Chapter 1

Introduction

Consider the following three experiments.

- **Experiment 1: Fisher's tea lady**
  The tea lady claims to know whether milk or tea is poured in first: for 10 pairs of cups of tea she makes the correct choice each time.

- **Experiment 2: Music expert**
  The expert claims he can distinguish between a page from a Haydn score and a page from a Mozart score: he does so correctly 10 times.

- **Experiment 3: The Drunk**
  A somewhat inebriated friend at a party claims they can predict the outcome of the toss of a coin: they do so correctly 10 times.

Consider the following three experiments.

- **Experiment 1: Fisher's tea lady**
  The tea lady claims to know whether milk or tea is poured in first: for 10 pairs of cups of tea she makes the correct choice each time.

- **Experiment 2: Music expert**
  The expert claims he can distinguish between a page from a Haydn score and a page from a Mozart score: he does so correctly 10 times.

- **Experiment 3: The Drunk**
  A somewhat inebriated friend at a party claims they can predict the outcome of the toss of a coin: they do so correctly 10 times.

Consider the following three experiments.

- **Experiment 1: Fisher's tea lady**
  The tea lady claims to know whether milk or tea is poured in first: for 10 pairs of cups of tea she makes the correct choice each time.
- **Experiment 2: Music expert**
  The expert claims he can distinguish between a page from a Haydn score and a page from a Mozart score: he does so correctly 10 times.
- **Experiment 3: The Drunk**
  A somewhat inebriated friend at a party claims they can predict the outcome of the toss of a coin: they do so correctly 10 times.

Introduction

Consider the following three experiments.

- **Experiment 1: Fisher's tea lady**
  The tea lady claims to know whether milk or tea is poured in first: for 10 pairs of cups of tea she makes the correct choice each time.
- **Experiment 2: Music expert**
  The expert claims he can distinguish between a page from a Haydn score and a page from a Mozart score: he does so correctly 10 times.
- **Experiment 3: The Drunk**
  A somewhat inebriated friend at a party claims they can predict the outcome of the toss of a coin: they do so correctly 10 times.

Let $\theta = \text{Pr(correct choice)}$.

Let's suppose the tea lady, the music expert and the drunk *cannot* do as they claim.

Then, just by guessing, we could expect each of them to 'get it right' 5 times out of 10, i.e. $\theta = 1/2$.

Let $\theta =$ Pr(correct choice).

Let's suppose the tea lady, the music expert and the drunk *cannot* do as they claim.

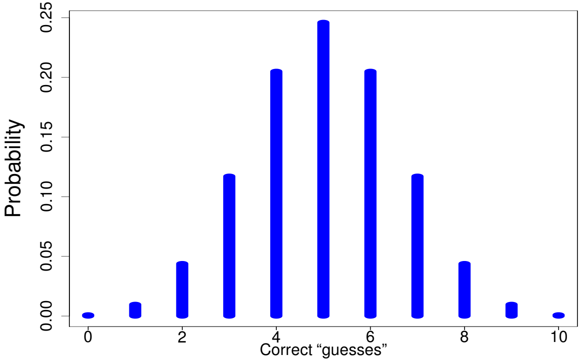
Then, just by guessing, we could expect each of them to 'get it right' 5 times out of 10, i.e. $\theta = 1/2$.

We could then test the **null hypothesis**

$$H_0 : \text{they were guessing}, \qquad \text{i.e.}$$

$$H_0 : \theta = 1/2;$$

Graphically, assuming H_0 is true, this gives:

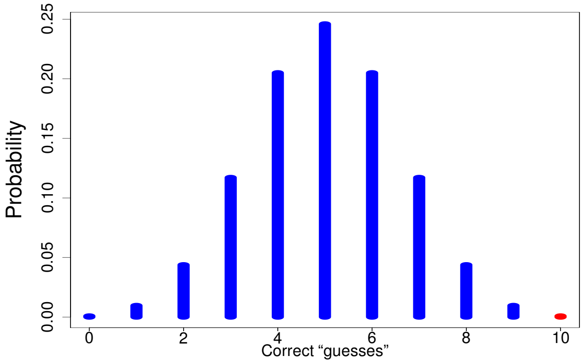


We could then test the **null hypothesis**

$$H_0 : \text{they were guessing}, \qquad \text{i.e.}$$

$$H_0 : \theta = 1/2;$$

Graphically, assuming H_0 is true, this gives:



Introduction

This gives us:

$$p\text{–value} = \left(\frac{1}{2}\right)^{10} = 0.00098 < 0.1\%.$$

From this p–value, we would:

- conclude that we had **very strong evidence** against H_0
- conclude that the choices were **not just guesses**
- perhaps feel justified in **validating each claim**

But does this make sense?

Surely, we have some additional information about what values of θ are **plausible** for each experiment?

Introduction

This gives us:

$$p\text{−value} = \left(\frac{1}{2}\right)^{10} = 0.00098 < 0.1\%.$$

From this p−value, we would:

- conclude that we had **very strong evidence** against H_0
- conclude that the choices were **not just guesses**
- perhaps feel justified in **validating each claim**

But does this make sense?

Surely, we have some additional information about what values of θ are **plausible** for each experiment?

This gives us:

$$p\text{−value} = \left(\frac{1}{2}\right)^{10} = 0.00098 < 0.1\%.$$

From this p−value, we would:

- conclude that we had **very strong evidence** against H_0
- conclude that the choices were **not just guesses**
- perhaps feel justified in **validating each claim**

But does this make sense?

Surely, we have some additional information about what values of θ are **plausible** for each experiment?

Introduction

This gives us:

$$p\text{−value} = \left(\frac{1}{2}\right)^{10} = 0.00098 < 0.1\%.$$

From this p−value, we would:

- conclude that we had **very strong evidence** against H_0
- conclude that the choices were **not just guesses**
- perhaps feel justified in **validating each claim**

But does this make sense?

Surely, we have some additional information about what values of θ are **plausible** for each experiment?

This gives us:

$$p\text{−value} = \left(\frac{1}{2}\right)^{10} = 0.00098 < 0.1\%.$$

From this p−value, we would:

- conclude that we had **very strong evidence** against H_0
- conclude that the choices were **not just guesses**
- perhaps feel justified in **validating each claim**

But does this make sense?

Surely, we have some additional information about what values of θ are **plausible** for each experiment?

Introduction

This gives us:

$$p\text{–value} = \left(\frac{1}{2}\right)^{10} = 0.00098 < 0.1\%.$$

From this p–value, we would:

- conclude that we had **very strong evidence** against H_0
- conclude that the choices were **not just guesses**
- perhaps feel justified in **validating each claim**

But does this make sense?

Surely, we have some additional information about what values of θ are **plausible** for each experiment?

This gives us:

$$p\text{−value} = \left(\frac{1}{2}\right)^{10} = 0.00098 < 0.1\%.$$

From this $p-$value, we would:

- conclude that we had **very strong evidence** against H_0
- conclude that the choices were **not just guesses**
- perhaps feel justified in **validating each claim**

But does this make sense?

Surely, we have some additional information about what values of θ are **plausible** for each experiment?

Prior to each experiment, our beliefs about θ may be

- **Experiment 1:** $\theta > 0.5$ – folklore (and science!) suggests this may be possible;
- **Experiment 2:** $0.9 < \theta < 1.0$ – we expect an "expert" to be correct;
- **Experiment 3:** $\theta = 0.5$ – no way of guessing correctly with a "fair" coin.

Prior to each experiment, our beliefs about θ may be

- **Experiment 1:** $\theta > 0.5$ – folklore (and science!) suggests this may be possible;
- **Experiment 2:** $0.9 < \theta < 1.0$ – we expect an "expert" to be correct;
- **Experiment 3:** $\theta = 0.5$ – no way of guessing correctly with a "fair" coin.

Prior to each experiment, our beliefs about θ may be

- **Experiment 1:** $\theta > 0.5$ – folklore (and science!) suggests this may be possible;
- **Experiment 2:** $0.9 < \theta < 1.0$ – we expect an "expert" to be correct;
- **Experiment 3:** $\theta = 0.5$ – no way of guessing correctly with a "fair" coin.

Prior to each experiment, our beliefs about θ may be

- **Experiment 1:** $\theta > 0.5$ – folklore (and science!) suggests this may be possible;
- **Experiment 2:** $0.9 < \theta < 1.0$ – we expect an "expert" to be correct;
- **Experiment 3:** $\theta = 0.5$ – no way of guessing correctly with a "fair" coin.

The traditional approach to Statistics, sometimes called **Frequentist Statistics** or **Classical Statistics**, may try to take this prior information into account:

For example, in Experiment 2, the test may be of $H_0 : \theta \geq 0.9$ against $H_1 : \theta < 0.9$.

However, in **Bayesian Statistics**, we attempt to calibrate our prior information about unknown quantities.

We do this by constructing a **probability distribution** which describes how likely we believe different values are to occur.

Introduction

The traditional approach to Statistics, sometimes called **Frequentist Statistics** or **Classical Statistics**, may try to take this prior information into account:

For example, in Experiment 2, the test may be of $H_0 : \theta \geq 0.9$ against $H_1 : \theta < 0.9$.

However, in **Bayesian Statistics**, we attempt to calibrate our prior information about unknown quantities.

We do this by constructing a **probability distribution** which describes how likely we believe different values are to occur.

The traditional approach to Statistics, sometimes called **Frequentist Statistics** or **Classical Statistics**, may try to take this prior information into account:

For example, in Experiment 2, the test may be of $H_0 : \theta \geq 0.9$ against $H_1 : \theta < 0.9$.

However, in **Bayesian Statistics**, we attempt to calibrate our prior information about unknown quantities.

We do this by constructing a **probability distribution** which describes how likely we believe different values are to occur.

The traditional approach to Statistics, sometimes called **Frequentist Statistics** or **Classical Statistics**, may try to take this prior information into account:

For example, in Experiment 2, the test may be of $H_0 : \theta \geq 0.9$ against $H_1 : \theta < 0.9$.

However, in **Bayesian Statistics**, we attempt to calibrate our prior information about unknown quantities.

We do this by constructing a **probability distribution** which describes how likely we believe different values are to occur.

This **prior information** is then combined with that from **experimental data** using Bayes Theorem.

The key ingredients of a Bayesian analysis are

- a statistical model for the experimental data;
- quantifiable prior information about any unknown parameters.

Before we consider any detailed descriptions of Bayesian analyses, we recap the various interpretations of probability and highlight the **subjective approach**.

This **prior information** is then combined with that from **experimental data** using Bayes Theorem.

The key ingredients of a Bayesian analysis are

- a statistical model for the experimental data;
- quantifiable prior information about any unknown parameters.

Before we consider any detailed descriptions of Bayesian analyses, we recap the various interpretations of probability and highlight the **subjective approach**.

This **prior information** is then combined with that from **experimental data** using Bayes Theorem.

The key ingredients of a Bayesian analysis are

- a statistical model for the experimental data;
- quantifiable prior information about any unknown parameters.

Before we consider any detailed descriptions of Bayesian analyses, we recap the various interpretations of probability and highlight the **subjective approach**.

This **prior information** is then combined with that from **experimental data** using Bayes Theorem.

The key ingredients of a Bayesian analysis are

- a statistical model for the experimental data;
- quantifiable prior information about any unknown parameters.

Before we consider any detailed descriptions of Bayesian analyses, we recap the various interpretations of probability and highlight the **subjective approach**.

This **prior information** is then combined with that from **experimental data** using Bayes Theorem.

The key ingredients of a Bayesian analysis are

- a statistical model for the experimental data;
- quantifiable prior information about any unknown parameters.

