WARWICK

WARWICK
THE UNIVERSITY OF WARWICK

Bayesian Estimation and Inference in R

Yun-Xiao Li

Fri 13^{th} Jun, 2025

What?

What are Bayesian estimation and Bayesian inference?

What?

WARWICK

What are Bayesian estimation and Bayesian inference?

- **Bayesian estimation** is a way of using data and prior knowledge to estimate unknown values, like the mean or proportion in a population;

What?

What are Bayesian estimation and Bayesian inference?

- **Bayesian estimation** is a way of using data and prior knowledge to estimate unknown values, like the mean or proportion in a population;
- **Bayesian inference** is a broader term. It's the whole approach of using Bayes' Theorem to update beliefs and make decisions in the face of uncertainty.

Why?

Why should I learn the Bayesian method?

Why?

Why should I learn the Bayesian method?

- Mirrors real-life thinking and is easy to understand (to some extent);

Why?

WARWICK

Why should I learn the Bayesian method?

- Mirrors real-life thinking and is easy to understand (to some extent);
- Makes use of prior knowledge;

Why?

WARWICK

Why should I learn the Bayesian method?

- Mirrors real-life thinking and is easy to understand (to some extent);
- Makes use of prior knowledge;
- Gives direct answers to real questions (not like p-values);

How?

How can I learn the Bayesian method?

How?

How can I learn the Bayesian method?

- Level 1: get the intuitive idea of what's going on;

How?

How can I learn the Bayesian method?

- Level 1: get the intuitive idea of what's going on;
- Level 2: know how to implement it (in R);

How?

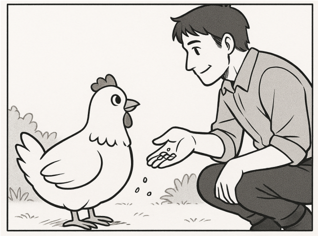
How can I learn the Bayesian method?

- Level 1: get the intuitive idea of what's going on;
- Level 2: know how to implement it (in R);
- Level 3: understand the mathematical principle;

Outline

Bayes Rule

Bayes Rule
Example



Bayes Rule
Example

Bayes Rule
Example

WARWICK

- Two fundamental perceptions of Bayesian inference:

- Two fundamental perceptions of Bayesian inference:
  - You have a prior belief;

WARWICK

- Two fundamental perceptions of Bayesian inference:
  - You have a prior belief;
  - New observations update this belief;

WARWICK

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

WARWICK

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

In the example above:

WARWICK

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

In the example above:

- $p(A|D)$: The chance of the man killing the chicken today if the chicken sees him with an axe;

WARWICK

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

In the example above:

- $p(A|D)$: The chance of the man killing the chicken today if the chicken sees him with an axe;
- $p(D)$: The chance of observing the man holding an axe;

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

In the example above:

- $p(A|D)$: The chance of the man killing the chicken today if the chicken sees him with an axe;
- $p(D)$: The chance of observing the man holding an axe;
- $P(AD)$: The chance of observing the man with an axe and the chicken being killed;

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

In the example above:

- $p(A|D)$: The chance of the man killing the chicken today if the chicken sees him with an axe;
- $p(D)$: The chance of observing the man holding an axe;
- $P(AD)$: The chance of observing the man with an axe and the chicken being killed;
- What is the value of $P(AD)$ and $P(D)$?

Let's warm up the brain!

Let's warm up the brain!

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

WARWICK

Let's warm up the brain!

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

$$p(A|D)p(D) = p(AD)$$

WARWICK

Let's warm up the brain!

Definition (Conditional probability)

Let A stands for an event or a hypothesis, and D stands for the observation, then the probability of event A occurring given that D is observed is

$$p(A|D) = \frac{p(AD)}{p(D)}.$$

$$p(A|D)p(D) = p(AD)$$

$$p(D|A)p(A) = p(DA)$$

WARWICK

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

- $P(A)$ is the **prior** belief of the chance that the man will kill the chicken;
- $P(D|A)$ is the **likelihood** of observing the man with an axe if the man decides to kill the chicken.

Let's decompose $P(D)$!

WARWICK

Let's decompose $P(D)$!

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

Let's decompose $P(D)$!

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

$$\begin{tikzcd}[column sep=small] & p(D) \arrow[dl, "\text{If } A \text{ happens}"'] \arrow[dr, "\text{If } A \text{ does not happen}"] & \\ p(D|A) & & p(D|\neg A) \end{tikzcd}$$

Bayes Rule
Bayes' Theorem

WARWICK

Let's decompose $P(D)$!

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

$$\begin{tikzcd} & p(D) \arrow[dl, "\text{If } A \text{ happens}"'] \arrow[dr, "\text{If } A \text{ does not happen}"] & \\ p(D|A) & & p(D|\neg A) \end{tikzcd}$$

$$p(D) = p(D|A)p(A) + p(D|\neg A)p(\neg A)$$

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Bayes Rule
Bayes' Theorem

WARWICK

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

WARWICK

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*

WARWICK

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*

Hints:

- $P(A)$ is the chance of having the disease;
- $p(D|A)$ is the chance of having a positive result if the person has the disease;

Bayes Rule

Bayes' Theorem

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*

Bayes Rule

Bayes' Theorem

WARWICK

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*

- $P(A) = \frac{1}{1000}$;
- $P(D|A) = 0.95$;
- $P(D|\neg A) = 0.01$.

WARWICK

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*

- $P(A) = \frac{1}{1000}$;
- $P(D|A) = 0.95$;
- $P(D|\neg A) = 0.01$.

$$P(A|D) = \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.01 \times 0.999} \approx 0.087$$

WARWICK

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*
- *Bob got a negative result in the second test.*

WARWICK

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*
- *Bob got a negative result in the second test.*

$$P(A|D) = \frac{0.05 \times 0.087}{0.05 \times 0.087 + 0.99 \times 0.913} \approx 0.0048$$

$$p(A|D) = \frac{p(D|A)p(A)}{p(D|A)p(A) + p(D|\neg A)p(\neg A)}.$$

Example

Assume the incidence rate of a disease is 1/1000. A particular test for whether someone has this disease is 95% sensitive, meaning that if the patient has the disease, this test has a 95% probability of being positive. This test, however, has a 1% probability of being positive even though the patient does not suffer from this disease.

- *Bob got a positive result for this test.*
- *Bob got a negative result in the second test.*

An alternative solution:

$$P(A|D) = \frac{0.95 \times 0.05 \times 0.001}{0.95 \times 0.05 \times 0.001 + 0.01 \times 0.99 \times 0.999} \approx 0.0048$$

What if A is not binary?

WARWICK

What if A is not binary?

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

WARWICK

What if A is not binary?

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

$$A = \{A_1, A_2, A_3\}$$

Bayes Rule

Bayes' Theorem

WARWICK

What if A is not binary?

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

$$A = \{A_1, A_2, A_3\}$$

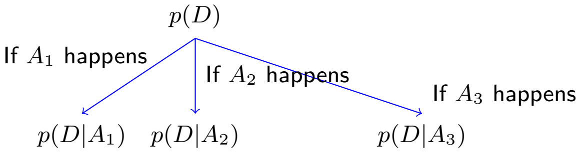
$$\begin{tikzcd} & p(D) \arrow[dl, "\text{If } A_1 \text{ happens}"'] \arrow[d, "\text{If } A_2 \text{ happens}"] \arrow[dr, "\text{If } A_3 \text{ happens}"] & \\ p(D|A_1) & p(D|A_2) & p(D|A_3) \end{tikzcd}$$

What if A is not binary?

