

If $X_1, X_2, \dots, X_n$ are independent random variables with probability density function

$$f(x|\theta) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, \dots,$$

- (a) Form the likelihood function $f(\mathbf{x}|\theta)$;
- (b) Use the Factorisation Theorem to obtain a sufficient statistic for θ .

Quick quiz: solution

The joint probability function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^n \theta(1-\theta)^{x_i-1} \\&= \theta^n(1-\theta)^{x_1-1}(1-\theta)^{x_2-1} \times \dots \times (1-\theta)^{x_n-1} \\&= \theta^n(1-\theta)^{\sum x_i - n} \\&= 1 \times \theta^n(1-\theta)^{\sum x_i - n} \\&= h(\mathbf{x}) g\left(\sum x_i, \theta\right)\end{aligned}$$

where $h(\mathbf{x}) = 1$ and $g(t, \theta) = \theta^n(1-\theta)^{t-n}$. Therefore, by the Factorisation Theorem, $T = \sum X_i$ is sufficient for θ .

Quick quiz: solution

The joint probability function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} \\&= \theta^n(1 - \theta)^{x_1-1}(1 - \theta)^{x_2-1} \times \dots \times (1 - \theta)^{x_n-1} \\&= \theta^n(1 - \theta)^{\sum x_i - n} \\&= 1 \times \theta^n(1 - \theta)^{\sum x_i - n} \\&= h(\mathbf{x}) g\left(\sum x_i, \theta\right)\end{aligned}$$

where $h(\mathbf{x}) = 1$ and $g(t, \theta) = \theta^n(1 - \theta)^{t-n}$. Therefore, by the Factorisation Theorem, $T = \sum X_i$ is sufficient for θ .

Quick quiz: solution

The joint probability function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} \\&= \theta^n(1 - \theta)^{x_1-1}(1 - \theta)^{x_2-1} \times \dots \times (1 - \theta)^{x_n-1} \\&= \theta^n(1 - \theta)^{\sum x_i - n} \\&= 1 \times \theta^n(1 - \theta)^{\sum x_i - n} \\&= h(\mathbf{x}) g\left(\sum x_i, \theta\right)\end{aligned}$$

where $h(\mathbf{x}) = 1$ and $g(t, \theta) = \theta^n(1 - \theta)^{t-n}$. Therefore, by the Factorisation Theorem, $T = \sum X_i$ is sufficient for θ .

Quick quiz: solution

The joint probability function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^n \theta(1-\theta)^{x_i-1} \\&= \theta^n(1-\theta)^{x_1-1}(1-\theta)^{x_2-1} \times \dots \times (1-\theta)^{x_n-1} \\&= \theta^n(1-\theta)^{\sum x_i - n} \\&= 1 \times \theta^n(1-\theta)^{\sum x_i - n} \\&= h(\mathbf{x}) g\left(\sum x_i, \theta\right)\end{aligned}$$

where $h(\mathbf{x}) = 1$ and $g(t, \theta) = \theta^n(1-\theta)^{t-n}$. Therefore, by the Factorisation Theorem, $T = \sum X_i$ is sufficient for θ .

Quick quiz: solution

The joint probability function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} \\&= \theta^n(1 - \theta)^{x_1-1}(1 - \theta)^{x_2-1} \times \dots \times (1 - \theta)^{x_n-1} \\&= \theta^n(1 - \theta)^{\sum x_i - n} \\&= 1 \times \theta^n(1 - \theta)^{\sum x_i - n} \\&= h(\mathbf{x}) g\left(\sum x_i, \theta\right)\end{aligned}$$

where $h(\mathbf{x}) = 1$ and $g(t, \theta) = \theta^n(1 - \theta)^{t-n}$. Therefore, by the Factorisation Theorem, $T = \sum X_i$ is sufficient for θ .