Before we consider any detailed descriptions of Bayesian analyses, we recap the various interpretations of probability and highlight the **subjective approach**.

The concept of probability (chance) has been around for a very long time, particularly in the area of **gambling**.

Games of chance have been played since about **3500 B.C.**

The Egyptians started using cubical dice around **2000 B.C.**

The mathematical theory of probability was started around the 17th century by **Galilei**, **Pascal** and **Fermat** to solve (again) gambling problems.

There are three main ways of understanding and thinking about probability.

The concept of probability (chance) has been around for a very long time, particularly in the area of **gambling**.

Games of chance have been played since about **3500 B.C.**

The Egyptians started using cubical dice around **2000 B.C.**

The mathematical theory of probability was started around the 17th century by **Galilei**, **Pascal** and **Fermat** to solve (again) gambling problems.

There are three main ways of understanding and thinking about probability.

The concept of probability (chance) has been around for a very long time, particularly in the area of **gambling**.

Games of chance have been played since about **3500 B.C.**

The Egyptians started using cubical dice around **2000 B.C.**

The mathematical theory of probability was started around the 17th century by **Galilei**, **Pascal** and **Fermat** to solve (again) gambling problems.

There are three main ways of understanding and thinking about probability.

The concept of probability (chance) has been around for a very long time, particularly in the area of **gambling**.

Games of chance have been played since about **3500 B.C.**

The Egyptians started using cubical dice around **2000 B.C.**

The mathematical theory of probability was started around the 17th century by **Galilei**, **Pascal** and **Fermat** to solve (again) gambling problems.

There are three main ways of understanding and thinking about probability.

The concept of probability (chance) has been around for a very long time, particularly in the area of **gambling**.

Games of chance have been played since about **3500 B.C.**

The Egyptians started using cubical dice around **2000 B.C.**

The mathematical theory of probability was started around the 17th century by **Galilei**, **Pascal** and **Fermat** to solve (again) gambling problems.

There are three main ways of understanding and thinking about probability.

Frequency interpretation

The probability of an outcome is the **relative frequency** with which the outcome would be obtained if the experiment were repeated a large number of times *under similar conditions*.

For example, if a coin is tossed 1,000,000 times and a head appears n times then

$$\Pr(\text{Head}) = \frac{n}{1,000,000}.$$

We would expect this probability to be about 0.5.

Frequency interpretation

The probability of an outcome is the **relative frequency** with which the outcome would be obtained if the experiment were repeated a large number of times *under similar conditions*.

For example, if a coin is tossed 1,000,000 times and a head appears n times then

$$\Pr(\text{Head}) = \frac{n}{1,000,000}.$$

We would expect this probability to be about 0.5.

Frequency interpretation

The probability of an outcome is the **relative frequency** with which the outcome would be obtained if the experiment were repeated a large number of times *under similar conditions*.

For example, if a coin is tossed 1,000,000 times and a head appears n times then

$$\text{Pr(Head)} = \frac{n}{1,000,000}.$$

We would expect this probability to be about 0.5.

This is based on the concept of **equally likely outcomes** resulting from ideas of **symmetry**.

If the outcome of an experiment must be one of n different outcomes and these n outcomes are equally likely, then the probability of each outcome is $1/n$.

Classical interpretation

This is based on the concept of **equally likely outcomes** resulting from ideas of **symmetry**.

If the outcome of an experiment must be one of n different outcomes and these n outcomes are equally likely, then the probability of each outcome is $1/n$.

Your **subjective probability** for an outcome A represents your own judgement of the likelihood that the outcome will occur.

This judgement will be based on the beliefs and information H you have at the time.

One way of determining (or quantifying) a subjective value for $\Pr_H(A)$ is to consider a series of possible bets with outcome

win £c if A occurs and £0 if A^c occurs.

How much would you be prepared to pay (stake) for placing such a bet?

Your **subjective probability** for an outcome A represents your own judgement of the likelihood that the outcome will occur.

This judgement will be based on the beliefs and information H you have at the time.

One way of determining (or quantifying) a subjective value for $\Pr_H(A)$ is to consider a series of possible bets with outcome

win £c if A occurs and £0 if A^c occurs.

How much would you be prepared to pay (stake) for placing such a bet?

Subjective interpretation

Your **subjective probability** for an outcome A represents your own judgement of the likelihood that the outcome will occur.

This judgement will be based on the beliefs and information H you have at the time.

One way of determining (or quantifying) a subjective value for $\Pr_H(A)$ is to consider a series of possible bets with outcome

$$\text{win £}c \text{ if } A \text{ occurs and £0 if } A^c \text{ occurs.}$$

How much would you be prepared to pay (stake) for placing such a bet?

Your **subjective probability** for an outcome A represents your own judgement of the likelihood that the outcome will occur.

This judgement will be based on the beliefs and information H you have at the time.

One way of determining (or quantifying) a subjective value for $\Pr_H(A)$ is to consider a series of possible bets with outcome

$$\text{win £}c \text{ if } A \text{ occurs and £0 if } A^c \text{ occurs.}$$

How much would you be prepared to pay (stake) for placing such a bet?

In terms of expected winnings, you should be prepared to stake £cp if you believe that $\Pr_H(A) = p$. Why?

Suppose you would win £10 if A occurs and £0 if it does not. Suppose further that you believe A is likely to occur with probability $p = 0.3$.

Then the **expected monetary value** of the bet is

$$£10 \times 0.3 = £3;$$

that is, on average, we can expect to win £3 from this bet. Thus, staking any more than £3 would not be wise!

In terms of expected winnings, you should be prepared to stake £cp if you believe that $\Pr_H(A) = p$. Why?

Suppose you would win £10 if A occurs and £0 if it does not. Suppose further that you believe A is likely to occur with probability $p = 0.3$.

Then the **expected monetary value** of the bet is
$£10 \times 0.3 = £3$;

that is, on average, we can expect to win £3 from this bet. Thus, staking any more than £3 would not be wise!

In terms of expected winnings, you should be prepared to stake £cp if you believe that $\Pr_H(A) = p$. Why?

Suppose you would win £10 if A occurs and £0 if it does not. Suppose further that you believe A is likely to occur with probability $p = 0.3$.

Then the **expected monetary value** of the bet is
£$10 \times 0.3 =$ £3;

that is, on average, we can expect to win £3 from this bet. Thus, staking any more than £3 would not be wise!

In terms of expected winnings, you should be prepared to stake £cp if you believe that $\Pr_H(A) = p$. Why?

Suppose you would win £10 if A occurs and £0 if it does not. Suppose further that you believe A is likely to occur with probability $p = 0.3$.

Then the **expected monetary value** of the bet is
$£10 \times 0.3 = £3$;

that is, on average, we can expect to win £3 from this bet. Thus, staking any more than £3 would not be wise!

In terms of expected winnings, you should be prepared to stake £cp if you believe that $\Pr_H(A) = p$. Why?

Suppose you would win £10 if A occurs and £0 if it does not. Suppose further that you believe A is likely to occur with probability $p = 0.3$.

Then the **expected monetary value** of the bet is
$\pounds 10 \times 0.3 = \pounds 3$;

that is, on average, we can expect to win £3 from this bet. Thus, staking any more than £3 would not be wise!

One problem with this approach is that, in general, p will depend on c.

A person who is willing to bet £1 on the spin of a coin to win £2 if it lands heads may refuse to bet if the stakes are raised to £1000: most people are **risk–averse**.

Therefore, we shall restrict our attention to the $c = 1$ case: pay £p for the bet

win £1 if A occurs and £0 if A^c occurs.

Subjective interpretation

One problem with this approach is that, in general, p will depend on c.

A person who is willing to bet £1 on the spin of a coin to win £2 if it lands heads may refuse to bet if the stakes are raised to £1000: most people are **risk–averse**.

Therefore, we shall restrict our attention to the $c = 1$ case: pay £p for the bet

win £1 if A occurs and £0 if A^c occurs.

One problem with this approach is that, in general, p will depend on c.

A person who is willing to bet £1 on the spin of a coin to win £2 if it lands heads may refuse to bet if the stakes are raised to £1000: most people are **risk–averse**.

Therefore, we shall restrict our attention to the $c = 1$ case: pay £p for the bet

win £1 if A occurs and £0 if A^c occurs.

One problem with this approach is that, in general, p will depend on c.

A person who is willing to bet £1 on the spin of a coin to win £2 if it lands heads may refuse to bet if the stakes are raised to £1000: most people are **risk–averse**.

Therefore, we shall restrict our attention to the $c = 1$ case: pay £p for the bet

win £1 if A occurs and £0 if A^c occurs.

Subjective interpretation

You can make sure that your bet is "honest" by randomising between whether you "**host**" the bet or "**place**" the bet.

For example, suppose you believe that $\Pr_H(A) = 0.5$.

An "honest" bet would mean that you would buy the bet for a maximum stake of £0.50.

However, if you weren't honest you might try to buy the bet for any amount less than £0.50, say £0.20.

If you were hosting the bet, you would take the bet for any amount more than £p, say for £0.80.

These conflicting interests can be offset if, when choosing p, it is equally likely that you are hosting the bet or placing the bet.

You can make sure that your bet is "honest" by randomising between whether you "**host**" the bet or "**place**" the bet.

For example, suppose you believe that $\Pr_H(A) = 0.5$.

An "honest" bet would mean that you would buy the bet for a maximum stake of £0.50.

However, if you weren't honest you might try to buy the bet for any amount less than £0.50, say £0.20.

If you were hosting the bet, you would take the bet for any amount more than £p, say for £0.80.

These conflicting interests can be offset if, when choosing p, it is equally likely that you are hosting the bet or placing the bet.

Subjective interpretation

You can make sure that your bet is "honest" by randomising between whether you "**host**" the bet or "**place**" the bet.

For example, suppose you believe that $\Pr_H(A) = 0.5$.

An "honest" bet would mean that you would buy the bet for a maximum stake of £0.50.

However, if you weren't honest you might try to buy the bet for any amount less than £0.50, say £0.20.

If you were hosting the bet, you would take the bet for any amount more than £p, say for £0.80.

These conflicting interests can be offset if, when choosing p, it is equally likely that you are hosting the bet or placing the bet.

You can make sure that your bet is "honest" by randomising between whether you "**host**" the bet or "**place**" the bet.

For example, suppose you believe that $\Pr_H(A) = 0.5$.

An "honest" bet would mean that you would buy the bet for a maximum stake of £0.50.

However, if you weren't honest you might try to buy the bet for any amount less than £0.50, say £0.20.

If you were hosting the bet, you would take the bet for any amount more than £p, say for £0.80.

These conflicting interests can be offset if, when choosing p, it is equally likely that you are hosting the bet or placing the bet.

Subjective interpretation

You can make sure that your bet is "honest" by randomising between whether you "**host**" the bet or "**place**" the bet.

For example, suppose you believe that $\Pr_H(A) = 0.5$.

An "honest" bet would mean that you would buy the bet for a maximum stake of £0.50.

However, if you weren't honest you might try to buy the bet for any amount less than £0.50, say £0.20.

If you were hosting the bet, you would take the bet for any amount more than £p, say for £0.80.

These conflicting interests can be offset if, when choosing p, it is equally likely that you are hosting the bet or placing the bet.

You can make sure that your bet is "honest" by randomising between whether you "**host**" the bet or "**place**" the bet.

For example, suppose you believe that $\Pr_H(A) = 0.5$.

An "honest" bet would mean that you would buy the bet for a maximum stake of £0.50.

However, if you weren't honest you might try to buy the bet for any amount less than £0.50, say £0.20.

If you were hosting the bet, you would take the bet for any amount more than £p, say for £0.80.

These conflicting interests can be offset if, when choosing p, it is equally likely that you are hosting the bet or placing the bet.