Theorem (Bayes' Theorem)

$$p(A|D) = \frac{p(D|A)p(A)}{p(D)}.$$

$$A = \{A_1, A_2, A_3\}$$



WARWICK

Theorem (Bayes' Theorem)

If $A = \{A_1, A_2, \cdots, A_N\}$ *is an event set, then the probability of event* $A_i \in A$ *occurring given that* D *is observed is*

$$p(A_i|D) = \frac{p(D|A_i)p(A_i)}{\sum_{i=1}^{N} p(D|A_i)p(A_i)}.$$

Bayes Rule

Summary

- The main idea of the Bayes rule is: Prior $+$ Observation (Likelihood) $=$ Posterior

- The main idea of the Bayes rule is: Prior $+$ Observation (Likelihood) $=$ Posterior
- The mathematical form of this formula is

$$p(A_i|D) = \frac{p(D|A_i)p(A_i)}{\sum_{i=1}^{N} p(D|A_i)p(A_i)}$$

Bayesian Estimation: Examples of Coins

Example

We flipped a coin six times and got the observation $\{1, 0, 1, 1, 0, 1\}$*, where 1 represents heads, while 0 means tails. What is the probability that the coin will come up heads?*

- Find a probabilistic description of the observation. — Likelihood function

Bayesian Estimation: Examples of Coins

A Discrete Prior

- Find a probabilistic description of the observation. — Likelihood function

Denote The underlying probability of heads as θ. This idea can be written formally as

$$p(y = 1|\theta) = \theta.$$

- Find a probabilistic description of the observation. — Likelihood function

Denote The underlying probability of heads as θ. This idea can be written formally as

$$p(y = 1|\theta) = \theta.$$

Therefore,

$$p(y = 0|\theta) = 1 - \theta.$$

- Find a probabilistic description of the observation. — Likelihood function

Denote The underlying probability of heads as θ. This idea can be written formally as

$$p(y = 1|\theta) = \theta.$$

Therefore,

$$p(y = 0|\theta) = 1 - \theta.$$

The probability of the observation $Y = \{1, 0, 1, 1, 0, 1\}$ (i.e., the likelihood function) is

$$p(Y|\theta) = \theta^4(1 - \theta)^2.$$

Bayesian Estimation: Examples of Coins

A Discrete Prior

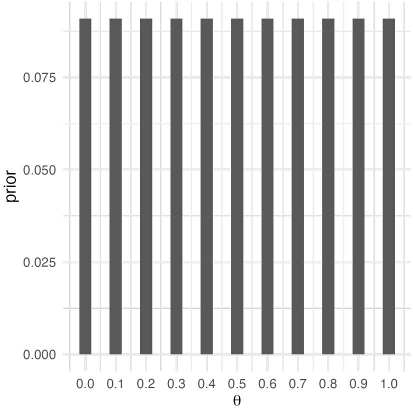
- Find a probability distribution of the parameter(s). — Prior distribution

Bayesian Estimation: Examples of Coins

A Discrete Prior

- Find a probability distribution of the parameter(s). — Prior distribution

We start with a prior that θ is equally likely to be 0, 0.1, 0.2, ..., 0.9, 1.



Bayesian Estimation: Examples of Coins

A Discrete Prior

WARWICK

- Following Bayes' theorem, we can reallocate credibility across the values of the parameter(s). — Posterior Distribution

Bayesian Estimation: Examples of Coins

A Discrete Prior

WARWICK

- Following Bayes' theorem, we can reallocate credibility across the values of the parameter(s). — Posterior Distribution

Theorem (Bayes' Theorem)

If $A = \{A_1, A_2, \cdots, A_N\}$ is an event set, then the probability of event $A_i \in A$ occurring given that D is observed is

$$p(A_i|D) = \frac{p(D|A_i)p(A_i)}{\sum_{i=1}^{N} p(D|A_i)p(A_i)}.$$

- Following Bayes' theorem, we can reallocate credibility across the values of the parameter(s). — Posterior Distribution

Theorem (Bayes' Theorem)

If $A = \{A_1, A_2, \cdots, A_N\}$ is an event set, then the probability of event $A_i \in A$ occurring given that D is observed is

$$p(A_i|D) = \frac{p(D|A_i)p(A_i)}{\sum_{i=1}^{N} p(D|A_i)p(A_i)}.$$

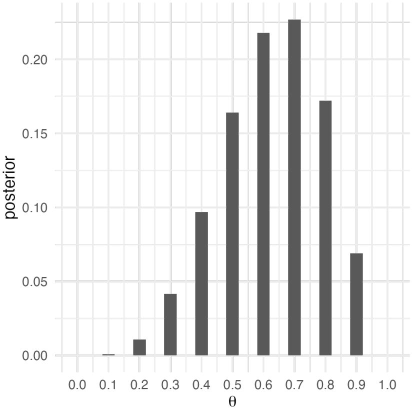
$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\sum_i^N p(Y|\theta_i)p(\theta_i)}$$

Bayesian Estimation: Examples of Coins

A Discrete Prior

WARWICK

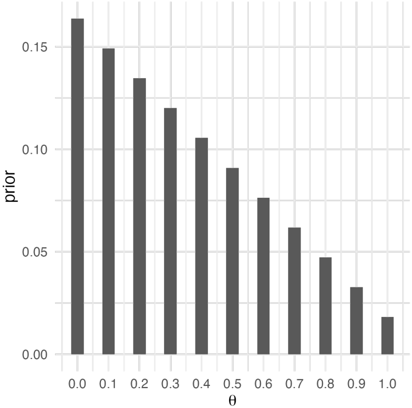
- Following Bayes' theorem, we can reallocate credibility across the values of the parameter(s). — Posterior Distribution



Bayesian estimation follows three steps:

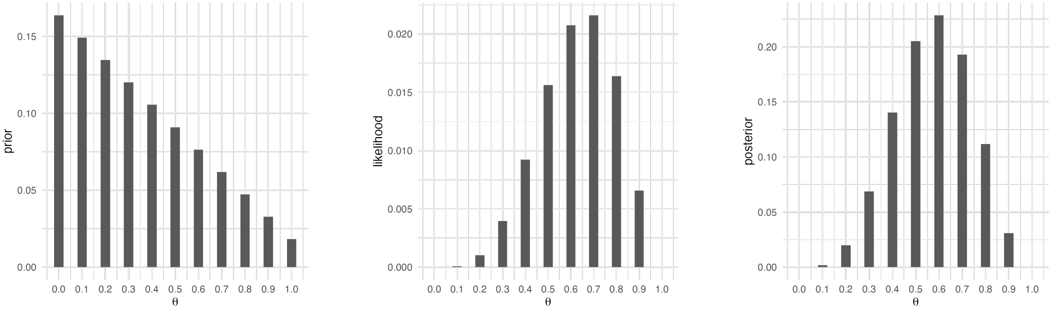
1. Find a probabilistic description of the observation. — Likelihood function;
2. Find a probability distribution of the parameter(s). — Prior distribution;
3. Following Bayes' theorem, we can reallocate credibility across the values of the parameter(s). — Posterior Distribution.