Quick quiz: solution

The joint probability function is

$$\begin{aligned}f(\mathbf{x}|\theta) &= \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} \\&= \theta^n(1 - \theta)^{x_1-1}(1 - \theta)^{x_2-1} \times \dots \times (1 - \theta)^{x_n-1} \\&= \theta^n(1 - \theta)^{\sum x_i - n} \\&= 1 \times \theta^n(1 - \theta)^{\sum x_i - n} \\&= h(\mathbf{x}) g\left(\sum x_i, \theta\right)\end{aligned}$$

where $h(\mathbf{x}) = 1$ and $g(t, \theta) = \theta^n(1 - \theta)^{t-n}$. Therefore, by the Factorisation Theorem, $T = \sum X_i$ is sufficient for θ .

Chapter 2

Bayes' Theorem for Distributions

Bayes' Theorem for Distributions

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

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

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

Bayes' Theorem for Distributions

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

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

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

Bayes' Theorem for Distributions

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

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

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

Bayes' Theorem for Distributions

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

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

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

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

where

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

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

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

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

Bayes' Theorem for Distributions

The posterior probability (density) function for θ is

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

where

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

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

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

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

Bayes' Theorem for Distributions

Thus, to obtain the posterior distribution, we need:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Also

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

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

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

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

It can be shown that

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

Also

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

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

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

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

It can be shown that

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

Also

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

Bayes Theorem for distributions in action

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

Recall that we have

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

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

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

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

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

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

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

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

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

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

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

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 $\theta = \text{Pr}(\text{Head})$.

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

$$\pi(\theta) = 1, \quad 0 < \theta < 1. \quad (2.4)$$

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 $\theta = \text{Pr}(\text{Head})$.

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

$$\pi(\theta) = 1, \quad 0 < \theta < 1. \quad (2.4)$$

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 $\theta = \text{Pr}(\text{Head})$.

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

$$\pi(\theta) = 1, \quad 0 < \theta < 1. \quad (2.4)$$

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 $\theta = \text{Pr}(\text{Head})$.

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

$$\pi(\theta) = 1, \quad 0 < \theta < 1. \quad (2.4)$$

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 $\theta = \text{Pr}(\text{Head})$.

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

$$\pi(\theta) = 1, \quad 0 < \theta < 1. \quad (2.4)$$

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 $X|\theta \sim \text{Bin}(5, \theta)$. This gives a likelihood function

$$\begin{aligned} f(x = 1|\theta) &= {}^5C_1\theta^1(1 - \theta)^4 \\ &= 5\theta(1 - \theta)^4 \end{aligned} \tag{2.5}$$

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 $X|\theta \sim \text{Bin}(5, \theta)$. This gives a likelihood function

$$\begin{aligned} f(x = 1|\theta) &= {}^5C_1\theta^1(1 - \theta)^4 \\ &= 5\theta(1 - \theta)^4 \end{aligned} \tag{2.5}$$

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 $X|\theta \sim \text{Bin}(5, \theta)$. This gives a likelihood function

$$\begin{aligned} f(x = 1|\theta) &= {}^5C_1\theta^1(1 - \theta)^4 \\ &= 5\theta(1 - \theta)^4 \end{aligned} \tag{2.5}$$

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 (2/4)

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

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

where

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

“Full” solution to Example 2.1 (2/4)

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

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

where

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

“Full” solution to Example 2.1 (2/4)

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

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

where

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

“Full” solution to Example 2.1 (3/4)

First the denominator:

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

which, using integration by parts, gives

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

“Full” solution to Example 2.1 (3/4)

First the denominator:

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

which, using integration by parts, gives

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

“Full” solution to Example 2.1 (3/4)

First the denominator:

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

which, using integration by parts, gives

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

“Full” solution to Example 2.1 (4/4)

Therefore, the posterior density is

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

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

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

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

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

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

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

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

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

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

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

as

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

and so the posterior distribution is $\theta|x=1 \sim \text{Be}(2,5)$ – 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

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

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

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

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 $f(\mathbf{x})$.

