribution takes the form $k\exp\{-D(\mu - B)^2/2\}$, $-\infty < \mu < \infty$ and so must be a normal distribution: we have $\mu|\boldsymbol{x} \sim N(B, 1/D)$.

Solution to Example 2.6 (5/5)

Let

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} \qquad \text{and} \qquad D = d + n\tau. \tag{2.21}$$

Then

$$\pi(\mu|\boldsymbol{x}) = k_1 \exp\left\{-\frac{D}{2}(\mu - B)^2 - \frac{c}{2}\right\}$$

$$= k \exp\left\{-\frac{D}{2}(\mu - B)^2\right\}, \tag{2.22}$$

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

Therefore, the posterior distribution takes the form $k\exp\{-D(\mu - B)^2/2\}$, $-\infty < \mu < \infty$ and so must be a normal distribution: we have $\mu|\boldsymbol{x} \sim N(B, 1/D)$.

Solution to Example 2.6 (5/5)

Let

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} \qquad \text{and} \qquad D = d + n\tau. \tag{2.21}$$

Then

$$\pi(\mu|\boldsymbol{x}) = k_1 \exp\left\{-\frac{D}{2}(\mu - B)^2 - \frac{c}{2}\right\}$$

$$= k \exp\left\{-\frac{D}{2}(\mu - B)^2\right\}, \tag{2.22}$$

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

Therefore, the posterior distribution takes the form $k \exp\{-D(\mu - B)^2/2\}$, $-\infty < \mu < \infty$ and so must be a normal distribution: we have $\mu|\boldsymbol{x} \sim N(B, 1/D)$.

Example 2.6: Summary

If we have a random sample from a $N(\mu, 1/\tau)$ distribution (with τ known) and our prior beliefs about μ follow a $N(b, 1/d)$ distribution then, after incorporating the data, our (posterior) beliefs about μ follow a $N(B, 1/D)$ distribution.

Example 2.6: Summary

If we have a random sample from a $N(\mu, 1/\tau)$ distribution (with τ known) and our prior beliefs about μ follow a $N(b, 1/d)$ distribution then, after incorporating the data, our (posterior) beliefs about μ follow a $N(B, 1/D)$ distribution.

The changes in our beliefs about μ are summarised in Table 2.6.

	Prior (2.19)	Likelihood (2.20)	Posterior (2.21)
$Mode(\mu)$	b	$\bar{x}$	$(db + n\tau\bar{x})/(d + n\tau)$
$E(\mu)$	b	–	$(db + n\tau\bar{x})/(d + n\tau)$
$Precision(\mu)$	d	–	$d + n\tau$

Example 2.6: Summary

The changes in our beliefs about μ are summarised in Table 2.6.

	Prior (2.19)	Likelihood (2.20)	Posterior (2.21)
$Mode(\mu)$	b	$\bar{x}$	$(db + n\tau\bar{x})/(d + n\tau)$
$E(\mu)$	b	–	$(db + n\tau\bar{x})/(d + n\tau)$
$Precision(\mu)$	d	–	$d + n\tau$

Example 2.6: Summary

Notice that the way prior information and observed data combine is through the parameters of the normal distribution:

$$b \to \frac{db + n\tau\bar{x}}{d + n\tau} \qquad \text{and} \qquad d \to d + n\tau.$$

Example 2.6: Summary

Notice that the way prior information and observed data combine is through the parameters of the normal distribution:

$$b \rightarrow \frac{db + n\tau\bar{x}}{d + n\tau} \qquad \text{and} \qquad d \rightarrow d + n\tau.$$

Example 2.6: Summary

Notice also that the posterior variance (and precision) does not depend on the data, and the posterior mean is a convex combination of the prior and sample means, that is,

$$
\begin{aligned}
B &= \frac{db + n\tau\bar{x}}{d + n\tau} \\
&= \frac{db}{d + n\tau} + \frac{n\tau\bar{x}}{d + n\tau} \\
&= \frac{d}{d + n\tau} b + \frac{n\tau}{d + n\tau} \bar{x} \\
&= \left(\frac{d}{d + n\tau}\right) b + \left(\frac{d + n\tau - d}{d + n\tau}\right) \bar{x} \\
&= \alpha b + (1 - \alpha)\bar{x}, \qquad \alpha = \frac{d}{d + n\tau}.
\end{aligned}
$$

