浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

Bayesian Statistics and Bayesian Inference

ADS 2, Lecture 2.8

Zhaoyuan Fang - zhaoyuan@intl.zju.edu.cn

April, 2024

All hypothesis tests are the same . . .

- Formulate the Null Hypothesis, Alternative Hypothesis.
- Design your experiment and collect data
- Summarise and describe your data.
- Think about what you would expect if H0 was true.
- Could your data be explained by the Null Hypothesis? Determine the probability of your data given H0.
- Interpret your p value and make a decision.
- Be aware that hypothesis tests are not perfect.

All hypothesis tests are the same . . .

- Formulate the Null Hypothesis, Alternative Hypothesis.
- Design your experiment and collect data
- Summarise and describe your data.
- Think about what you would expect if H0 was true.
- Could your data be explained by the Null Hypothesis? Determine the probability of your data given H0.
- Interpret your p value and make a decision.
- Be aware that hypothesis tests are not perfect.

Or are they?

Has this ever felt weird to you?

- Formulate the Null Hypothesis, Alternative Hypothesis.
- Design your experiment and collect data
- Summarise and describe your data.
- Think about what you would expect if H0 was true.
- Could your data be explained by the Null Hypothesis? Determine the probability of your data given H0.
- Interpret your p value and make a decision.
- Be aware that hypothesis tests are not perfect.

Problems with p values

(And this type of hypothesis testing in general)

- The $\alpha = 0.05$ threshold is arbitrary. Who says 0.05 is better than 0.04 or 0.06?

Problems with p values

(And this type of hypothesis testing in general)

- The $\alpha = 0.05$ threshold is arbitrary. Who says 0.05 is better than 0.04 or 0.06?
- Your chosen α value imposes a strict binary threshold. A result with $p = 0.049$ is statistically significant, a result with $p = 0.051$ is not. This feels unsatisfactory to many people.

Problems with p values

(And this type of hypothesis testing in general)

- The $\alpha = 0.05$ threshold is arbitrary. Who says 0.05 is better than 0.04 or 0.06?
- Your chosen α value imposes a strict binary threshold. A result with $p = 0.049$ is statistically significant, a result with $p = 0.051$ is not. This feels unsatisfactory to many people.
- People often think that a p value of 0.000001 is somehow more convincing than a p value of 0.01. But there is no way of quantifying this "more convincing-ness". Within this framework, a result is statistically significant, or it is not.

Problems with p values

(And this type of hypothesis testing in general)

- The $\alpha = 0.05$ threshold is arbitrary. Who says 0.05 is better than 0.04 or 0.06?
- Your chosen α value imposes a strict binary threshold. A result with $p = 0.049$ is statistically significant, a result with $p = 0.051$ is not. This feels unsatisfactory to many people.
- People often think that a p value of 0.000001 is somehow more convincing than a p value of 0.01. But there is no way of quantifying this "more convincing-ness". Within this framework, a result is statistically significant, or it is not.
- Two statisticians analysing the same data set may choose two different tests (e.g. a t-test or a Wilcoxon rank sum test). Both tests are equally valid, but one may give a statistically significant result and the other doesn't.

Problems with p values

(And this type of hypothesis testing in general)

- The $\alpha = 0.05$ threshold is arbitrary. Who says 0.05 is better than 0.04 or 0.06?
- Your chosen α value imposes a strict binary threshold. A result with $p = 0.049$ is statistically significant, a result with $p = 0.051$ is not. This feels unsatisfactory to many people.
- People often think that a p value of 0.000001 is somehow more convincing than a p value of 0.01. But there is no way of quantifying this "more convincing-ness". Within this framework, a result is statistically significant, or it is not.
- Two statisticians analysing the same data set may choose two different tests (e.g. a t-test or a Wilcoxon rank sum test). Both tests are equally valid, but one may give a statistically significant result and the other doesn't.

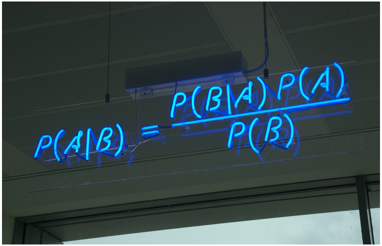
Is there an alternative?

Yes, there is!

Bayesian inference

Key ideas:

- Instead of making a yes/no decision, quantify the strength of our belief.
- Beliefs exist even before we have data, and we take those into account.
- Beliefs are updated as we collect more data.