Frequency interpretation: drawbacks

1. It does not say how many times the experiment should be repeated
2. "Similar conditions" is a vague concept
3. It is not appropriate for many probability calculations of one-off events
4. Standard statistical methods using the frequentist approach are not totally objective since they require subjective judgements about the validity of probability models, choice of hypotheses and interpretation of results (for instance, see BMI example in the preface to these lecture notes)

Frequency interpretation: drawbacks

1. It does not say how many times the experiment should be repeated
2. "Similar conditions" is a vague concept
3. It is not appropriate for many probability calculations of one-off events
4. Standard statistical methods using the frequentist approach are not totally objective since they require subjective judgements about the validity of probability models, choice of hypotheses and interpretation of results (for instance, see BMI example in the preface to these lecture notes)

Frequency interpretation: drawbacks

1. It does not say how many times the experiment should be repeated
2. “Similar conditions” is a vague concept
3. It is not appropriate for many probability calculations of one-off events
4. Standard statistical methods using the frequentist approach are not totally objective since they require subjective judgements about the validity of probability models, choice of hypotheses and interpretation of results (for instance, see BMI example in the preface to these lecture notes)

Frequency interpretation: drawbacks

1. It does not say how many times the experiment should be repeated
2. “Similar conditions” is a vague concept
3. It is not appropriate for many probability calculations of one-off events
4. Standard statistical methods using the frequentist approach are not totally objective since they require subjective judgements about the validity of probability models, choice of hypotheses and interpretation of results (for instance, see BMI example in the preface to these lecture notes)

Classical interpretation: drawbacks

1. Only applies to equally likely outcomes
2. Depends on a subjective assessment of whether symmetry arguments apply
3. It is not appropriate for many probability calculations of one–off events

Classical interpretation: drawbacks

1. Only applies to equally likely outcomes
2. Depends on a subjective assessment of whether symmetry arguments apply
3. It is not appropriate for many probability calculations of one–off events

Classical interpretation: drawbacks

1. Only applies to equally likely outcomes
2. Depends on a subjective assessment of whether symmetry arguments apply
3. It is not appropriate for many probability calculations of one–off events

Subjective interpretation: drawbacks

1. It is not objective – but perhaps it is more obvious (honest) about when subjective beliefs are used
2. It requires people to be **coherent**: they will not make any wagers which they are certain to lose; also, they will not prefer to suffer a given penalty when there is the option of another penalty which is certainly smaller

   Being coherent results in, *inter alia*, that

$$\Pr(A_1|H) > \Pr(A_2|H) \quad \text{and} \quad \Pr(A_2|H) > \Pr(A_3|H) \\ \implies \quad \Pr(A_1|H) > \Pr(A_3|H).$$

Subjective interpretation: drawbacks

1. It is not objective – but perhaps it is more obvious (honest) about when subjective beliefs are used

2. It requires people to be **coherent**: they will not make any wagers which they are certain to lose; also, they will not prefer to suffer a given penalty when there is the option of another penalty which is certainly smaller

   Being coherent results in, *inter alia*, that

$$\Pr(A_1|H) > \Pr(A_2|H) \quad \text{and} \quad \Pr(A_2|H) > \Pr(A_3|H) \implies \Pr(A_1|H) > \Pr(A_3|H).$$

Subjective interpretation: drawbacks

1. It is not objective – but perhaps it is more obvious (honest) about when subjective beliefs are used
2. It requires people to be **coherent**: they will not make any wagers which they are certain to lose; also, they will not prefer to suffer a given penalty when there is the option of another penalty which is certainly smaller

   Being coherent results in, *inter alia*, that

$$\Pr(A_1|H) > \Pr(A_2|H) \quad \text{and} \quad \Pr(A_2|H) > \Pr(A_3|H)$$
$$\implies \quad \Pr(A_1|H) > \Pr(A_3|H).$$

Subjective interpretation: drawbacks

1. It is not objective – but perhaps it is more obvious (honest) about when subjective beliefs are used

2. It requires people to be **coherent**: they will not make any wagers which they are certain to lose; also, they will not prefer to suffer a given penalty when there is the option of another penalty which is certainly smaller

   Being coherent results in, *inter alia*, that

$$\Pr(A_1|H) > \Pr(A_2|H) \quad \text{and} \quad \Pr(A_2|H) > \Pr(A_3|H) \\ \implies \quad \Pr(A_1|H) > \Pr(A_3|H).$$

Each of these interpretations use quite different methods of reasoning.

In this course – unlike any other course you have taken so far – we will concentrate on the subjective interpretation and describe how, if carefully used, it can be a more useful approach than the other two methods.

Everything we do rests on **Bayes' Theorem**, and we review this now.

Bayes' Theorem

Before we state **Bayes' Theorem**, we need a recap of **conditional probability**.

Definition (Conditional Probability)

Consider two events E and F, where $\Pr(F) > 0$.

The **conditional probability** of E given that F has occurred is

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)}.$$

Before we state **Bayes' Theorem**, we need a recap of **conditional probability**.

Definition (Conditional Probability)

Consider two events E and F, where $\Pr(F) > 0$.

The **conditional probability** of E given that F has occurred is

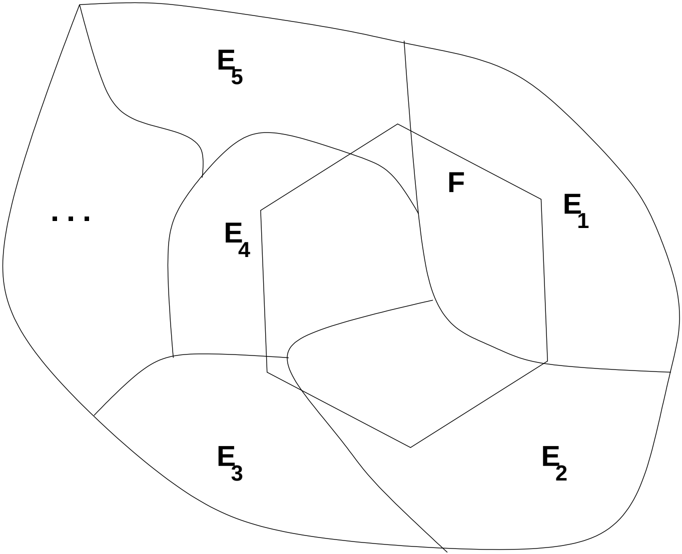
$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)}.$$

Definition (Parition)

The events $E_1, E_2, \ldots, E_n$ form a **partition** of the sample space $\mathcal{S}$ if they are disjoint events ($E_i \cap E_j = \emptyset$, $i \neq j$) with $\Pr(E_i) > 0$, $i = 1, 2, \ldots, n$, and $\cup_{i=1}^{n} E_i = \mathcal{S}$.

Figure 1.1 gives a diagram of a typical partition with an additional event F.

Bayes' Theorem



Fact (Law of Total Probability)

If $E_1, E_2, \ldots, E_n$ are a **partition** *of $\mathcal{S}$ and F is any event then*

$$\Pr(F) = \sum_{i=1}^{n} \Pr(F|E_i)\Pr(E_i).$$

Law of Total Probability: Proof

As $E_1, E_2, \ldots, E_n$ are a **partition** of $\mathcal{S}$, we have

$$\Pr(F) = \Pr(F \cap E_1) + \Pr(F \cap E_2) + \ldots + \Pr(F \cap E_n)$$

$$= \Pr(F|E_1)\Pr(E_1) + \Pr(F|E_2)\Pr(E_2) + \ldots + \Pr(F|E_n)\Pr(E_n)$$

(by conditional probability)

$$= \sum_{i=1}^{n} \Pr(F|E_i)\Pr(E_i).$$

Bayes' Theorem

Theorem (Bayes' Theorem)

If $E_1, E_2, \ldots, E_n$ are a **partition** *of $\mathcal{S}$ and F is any event with $\Pr(F) > 0$ then*

$$\Pr(E_i|F) = \frac{\Pr(F|E_i)\Pr(E_i)}{\sum_{j=1}^{n} \Pr(F|E_j)\Pr(E_j)}, \qquad i = 1, 2, \ldots, n.$$

Bayes' Theorem: Proof (slide 1/1)

Using the definition of conditional probability, for $i = 1, 2, \ldots, n$,

$$\Pr(E_i|F) = \frac{\Pr(E_i \cap F)}{\Pr(F)}$$

$$= \frac{\Pr(F \cap E_i)}{\Pr(F)}$$

$$= \frac{\Pr(F|E_i)\Pr(E_i)}{\Pr(F)} \quad \text{again using the definition}$$

$$= \frac{\Pr(F|E_i)\Pr(E_i)}{\sum_{j=1}^{n} \Pr(F|E_j)\Pr(E_j)} \quad \text{using the Law of Total Probability.}$$

Using the definition of conditional probability, for $i = 1, 2, \ldots, n$,

$$\Pr(E_i|F) = \frac{\Pr(E_i \cap F)}{\Pr(F)}$$

$$= \frac{\Pr(F \cap E_i)}{\Pr(F)}$$

$$= \frac{\Pr(F|E_i)\Pr(E_i)}{\Pr(F)}$$ again using the definition

$$= \frac{\Pr(F|E_i)\Pr(E_i)}{\sum_{j=1}^{n} \Pr(F|E_j)\Pr(E_j)}$$ using the Law of Total Probability.

Example 1.2

A laboratory blood test is 95% effective in detecting a certain disease when it is present.

However, the test also yields a “false positive” result for 1% of healthy people tested.

Also, 0.5% of the population actually have the disease.

(a) Calculate the probability that a person who tests positive actually has the disease.

(b) Find the probability that a person who tests negative does *not* have the disease.

Solution to Example 1.2 (slide 1/3)

Let D = person has the disease, +ve = result is postive and –ve = result is negative.

We require (a) $\Pr(D|+\text{ve})$, and (b) $\Pr(D^c|-\text{ve})$.

From the question we have

- $\Pr(+\text{ve}|D) = 0.95$
- $\Pr(+\text{ve}|D^c) = 0.01$
- $\Pr(D) = 0.005$ and
- $\Pr(D^c) = 0.995$

Solution to Example 1.2 (slide 1/3)

Let $D =$ person has the disease, +ve $=$ result is postive and –ve $=$ result is negative.

We require (a) $\Pr(D|\text{+ve})$, and (b) $\Pr(D^c|\text{–ve})$.

From the question we have

- $\Pr(\text{+ve}|D) = 0.95$
- $\Pr(\text{+ve}|D^c) = 0.01$
- $\Pr(D) = 0.005$ and
- $\Pr(D^c) = 0.995$

Let $D =$ person has the disease, +ve $=$ result is postive and –ve $=$ result is negative.

We require (a) $\Pr(D|+\text{ve})$, and (b) $\Pr(D^c|-\text{ve})$.

From the question we have

- $\Pr(+\text{ve}|D) = 0.95$
- $\Pr(+\text{ve}|D^c) = 0.01$
- $\Pr(D) = 0.005$ and
- $\Pr(D^c) = 0.995$

Solution to Example 1.2 (slide 1/3)

Let D = person has the disease, +ve = result is postive and –ve = result is negative.

We require (a) $\Pr(D|\text{+ve})$, and (b) $\Pr(D^c|\text{–ve})$.

From the question we have

- $\Pr(\text{+ve}|D) = 0.95$
- $\Pr(\text{+ve}|D^c) = 0.01$
- $\Pr(D) = 0.005$ and
- $\Pr(D^c) = 0.995$

Solution to Example 1.2 (slide 2/3)

Also, D and D^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem

$$\Pr(D|+\text{ve}) = \frac{\Pr(+\text{ve}|D)\Pr(D)}{\Pr(+\text{ve}|D)\Pr(D) + \Pr(+\text{ve}|D^c)\Pr(D^c)}$$

$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 \, + \, 0.01 \times 0.995}$$

$$\simeq 0.323.$$

Thus, only 32.3% of people who test positive actually have the disease – 67.7% of people who test positive do *not* have the disease!