Example 2.6: Summary

Notice also that the posterior variance (and precision) does not depend on the data, and the posterior mean is a convex combination of the prior and sample means, that is,

$$
\begin{aligned}
B &= \frac{db + n\tau\bar{x}}{d + n\tau} \\
&= \frac{db}{d + n\tau} + \frac{n\tau\bar{x}}{d + n\tau} \\
&= \frac{d}{d + n\tau}b + \frac{n\tau}{d + n\tau}\bar{x} \\
&= \left(\frac{d}{d + n\tau}\right)b + \left(\frac{d + n\tau - d}{d + n\tau}\right)\bar{x} \\
&= \alpha b + (1 - \alpha)\bar{x}, \qquad \alpha = \frac{d}{d + n\tau}.
\end{aligned}
$$

Example 2.6: Summary

Notice also that the posterior variance (and precision) does not depend on the data, and the posterior mean is a convex combination of the prior and sample means, that is,

$$
\begin{aligned}
B &= \frac{db + n\tau\bar{x}}{d + n\tau} \\
&= \frac{db}{d + n\tau} + \frac{n\tau\bar{x}}{d + n\tau} \\
&= \frac{d}{d + n\tau} b + \frac{n\tau}{d + n\tau}\bar{x} \\
&= \left(\frac{d}{d + n\tau}\right) b + \left(\frac{d + n\tau - d}{d + n\tau}\right)\bar{x} \\
&= \alpha b + (1 - \alpha)\bar{x}, \qquad \alpha = \frac{d}{d + n\tau}.
\end{aligned}
$$

Example 2.6: Summary

Notice also that the posterior variance (and precision) does not depend on the data, and the posterior mean is a convex combination of the prior and sample means, that is,

$$
\begin{aligned}
B &= \frac{db + n\tau\bar{x}}{d + n\tau} \\
&= \frac{db}{d + n\tau} + \frac{n\tau\bar{x}}{d + n\tau} \\
&= \frac{d}{d + n\tau} b + \frac{n\tau}{d + n\tau}\bar{x} \\
&= \left(\frac{d}{d + n\tau}\right) b + \left(\frac{d + n\tau - d}{d + n\tau}\right)\bar{x} \\
&= \alpha b + (1-\alpha)\bar{x}, \qquad \alpha = \frac{d}{d + n\tau}.
\end{aligned}
$$

Example 2.6: Summary

Notice also that the posterior variance (and precision) does not depend on the data, and the posterior mean is a convex combination of the prior and sample means, that is,

$$\begin{aligned} B &= \frac{db + n\tau\bar{x}}{d + n\tau} \\ &= \frac{db}{d + n\tau} + \frac{n\tau\bar{x}}{d + n\tau} \\ &= \frac{d}{d + n\tau}b + \frac{n\tau}{d + n\tau}\bar{x} \\ &= \left(\frac{d}{d + n\tau}\right)b + \left(\frac{d + n\tau - d}{d + n\tau}\right)\bar{x} \\ &= \alpha b + (1 - \alpha)\bar{x}, \qquad \alpha = \frac{d}{d + n\tau}. \end{aligned}$$

This equation for the posterior mean, which can be rewritten as

$$E(\mu|\boldsymbol{x}) = \alpha E(\mu) + (1 - \alpha)\bar{x},$$

arises in other models and is known as the **Bayes linear rule** (see Problems Sheet 2, question 3).

Example 2.6: Summary

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

$$E(\mu|\boldsymbol{x}) > E(\mu) \quad \Longleftrightarrow \quad Mode[f(\boldsymbol{x}|\mu)] > E(\mu).$$

Also, the standard deviation of the posterior distribution is smaller than that of the prior distribution.

Example 2.6: Summary

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

$$E(\mu|\boldsymbol{x}) > E(\mu) \quad \Longleftrightarrow \quad Mode[f(\boldsymbol{x}|\mu)] > E(\mu).$$

Also, the standard deviation of the posterior distribution is smaller than that of the prior distribution.