Let's try a new discrete prior!



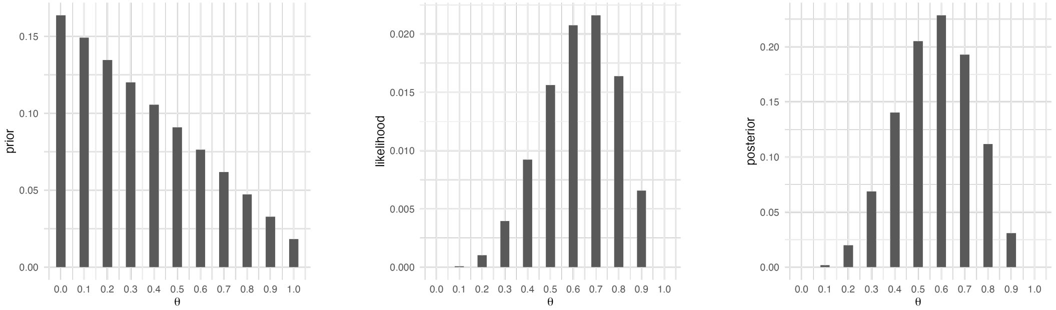
Bayesian Estimation: Examples of Coins

A Discrete Prior



Bayesian Estimation: Examples of Coins

A Discrete Prior



- The shape of the posterior is the combination of the shape of the prior and the likelihood.

Let's think about a more reasonable prior: a Beta distribution.

Let's think about a more reasonable prior: a Beta distribution.

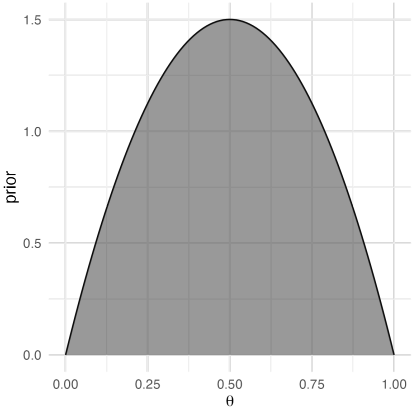


Fig. 1: The density function of $Beta(2,2)$.

Bayesian Estimation: Examples of Coins

A Continuous Prior

WARWICK

- Prior: $\theta \sim Beta(2, 2)$;
- Likelihood:

Bayesian Estimation: Examples of Coins

A Continuous Prior

WARWICK

- Prior: $\theta \sim Beta(2,2)$;
- Likelihood: $p(Y|\theta) = \theta^4(1-\theta)^2$;

Bayesian Estimation: Examples of Coins

A Continuous Prior

- Prior: $\theta \sim Beta(2, 2)$;
- Likelihood: $p(Y|\theta) = \theta^4(1-\theta)^2$;
- Bayes' Theorem:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\sum_i^N p(Y|\theta_i)p(\theta_i)}$$

Bayesian Estimation: Examples of Coins

A Continuous Prior

- Prior: $\theta \sim Beta(2,2)$;
- Likelihood: $p(Y|\theta) = \theta^4(1-\theta)^2$;
- Bayes' Theorem:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

- Prior: $\theta \sim Beta(2, 2)$;
- Likelihood: $p(Y|\theta) = \theta^4(1-\theta)^2$;
- Bayes' Theorem:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

- To "calculate" the posterior distribution, we used the Markov Chain Monte Carlo (MCMC) method.

Bayesian Estimation: Examples of Coins

A Continuous Prior

Target:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

Target:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

A rough idea of the process of MCMC is:

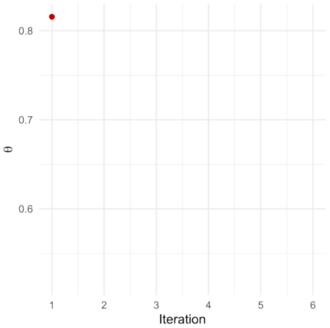
Bayesian Estimation: Examples of Coins

A Continuous Prior

Target:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

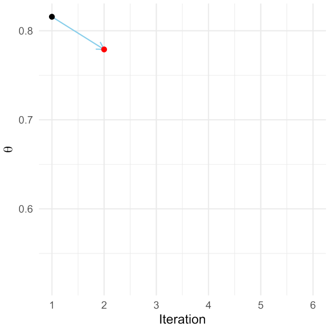
A rough idea of the process of MCMC is:



Target:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

A rough idea of the process of MCMC is:



Bayesian Estimation: Examples of Coins

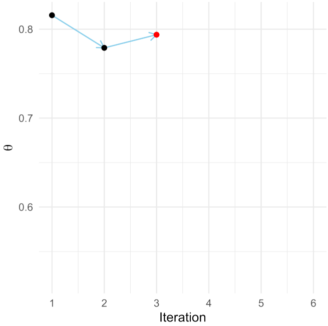
A Continuous Prior

WARWICK

Target:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

A rough idea of the process of MCMC is:



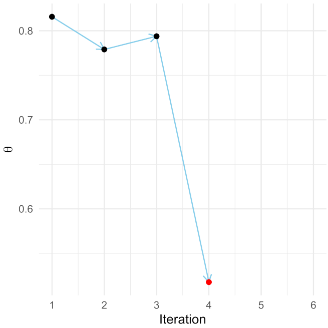
Bayesian Estimation: Examples of Coins

A Continuous Prior

Target:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

A rough idea of the process of MCMC is:



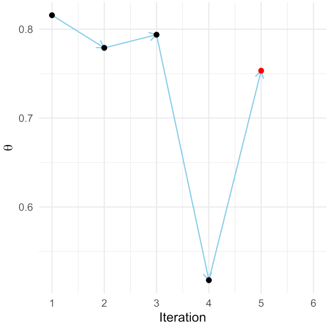
Bayesian Estimation: Examples of Coins

A Continuous Prior

Target:

$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

A rough idea of the process of MCMC is:



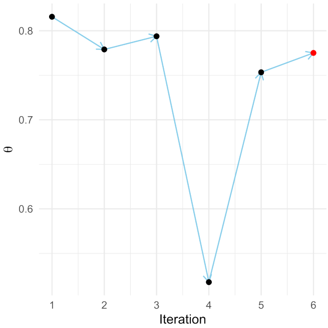
Bayesian Estimation: Examples of Coins

A Continuous Prior

Target:

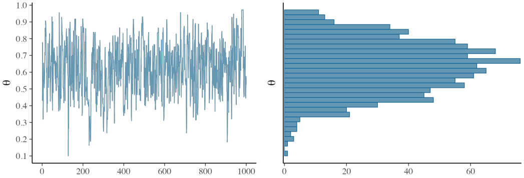
$$p(\theta_i|Y) = \frac{p(Y|\theta_i)p(\theta_i)}{\int p(Y|\theta)p(\theta)d\theta}$$

A rough idea of the process of MCMC is:



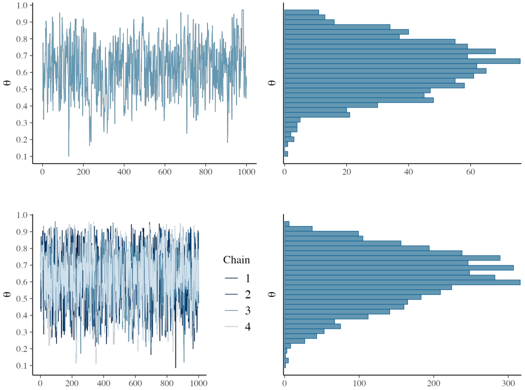
Bayesian Estimation: Examples of Coins

A Continuous Prior



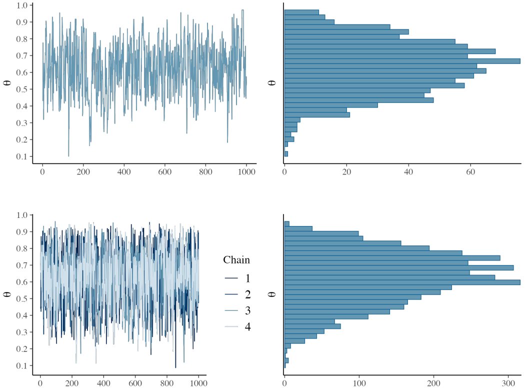
Bayesian Estimation: Examples of Coins

A Continuous Prior



Bayesian Estimation: Examples of Coins

A Continuous Prior



- The MCMC is implemented in a `.stan` file.

WARWICK

What is a `.stan` file?

What is a `.stan` file?

```
data {
 // The observations
}

parameters {
 // The parameters
}

model {
 // The prior and likelihood function
}
```

Stan uses the syntax of `C++`. Thus, you should always keep three things in mind:

WARWICK

Stan uses the syntax of `C++`. Thus, you should always keep three things in mind:

- `C++` is sensitive to integer and float values: 1 is different from 1.0!

Stan uses the syntax of `C++`. Thus, you should always keep three things in mind:

- `C++` is sensitive to integer and float values: 1 is different from 1.0!
- Each line should end with a semicolon ";"

Stan uses the syntax of `C++`. Thus, you should always keep three things in mind:

- `C++` is sensitive to integer and float values: 1 is different from 1.0!
- Each line should end with a semicolon ";"
- Comments are indicated by a double slash "//" instead of by a number sign "#."

We will write two `.stan` files:

- A “Hello world” example (See *hello_world.stan*);
- The Bayesian estimation of the coin example (See *coin.stan*).

- The shape of the posterior distribution is the combination of the shape of the prior and the likelihood;

- The shape of the posterior distribution is the combination of the shape of the prior and the likelihood;
- The continuous version of Bayes' Theorem:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}.$$

- The shape of the posterior distribution is the combination of the shape of the prior and the likelihood;
- The continuous version of Bayes' Theorem:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}.$$

- The posterior distribution $p(\theta|D)$ can be approximated by MCMC, which is implemented in Stan;

Bayesian one-sample t-test

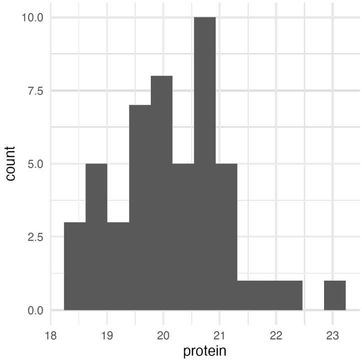
Bayesian one-sample t-test

An Energy Bar Example

WARWICK

Example

There is a type of energy bar that claims each bar contains 20 grams of protein. We collected 50 samples and tested the protein content of this product. We want to know whether the label is accurate.



- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

- Calculate the t value:

$$T = \frac{|\overline{X} - 20|}{s/\sqrt{n}} = 1.06, \quad df = 49$$

WARWICK

- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

- Calculate the t value:

$$T = \frac{|\overline{X} - 20|}{s/\sqrt{n}} = 1.06, \quad df = 49$$

- Find the p-value:

$$p = 0.29 > 0.05$$

Bayesian one-sample t-test

The Frequentist t-test

- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

- Calculate the *t* value:

$$T = \frac{|\overline{X} - 20|}{s/\sqrt{n}} = 1.06, \quad df = 49$$

- Find the *p*-value:

$$p = 0.29 > 0.05$$

- we **cannot reject** the null hypothesis. (Does this mean the label of protein is accurate?)

- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

- Calculate the *t* value:

$$T = \frac{|\overline{X} - 20|}{s/\sqrt{n}} = 1.06, \quad df = 49$$

- Find the *p*-value:

$$p = 0.29 > 0.05$$

- we **cannot reject** the null hypothesis. (Does this mean the label of protein is accurate?)

How to *find* the *p*-value?

- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

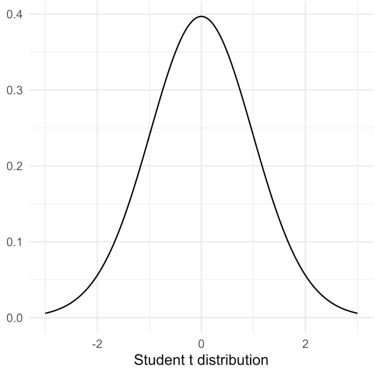
- Calculate the t value:

$$T = \frac{|\overline{X} - 20|}{s/\sqrt{n}} = 1.06, \quad df = 49$$

- Find the p-value:

$$p = 0.29 > 0.05$$

- we **cannot reject** the null hypothesis. (Does this mean the label of protein is accurate?)



WARWICK

- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

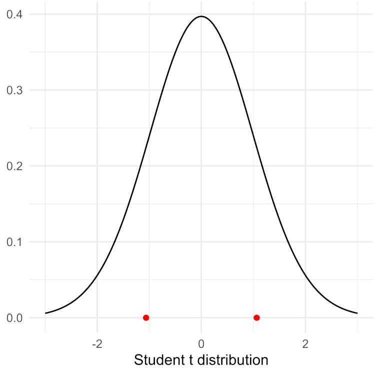
- Calculate the t value:

$$T = \frac{|\overline{X} - 20|}{s/\sqrt{n}} = 1.06, \quad df = 49$$

- Find the p-value:

$$p = 0.29 > 0.05$$

- we **cannot reject** the null hypothesis. (Does this mean the label of protein is accurate?)



WARWICK

- The null and alternative hypotheses:

$$H_0 : \mu = 20$$
$$H_1 : \mu \neq 20$$

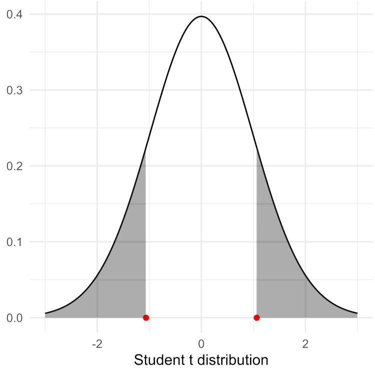
- Calculate the t value:

$$T = \frac{|\overline{X} - 20|}{s/\sqrt{n}} = 1.06, \quad df = 49$$

- Find the p-value:

$$p = 0.29 > 0.05$$

- we **cannot reject** the null hypothesis. (Does this mean the label of protein is accurate?)



But what does the Student's t-distribution mean?