Solution to Example 1.2 (slide 2/3)

Also, D and D^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem

$$\Pr(D|\text{+ve}) = \frac{\Pr(\text{+ve}|D)\Pr(D)}{\Pr(\text{+ve}|D)\Pr(D) + \Pr(\text{+ve}|D^c)\Pr(D^c)}$$

$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 \ + \ 0.01 \times 0.995}$$

$$\simeq 0.323.$$

Thus, only 32.3% of people who test positive actually have the disease – 67.7% of people who test positive do *not* have the disease!

Solution to Example 1.2 (slide 2/3)

Also, D and D^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem

$$\Pr(D|\text{+ve}) = \frac{\Pr(\text{+ve}|D)\Pr(D)}{\Pr(\text{+ve}|D)\Pr(D) + \Pr(\text{+ve}|D^c)\Pr(D^c)}$$

$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 \; + \; 0.01 \times 0.995}$$

$$\simeq 0.323.$$

Thus, only 32.3% of people who test positive actually have the disease – 67.7% of people who test positive do *not* have the disease!

Solution to Example 1.2 (slide 2/3)

Also, D and D^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem

$$\Pr(D|+\text{ve}) = \frac{\Pr(+\text{ve}|D)\Pr(D)}{\Pr(+\text{ve}|D)\Pr(D) + \Pr(+\text{ve}|D^c)\Pr(D^c)}$$

$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 \; + \; 0.01 \times 0.995}$$

$$\simeq 0.323.$$

Thus, only 32.3% of people who test positive actually have the disease – 67.7% of people who test positive do *not* have the disease!

Solution to Example 1.2 (slide 3/3)

Now for part (b),

$$\Pr(D^c|\text{–ve}) = \frac{\Pr(\text{–ve}|D^c)\Pr(D^c)}{\Pr(\text{–ve}|D^c)\Pr(D^c) + \Pr(\text{–ve}|D)\Pr(D)}$$

$$= \frac{0.99 \times 0.995}{0.99 \times 0.995 \ + \ 0.05 \times 0.005}$$

$$\simeq 0.9997.$$

Therefore, nearly everyone who tests negative does not have the disease – very few people who test negative *will* have the disease.

Now for part (b),

$$\Pr(D^c|\text{–ve}) = \frac{\Pr(\text{–ve}|D^c)\Pr(D^c)}{\Pr(\text{–ve}|D^c)\Pr(D^c) + \Pr(\text{–ve}|D)\Pr(D)}$$

$$= \frac{0.99 \times 0.995}{0.99 \times 0.995 \; + \; 0.05 \times 0.005}$$

$$\simeq 0.9997.$$

Therefore, nearly everyone who tests negative does not have the disease – very few people who test negative *will* have the disease.

Solution to Example 1.2 (slide 3/3)

Now for part (b),

$$\Pr(D^c|\text{–ve}) = \frac{\Pr(\text{–ve}|D^c)\Pr(D^c)}{\Pr(\text{–ve}|D^c)\Pr(D^c) + \Pr(\text{–ve}|D)\Pr(D)}$$

$$= \frac{0.99 \times 0.995}{0.99 \times 0.995 \; + \; 0.05 \times 0.005}$$

$$\simeq 0.9997.$$

Therefore, nearly everyone who tests negative does not have the disease – very few people who test negative *will* have the disease.

Quick recap

- General intro to Bayesian methods
  - Deciphering Enigma messages
  - Air France 447
  - Body Mass Index
- Recap of probability
  - Frequency, classical and subjective interpretations
  - Limitations
  - Rules: conditional probability, total probability, Bayes
  - Examples: blood test, **car diagnosis**, multiple choice exam

Quick recap

- General intro to Bayesian methods
  - Deciphering Enigma messages
  - Air France 447
  - Body Mass Index
- Recap of probability
  - Frequency, classical and subjective interpretations
  - Limitations
  - Rules: conditional probability, total probability, Bayes
  - Examples: blood test, **car diagnosis**, multiple choice exam

Example 1.3

Suppose that your car suffers from two intermittent problems, one caused by a fault in the engine (θ_1) and the other due to a fault in the gearbox (θ_2).

These occur with probabilities 0.4 and 0.6 respectively.

When examined your car exhibits one of the following symptoms

$$\begin{aligned} x_1 &: \text{overheating only,} \\ x_2 &: \text{irregular traction only,} \\ x_3 &: \text{both symptoms.} \end{aligned}$$

Example 1.3

Suppose it is known in the garage trade that these symptoms occur with probabilities that depend on the fault.

The probabilities $\Pr(X = x|\theta)$ are given in Table 1.1.

	O/H x_1	I/T x_2	Both x_3
θ_1: fault in engine	0.1	0.4	0.5
θ_2: fault in gearbox	0.5	0.3	0.2

Construct a diagnostic rule for these symptoms and determine the probability of misdiagnosis.

Solution to Example 1.3 (1/3)

First we must calculate the **posterior probabilities** $\Pr(\theta_1|x)$ and $\Pr(\theta_2|x)$ for $x = x_1, x_2, x_3$.

Since θ_1 and θ_2 form a partition, we can use Bayes' Theorem as follows.

We have

$$\Pr(\theta_1|x_1) = \frac{\Pr(X = x_1|\theta_1)\Pr(\theta_1)}{\Pr(X = x_1|\theta_1)\Pr(\theta_1) + \Pr(X = x_1|\theta_2)\Pr(\theta_2)}$$

$$= \frac{0.1 \times 0.4}{0.1 \times 0.4 + 0.5 \times 0.6}$$

$$= \frac{4}{34} = 0.118.$$

Solution to Example 1.3 (1/3)

First we must calculate the **posterior probabilities** $\Pr(\theta_1|x)$ and $\Pr(\theta_2|x)$ for $x = x_1, x_2, x_3$.

Since θ_1 and θ_2 form a partition, we can use Bayes' Theorem as follows.

We have

$$\Pr(\theta_1|x_1) = \frac{\Pr(X = x_1|\theta_1)\Pr(\theta_1)}{\Pr(X = x_1|\theta_1)\Pr(\theta_1) + \Pr(X = x_1|\theta_2)\Pr(\theta_2)}$$

$$= \frac{0.1 \times 0.4}{0.1 \times 0.4 + 0.5 \times 0.6}$$

$$= \frac{4}{34} = 0.118.$$

Solution to Example 1.3 (1/3)

First we must calculate the **posterior probabilities** $\Pr(\theta_1|x)$ and $\Pr(\theta_2|x)$ for $x = x_1, x_2, x_3$.

Since θ_1 and θ_2 form a partition, we can use Bayes' Theorem as follows.

We have

$$\Pr(\theta_1|x_1) = \frac{\Pr(X = x_1|\theta_1)\Pr(\theta_1)}{\Pr(X = x_1|\theta_1)\Pr(\theta_1) + \Pr(X = x_1|\theta_2)\Pr(\theta_2)}$$

$$= \frac{0.1 \times 0.4}{0.1 \times 0.4 + 0.5 \times 0.6}$$

$$= \frac{4}{34} = 0.118.$$

Also,

$$\Pr(\theta_1|x_2) = \frac{\Pr(X = x_2|\theta_1)\Pr(\theta_1)}{\Pr(X = x_2|\theta_1)\Pr(\theta_1) + \Pr(X = x_2|\theta_2)\Pr(\theta_2)}$$

$$= \frac{0.4 \times 0.4}{0.4 \times 0.4 + 0.3 \times 0.6}$$

$$= \frac{16}{34} = 0.471.$$

Solution to Example 1.3 (2/3)

Also,

$$\Pr(\theta_1|x_2) = \frac{\Pr(X = x_2|\theta_1)\Pr(\theta_1)}{\Pr(X = x_2|\theta_1)\Pr(\theta_1) + \Pr(X = x_2|\theta_2)\Pr(\theta_2)}$$

$$= \frac{0.4 \times 0.4}{0.4 \times 0.4 + 0.3 \times 0.6}$$

$$= \frac{16}{34} = 0.471.$$

And

$$\Pr(\theta_1|x_3) = \frac{\Pr(X = x_3|\theta_1)\Pr(\theta_1)}{\Pr(X = x_3|\theta_1)\Pr(\theta_1) + \Pr(X = x_3|\theta_2)\Pr(\theta_2)}$$

$$= \frac{0.5 \times 0.4}{0.5 \times 0.4 + 0.2 \times 0.6}$$

$$= \frac{20}{32} = 0.625.$$

Also, $\Pr(\theta_2|x_i) = 1 - \Pr(\theta_1|x_i)$, $i = 1, 2, 3$, and so we obtain the **posterior distributions** $\Pr(\theta|x)$ given in Table 1.2.

Solution to Example 1.3 (3/3)

And

$$\Pr(\theta_1|x_3) = \frac{\Pr(X = x_3|\theta_1)\Pr(\theta_1)}{\Pr(X = x_3|\theta_1)\Pr(\theta_1) + \Pr(X = x_3|\theta_2)\Pr(\theta_2)}$$

$$= \frac{0.5 \times 0.4}{0.5 \times 0.4 + 0.2 \times 0.6}$$

$$= \frac{20}{32} = 0.625.$$

Also, $\Pr(\theta_2|x_i) = 1 - \Pr(\theta_1|x_i)$, $i = 1, 2, 3$, and so we obtain the **posterior distributions** $\Pr(\theta|x)$ given in Table 1.2.

Example 1.3

	O/H x_1	I/T x_2	Both x_3
θ_1: fault in engine	0.118	0.471	0.625
θ_2: fault in gearbox	0.882	0.529	0.375

This table is very informative. For example, it shows that if both symptoms (x_3) are observed, then the probability that the fault is in the engine (θ_1) changes from 0.4 to 0.625.

In terms of **odds**:

$$\text{Prior odds} : \frac{\Pr(\theta_1)}{\Pr(\theta_2)} = \frac{0.4}{0.6} = \frac{2}{3} \quad \text{or 3:2 in favour of } \theta_2$$

$$\text{Posterior odds} : \frac{\Pr(\theta_1|x_3)}{\Pr(\theta_2|x_3)} = \frac{0.625}{0.375} = \frac{5}{3} \quad \text{or 5:3 in favour of } \theta_1.$$

	O/H x_1	I/T x_2	Both x_3
θ_1: fault in engine	0.118	0.471	0.625
θ_2: fault in gearbox	0.882	0.529	0.375

This table is very informative. For example, it shows that if both symptoms (x_3) are observed, then the probability that the fault is in the engine (θ_1) changes from 0.4 to 0.625.

In terms of **odds**:

$$\text{Prior odds} : \frac{\Pr(\theta_1)}{\Pr(\theta_2)} = \frac{0.4}{0.6} = \frac{2}{3} \quad \text{or 3:2 in favour of } \theta_2$$

$$\text{Posterior odds} : \frac{\Pr(\theta_1|x_3)}{\Pr(\theta_2|x_3)} = \frac{0.625}{0.375} = \frac{5}{3} \quad \text{or 5:3 in favour of } \theta_1.$$

Example 1.3

	O/H x_1	I/T x_2	Both x_3
θ_1: fault in engine	0.118	0.471	0.625
θ_2: fault in gearbox	0.882	0.529	0.375

This table is very informative. For example, it shows that if both symptoms (x_3) are observed, then the probability that the fault is in the engine (θ_1) changes from 0.4 to 0.625.