Review of Bayes' theorem

Given two events A and B, we have

$$P(A) + P(\neg A) \equiv 1$$

$$P(B) + P(\neg B) \equiv 1$$

Conditional probability

$$P(A|B) = \frac{P(A\&B)}{P(B)}$$

$$P(B|A) = \frac{P(A\&B)}{P(A)}$$

Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B\&A) + P(B\&\neg A)}$$

Review of Bayes' theorem

Thus

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

Similarly

$$P(\neg A|B) = \frac{P(B|\neg A)P(\neg A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

Bayesian hypothesis testing

We have two hypothesis H_0 and H_1, and some data D, then

$$P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0) + P(D|H_1)P(H_1)} = \frac{P(D|H_0)P(H_0)}{P(D)}$$

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{P(D|H_0)P(H_0) + P(D|H_1)P(H_1)} = \frac{P(D|H_1)P(H_1)}{P(D)}$$

Since the two sums to 1, we usually only need to calculate $P(H_0|D)$. Other times we could compare the two using Posterior ratio $\frac{P(H_1|D)}{P(H_0|D)}$, which is the product of Likelihood ratio (Bayes factor) and Prior ratio.

Bayesian hypothesis testing

Let's take more look at this,

$$P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0) + P(D|H_1)P(H_1)} = \frac{P(D|H_0)P(H_0)}{P(D)}$$

Here we have names for each term,

$P(H_0|D)$: Posterior (what we want)
$P(D)$: Normalization factor (just for scaling)
$P(D|H_0)$: Likelihood (connection strength between data and hypothesis)
$P(H_0)$: Prior (what we know before data - "Belief")

Bayesian hypothesis testing

Let's take more look at this,

$$P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0) + P(D|H_1)P(H_1)} = \frac{P(D|H_0)P(H_0)}{P(D)}$$

Here we have names for each term,

$P(H_0|D)$: Posterior (what we want)
$P(D)$: Normalization factor (just for scaling)
$P(D|H_0)$: Likelihood (connection strength between data and hypothesis)
$P(H_0)$: Prior (what we know before data - "Belief")
Is that really possible?

Learning Objectives

After this lecture, you should be able to do the following:

- Explain what “belief” means to a Bayesian statistician
- Define the terms “prior” and “posterior”
- Specify prior probabilities and distributions
- Use Bayes’ theorem to determine posterior probabilities
- Describe the difference between Bayesian and Frequentist statistics

Outline

1. The role of belief in Bayesian statistics
2. Priors – what are they? How do we define them?
3. How to calculate posterior probabilities
4. Bayesian vs Frequentist statistic

We all hold beliefs

- Our beliefs are subjective and influenced by opinion, e.g. do you believe in ghosts?

We all hold beliefs

- Our beliefs are subjective and influenced by opinion, e.g. do you believe in ghosts?
- Beliefs change based on experience, e.g. if you have never seen a ghost, are you more/less likely to believe?

We all hold beliefs

- Our beliefs are subjective and influenced by opinion, e.g. do you believe in ghosts?
- Beliefs change based on experience, e.g. if you have never seen a ghost, are you more/less likely to believe?
- Belief determines scientific hypotheses, interesting experiments, direction of research...

It's not just ghosts!

It's not just ghosts!

Example

Dr. Li is studying the effects of childhood exposure to toxic chemicals on growth. For this she measures the height of two groups of adults: Those who have grown up near a chemical plant and those who haven't.
What beliefs may Dr. Li have that are relevant to this study?

It's not just ghosts!

Example

Dr. Li is studying the effects of childhood exposure to toxic chemicals on growth. For this she measures the height of two groups of adults: Those who have grown up near a chemical plant and those who haven't.
What beliefs may Dr. Li have that are relevant to this study?

- Adult human height is bounded.
- Adult human height is (essentially) constant.
- Adult human height can be measured.
- Events during childhood can influence adult height.
- Living near a chemical plant increases a person's exposure to toxic chemicals.
- People do not usually lie about where they grew up.
- ...

What do I mean by belief?

- There are beliefs that with different levels of evidences/reliability. But all of them could be used as Prior in Bayesian statistics.
  1. Something you believe without any data, e.g. ghosts
  2. Something you believe with very weak data, e.g. the most distant edge of our universe
  3. Something you believe with limited or partial data, e.g. the average height of UK people

- Prior is "before data" - this only means that it is before the new specific data your are using to infer your hypothesis, but there could already be some old data that can help you to determine Prior.