But what does the Student's t-distribution mean?

- Now, imagine we repeat this process over and over—

WARWICK

But what does the Student's t-distribution mean?

- Now, imagine we repeat this process over and over—buying 50 bars each time, measuring their protein, and calculating a new average and T value every time.

WARWICK

But what does the Student's t-distribution mean?

- Now, imagine we repeat this process over and over—buying 50 bars each time, measuring their protein, and calculating a new average and T value every time. If the true mean really is 20 grams, the values of T we get from all these repeated samples will follow what's called the Student's t-distribution.

But what does the Student's t-distribution mean?

- Now, imagine we repeat this process over and over—buying 50 bars each time, measuring their protein, and calculating a new average and T value every time. If the true mean really is 20 grams, the values of T we get from all these repeated samples will follow what's called the Student's t-distribution.
- The t-distribution shows us what kind of T values we'd expect to see from random sampling, if the energy bar label is completely accurate.

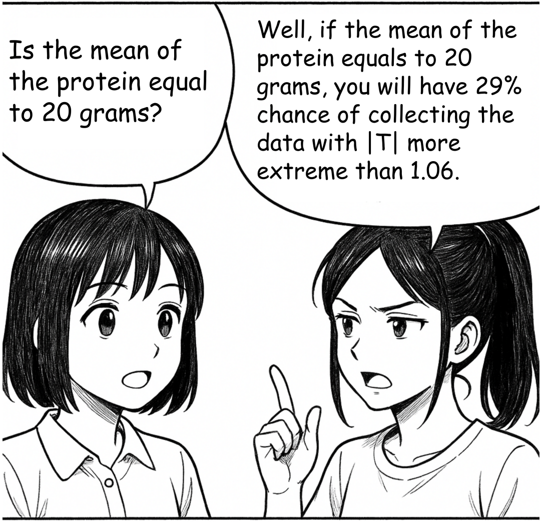
The p-value means the probability of obtaining the results as extreme as (or more extreme than) those observed, **assuming the null hypothesis is true**.

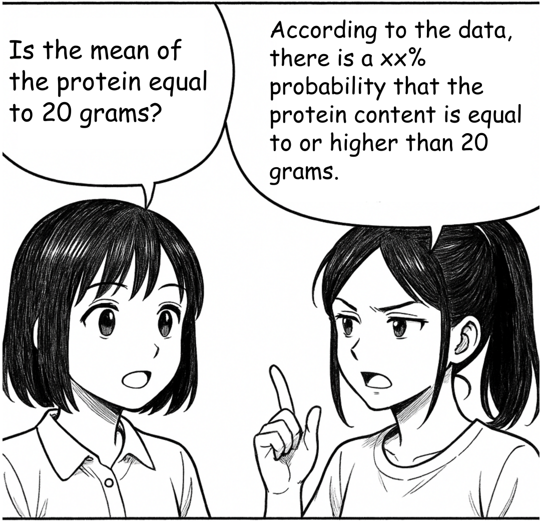
The *p*-value means the probability of obtaining the results as extreme as (or more extreme than) those observed, **assuming the null hypothesis is true**.

$$p = P(t < -|T| \;|\mu = 20) + P(t > |T| \;|\mu = 20)$$

Bayesian one-sample t-test

The Frequentist t-test





Is the mean of the protein equal to 20 grams?
According to the data, there is a xx% probability that the protein content is equal to or higher than 20 grams.

- Likelihood function:

- Prior:

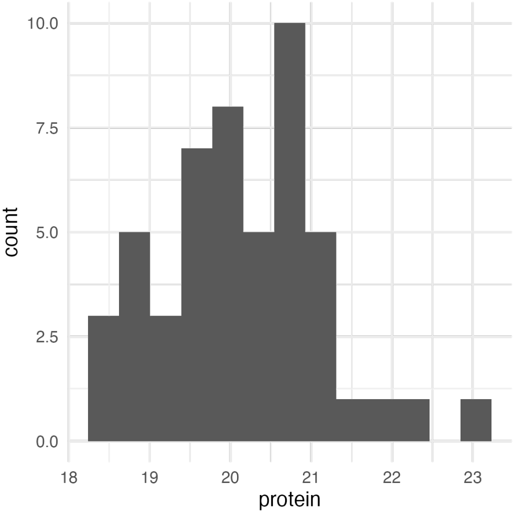
Bayesian one-sample t-test

The Bayesian Estimation

WARWICK

- Likelihood function:

- Prior:



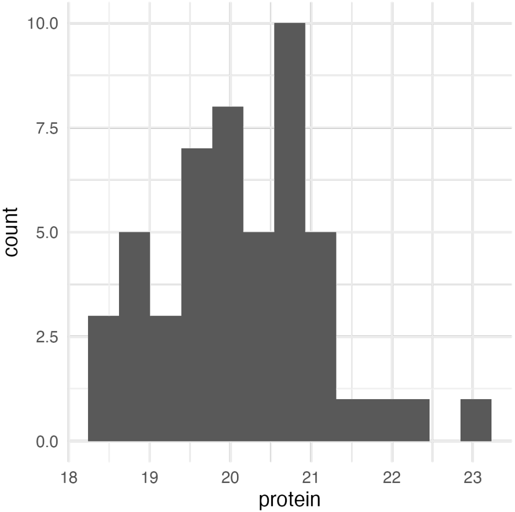
Bayesian one-sample t-test

The Bayesian Estimation

- Likelihood function:

$$L(D|\mu, \sigma) \sim N(\mu, \sigma^2)$$

- Prior:



Bayesian one-sample t-test

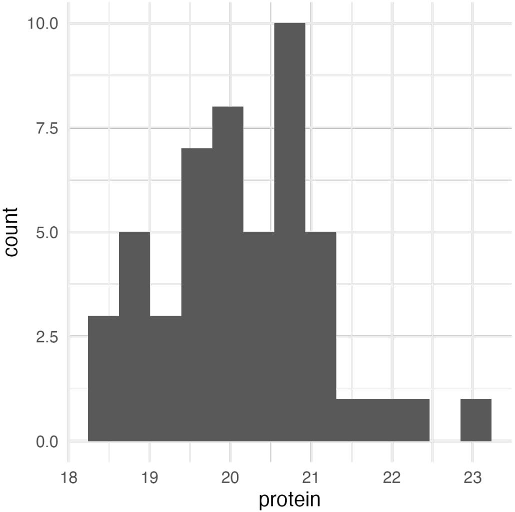
The Bayesian Estimation

- Likelihood function:

$$L(D|\mu, \sigma) \sim N(\mu, \sigma^2)$$

- Prior:

$$\mu \sim N(20, 10)$$
$$\sigma \sim N(0, 10)T[0, \infty]$$

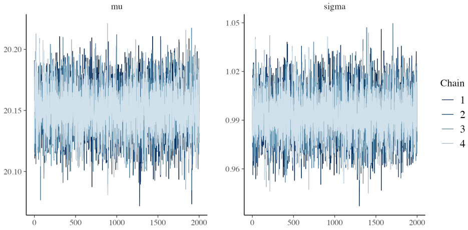


Bayesian one-sample t-test

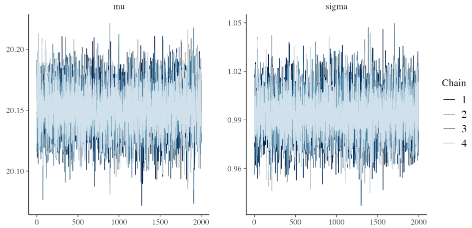
The Bayesian Estimation

WARWICK

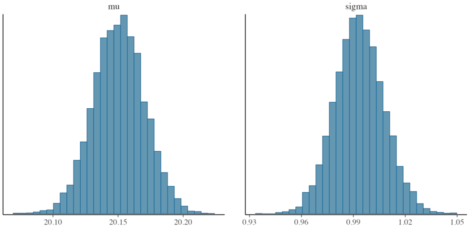
- The trace plot:



- The trace plot:

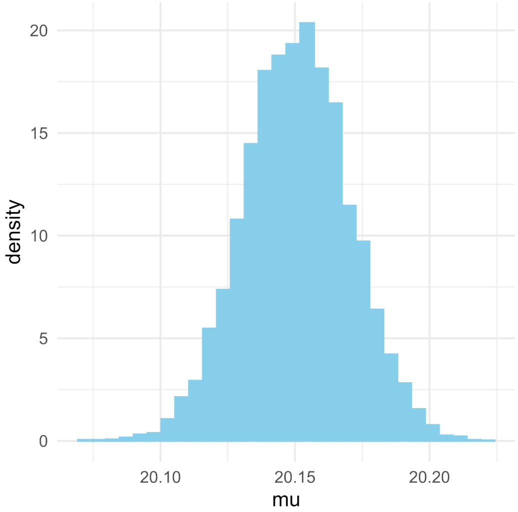


- The posterior distribution:



Bayesian one-sample t-test

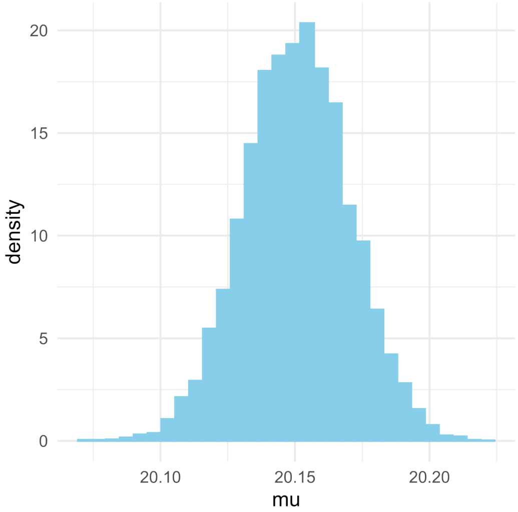
The Bayesian Estimation



How do we summarise the posterior distribution?

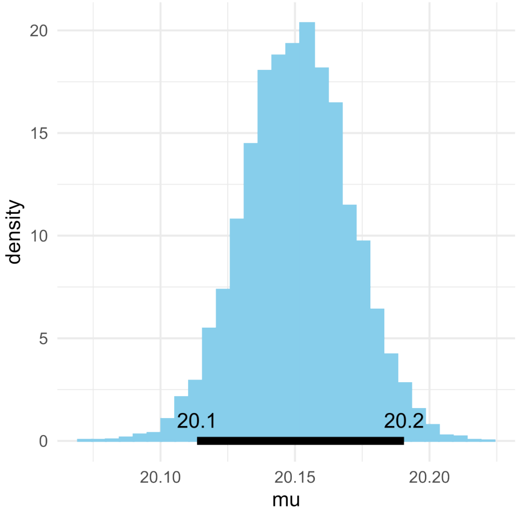
Bayesian one-sample t-test

The Bayesian Estimation



How do we summarise the posterior distribution?

- HDI (highest density interval): A 95% HDI is the shortest interval that covers the 95% mass of the distribution.



How do we summarise the posterior distribution?

- HDI (highest density interval): A 95% HDI is the shortest interval that covers the 95% mass of the distribution.

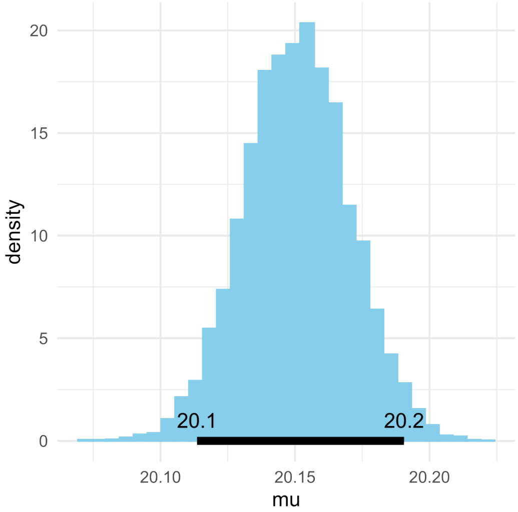
What's the difference between the confidence interval (CI) and the highest density interval (HDI)?

What's the difference between the confidence interval (CI) and the highest density interval (HDI)?

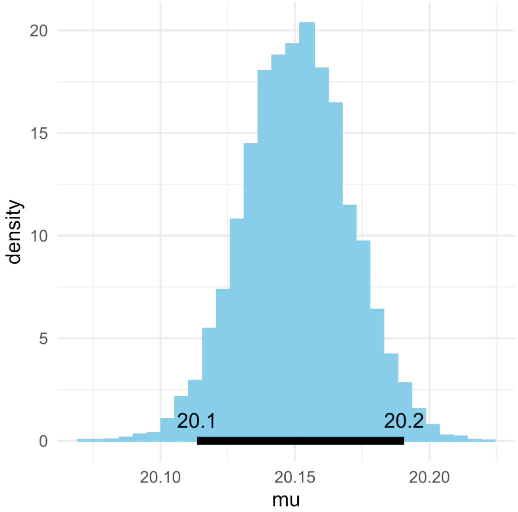
- The 95% CI means that if we did this experiment over and over, 95% of those intervals would include the true mean.

What's the difference between the confidence interval (CI) and the highest density interval (HDI)?

- The 95% CI means that if we did this experiment over and over, 95% of those intervals would include the true mean.
- The 95% HDI means that there is a 95% probability that the true mean lies in this interval, given our data and assumptions.

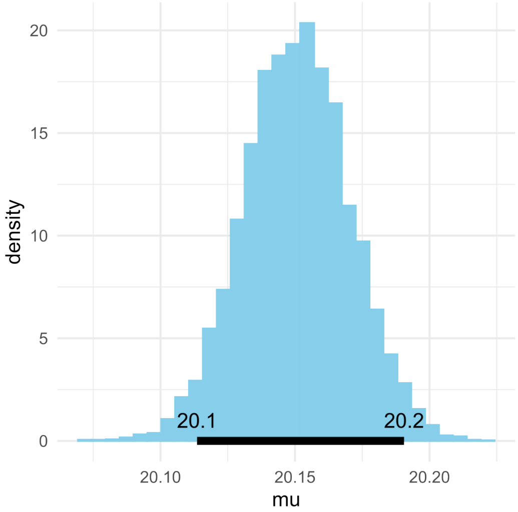


How do we conclude?