Example 2.7

The ages of *Ennerdale granophyre* rocks can be determined using the relative proportions of rubidium–87 and strontium–87 in the rock.

An expert in the field suggests that the ages of such rocks (in millions of years) $X|\mu \sim N(\mu, 8^2)$ and that a prior distribution $\mu \sim N(370, 20^2)$ is appropriate.

A rock is found whose chemical analysis yields $x = 421$. What is the posterior distribution for μ and what is the probability that the rock will be older than 400 million years?

Solution to Example 2.7 (1/2)

We have $n = 1$, $\bar{x} = x = 421$, $\tau = 1/64$, $b = 370$ and $d = 1/400$. Therefore, using the results in Example 2.6,

$$
\begin{aligned}
B &= \frac{db + n\tau\bar{x}}{d + n\tau} \\
&= \frac{370/400 + 421/64}{1/400 + 1/64} = 414.0 \qquad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
D &= d + n\tau \\
&= 1/400 + 1/64 = 1/7.43^2.
\end{aligned}
$$

Solution to Example 2.7 (1/2)

We have $n = 1$, $\bar{x} = x = 421$, $\tau = 1/64$, $b = 370$ and $d = 1/400$. Therefore, using the results in Example 2.6,

$$B = \frac{db + n\tau\bar{x}}{d + n\tau}$$

$$= \frac{370/400 + 421/64}{1/400 + 1/64} = 414.0 \qquad \text{and}$$

$$D = d + n\tau$$

$$= 1/400 + 1/64 = 1/7.43^2.$$

Solution to Example 2.7 (1/2)

We have $n = 1$, $\bar{x} = x = 421$, $\tau = 1/64$, $b = 370$ and $d = 1/400$. Therefore, using the results in Example 2.6,

$$
\begin{aligned}
B &= \frac{db + n\tau\bar{x}}{d + n\tau} \\
&= \frac{370/400 + 421/64}{1/400 + 1/64} = 414.0 \qquad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
D &= d + n\tau \\
&= 1/400 + 1/64 = 1/7.43^2.
\end{aligned}
$$

Solution to Example 2.7 (1/2)

We have $n = 1$, $\bar{x} = x = 421$, $\tau = 1/64$, $b = 370$ and $d = 1/400$. Therefore, using the results in Example 2.6,

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} = \frac{370/400 + 421/64}{1/400 + 1/64} = 414.0 \qquad \text{and}$$

$$D = d + n\tau = 1/400 + 1/64 = 1/7.43^2.$$

Solution to Example 2.7 (1/2)

We have $n = 1$, $\bar{x} = x = 421$, $\tau = 1/64$, $b = 370$ and $d = 1/400$. Therefore, using the results in Example 2.6,

$$B = \frac{db + n\tau\bar{x}}{d + n\tau} = \frac{370/400 + 421/64}{1/400 + 1/64} = 414.0 \qquad \text{and}$$

$$D = d + n\tau = 1/400 + 1/64 = 1/7.43^2.$$

Solution to Example 2.7 (2/2)

Thus, the posterior distribution is $\mu|x = 421 \sim N(414.0, 7.43^2)$.

The (posterior) probability that the rock will be older than 400 million years is

$$\Pr(\mu > 400|x = 421) = 0.9702$$

calculated using the R commands `1-pnorm(400,414,7.43)` or `1-pnorm(-1.884)` (or indeed by hand using tables).

Without the chemical analysis, the only basis for determining the age of the rock is via the prior distribution.

The (prior) probability that the rock will be older than 400 million years is $\Pr(\mu > 400) = 0.0668$ calculated using the R command `1-pnorm(400,370,20)`.

Solution to Example 2.7 (2/2)

Thus, the posterior distribution is $\mu|x = 421 \sim N(414.0, 7.43^2)$.

The (posterior) probability that the rock will be older than 400 million years is

$$\Pr(\mu > 400|x = 421) = 0.9702$$

calculated using the R commands `1-pnorm(400,414,7.43)` or `1-pnorm(-1.884)` (or indeed by hand using tables).

Without the chemical analysis, the only basis for determining the age of the rock is via the prior distribution.