In this example, we can calculate the posterior distribution as

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

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

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

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

$$\pi(\theta|\mathbf{x}) \propto \pi(\theta)f(x=1|\theta)$$

$$\propto 1 \times 5\theta(1-\theta)^4, \quad 0 < \theta < 1$$

$$= k\theta(1-\theta)^4, \quad 0 < \theta < 1.$$

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

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

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 $f(\mathbf{x})$.

In this example, we can calculate the posterior distribution as

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

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

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

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

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

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

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

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

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 $f(\mathbf{x})$.

In this example, we can calculate the posterior distribution as

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 2.1: Summary

It is possible that we have a biased coin.

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

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

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

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

Thus:

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

Example 2.1: Summary

It is possible that we have a biased coin.

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

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

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

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

Thus:

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

Example 2.1: Summary

It is possible that we have a biased coin.

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

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

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

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

Thus:

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

Example 2.1: Summary

It is possible that we have a biased coin.

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

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

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

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

Thus:

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

Example 2.1: Summary

It is possible that we have a biased coin.

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

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

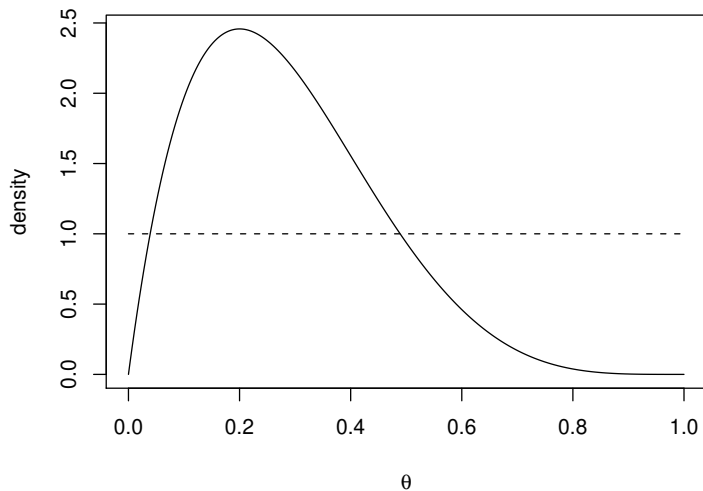
Also, if $Y \sim \text{Be}(a, b)$,

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

Thus:

	Prior $U(0, 1)$	Observed $\theta = 0.2$	Posterior $\text{Be}(2, 5)$
Mean	0.5	$\longrightarrow$	0.286
St. dev.	0.289	$\longrightarrow$	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 $\theta = \text{Pr}(\text{correct choice})$.

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 $\theta \sim \text{Be}(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

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

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

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

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

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

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 $\theta \sim \text{Be}(77, 5)$ (see Example 3.2, Chapter 3), with probability density function

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

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

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

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.

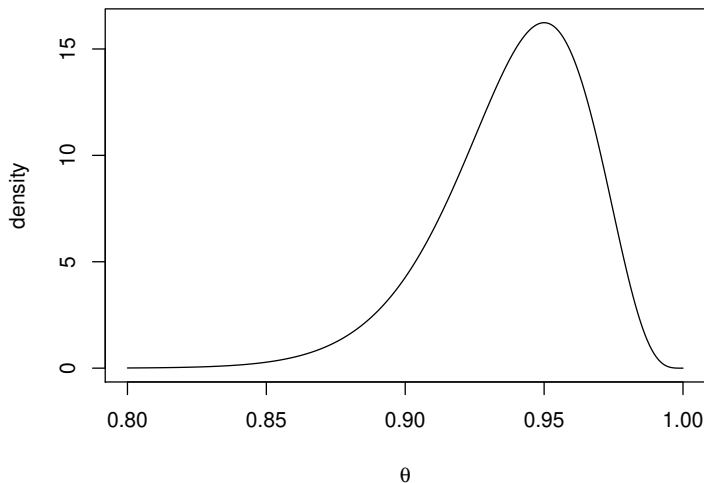
In fact, it is suggested that we should have a prior distribution which has a mode around $\theta = 0.95$ and for which $\Pr(\theta < 0.8)$ is very small.