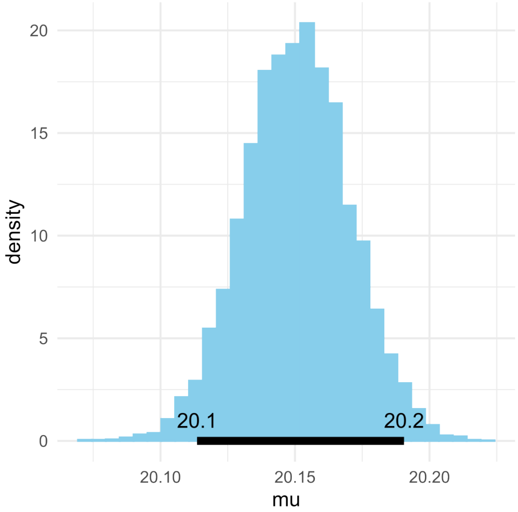
How do we conclude?

The value of 20 grams is excluded from the 95% HDI.



How do we conclude?

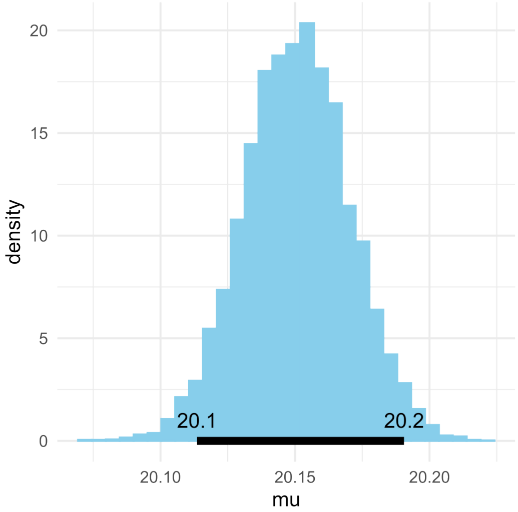
The value of 20 grams is excluded from the 95% HDI. Can we reject the null hypothesis now?



How do we conclude?

The value of 20 grams is excluded from the 95% HDI. Can we reject the null hypothesis now?

Not yet.



How do we conclude?

The value of 20 grams is excluded from the 95% HDI. Can we reject the null hypothesis now?

Not yet. We also want to know whether the difference between the posterior distribution and the null hypothesis is big *enough*.


One more thing...
Adam Tow | AllThingsD.com

ROPE (region of practical equivalence):

ROPE (region of practical equivalence):

- It's a range of parameter values that are considered to be practically equivalent to the null value.

ROPE (region of practical equivalence):

- It's a range of parameter values that are considered to be practically equivalent to the null value.
- In hypothesis testing, we usually employ the effect size as the measure of equivalence (Kruschke, 2012).

$$d = \frac{|\mu - \mu_0|}{\sigma}.$$

ROPE (region of practical equivalence):

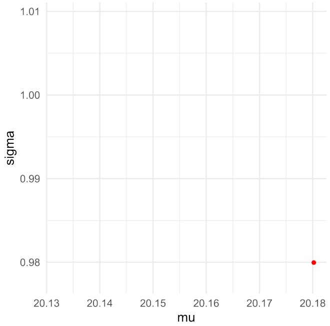
- It's a range of parameter values that are considered to be practically equivalent to the null value.
- In hypothesis testing, we usually employ the effect size as the measure of equivalence (Kruschke, 2012).

$$d = \frac{|\mu - \mu_0|}{\sigma}.$$

- Cohen (1988) noted that the effect is "small" when $d = 0.2$. In other words, when $d \leq 0.2$, the estimate of μ is practically equivalent to the null value μ_0.

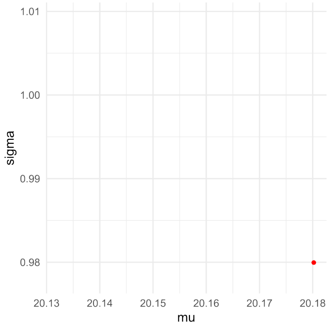
Bayesian one-sample t-test

The Bayesian Inference



Bayesian one-sample t-test

The Bayesian Inference



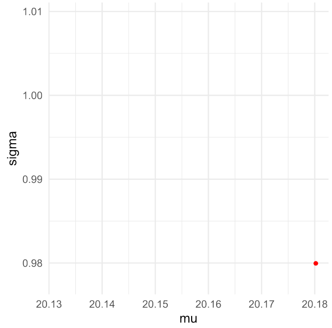
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$

Bayesian one-sample t-test

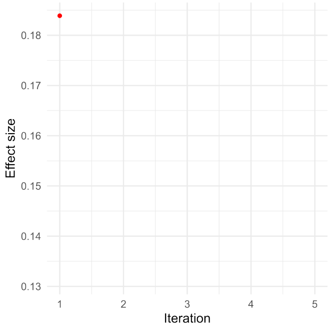
The Bayesian Inference

WARWICK



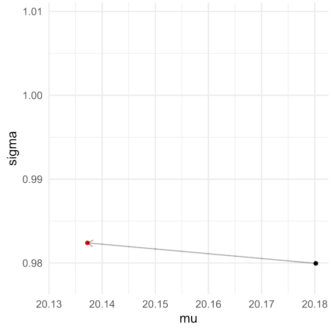
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$



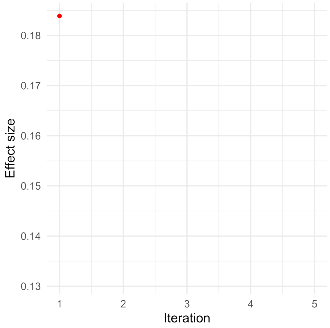
Bayesian one-sample t-test

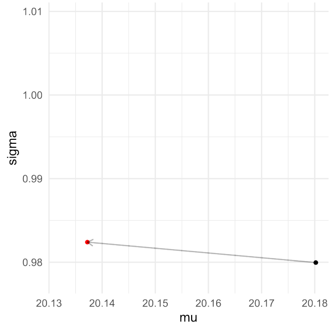
The Bayesian Inference



$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

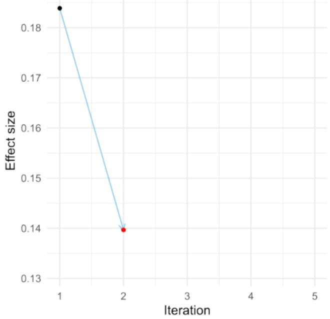
$$\Rightarrow$$

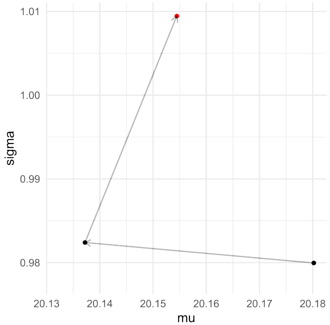




$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

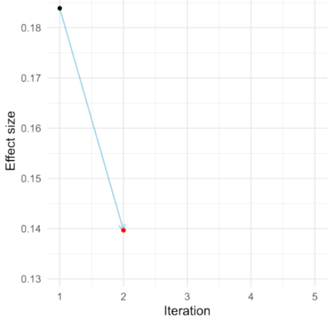
$$\Rightarrow$$





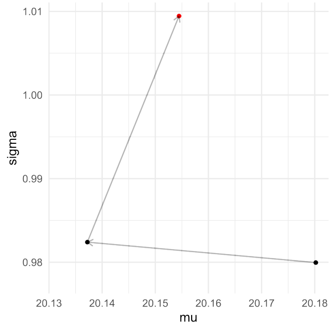
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$