The (prior) probability that the rock will be older than 400 million years is $\Pr(\mu > 400) = 0.0668$ calculated using the R command `1-pnorm(400,370,20)`.

Solution to Example 2.7 (2/2)

Thus, the posterior distribution is $\mu|x = 421 \sim N(414.0, 7.43^2)$.

The (posterior) probability that the rock will be older than 400 million years is

$$\Pr(\mu > 400|x = 421) = 0.9702$$

calculated using the R commands `1-pnorm(400,414,7.43)` or `1-pnorm(-1.884)` (or indeed by hand using tables).

Without the chemical analysis, the only basis for determining the age of the rock is via the prior distribution.

The (prior) probability that the rock will be older than 400 million years is $\Pr(\mu > 400) = 0.0668$ calculated using the R command `1-pnorm(400,370,20)`.

Solution to Example 2.7 (2/2)

Thus, the posterior distribution is $\mu|x = 421 \sim N(414.0, 7.43^2)$.

The (posterior) probability that the rock will be older than 400 million years is

$$\Pr(\mu > 400|x = 421) = 0.9702$$

calculated using the R commands `1-pnorm(400,414,7.43)` or `1-pnorm(-1.884)` (or indeed by hand using tables).

Without the chemical analysis, the only basis for determining the age of the rock is via the prior distribution.

The (prior) probability that the rock will be older than 400 million years is $\Pr(\mu > 400) = 0.0668$ calculated using the R command `1-pnorm(400,370,20)`.

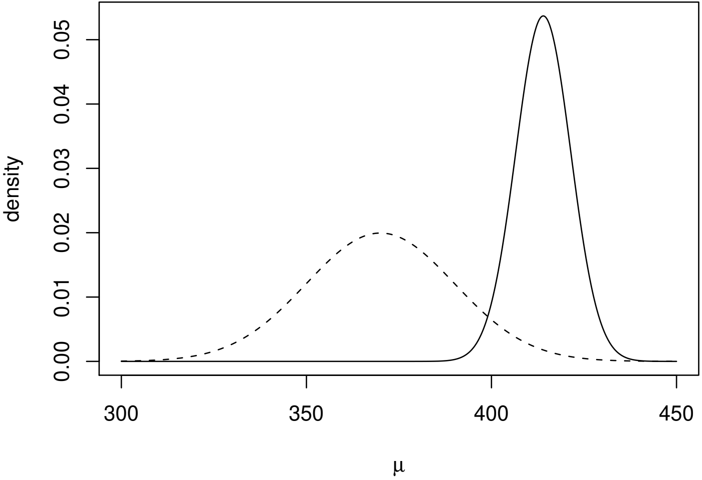
Example 2.7: Summary

This highlights the benefit of taking the chemical measurements.

Note that the large difference between these probabilities is not necessarily due to the expert’s prior distribution being inaccurate, *per se*.

It is probably due to the large prior uncertainty about rock ages, as shown in Figure 2.10.

Example 2.7: Summary



Posterior distributions and sufficient statistics

We have already met the concept of sufficient statistics. Not surprisingly they also play a role in Bayesian Inference.

Suppose that we have data $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ available and we want to make inferences about the parameter θ in the statistical model $f(\boldsymbol{x}|\theta)$.

If T is a sufficient statistic then by the Factorisation Theorem

$$f(\boldsymbol{x}|\theta) = h(\boldsymbol{x})\, g(t, \theta).$$

Posterior distributions and sufficient statistics

We have already met the concept of sufficient statistics. Not surprisingly they also play a role in Bayesian Inference.

Suppose that we have data $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ available and we want to make inferences about the parameter θ in the statistical model $f(\boldsymbol{x}|\theta)$.

If T is a sufficient statistic then by the Factorisation Theorem

$$f(\boldsymbol{x}|\theta) = h(\boldsymbol{x})\, g(t, \theta).$$

Posterior distributions and sufficient statistics

We have already met the concept of sufficient statistics. Not surprisingly they also play a role in Bayesian Inference.

Suppose that we have data $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ available and we want to make inferences about the parameter θ in the statistical model $f(\boldsymbol{x}|\theta)$.