Outline

1. The role of belief in Bayesian statistics
2. Priors – what are they? How do we define them?
3. How to calculate posterior probabilities
4. Bayesian vs Frequentist statistic

What is a prior?

The probability (or probability distribution) which describes our belief about the world before we collect any specific data.

What is a prior?

The probability (or probability distribution) which describes our belief about the world before we collect any specific data.

Examples:

- Adult human height is bounded between 50 cm and 300 cm.
- Adult human height is normally distributed with a mean around 170 cm
- An enzymatic reaction is unlikely to proceed with a $\mathrm{k_{cat}}$ faster than $10^6\,s^{-1}$ (the $\mathrm{k_{cat}}$) of carbonic anhydrase
- A coin has a 50 % chance of landing on "heads".
- For a specific somatic gene, if two heterozygous mice (each carrying one mutated and one wildtype allele) mate, 25 % of their offspring will be homozygous for the mutated allele.
- ...

Priors: Example

Example:

- Hypothesis: Rob's date wants to see Avengers: Endgame.

Priors: Example

Example:

- Hypothesis: Rob's date wants to see Avengers: Endgame.
- What is the prior probability for this hypothesis?

Priors: Example

Example:

- Hypothesis: Rob's date wants to see Avengers: Endgame.
- What is the prior probability for this hypothesis?
- Rob knows (from talking to his friends) that about 60 % of them want to watch Endgame

Priors: Example

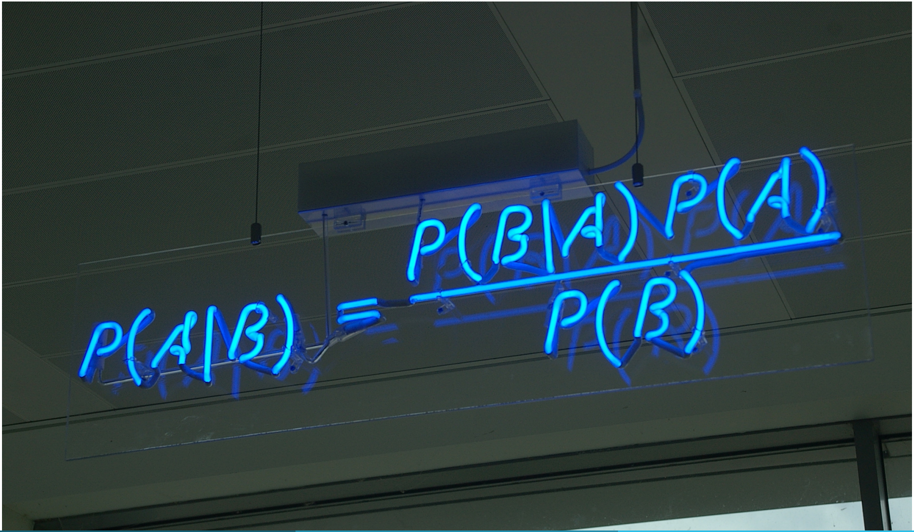
Example:

- Hypothesis: Rob's date wants to see Avengers: Endgame.
- What is the prior probability for this hypothesis?
- Rob knows (from talking to his friends) that about 60 % of them want to watch Endgame
- Since we haven't collected any data on Rob's date yet, we are going to use this as our prior for the probability of the hypothesis that she wants to watch Endgame:

$$P(H) = 0.6$$

Key idea of Bayesian statistics

We state our beliefs before we do the experiment (prior). We then collect data and use that to update our beliefs (posterior).



Outline

1. The role of belief in Bayesian statistics
2. Priors – what are they? How do we define them?
3. How to calculate posterior probabilities
4. Bayesian vs Frequentist statistic

Example:

- Hypothesis: Rob's date wants to see Avengers: Endgame.

Posteriors: Example

Example:

- Hypothesis: Rob's date wants to see Avengers: Endgame.
- Our prior (before collecting data) was

$$P(H) = 0.6$$

Posteriors: Example

Example:

- Hypothesis: Rob's date wants to see Avengers: Endgame.
- Our prior (before collecting data) was

$$P(H) = 0.6$$

- Now, let's collect some data. Rob does this by asking her if she has seen Avengers: Infinity War. She says yes.
- *How does this change our belief in the hypothesis?*

Posteriors: Example

- Hypothesis: Rob's date wants to see Avengers: Endgame.
  Data: She has seen Avengers: Infinity war.