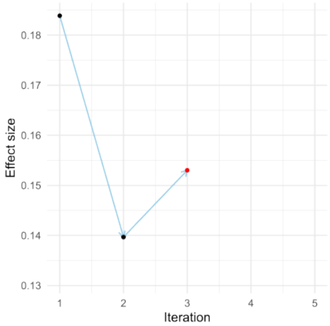
Bayesian one-sample t-test

The Bayesian Inference



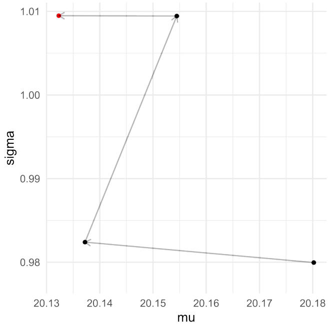
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$



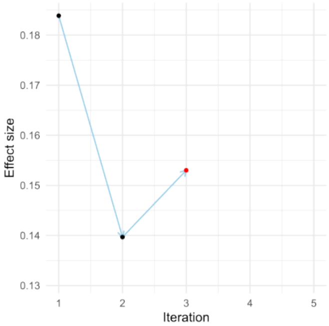
Bayesian one-sample t-test

The Bayesian Inference



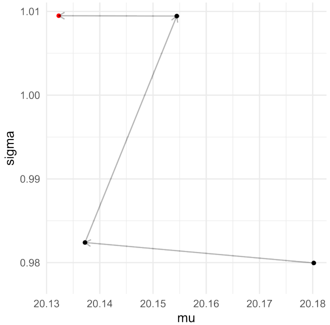
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$



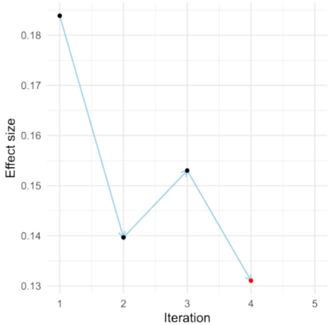
Bayesian one-sample t-test

The Bayesian Inference



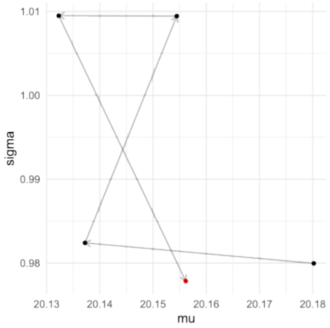
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$



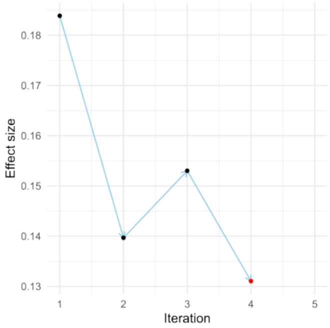
Bayesian one-sample t-test

The Bayesian Inference



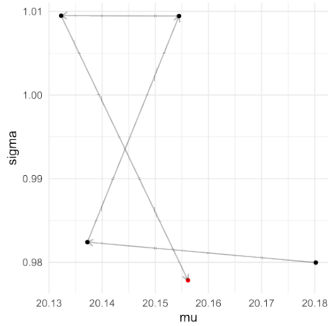
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$



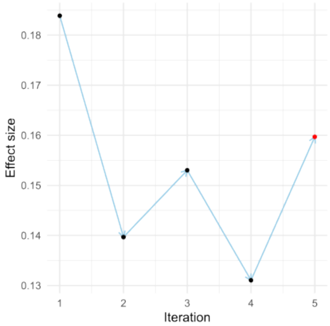
Bayesian one-sample t-test

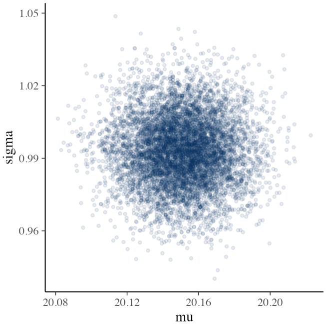
The Bayesian Inference



$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

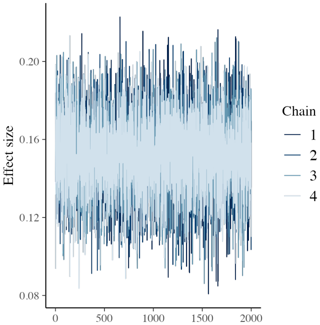
$$\Rightarrow$$





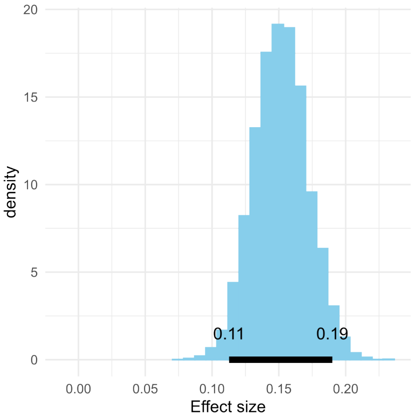
$$\text{Effect size} = \frac{|\mu - 20|}{\sigma}$$

$$\Rightarrow$$



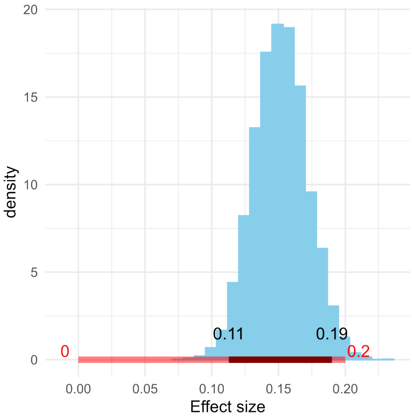
Bayesian one-sample t-test

The Bayesian Inference



Bayesian one-sample t-test

The Bayesian Inference



The hypothesis test via Bayesian estimation:

- 95% HDI lies entirely inside the ROPE: the parameter value is declared to be accepted;

The hypothesis test via Bayesian estimation:

- 95% HDI lies entirely inside the ROPE: the parameter value is declared to be accepted;
- 95% HDI lies outside the ROPE: the parameter value is declared to be not credible, or rejected.

The hypothesis test via Bayesian estimation:

- 95% HDI lies entirely inside the ROPE: the parameter value is declared to be accepted;
- 95% HDI lies outside the ROPE: the parameter value is declared to be not credible, or rejected.
- 95% HDI and ROPE overlap: current data are insufficient to yield a clear conclusion.

- The idea of the p-values does not answer the question we are interested in;

- The idea of the p-values does not answer the question we are interested in;
- Bayesian estimation can provide a more direct answer to our question;

- The idea of the *p*-values does not answer the question we are interested in;
- Bayesian estimation can provide a more direct answer to our question;
- (Advanced topic) Practically, we need 95% HDI and ROPE to test the hypotheses.

Thanks!

References I

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Kruschke, J. K. (2012).Bayesian estimation supersedes the t test.. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146