If T is a sufficient statistic then by the Factorisation Theorem

$$f(\boldsymbol{x}|\theta) = h(\boldsymbol{x})\, g(t, \theta).$$

Therefore, using Bayes Theorem,

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)\, f(\boldsymbol{x}|\theta) \\ &\propto \pi(\theta)\, h(\boldsymbol{x})\, g(t,\theta) \\ &\propto \pi(\theta)\, g(t,\theta). \end{aligned}$$

It can be shown that, up to a constant not depending on θ, $g(t,\theta)$ is equal to the probability (density) function of T; that is

$$g(t,\theta) \propto f_T(t|\theta).$$

Hence

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)\, f_T(t|\theta).$$

Posterior distributions and sufficient statistics

Therefore, using Bayes Theorem,

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)\, f(\boldsymbol{x}|\theta) \\ &\propto \pi(\theta)\, h(\boldsymbol{x})\, g(t,\theta) \\ &\propto \pi(\theta)\, g(t,\theta).\end{aligned}$$

It can be shown that, up to a constant not depending on θ, $g(t,\theta)$ is equal to the probability (density) function of T; that is

$$g(t,\theta) \propto f_T(t|\theta).$$

Hence

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)\, f_T(t|\theta).$$

Posterior distributions and sufficient statistics

Therefore, using Bayes Theorem,

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)\, f(\boldsymbol{x}|\theta)\\ &\propto \pi(\theta)\, h(\boldsymbol{x})\, g(t,\theta)\\ &\propto \pi(\theta)\, g(t,\theta).\end{aligned}$$

It can be shown that, up to a constant not depending on θ, $g(t,\theta)$ is equal to the probability (density) function of T; that is

$$g(t,\theta) \propto f_T(t|\theta).$$

Hence

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)\, f_T(t|\theta).$$

Posterior distributions and sufficient statistics

Therefore, using Bayes Theorem,

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta)\, f(\boldsymbol{x}|\theta) \\ &\propto \pi(\theta)\, h(\boldsymbol{x})\, g(t,\theta) \\ &\propto \pi(\theta)\, g(t,\theta).\end{aligned}$$

It can be shown that, up to a constant not depending on θ, $g(t,\theta)$ is equal to the probability (density) function of T; that is

$$g(t,\theta) \propto f_T(t|\theta).$$

Hence

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta)\, f_T(t|\theta).$$

Posterior distributions and sufficient statistics

However, applying Bayes Theorem to the data t gives

$$\pi(\theta|t) \quad \propto \quad \pi(\theta)\, f_T(t|\theta)$$

and so, since

$$\pi(\theta|\boldsymbol{x}) \quad \propto \quad \pi(\theta|t)$$

and both are probability (density) functions, we have

$$\pi(\theta|\boldsymbol{x}) \quad = \quad \pi(\theta|t).$$

Posterior distributions and sufficient statistics

However, applying Bayes Theorem to the data t gives

$$\pi(\theta|t) \quad \propto \quad \pi(\theta)\, f_T(t|\theta)$$

and so, since

$$\pi(\theta|\boldsymbol{x}) \quad \propto \quad \pi(\theta|t)$$

and both are probability (density) functions, we have

$$\pi(\theta|\boldsymbol{x}) \quad = \quad \pi(\theta|t).$$

Posterior distributions and sufficient statistics

However, applying Bayes Theorem to the data t gives

$$\pi(\theta|t) \quad \propto \quad \pi(\theta)\, f_T(t|\theta)$$

and so, since

$$\pi(\theta|\boldsymbol{x}) \quad \propto \quad \pi(\theta|t)$$

and both are probability (density) functions, we have

$$\pi(\theta|\boldsymbol{x}) \quad = \quad \pi(\theta|t).$$

Posterior distributions and sufficient statistics

Therefore, our (posterior) beliefs about θ having observed the full data $\boldsymbol{x}$ are the same as if we had observed only the sufficient statistic T.

This is what we would expect if all the information about θ in the data were contained in the sufficient statistic.

Posterior distributions and sufficient statistics

Therefore, our (posterior) beliefs about θ having observed the full data $\boldsymbol{x}$ are the same as if we had observed only the sufficient statistic T.

This is what we would expect if all the information about θ in the data were contained in the sufficient statistic.