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

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

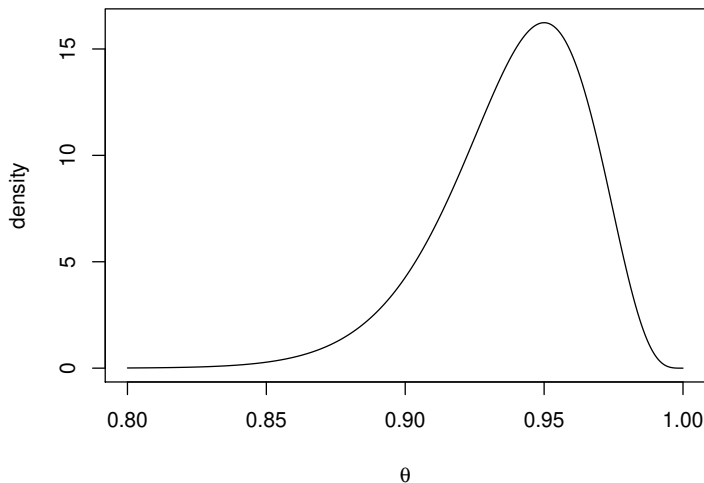
Example 2.2

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 $X|\theta \sim \text{Bin}(10, \theta)$.

This gives a likelihood function of

$$\begin{aligned} f(x = 9|\theta) &= {}^{10}C_9 \theta^9 (1 - \theta)^1 \\ &= 10\theta^9(1 - \theta) \end{aligned} \tag{2.8}$$

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 $X|\theta \sim \text{Bin}(10, \theta)$.

This gives a likelihood function of

$$\begin{aligned} f(x = 9|\theta) &= {}^{10}\text{C}_9 \theta^9 (1 - \theta)^1 \\ &= 10\theta^9 (1 - \theta) \end{aligned} \tag{2.8}$$

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 $X|\theta \sim \text{Bin}(10, \theta)$.

This gives a likelihood function of

$$\begin{aligned} f(x = 9|\theta) &= {}^{10}\text{C}_9 \theta^9 (1 - \theta)^1 \\ &= 10\theta^9 (1 - \theta) \end{aligned} \tag{2.8}$$

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 $X|\theta \sim \text{Bin}(10, \theta)$.

This gives a likelihood function of

$$\begin{aligned} f(x = 9|\theta) &= {}^{10}\text{C}_9 \theta^9 (1 - \theta)^1 \\ &= 10\theta^9 (1 - \theta) \end{aligned} \tag{2.8}$$

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

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

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

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

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

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

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

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

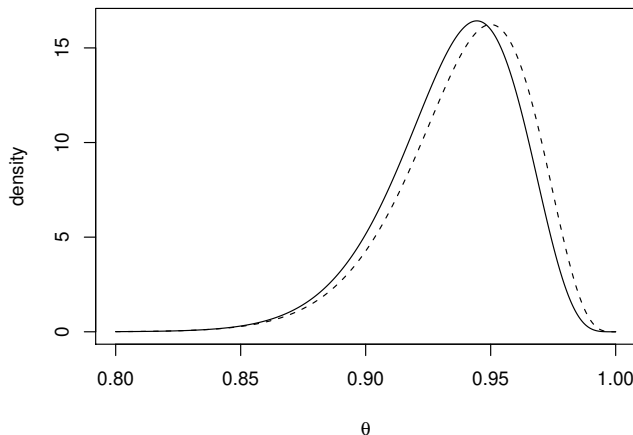
$$= 128107980 \theta^{76}(1 - \theta)^4 \times 10 \theta^9(1 - \theta), \quad 0 < \theta < 1$$

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

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)
$Mode(\theta)$	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 2.2: 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.