In terms of **odds**:

$$\text{Prior odds}: \frac{\Pr(\theta_1)}{\Pr(\theta_2)} = \frac{0.4}{0.6} = \frac{2}{3} \quad \text{or 3:2 in favour of } \theta_2$$

$$\text{Posterior odds}: \frac{\Pr(\theta_1|x_3)}{\Pr(\theta_2|x_3)} = \frac{0.625}{0.375} = \frac{5}{3} \quad \text{or 5:3 in favour of } \theta_1.$$

Example 1.3

	O/H x_1	I/T x_2	Both x_3
θ_1: fault in engine	0.118	0.471	0.625
θ_2: fault in gearbox	0.882	0.529	0.375

This table is very informative. For example, it shows that if both symptoms (x_3) are observed, then the probability that the fault is in the engine (θ_1) changes from 0.4 to 0.625.

In terms of **odds**:

$$\text{Prior odds}: \frac{\Pr(\theta_1)}{\Pr(\theta_2)} = \frac{0.4}{0.6} = \frac{2}{3} \quad \text{or 3:2 in favour of } \theta_2$$

$$\text{Posterior odds}: \frac{\Pr(\theta_1|x_3)}{\Pr(\theta_2|x_3)} = \frac{0.625}{0.375} = \frac{5}{3} \quad \text{or 5:3 in favour of } \theta_1.$$

Example 1.3

We are now in a position to design our **diagnostic rule**.

This is simply a rule which diagnoses a symptom (x) as being due to some particular fault (θ).

Consider first that we observe overheating only (x_1). The posterior probabilities are in favour of declaring the fault as in the gearbox (θ_2) since $\Pr(\theta_2|x_1) > \Pr(\theta_1|x_1)$.

In the same way, we can determine the most likely diagnosis having observed irregular traction only (x_2) and both symptoms (x_3), giving the diagnostic rule in Table 1.3.

Symptom	Diagnosis
overheating only (x_1)	fault in gearbox (θ_2)
irregular traction only (x_2)	fault in gearbox (θ_2)
both symptoms (x_3)	fault in engine (θ_1)

Example 1.3

We are now in a position to design our **diagnostic rule**.

This is simply a rule which diagnoses a symptom (x) as being due to some particular fault (θ).

Consider first that we observe overheating only (x_1). The posterior probabilities are in favour of declaring the fault as in the gearbox (θ_2) since $\Pr(\theta_2|x_1) > \Pr(\theta_1|x_1)$.

In the same way, we can determine the most likely diagnosis having observed irregular traction only (x_2) and both symptoms (x_3), giving the diagnostic rule in Table 1.3.

Symptom	Diagnosis
overheating only (x_1)	fault in gearbox (θ_2)
irregular traction only (x_2)	fault in gearbox (θ_2)
both symptoms (x_3)	fault in engine (θ_1)

Example 1.3

We are now in a position to design our **diagnostic rule**.

This is simply a rule which diagnoses a symptom (x) as being due to some particular fault (θ).

Consider first that we observe overheating only (x_1). The posterior probabilities are in favour of declaring the fault as in the gearbox (θ_2) since $\Pr(\theta_2|x_1) > \Pr(\theta_1|x_1)$.

In the same way, we can determine the most likely diagnosis having observed irregular traction only (x_2) and both symptoms (x_3), giving the diagnostic rule in Table 1.3.

Symptom	Diagnosis
overheating only (x_1)	fault in gearbox (θ_2)
irregular traction only (x_2)	fault in gearbox (θ_2)
both symptoms (x_3)	fault in engine (θ_1)

Example 1.3

We are now in a position to design our **diagnostic rule**.

This is simply a rule which diagnoses a symptom (x) as being due to some particular fault (θ).

Consider first that we observe overheating only (x_1). The posterior probabilities are in favour of declaring the fault as in the gearbox (θ_2) since $\Pr(\theta_2|x_1) > \Pr(\theta_1|x_1)$.

In the same way, we can determine the most likely diagnosis having observed irregular traction only (x_2) and both symptoms (x_3), giving the diagnostic rule in Table 1.3.

Symptom	Diagnosis
overheating only (x_1)	fault in gearbox (θ_2)
irregular traction only (x_2)	fault in gearbox (θ_2)
both symptoms (x_3)	fault in engine (θ_1)

Example 1.3

We are now in a position to design our **diagnostic rule**.

This is simply a rule which diagnoses a symptom (x) as being due to some particular fault (θ).

Consider first that we observe overheating only (x_1). The posterior probabilities are in favour of declaring the fault as in the gearbox (θ_2) since $\Pr(\theta_2|x_1) > \Pr(\theta_1|x_1)$.

In the same way, we can determine the most likely diagnosis having observed irregular traction only (x_2) and both symptoms (x_3), giving the diagnostic rule in Table 1.3.

Symptom	Diagnosis
overheating only (x_1)	fault in gearbox (θ_2)
irregular traction only (x_2)	fault in gearbox (θ_2)
both symptoms (x_3)	fault in engine (θ_1)

Solution to Example 1.3 (1/1)

If we were to carry out this diagnostic rule repeatedly then the probability of misdiagnosing a fault is

$$
\begin{aligned}
\Pr(\text{Misdiagnosis}) \\
&= \Pr(\theta_1, X = x_1) + \Pr(\theta_1, X = x_2) + \Pr(\theta_2, X = x_3) \\
&= \Pr(X = x_1|\theta_1)\Pr(\theta_1) + \Pr(X = x_2|\theta_1)\Pr(\theta_1) \\
&\qquad + \Pr(X = x_3|\theta_2)\Pr(\theta_2) \\
&= (0.1 \times 0.4) + (0.4 \times 0.4) + (0.2 \times 0.6) \\
&= 0.32.
\end{aligned}
$$

Therefore, in repeated use of this rule, around a third of the diagnoses will be wrong.

Solution to Example 1.3 (1/1)

If we were to carry out this diagnostic rule repeatedly then the probability of misdiagnosing a fault is

$$
\begin{aligned}
\Pr(\text{Misdiagnosis}) & \\
&= \Pr(\theta_1, X = x_1) + \Pr(\theta_1, X = x_2) + \Pr(\theta_2, X = x_3) \\
&= \Pr(X = x_1|\theta_1)\Pr(\theta_1) + \Pr(X = x_2|\theta_1)\Pr(\theta_1) \\
&\qquad + \Pr(X = x_3|\theta_2)\Pr(\theta_2) \\
&= (0.1 \times 0.4) + (0.4 \times 0.4) + (0.2 \times 0.6) \\
&= 0.32.
\end{aligned}
$$

Therefore, in repeated use of this rule, around a third of the diagnoses will be wrong.

Solution to Example 1.3 (1/1)

If we were to carry out this diagnostic rule repeatedly then the probability of misdiagnosing a fault is

$$
\begin{aligned}
&\Pr(\text{Misdiagnosis}) \\
&= \Pr(\theta_1, X = x_1) + \Pr(\theta_1, X = x_2) + \Pr(\theta_2, X = x_3) \\
&= \Pr(X = x_1|\theta_1)\Pr(\theta_1) + \Pr(X = x_2|\theta_1)\Pr(\theta_1) \\
&\qquad + \Pr(X = x_3|\theta_2)\Pr(\theta_2) \\
&= (0.1 \times 0.4) + (0.4 \times 0.4) + (0.2 \times 0.6) \\
&= 0.32.
\end{aligned}
$$

Therefore, in repeated use of this rule, around a third of the diagnoses will be wrong.

Example 1.4

A student sits a multiple choice exam in which there are *m* alternative answers to each question.

The student either knows the answer (with probability θ) or guesses randomly (with probability $1 - \theta$).

What is the probability that the student actually knew the answer to a question they answered correctly?

Solution to Example 1.4 (1/2)

Let

$$C = \text{student answers question correctly}$$
$$K = \text{student knows answer.}$$

We require $\Pr(K|C)$. From the question we have

- $\Pr(K) = \theta$
- $\Pr(K^c) = 1 - \theta$
- $\Pr(C|K) = 1$ and
- $\Pr(C|K^c) = \frac{1}{m}$.

Let

$$C = \text{student answers question correctly}$$
$$K = \text{student knows answer.}$$

We require $\Pr(K|C)$. From the question we have

- $\Pr(K) = \theta$
- $\Pr(K^c) = 1 - \theta$
- $\Pr(C|K) = 1$ and
- $\Pr(C|K^c) = \frac{1}{m}$.

Solution to Example 1.4 (1/2)

Let

$$C = \text{student answers question correctly}$$
$$K = \text{student knows answer.}$$

We require $\Pr(K|C)$. From the question we have

- $\Pr(K) = \theta$
- $\Pr(K^c) = 1 - \theta$
- $\Pr(C|K) = 1$ and
- $\Pr(C|K^c) = \frac{1}{m}$.

Solution to Example 1.4 (1/2)

Let

$$C = \text{student answers question correctly}$$
$$K = \text{student knows answer.}$$

We require $\Pr(K|C)$. From the question we have

- $\Pr(K) = \theta$
- $\Pr(K^c) = 1 - \theta$
- $\Pr(C|K) = 1$ and
- $\Pr(C|K^c) = \frac{1}{m}$.

Solution to Example 1.4 (1/2)

Let

$$C = \text{student answers question correctly}$$
$$K = \text{student knows answer.}$$

We require $\Pr(K|C)$. From the question we have

- $\Pr(K) = \theta$
- $\Pr(K^c) = 1 - \theta$
- $\Pr(C|K) = 1$ and
- $\Pr(C|K^c) = \frac{1}{m}$.

Also, K and K^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem,

$$\begin{aligned}
\Pr(K|C) &= \frac{\Pr(C|K)\Pr(K)}{\Pr(C|K)\Pr(K) + \Pr(C|K^c)\Pr(K^c)} \\
&= \frac{1 \times \theta}{1 \times \theta \; + \; \frac{1}{m} \times (1-\theta)} \\
&= \frac{\theta}{\frac{m\theta + (1-\theta)}{m}} \\
&= \frac{m\theta}{1 + (m-1)\theta}.
\end{aligned}$$

Also, K and K^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem,

$$\begin{aligned}
\Pr(K|C) &= \frac{\Pr(C|K)\Pr(K)}{\Pr(C|K)\Pr(K) + \Pr(C|K^c)\Pr(K^c)} \\
&= \frac{1 \times \theta}{1 \times \theta \ + \ \frac{1}{m} \times (1-\theta)} \\
&= \frac{\theta}{\frac{m\theta + (1-\theta)}{m}} \\
&= \frac{m\theta}{1 + (m-1)\theta}.
\end{aligned}$$

Also, K and K^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem,

$$\begin{aligned}
\Pr(K|C) &= \frac{\Pr(C|K)\Pr(K)}{\Pr(C|K)\Pr(K) + \Pr(C|K^c)\Pr(K^c)} \\
&= \frac{1 \times \theta}{1 \times \theta \; + \; \frac{1}{m} \times (1 - \theta)} \\
&= \frac{\theta}{\frac{m\theta + (1-\theta)}{m}} \\
&= \frac{m\theta}{1 + (m-1)\theta}.
\end{aligned}$$

Solution to Example 1.4 (2/2)

Also, K and K^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem,

$$\begin{aligned}
\Pr(K|C) &= \frac{\Pr(C|K)\Pr(K)}{\Pr(C|K)\Pr(K) + \Pr(C|K^c)\Pr(K^c)} \\
&= \frac{1 \times \theta}{1 \times \theta \ + \ \frac{1}{m} \times (1-\theta)} \\
&= \frac{\theta}{\frac{m\theta + (1-\theta)}{m}} \\
&= \frac{m\theta}{1 + (m-1)\theta}.
\end{aligned}$$

Solution to Example 1.4 (2/2)

Also, K and K^c form a partition of $\mathcal{S}$. Therefore, by Bayes' Theorem,

$$\begin{aligned}
\Pr(K|C) &= \frac{\Pr(C|K)\Pr(K)}{\Pr(C|K)\Pr(K) + \Pr(C|K^c)\Pr(K^c)} \\
&= \frac{1 \times \theta}{1 \times \theta \; + \; \frac{1}{m} \times (1 - \theta)} \\
&= \frac{\theta}{\frac{m\theta + (1-\theta)}{m}} \\
&= \frac{m\theta}{1 + (m-1)\theta}.
\end{aligned}$$

Example 1.4

Suppose that there are $m = 5$ alternative answers for each question. Then

$$\Pr(K|C) = \frac{5\theta}{1+4\theta}.$$

We can see the effect of observing a correct answer on our belief that the student actually knows the answer by calculating $\Pr(K|C)$ for various θ – see Table 1.4.