Example 2.8

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

Determine a sufficient statistic T for θ and verify that $\pi(\theta|\boldsymbol{x}) = \pi(\theta|t)$.

Solution to Example 2.8 (1/2)

The density of the data is

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= 1 \times \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= h(\boldsymbol{x})\, g(\Sigma x_i, \theta)
\end{aligned}
$$

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

Solution to Example 2.8 (1/2)

The density of the data is

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= 1 \times \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= h(\boldsymbol{x})\, g(\Sigma x_i, \theta)
\end{aligned}
$$

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

Solution to Example 2.8 (1/2)

The density of the data is

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= 1 \times \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= h(\boldsymbol{x})\, g(\Sigma x_i, \theta)
\end{aligned}
$$

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

Solution to Example 2.8 (1/2)

The density of the data is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\ &= \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\ &= 1 \times \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\ &= h(\boldsymbol{x})\, g(\Sigma x_i, \theta) \end{aligned}$$

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

Solution to Example 2.8 (1/2)

The density of the data is

$$\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= 1 \times \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \\
&= h(\boldsymbol{x})\, g(\Sigma x_i, \theta)
\end{aligned}$$

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

Solution to Example 2.8 (2/2)

Now $T|\theta \sim Ga(n, \theta)$ and so

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \quad \theta > 0.$$

Also

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

Therefore, by Bayes Theorem,

$$\begin{aligned} \pi(\theta|t) &\propto \pi(\theta) f(t|\theta) \\ &\propto \frac{h^g \, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \qquad \theta > 0 \\ &\propto \theta^{g+n-1} e^{-(h+t)\theta}, \qquad \theta > 0 \end{aligned}$$

and so the posterior distribution is $\theta|t \sim Ga(g+n, h+t)$. This is the same as the result we obtained previously for $\theta|\boldsymbol{x}$.

Solution to Example 2.8 (2/2)

Now $T|\theta \sim Ga(n, \theta)$ and so

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \quad \theta > 0.$$

Also

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

Therefore, by Bayes Theorem,

$$\begin{aligned}
\pi(\theta|t) &\propto \pi(\theta) f(t|\theta) \\
&\propto \frac{h^g \, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \qquad \theta > 0 \\
&\propto \theta^{g+n-1} e^{-(h+t)\theta}, \qquad \theta > 0
\end{aligned}$$

and so the posterior distribution is $\theta|t \sim Ga(g+n, h+t)$. This is the same as the result we obtained previously for $\theta|\boldsymbol{x}$.

Solution to Example 2.8 (2/2)

Now $T|\theta \sim Ga(n, \theta)$ and so

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \quad \theta > 0.$$

Also

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

Therefore, by Bayes Theorem,

$$\begin{aligned}
\pi(\theta|t) &\propto \pi(\theta) f(t|\theta) \\
&\propto \frac{h^g \, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \qquad \theta > 0 \\
&\propto \theta^{g+n-1} e^{-(h+t)\theta}, \qquad \theta > 0
\end{aligned}$$

and so the posterior distribution is $\theta|t \sim Ga(g+n, h+t)$. This is the same as the result we obtained previously for $\theta|\boldsymbol{x}$.

Solution to Example 2.8 (2/2)

Now $T|\theta \sim Ga(n, \theta)$ and so

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \quad \theta > 0.$$

Also

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

Therefore, by Bayes Theorem,

$$\begin{aligned} \pi(\theta|t) &\propto \pi(\theta) f(t|\theta) \\ &\propto \frac{h^g \, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \qquad \theta > 0 \\ &\propto \theta^{g+n-1} e^{-(h+t)\theta}, \qquad \theta > 0 \end{aligned}$$

and so the posterior distribution is $\theta|t \sim Ga(g+n, h+t)$. This is the same as the result we obtained previously for $\theta|\boldsymbol{x}$.

Solution to Example 2.8 (2/2)

Now $T|\theta \sim Ga(n, \theta)$ and so

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \quad \theta > 0.$$

Also

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

Therefore, by Bayes Theorem,

$$\begin{aligned}
\pi(\theta|t) &\propto \pi(\theta) f(t|\theta) \\
&\propto \frac{h^g \, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \qquad \theta > 0 \\
&\propto \theta^{g+n-1} e^{-(h+t)\theta}, \qquad \theta > 0
\end{aligned}$$

and so the posterior distribution is $\theta|t \sim Ga(g+n, h+t)$. This is the same as the result we obtained previously for $\theta|\boldsymbol{x}$.