Posteriors: Example

- Hypothesis: Rob's date wants to see Avengers: Endgame.
  Data: She has seen Avengers: Infinity war.
- We need our posterior given that new data:

$$P(H|D)$$

Posteriors: Example

- Hypothesis: Rob's date wants to see Avengers: Endgame.
  Data: She has seen Avengers: Infinity war.
- We need our posterior given that new data:

$$P(H|D)$$

- This is where Bayes comes in!

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Posteriors: Example

- Hypothesis: Rob's date wants to see Avengers: Endgame.
  Data: She has seen Avengers: Infinity War.

Posteriors: Example

- Hypothesis: Rob's date wants to see Avengers: Endgame.
  Data: She has seen Avengers: Infinity War.
- We need

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

- $P(H)$ ... prior
  $P(D)$ ... proportion of people who have seen Infinty War.
  $P(D|H)$ ... proportion of people wanting to see Avengers: Endgame who have seen Infinity War.

Posteriors: Example

- Hypothesis: Rob's date wants to see Avengers: Endgame.
  Data: She has seen Avengers: Infinity War.
- We need

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

- $P(H)$ ... prior
  $P(D)$ ... proportion of people who have seen Infinty War.
  $P(D|H)$ ... proportion of people wanting to see Avengers: Endgame who have seen Infinity War.
- Assume $P(D) = 0.6$ and $P(D|H) = 0.8$. What is $P(H|D)$?

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} =$$

$$= \frac{0.8 \times 0.6}{0.6} =$$

$$= 0.8$$

- Our posterior is therefore 0.8.
- We have therefore updated our belief that Rob's date wants to see Avengers: Endgame from a probability of 0.6 to a probability of 0.8.

Posteriors: Example

- OK, so our posterior is 0.8. But what if we collect additional new data?
- For instance: Rob's date has seen every single film starring Robert Downey Jr.
- This is where the magic happens: We can go through the updating process again, this time using the posterior we found previously as the new prior!

Sometimes we want to compare two hypotheses

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker
Instead of the posterior $P(H|D)$ we are now interested in the ratio of posteriors:

$$\frac{P(H_1|D)}{P(H_2|D)} =$$

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker
Instead of the posterior $P(H|D)$ we are now interested in the ratio of posteriors:

$$\frac{P(H_1|D)}{P(H_2|D)} =$$

$$= \frac{\frac{P(D|H_1)P(H_1)}{P(D)}}{\frac{P(D|H_2)P(H_2)}{P(D)}} =$$

$$= \frac{P(D|H_1)}{P(D|H_2)} \frac{P(H_1)}{P(H_2)}$$

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)}{P(D|H_2)}\frac{P(H_1)}{P(H_2)}$$

We call the Likelihood ratio $\frac{P(D|H_1)}{P(D|H_2)}$ the "Bayes Factor".
(How much more likely are we to see this data if H_1 is true than if H_2 is true?)

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)}{P(D|H_2)}\frac{P(H_1)}{P(H_2)}$$

We call the Likelihood ratio $\frac{P(D|H_1)}{P(D|H_2)}$ the "Bayes Factor".
(How much more likely are we to see this data if H_1 is true than if H_2 is true?)
Therefore:
Ratio of posteriors = Bayes factor $\times$ Ratio of Priors

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker
Smokers are about 25 times more likely to get lung cancer than non-smokers
20 % of people smoke

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)}{P(D|H_2)} \frac{P(H_1)}{P(H_2)}$$

Sometimes we want to compare two hypotheses

Data: A patient gets admitted to hospital with lung cancer.
Hypothesis 1: The patient is a smoker
Hypothesis 2: The patient is not a smoker
Smokers are about 25 times more likely to get lung cancer than non-smokers
20 % of people smoke

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)}{P(D|H_2)}\frac{P(H_1)}{P(H_2)}$$

$$\frac{P(H_1|D)}{P(H_2|D)} = 25\frac{0.2}{0.8} =$$

$$= 6.25$$

Some things you may worry about . . .

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} =$$

Some things you may worry about ...

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} =$$

- My prior is just a guess. What if it's a bad guess?
  That's OK. The priors are just a starting point. The whole point is that we are updating our beliefs when we get new information.

Some things you may worry about . . .