Example 1.4

Suppose that there are $m = 5$ alternative answers for each question. Then

$$\Pr(K|C) = \frac{5\theta}{1 + 4\theta}.$$

We can see the effect of observing a correct answer on our belief that the student actually knows the answer by calculating $\Pr(K|C)$ for various θ – see Table 1.4.

Example 1.4

Suppose that there are $m = 5$ alternative answers for each question. Then

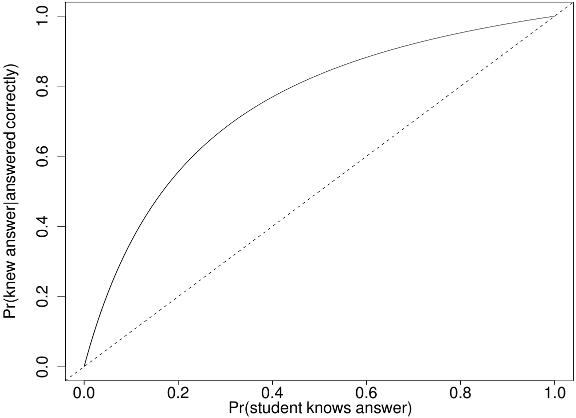
$$\text{Pr}(K|C) = \frac{5\theta}{1 + 4\theta}.$$

We can see the effect of observing a correct answer on our belief that the student actually knows the answer by calculating $\text{Pr}(K|C)$ for various θ – see Table 1.4.

Example 1.4

$\Pr(K)$ $= \theta$	$\Pr(K\|C)$ $= 5\theta/(1+4\theta)$
0.0	0.000
0.1	0.357
0.2	0.556
0.3	0.682
0.4	0.769
0.5	0.833
0.6	0.882
0.7	0.921
0.8	0.952
0.9	0.978
1.0	1.000

Example 1.4



Pr(knew answer|answered correctly)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Pr(student knows answer)

Likelihood

Suppose that an experiment results in data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ and we decide to model the data using a probability (density) function $f(\boldsymbol{x}|\theta)$.

This p(d)f describes how likely different data $\boldsymbol{x}$ are to occur given a value of the (unknown) parameter θ.

However, once we have observed the data, $f(\boldsymbol{x}|\theta)$ tells us how likely different values of the parameters θ are: it is then known as the **likelihood function** for θ.

In other courses you may have seen it written as $L(\theta|\boldsymbol{x})$ or $L(\theta)$ but, whatever the notation used for the likelihood function, it is simply the joint probability (density) function of the data, $f(\boldsymbol{x}|\theta)$, regarded as a function of θ rather than of $\boldsymbol{x}$.

The likelihood function can be simplified if we have further structure in the data.

For example, we may have independent observations, in which case

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f_{X_i}(x_i|\theta), \tag{1.1}$$

or independent and identically distributed observations (random sample), so that

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta). \tag{1.2}$$

Example 1.5

Suppose we have a random sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ of radioactive particle counts.

A typical model for such data would be $X_i|\theta \sim Poisson(\theta)$, usually abbreviated $X_i|\theta \sim Po(\theta)$, (independent).

Determine the likelihood function for θ.

Example 1.5

Suppose we have a random sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ of radioactive particle counts.

A typical model for such data would be $X_i|\theta \sim Poisson(\theta)$, usually abbreviated $X_i|\theta \sim Po(\theta)$, (independent).

Determine the likelihood function for θ.

Example 1.5

Suppose we have a random sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ of radioactive particle counts.

A typical model for such data would be $X_i|\theta \sim Poisson(\theta)$, usually abbreviated $X_i|\theta \sim Po(\theta)$, (independent).

Determine the likelihood function for θ.

Solution to Example 1.5 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\
&= \frac{e^{-\theta} \times \ldots \times e^{-\theta} \times \theta^{x_1} \times \ldots \times \theta^{x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-\theta-\theta-\ldots-\theta}\theta^{x_1+\ldots+x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}.
\end{aligned}
$$

Solution to Example 1.5 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\
&= \frac{e^{-\theta} \times \ldots \times e^{-\theta} \times \theta^{x_1} \times \ldots \times \theta^{x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-\theta-\theta-\ldots-\theta}\theta^{x_1+\ldots+x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}.
\end{aligned}
$$

Solution to Example 1.5 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\
&= \frac{e^{-\theta} \times \ldots \times e^{-\theta} \times \theta^{x_1} \times \ldots \times \theta^{x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-\theta-\theta-\ldots-\theta}\theta^{x_1+\ldots+x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}.
\end{aligned}
$$

Solution to Example 1.5 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\
&= \frac{e^{-\theta} \times \ldots \times e^{-\theta} \times \theta^{x_1} \times \ldots \times \theta^{x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-\theta-\theta-\ldots-\theta}\theta^{x_1+\ldots+x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}.
\end{aligned}
$$

Solution to Example 1.5 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\
&= \frac{e^{-\theta} \times \ldots \times e^{-\theta} \times \theta^{x_1} \times \ldots \times \theta^{x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-\theta-\theta-\ldots-\theta}\theta^{x_1+\ldots+x_n}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}.
\end{aligned}
$$

Solution to Example 1.5 (1/1)

Using Equation (1.2), the likelihood function is

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\ &= \frac{e^{-\theta} \times \ldots \times e^{-\theta} \times \theta^{x_1} \times \ldots \times \theta^{x_n}}{\prod_{i=1}^{n} x_i!} \\ &= \frac{e^{-\theta-\theta-\ldots-\theta}\theta^{x_1+\ldots+x_n}}{\prod_{i=1}^{n} x_i!} \\ &= \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \\ &= \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}. \end{aligned}$$

Example 1.6

Suppose we have a random sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ of times between radioactive particle emissions.

If the emissions occur randomly in time then a plausible model for such data would be $X_i|\theta \sim \textit{Exponential}(\theta)$, usually abbreviated $X_i|\theta \sim \textit{Exp}(\theta)$, (independent).

Determine the likelihood function for θ.

Solution to Example 1.6 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta \times \ldots \times \theta \times e^{-\theta x_1} \times \ldots \times e^{-\theta x_n} \\
&= \theta^n e^{-\theta x_1 - \theta x_2 - \ldots - \theta x_n} \\
&= \theta^n e^{-\theta(x_1 + \ldots + x_n)} \\
&= \theta^n e^{-\theta n \bar{x}}.
\end{aligned}
$$

Solution to Example 1.6 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta \times \ldots \times \theta \times e^{-\theta x_1} \times \ldots \times e^{-\theta x_n} \\
&= \theta^n e^{-\theta x_1 - \theta x_2 - \ldots - \theta x_n} \\
&= \theta^n e^{-\theta(x_1 + \ldots + x_n)} \\
&= \theta^n e^{-\theta n \bar{x}}.
\end{aligned}
$$

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta \times \ldots \times \theta \times e^{-\theta x_1} \times \ldots \times e^{-\theta x_n} \\
&= \theta^n e^{-\theta x_1 - \theta x_2 - \ldots - \theta x_n} \\
&= \theta^n e^{-\theta(x_1 + \ldots + x_n)} \\
&= \theta^n e^{-\theta n \bar{x}}.
\end{aligned}
$$

Solution to Example 1.6 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta \times \ldots \times \theta \times e^{-\theta x_1} \times \ldots \times e^{-\theta x_n} \\
&= \theta^n e^{-\theta x_1 - \theta x_2 - \ldots - \theta x_n} \\
&= \theta^n e^{-\theta(x_1 + \ldots + x_n)} \\
&= \theta^n e^{-\theta n \bar{x}}.
\end{aligned}
$$

Solution to Example 1.6 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta \times \ldots \times \theta \times e^{-\theta x_1} \times \ldots \times e^{-\theta x_n} \\
&= \theta^n e^{-\theta x_1 - \theta x_2 - \ldots - \theta x_n} \\
&= \theta^n e^{-\theta(x_1 + \ldots + x_n)} \\
&= \theta^n e^{-\theta n \bar{x}}.
\end{aligned}
$$

Solution to Example 1.6 (1/1)

Using Equation (1.2), the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \theta e^{-\theta x_i} \\
&= \theta \times \ldots \times \theta \times e^{-\theta x_1} \times \ldots \times e^{-\theta x_n} \\
&= \theta^n e^{-\theta x_1 - \theta x_2 - \ldots - \theta x_n} \\
&= \theta^n e^{-\theta(x_1 + \ldots + x_n)} \\
&= \theta^n e^{-\theta n \bar{x}}.
\end{aligned}
$$

Example 1.7

Suppose we have a random sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ from a Normal distribution: $X_i|\mu, \sigma \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ (independent).

Determine the likelihood function for (μ, σ).

Solution to Example 1.7 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\}.
\end{aligned}
$$

Solution to Example 1.7 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\}.
\end{aligned}
$$

Solution to Example 1.7 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\}.
\end{aligned}
$$

Consider again the Poisson model in Example 1.5. The likelihood function is

$$f(\boldsymbol{x}|\theta) = \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$
$$= \left(\prod_{i=1}^{n} x_i!\right)^{-1} \times \theta^{n\bar{x}} e^{-n\theta}.$$

Notice that this likelihood function depends on the data only through $\prod_{i=1}^{n}(x_i!)^{-1}$ and $\bar{x}$.

Further, in $f(\boldsymbol{x}|\theta)$, θ only "interacts" with $\bar{x}$ — the other term simply scales $f(\boldsymbol{x}|\theta)$ — so that, for example, the point at which $f(\boldsymbol{x}|\theta)$ is maximized is determined only by $\bar{x}$.

Sufficiency

Consider again the Poisson model in Example 1.5. The likelihood function is

$$f(\boldsymbol{x}|\theta) = \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$
$$= \left(\prod_{i=1}^{n} x_i!\right)^{-1} \times \theta^{n\bar{x}} e^{-n\theta}.$$

Notice that this likelihood function depends on the data only through $\prod_{i=1}^{n}(x_i!)^{-1}$ and $\bar{x}$.

Further, in $f(\boldsymbol{x}|\theta)$, θ only "interacts" with $\bar{x}$ — the other term simply scales $f(\boldsymbol{x}|\theta)$ — so that, for example, the point at which $f(\boldsymbol{x}|\theta)$ is maximized is determined only by $\bar{x}$.

Sufficiency

Consider again the Poisson model in Example 1.5. The likelihood function is

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

$$= \left(\prod_{i=1}^{n} x_i!\right)^{-1} \times \theta^{n\bar{x}}e^{-n\theta}.$$

Notice that this likelihood function depends on the data only through $\prod_{i=1}^{n}(x_i!)^{-1}$ and $\bar{x}$.