Solution to Example 2.8 (2/2)

Now $T|\theta \sim Ga(n, \theta)$ and so

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \quad \theta > 0.$$

Also

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

Therefore, by Bayes Theorem,

$$\begin{aligned} \pi(\theta|t) &\propto \pi(\theta) f(t|\theta) \\ &\propto \frac{h^g \, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \qquad \theta > 0 \\ &\propto \theta^{g+n-1} e^{-(h+t)\theta}, \qquad \theta > 0 \end{aligned}$$

and so the posterior distribution is $\theta|t \sim Ga(g+n, h+t)$. This is the same as the result we obtained previously for $\theta|\boldsymbol{x}$.

Solution to Example 2.8 (2/2)

Now $T|\theta \sim Ga(n, \theta)$ and so

$$f_T(t|\theta) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \quad \theta > 0.$$

Also

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

Therefore, by Bayes Theorem,

$$\begin{aligned}
\pi(\theta|t) &\propto \pi(\theta) f(t|\theta) \\
&\propto \frac{h^g\, \theta^{g-1} e^{-h\theta}}{\Gamma(g)} \times \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)}, \qquad \theta > 0 \\
&\propto \theta^{g+n-1} e^{-(h+t)\theta}, \qquad \theta > 0
\end{aligned}$$

and so the posterior distribution is $\theta|t \sim Ga(g+n, h+t)$. This is the same as the result we obtained previously for $\theta|\boldsymbol{x}$.

Example 2.9

Suppose we have a random sample from a normal distribution with known variance and a normal prior distribution for the mean parameter, that is, $X_i|\mu \sim N(\mu, 1/\tau)$, $i = 1, 2, \ldots, n$ (independent) and $\mu \sim N(b, 1/d)$.

Determine a sufficient statistic T for μ and verify that $\pi(\mu|\boldsymbol{x}) = \pi(\mu|t)$.

Solution to Example 2.9 (1/3)

Recall from Equation (2.20) that

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2 + (\bar{x}-\mu)^2\right]\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau s^2}{2}\right\} \times \exp\left\{-\frac{n\tau}{2}(\bar{x}-\mu)^2\right\} \\
&= h(\boldsymbol{x})\, g(\bar{x},\mu)
\end{aligned}
$$

and therefore, by the Factorisation Theorem, $T = \bar{X}$ is sufficient for μ.

Recall from Equation (2.20) that

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2 + (\bar{x}-\mu)^2\right]\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau s^2}{2}\right\} \times \exp\left\{-\frac{n\tau}{2}(\bar{x}-\mu)^2\right\} \\
&= h(\boldsymbol{x})\, g(\bar{x},\mu)
\end{aligned}
$$

and therefore, by the Factorisation Theorem, $T = \bar{X}$ is sufficient for μ.

Solution to Example 2.9 (1/3)

Recall from Equation (2.20) that

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}|\mu) &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2 + (\bar{x}-\mu)^2\right]\right\} \\ &= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau s^2}{2}\right\} \times \exp\left\{-\frac{n\tau}{2}(\bar{x}-\mu)^2\right\} \\ &= h(\boldsymbol{x})\, g(\bar{x}, \mu) \end{aligned}$$

and therefore, by the Factorisation Theorem, $T = \bar{X}$ is sufficient for μ.