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} =$$

- My prior is just a guess. What if it's a bad guess?
  That's OK. The priors are just a starting point. The whole point is that we are updating our beliefs when we get new information.
- How do I get $P(D|H)$?
  Often this can be done using simulation. You have done this before!

Some things you may worry about ...

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} =$$

- My prior is just a guess. What if it's a bad guess?
  That's OK. The priors are just a starting point. The whole point is that we are updating our beliefs when we get new information.
- How do I get $P(D|H)$?
  Often this can be done using simulation. You have done this before!
- How do I know $P(D)$?
  If you are comparing two hypotheses, the nice thing is you don't need to know it!

Outline

1. The role of belief in Bayesian statistics
2. Priors – what are they? How do we define them?
3. How to calculate posterior probabilities
4. Bayesian vs Frequentist statistic

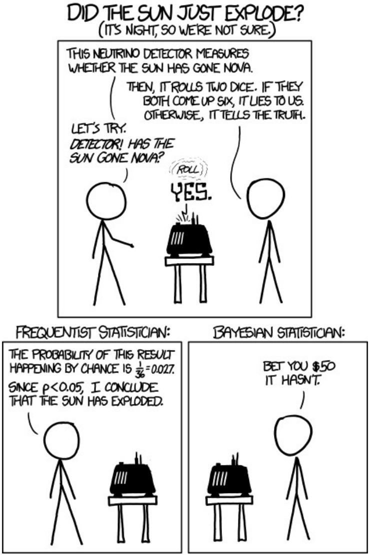
Are you a Bayesian?

- Up to now in ADS2, we have talked about frequentist statistics, e.g. t-tests, ANOVA, linear regression.
  - Parameters are fixed but unknown, e.g. true average population height, relationship between height and weight.
  - Our major aim was to test and/or falsify various hypotheses and make a decision.

Are you a Bayesian?

- Up to now in ADS2, we have talked about frequentist statistics, e.g. t-tests, ANOVA, linear regression.
  - Parameters are fixed but unknown, e.g. true average population height, relationship between height and weight.
  - Our major aim was to test and/or falsify various hypotheses and make a decision.
- Bayesian statistics is quite different.
  - Include probabilistic methods and reasoning to the parameters: want to estimate the chance something is true *and* the extent of our belief in the estimate.
  - Can be updated with new knowledge, just like scientists operate normally.
  - Truly explicit/scrutable - did you notice the lack of assumptions in today's lecture?

Are you a Bayesian?
Are you a Frequentist?


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36}=0.027$.
SINCE $p<0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

A lot of frequentists may secretly be Bayesians

A lot of frequentists may secretly be Bayesians

Examples of how non-significant results have been described in the literature (for a full list, see: `https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/`)

barely outside the range of significance ($p = 0.06$)
below (but verging on) the statistical significant level ($p > 0.05$)
bordered on being significant ($p > 0.07$)
borderline level of statistical significance ($p = 0.053$)
close to a marginally significant level ($p = 0.06$)
close to being significant ($p = 0.06$)
close to being statistically significant ($p = 0.055$)
close to borderline significance ($p = 0.072$)

near-to-significance ($p = 0.093$)
non-significant in the statistical sense ($p > 0.05$)
not absolutely significant but very probably so ($p > 0.05$)
not clearly significant ($p = 0.08$)
not completely significant ($p = 0.07$)
not conventionally significant ($p = 0.089$), but ...
not especially significant ($p > 0.05$)
not exactly significant ($p = 0.052$)
not formally significant ($p = 0.06$)
not insignificant ($p = 0.056$)
not markedly significant ($p = 0.06$)

Review

Now, you should be able to do the following:

- Explain what "belief" means to a Bayesian statistician
- Define the terms "prior" and "posterior"
- Specify prior probabilities and distributions
- Use Bayes' theorem to determine posterior probabilities
- Describe the difference between Bayesian and Frequentist statistics

Acknowledgments and Image Credits

This lecture uses materials from an ADS2 lecture from previous years by Rob Young (who is currently at the cinema watching Avengers). Where not otherwise indicated, images are also from that lecture.

Image credits:

- Bayes' Theorem in neon letters. By mattbuck (category) - Own work by mattbuck., `https://commons.wikimedia.org/w/index.php?curid=14658489`
- Fireworks. By Ijasmuhammed - Own work, `https://commons.wikimedia.org/w/index.php?curid=75435451`