Further, in $f(\boldsymbol{x}|\theta)$, θ only "interacts" with $\bar{x}$ — the other term simply scales $f(\boldsymbol{x}|\theta)$ — so that, for example, the point at which $f(\boldsymbol{x}|\theta)$ is maximized is determined only by $\bar{x}$.

Sufficiency

Consider again the Poisson model in Example 1.5. The likelihood function is

$$f(\boldsymbol{x}|\theta) = \frac{\theta^{n\bar{x}}e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$
$$= \left(\prod_{i=1}^{n} x_i!\right)^{-1} \times \theta^{n\bar{x}}e^{-n\theta}.$$

Notice that this likelihood function depends on the data only through $\prod_{i=1}^{n}(x_i!)^{-1}$ and $\bar{x}$.

Further, in $f(\boldsymbol{x}|\theta)$, θ only "interacts" with $\bar{x}$ — the other term simply scales $f(\boldsymbol{x}|\theta)$ — so that, for example, the point at which $f(\boldsymbol{x}|\theta)$ is maximized is determined only by $\bar{x}$.

Sufficiency

Consider again the Poisson model in Example 1.5. The likelihood function is

$$f(\boldsymbol{x}|\theta) = \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$
$$= \left(\prod_{i=1}^{n} x_i!\right)^{-1} \times \theta^{n\bar{x}} e^{-n\theta}.$$

Notice that this likelihood function depends on the data only through $\prod_{i=1}^{n}(x_i!)^{-1}$ and $\bar{x}$.

Further, in $f(\boldsymbol{x}|\theta)$, θ only "interacts" with $\bar{x}$ — the other term simply scales $f(\boldsymbol{x}|\theta)$ — so that, for example, the point at which $f(\boldsymbol{x}|\theta)$ is maximized is determined only by $\bar{x}$.

Sufficiency

Consider again the Poisson model in Example 1.5. The likelihood function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \\
&= \left(\prod_{i=1}^{n} x_i!\right)^{-1} \times \theta^{n\bar{x}} e^{-n\theta}.
\end{aligned}
$$

Notice that this likelihood function depends on the data only through $\prod_{i=1}^{n}(x_i!)^{-1}$ and $\bar{x}$.

Further, in $f(\boldsymbol{x}|\theta)$, θ only "interacts" with $\bar{x}$ — the other term simply scales $f(\boldsymbol{x}|\theta)$ — so that, for example, the point at which $f(\boldsymbol{x}|\theta)$ is maximized is determined only by $\bar{x}$.

Sufficiency

Informally, we think of all the information about θ in the data being contained in $\bar{x}$.

More formally, we can show that the distribution of the data given the value $\bar{x}$ does not depend on θ.

Sufficiency

Informally, we think of all the information about θ in the data being contained in $\bar{x}$.

More formally, we can show that the distribution of the data given the value $\bar{x}$ does not depend on θ.

Definition (Statistic)

A **statistic** is any function of the data (and not of unknown parameters).

Definition (Sufficiency)

The statistic $T(\boldsymbol{X})$ is **sufficient** for θ if $f(\boldsymbol{x}|T(\boldsymbol{X})=t)$ does not depend on θ.

Sufficiency

Definition (Statistic)

A **statistic** is any function of the data (and not of unknown parameters).

Definition (Sufficiency)

The statistic $T(\boldsymbol{X})$ is **sufficient** for θ if $f(\boldsymbol{x}|T(\boldsymbol{X}) = t)$ does not depend on θ.

Example 1.8

Consider again the Poisson model in Example 1.5.

Suppose we had just two observations.

Then $n = 2$ and $X_i|\theta \sim Po(\theta)$, $i = 1, 2$ (independent).

Show that $T = X_1 + X_2$ is sufficient for θ. Note that $T|\theta \sim Po(2\theta)$.

Example 1.8

Consider again the Poisson model in Example 1.5.

Suppose we had just two observations.

Then $n = 2$ and $X_i|\theta \sim Po(\theta)$, $i = 1, 2$ (independent).

Show that $T = X_1 + X_2$ is sufficient for θ. Note that $T|\theta \sim Po(2\theta)$.

Example 1.8

Consider again the Poisson model in Example 1.5.

Suppose we had just two observations.

Then $n = 2$ and $X_i|\theta \sim Po(\theta)$, $i = 1, 2$ (independent).

Show that $T = X_1 + X_2$ is sufficient for θ. Note that $T|\theta \sim Po(2\theta)$.

Example 1.8

Consider again the Poisson model in Example 1.5.

Suppose we had just two observations.

Then $n = 2$ and $X_i|\theta \sim Po(\theta)$, $i = 1, 2$ (independent).

Show that $T = X_1 + X_2$ is sufficient for θ. Note that $T|\theta \sim Po(2\theta)$.

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$\begin{aligned} f(\boldsymbol{x}|T(\boldsymbol{X}) = t) &= \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} \\ &= \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}. \end{aligned}$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned} \Pr(X_1 = x_1, X_2 = t - x_1|\theta) &= \Pr(X_1 = x_1|\theta)\Pr(X_2 = t - x_1|\theta) \\ &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t - x_1)!} \\ &= \frac{e^{-2\theta}\theta^{x_1+t-x_1}}{x_1!(t - x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t - x_1)!}. \end{aligned}$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$f(\boldsymbol{x}|T(\boldsymbol{X}) = t) = \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)}$$

$$= \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}.$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned}
\Pr(X_1 = x_1, X_2 = t - x_1|\theta) &= \Pr(X_1 = x_1|\theta)\Pr(X_2 = t - x_1|\theta) \\
&= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t - x_1)!} \\
&= \frac{e^{-2\theta}\theta^{x_1+t-x_1}}{x_1!(t - x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t - x_1)!}.
\end{aligned}$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$\begin{aligned} f(\boldsymbol{x}|T(\boldsymbol{X}) = t) &= \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} \\ &= \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}. \end{aligned}$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned} \Pr(X_1 = x_1, X_2 = t - x_1|\theta) &= \Pr(X_1 = x_1|\theta)\Pr(X_2 = t - x_1|\theta) \\ &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t-x_1)!} \\ &= \frac{e^{-2\theta}\theta^{x_1+t-x_1}}{x_1!(t-x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t-x_1)!}. \end{aligned}$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$\begin{aligned} f\left(\boldsymbol{x} | T(\boldsymbol{X}) = t\right) &= \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} \\ &= \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}. \end{aligned}$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned} \Pr(X_1 = x_1, X_2 = t - x_1 | \theta) &= \Pr(X_1 = x_1 | \theta) \Pr(X_2 = t - x_1 | \theta) \\ &= \frac{e^{-\theta} \theta^{x_1}}{x_1!} \times \frac{e^{-\theta} \theta^{t - x_1}}{(t - x_1)!} \\ &= \frac{e^{-2\theta} \theta^{x_1 + t - x_1}}{x_1!(t - x_1)!} = \frac{e^{-2\theta} \theta^{t}}{x_1!(t - x_1)!}. \end{aligned}$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$f(\boldsymbol{x}|T(\boldsymbol{X}) = t) = \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} = \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}.$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\Pr(X_1 = x_1, X_2 = t - x_1|\theta) = \Pr(X_1 = x_1|\theta)\Pr(X_2 = t - x_1|\theta)$$

$$= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t-x_1)!}$$

$$= \frac{e^{-2\theta}\theta^{x_1+t-x_1}}{x_1!(t-x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t-x_1)!}.$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$\begin{aligned} f\left(\boldsymbol{x} | T(\boldsymbol{X}) = t\right) &= \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} \\ &= \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}. \end{aligned}$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned} \Pr(X_1 = x_1, X_2 = t - x_1 | \theta) &= \Pr(X_1 = x_1 | \theta)\Pr(X_2 = t - x_1 | \theta) \\ &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t - x_1)!} \\ &= \frac{e^{-2\theta}\theta^{x_1+t-x_1}}{x_1!(t - x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t - x_1)!}. \end{aligned}$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$\begin{aligned} f\left(\boldsymbol{x} | T(\boldsymbol{X}) = t\right) &= \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} \\ &= \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}. \end{aligned}$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned} \Pr(X_1 = x_1, X_2 = t - x_1 | \theta) &= \Pr(X_1 = x_1 | \theta)\Pr(X_2 = t - x_1 | \theta) \\ &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t - x_1)!} \\ &= \frac{e^{-2\theta}\theta^{x_1+t-x_1}}{x_1!(t - x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t - x_1)!}. \end{aligned}$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$\begin{aligned} f\left(\boldsymbol{x} | T(\boldsymbol{X}) = t\right) &= \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} \\ &= \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}. \end{aligned}$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned} \Pr(X_1 = x_1, X_2 = t - x_1 | \theta) &= \Pr(X_1 = x_1 | \theta)\Pr(X_2 = t - x_1 | \theta) \\ &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t - x_1)!} \\ &= \frac{e^{-2\theta}\theta^{x_1 + t - x_1}}{x_1!(t - x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t - x_1)!}. \end{aligned}$$

Solution to Example 1.8 (1/2)

Consider the conditional distribution of the data $\boldsymbol{X}$ on $T = t$.

As X_1 and X_2 are discrete random variables, we have

$$f(\boldsymbol{x}|T(\boldsymbol{X}) = t) = \frac{\Pr(X_1 = x_1, X_2 = x_2, T = t)}{\Pr(T = t)} = \frac{\Pr(X_1 = x_1, X_2 = t - x_1)}{\Pr(T = t)}.$$

Now, as X_1 and X_2 are independent, for the numerator we have:

$$\begin{aligned}\Pr(X_1 = x_1, X_2 = t - x_1|\theta) &= \Pr(X_1 = x_1|\theta)\Pr(X_2 = t - x_1|\theta) \\ &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{t-x_1}}{(t - x_1)!} \\ &= \frac{e^{-2\theta}\theta^{x_1+t-x_1}}{x_1!(t - x_1)!} = \frac{e^{-2\theta}\theta^{t}}{x_1!(t - x_1)!}.\end{aligned}$$

For the denominator we know that

$$\Pr(T = t|\theta) \quad = \quad \frac{e^{-2\theta}(2\theta)^t}{t!} \qquad \text{(sum of two Poissons).}$$

Therefore

$$\begin{aligned} f(\boldsymbol{x}|T(\boldsymbol{X}) = t) &= \frac{e^{-2\theta}\theta^t}{x_1!(t-x_1)!} \Big/ \frac{e^{-2\theta}(2\theta)^t}{t!} \\ &= \frac{t!e^{-2\theta}\theta^t}{x_1!(t-x_1)!e^{-2\theta}(2\theta)^t} \\ &= \frac{t!}{2^t x_1!(t-x_1)!} \end{aligned}$$

which does not depend on θ. Hence, by Definition 1.4, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.8 (2/2)

For the denominator we know that

$$\Pr(T = t|\theta) \quad = \quad \frac{e^{-2\theta}(2\theta)^t}{t!} \qquad \text{(sum of two Poissons).}$$

Therefore

$$\begin{aligned} f(\boldsymbol{x}|T(\boldsymbol{X}) = t) &= \frac{e^{-2\theta}\theta^t}{x_1!(t-x_1)!} \bigg/ \frac{e^{-2\theta}(2\theta)^t}{t!} \\ &= \frac{t!e^{-2\theta}\theta^t}{x_1!(t-x_1)!e^{-2\theta}(2\theta)^t} \\ &= \frac{t!}{2^t x_1!(t-x_1)!} \end{aligned}$$

which does not depend on θ. Hence, by Definition 1.4, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.8 (2/2)

For the denominator we know that

$$\Pr(T = t|\theta) = \frac{e^{-2\theta}(2\theta)^t}{t!} \qquad \text{(sum of two Poissons).}$$