Solution to Example 2.9 (2/3)

Now $T|\mu \sim N(\mu, 1/(n\tau))$ (by the Central Limit Theorem) and so

$$f_T(t|\mu) \quad = \quad \left(\frac{n\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\tau}{2}(t-\mu)^2\right\}.$$

Also

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

Solution to Example 2.9 (2/3)

Now $T|\mu \sim N(\mu, 1/(n\tau))$ (by the Central Limit Theorem) and so

$$f_T(t|\mu) = \left(\frac{n\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\tau}{2}(t-\mu)^2\right\}.$$

Also

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

Solution to Example 2.9 (3/3)

Therefore, by Bayes' Theorem,

$$
\begin{aligned}
\pi(\mu|t) &\propto \pi(\mu) f(t|\mu) \\
&\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu-b)^2\right\} \\
&\qquad\qquad \times \left(\frac{n\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\tau}{2}(t-\mu)^2\right\} \\
&\propto \exp\left\{-\frac{d}{2}(\mu-b)^2 - \frac{n\tau}{2}(t-\mu)^2\right\} \\
&\;\;\vdots \\
&\propto \exp\left\{-\frac{D}{2}(\mu-B)^2\right\}
\end{aligned}
$$

where B and D are as in Equation (2.21), with t replacing $\bar{x}$; that is, $\mu|t \sim N(B, 1/D)$, the same distribution as $\mu|\boldsymbol{x}$.

Solution to Example 2.9 (3/3)

Therefore, by Bayes' Theorem,

$$\begin{aligned}
\pi(\mu|t) &\propto \pi(\mu) f(t|\mu) \\
&\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu - b)^2\right\} \\
&\qquad\qquad \times \left(\frac{n\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\tau}{2}(t - \mu)^2\right\} \\
&\propto \exp\left\{-\frac{d}{2}(\mu - b)^2 - \frac{n\tau}{2}(t - \mu)^2\right\} \\
&\;\;\vdots \\
&\propto \exp\left\{-\frac{D}{2}(\mu - B)^2\right\}
\end{aligned}$$

where B and D are as in Equation (2.21), with t replacing $\bar{x}$; that is, $\mu|t \sim N(B, 1/D)$, the same distribution as $\mu|\boldsymbol{x}$.

Solution to Example 2.9 (3/3)

Therefore, by Bayes' Theorem,

$$
\begin{aligned}
\pi(\mu|t) &\propto \pi(\mu) f(t|\mu) \\
&\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu-b)^2\right\} \\
&\qquad\qquad \times \left(\frac{n\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\tau}{2}(t-\mu)^2\right\} \\
&\propto \exp\left\{-\frac{d}{2}(\mu-b)^2 - \frac{n\tau}{2}(t-\mu)^2\right\} \\
&\;\;\vdots \\
&\propto \exp\left\{-\frac{D}{2}(\mu-B)^2\right\}
\end{aligned}
$$

where B and D are as in Equation (2.21), with t replacing $\bar{x}$; that is, $\mu|t \sim N(B, 1/D)$, the same distribution as $\mu|\boldsymbol{x}$.

Solution to Example 2.9 (3/3)

Therefore, by Bayes' Theorem,

$$
\begin{aligned}
\pi(\mu|t) &\propto \pi(\mu)f(t|\mu) \\
&\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu-b)^2\right\} \\
&\qquad\qquad \times \left(\frac{n\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\tau}{2}(t-\mu)^2\right\} \\
&\propto \exp\left\{-\frac{d}{2}(\mu-b)^2 - \frac{n\tau}{2}(t-\mu)^2\right\} \\
&\ \ \vdots \\
&\propto \exp\left\{-\frac{D}{2}(\mu-B)^2\right\}
\end{aligned}
$$

where B and D are as in Equation (2.21), with t replacing $\bar{x}$; that is, $\mu|t \sim N(B, 1/D)$, the same distribution as $\mu|\boldsymbol{x}$.

Solution to Example 2.9 (3/3)

Therefore, by Bayes' Theorem,

$$\begin{aligned}
\pi(\mu|t) &\propto \pi(\mu)f(t|\mu) \\
&\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu-b)^2\right\} \\
&\qquad\qquad \times \left(\frac{n\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\tau}{2}(t-\mu)^2\right\} \\
&\propto \exp\left\{-\frac{d}{2}(\mu-b)^2 - \frac{n\tau}{2}(t-\mu)^2\right\} \\
&\;\;\vdots \\
&\propto \exp\left\{-\frac{D}{2}(\mu-B)^2\right\}
\end{aligned}$$

where B and D are as in Equation (2.21), with t replacing $\bar{x}$; that is, $\mu|t \sim N(B, 1/D)$, the same distribution as $\mu|\boldsymbol{x}$.