Therefore

$$\begin{aligned} f(\boldsymbol{x}|T(\boldsymbol{X}) = t) &= \frac{e^{-2\theta}\theta^t}{x_1!(t-x_1)!} \Bigg/ \frac{e^{-2\theta}(2\theta)^t}{t!} \\ &= \frac{t!e^{-2\theta}\theta^t}{x_1!(t-x_1)!e^{-2\theta}(2\theta)^t} \\ &= \frac{t!}{2^t x_1!(t-x_1)!} \end{aligned}$$

which does not depend on θ. Hence, by Definition 1.4, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.8 (2/2)

For the denominator we know that

$$\Pr(T = t|\theta) \quad = \quad \frac{e^{-2\theta}(2\theta)^t}{t!} \qquad \text{(sum of two Poissons).}$$

Therefore

$$\begin{aligned} f(\boldsymbol{x}|T(\boldsymbol{X}) = t) &= \frac{e^{-2\theta}\theta^t}{x_1!(t-x_1)!} \Bigg/ \frac{e^{-2\theta}(2\theta)^t}{t!} \\ &= \frac{t!e^{-2\theta}\theta^t}{x_1!(t-x_1)!e^{-2\theta}(2\theta)^t} \\ &= \frac{t!}{2^t x_1!(t-x_1)!} \end{aligned}$$

which does not depend on θ. Hence, by Definition 1.4, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.8 (2/2)

For the denominator we know that

$$\Pr(T = t|\theta) \quad = \quad \frac{e^{-2\theta}(2\theta)^t}{t!} \qquad \text{(sum of two Poissons).}$$

Therefore

$$\begin{aligned} f(\boldsymbol{x}|T(\boldsymbol{X}) = t) &= \frac{e^{-2\theta}\theta^t}{x_1!(t-x_1)!} \Big/ \frac{e^{-2\theta}(2\theta)^t}{t!} \\ &= \frac{t!e^{-2\theta}\theta^t}{x_1!(t-x_1)!e^{-2\theta}(2\theta)^t} \\ &= \frac{t!}{2^t x_1!(t-x_1)!} \end{aligned}$$

which does not depend on θ. Hence, by Definition 1.4, $T = X_1 + X_2$ is sufficient for θ.

Sufficiency

Definition (Joint sufficiency)

The statistics $\boldsymbol{T}(\boldsymbol{X}) = \big(T_1(\boldsymbol{X}), T_2(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X})\big)^T$ are **jointly sufficient** for θ if $f(\boldsymbol{x}|\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})$ does not depend on θ.

Theorem (Factorisation Theorem)

Under certain regularity conditions

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

for some functions h and g.

Definition (Joint sufficiency)

The statistics $\boldsymbol{T}(\boldsymbol{X}) = \left(T_1(\boldsymbol{X}), T_2(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X})\right)^T$ are **jointly sufficient** for θ if $f(\boldsymbol{x}|\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})$ does not depend on θ.

Theorem (Factorisation Theorem)

Under certain regularity conditions

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

for some functions h and g.

Example 1.9

Consider again the Poisson model in Example 1.5 with $n = 2$: $X_i|\theta \sim Po(\theta)$, $i = 1, 2$ (independent). Determine a sufficient statistic for θ.

Solution to Example 1.9 (1/1)

The (joint) probability function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{x_2}}{x_2!} \\
&= \frac{e^{-2\theta}\theta^{x_1+x_2}}{x_1!x_2!} \\
&= \frac{1}{x_1!x_2!} \times e^{-2\theta}\theta^{x_1+x_2} \\
&= h(\boldsymbol{x})\, g(x_1+x_2,\theta),
\end{aligned}
$$

where $h(\boldsymbol{x}) = 1/(x_1!x_2!)$ and $g(t,\theta) = e^{-2\theta}\theta^{t}$.

Therefore, by the Factorisation Theorem, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.9 (1/1)

The (joint) probability function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{x_2}}{x_2!} \\
&= \frac{e^{-2\theta}\theta^{x_1+x_2}}{x_1!x_2!} \\
&= \frac{1}{x_1!x_2!} \times e^{-2\theta}\theta^{x_1+x_2} \\
&= h(\boldsymbol{x})\, g(x_1+x_2,\theta),
\end{aligned}
$$

where $h(\boldsymbol{x}) = 1/(x_1!x_2!)$ and $g(t,\theta) = e^{-2\theta}\theta^{t}$.

Therefore, by the Factorisation Theorem, $T = X_1 + X_2$ is sufficient for θ.

The (joint) probability function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{x_2}}{x_2!} \\
&= \frac{e^{-2\theta}\theta^{x_1+x_2}}{x_1!x_2!} \\
&= \frac{1}{x_1!x_2!} \times e^{-2\theta}\theta^{x_1+x_2} \\
&= h(\boldsymbol{x})\, g(x_1 + x_2, \theta),
\end{aligned}
$$

where $h(\boldsymbol{x}) = 1/(x_1!x_2!)$ and $g(t, \theta) = e^{-2\theta}\theta^{t}$.

Therefore, by the Factorisation Theorem, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.9 (1/1)

The (joint) probability function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{x_2}}{x_2!} \\
&= \frac{e^{-2\theta}\theta^{x_1+x_2}}{x_1!x_2!} \\
&= \frac{1}{x_1!x_2!} \times e^{-2\theta}\theta^{x_1+x_2} \\
&= h(\boldsymbol{x})\, g(x_1 + x_2, \theta),
\end{aligned}
$$

where $h(\boldsymbol{x}) = 1/(x_1!x_2!)$ and $g(t, \theta) = e^{-2\theta}\theta^{t}$.

Therefore, by the Factorisation Theorem, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.9 (1/1)

The (joint) probability function is

$$
\begin{aligned}
f(\boldsymbol{x}|\theta) &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{x_2}}{x_2!} \\
&= \frac{e^{-2\theta}\theta^{x_1+x_2}}{x_1!x_2!} \\
&= \frac{1}{x_1!x_2!} \times e^{-2\theta}\theta^{x_1+x_2} \\
&= h(\boldsymbol{x})\, g(x_1+x_2, \theta),
\end{aligned}
$$

where $h(\boldsymbol{x}) = 1/(x_1!x_2!)$ and $g(t,\theta) = e^{-2\theta}\theta^t$.

Therefore, by the Factorisation Theorem, $T = X_1 + X_2$ is sufficient for θ.

Solution to Example 1.9 (1/1)

The (joint) probability function is

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \times \frac{e^{-\theta}\theta^{x_2}}{x_2!} \\ &= \frac{e^{-2\theta}\theta^{x_1+x_2}}{x_1!x_2!} \\ &= \frac{1}{x_1!x_2!} \times e^{-2\theta}\theta^{x_1+x_2} \\ &= h(\boldsymbol{x})\,g(x_1+x_2,\theta), \end{aligned}$$

where $h(\boldsymbol{x}) = 1/(x_1!x_2!)$ and $g(t,\theta) = e^{-2\theta}\theta^t$.

Therefore, by the Factorisation Theorem, $T = X_1 + X_2$ is sufficient for θ.

Example 1.10

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

Determine a sufficient statistic for θ.

Solution to Example 1.10 (1/1)

The (joint) probability density function is

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

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

Solution to Example 1.10 (1/1)

The (joint) probability density function is

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

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

The (joint) probability density function is

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

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

Solution to Example 1.10 (1/1)

The (joint) probability density function is

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

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

Solution to Example 1.10 (1/1)

The (joint) probability density function is

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

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

Example 1.11

Suppose we have a random sample from a Normal distribution: $X_i|\mu,\sigma \sim N(\mu,\sigma^2)$, $i = 1,2,\ldots,n$ (independent).

Determine sufficient statistics for (μ,σ).

Solution to Example 1.11 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\} \\
&= h(\boldsymbol{x})\, g(\Sigma x_i, \Sigma x_i^2, \mu, \sigma),
\end{aligned}
$$

where $h(\boldsymbol{x}) = (2\pi)^{-n/2}$ and
$g(t_1, t_2, \mu, \sigma) = \sigma^{-n}\exp\{-(t_2 - 2\mu t_1 + n\mu^2)/(2\sigma^2)\}$.

Therefore, by the Factorisation Theorem, $T_1 = \sum_{i=1}^{n} X_i$ and $T_2 = \sum_{i=1}^{n} X_i^2$ are (jointly) sufficient for μ and σ.

Solution to Example 1.11 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\} \\
&= h(\boldsymbol{x})\, g(\Sigma\, x_i, \Sigma\, x_i^2, \mu, \sigma),
\end{aligned}
$$

where $h(\boldsymbol{x}) = (2\pi)^{-n/2}$ and
$g(t_1, t_2, \mu, \sigma) = \sigma^{-n}\exp\{-(t_2 - 2\mu t_1 + n\mu^2)/(2\sigma^2)\}$.

Therefore, by the Factorisation Theorem, $T_1 = \sum_{i=1}^{n} X_i$ and $T_2 = \sum_{i=1}^{n} X_i^2$ are (jointly) sufficient for μ and σ.

Solution to Example 1.11 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\} \\
&= h(\boldsymbol{x})\, g(\Sigma\, x_i, \Sigma\, x_i^2, \mu, \sigma),
\end{aligned}
$$

where $h(\boldsymbol{x}) = (2\pi)^{-n/2}$ and
$g(t_1, t_2, \mu, \sigma) = \sigma^{-n}\exp\{-(t_2 - 2\mu t_1 + n\mu^2)/(2\sigma^2)\}$.

Therefore, by the Factorisation Theorem, $T_1 = \sum_{i=1}^{n} X_i$ and $T_2 = \sum_{i=1}^{n} X_i^2$ are (jointly) sufficient for μ and σ.

Solution to Example 1.11 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\} \\
&= h(\boldsymbol{x})\, g(\Sigma\, x_i, \Sigma\, x_i^2, \mu, \sigma),
\end{aligned}
$$

where $h(\boldsymbol{x}) = (2\pi)^{-n/2}$ and
$g(t_1, t_2, \mu, \sigma) = \sigma^{-n}\exp\{-(t_2 - 2\mu t_1 + n\mu^2)/(2\sigma^2)\}$.

Therefore, by the Factorisation Theorem, $T_1 = \sum_{i=1}^{n} X_i$ and $T_2 = \sum_{i=1}^{n} X_i^2$ are (jointly) sufficient for μ and σ.

Solution to Example 1.11 (1/1)

The (joint) probability density function is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\} \\
&= h(\boldsymbol{x})\, g(\Sigma\, x_i, \Sigma\, x_i^2, \mu, \sigma),
\end{aligned}
$$

where $h(\boldsymbol{x}) = (2\pi)^{-n/2}$ and
$g(t_1, t_2, \mu, \sigma) = \sigma^{-n} \exp\{-(t_2 - 2\mu t_1 + n\mu^2)/(2\sigma^2)\}$.

Therefore, by the Factorisation Theorem, $T_1 = \sum_{i=1}^{n} X_i$ and $T_2 = \sum_{i=1}^{n} X_i^2$ are (jointly) sufficient for μ and σ.

Quick quiz

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

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

(a) Form the likelihood function $f(\boldsymbol{x}|\theta)$;

(b) Use the Factorisation Theorem to obtain a sufficient statistic for θ.

Quick quiz: solution

The joint probability function is

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

where $h(\boldsymbol{x}) = 1$ 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

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

where $h(\boldsymbol{x}) = 1$ 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

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

where $h(\boldsymbol{x}) = 1$ 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

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

where $h(\boldsymbol{x}) = 1$ 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

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

where $h(\boldsymbol{x}) = 1$ 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

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

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