

# *Introduction to Bayesian statistics*

## *Part 1 — Concepts*

Jorge N. Tendeiro

Office of Research and Academia-Government-Community Collaboration  
Education and Research Center for Artificial Intelligence and Data Innovation  
Hiroshima University, Japan

04 Fevereiro 2022

✉ [tendeiro@hiroshima-u.ac.jp](mailto:tendeiro@hiroshima-u.ac.jp)

🌐 [www.jorgetendeiro.com](http://www.jorgetendeiro.com)

🔗 <https://github.com/jorgetendeiro/Seminario-IntroBayesStats>

## Bayes rule

- ▶  $\mathcal{D}$  = data
- ▶  $\theta$  = unknown parameter

$$p(\theta|\mathcal{D}) = \frac{p(\theta)p(\mathcal{D}|\theta)}{p(\mathcal{D})}$$

In words,

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

The *evidence* does not depend on  $\theta$ ; let's hide it:

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

The symbol  $\propto$  means “proportional to”.

## Bayes rule

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

- ▶ *Prior*: Belief about the 'true' value of  $\theta$ , *before looking at the data*.
- ▶ *Likelihood*: The statistical model, linking  $\theta$  to data.
- ▶ *Posterior*: Updated knowledge about  $\theta$ , in light of the observed data.

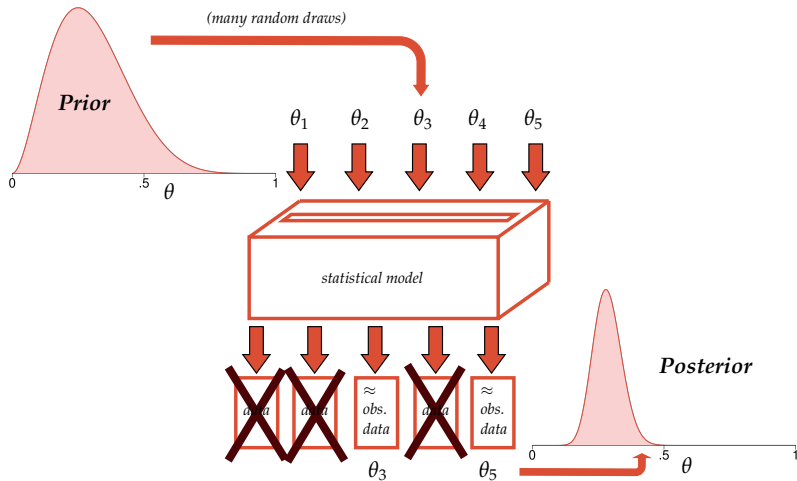
Let's look at the Bayes rule from various angles.

## Bayes rule – ABC

One useful way to think about the Bayes rule is by considering *Approximate Bayesian Computation* (ABC; see [Wiki](#)).

- ▶ ABC is actually computationally *very* inefficient.
- ▶ But, it is *conceptually* very clear!

## Bayes rule – ABC



## Bayes rule – ABC

The Bayes rule from the ABC perspective:

*Find the values of  $\theta$  that allow the model to predict data pretty much like our observed data.*

Humm...

Maximum likelihood estimation, anyone?

Bayesian inference can be *conceptually* thought of as an extension of MLE!

## Bayes rule – Inverse probability

Bayes rule allows reversing conditional probabilities.

$$p(\mathcal{A}|\mathcal{B}) = \frac{p(\mathcal{A})p(\mathcal{B}|\mathcal{A})}{p(\mathcal{B})}$$

Consider the canonical example:

- ▶  $\mathcal{A}$  : Have disease.
- ▶  $\mathcal{B}$  : Test positive.

Then:

- ▶  $p(\mathcal{B}|\mathcal{A})$  : Probability of testing positive given that one has the disease.  
*Test's sensitivity.*
- ▶  $p(\mathcal{A}|\mathcal{B})$  : Probability of having the disease given that (s)he tests positive.  
*What patients really want to know.*

# Bayes rule – Updating beliefs

Definition of probability:

- ▶ *Frequentist*: Long-run relative frequency.  
(Problem:  $p(\text{Pandemic is over in 2022})?$ ...)
- ▶ *Bayesian*: Degree of subjective belief.

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

- ▶ Bayes rule is a rational means of updating our current belief (*prior*) by means of observed data (*likelihood*).
- ▶ “Today’s posterior is tomorrow’s prior” – Lindley (1970)



## Bayes rule – Summary

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

Bayesian modelling requires three ingredients:

- ▶ Data.
- ▶ Priors, reflecting our subjective belief about the parameters.
- ▶ A statistical model, relating parameters to data.

Bayes rule is a mathematically rigorous means to combine prior information on *parameters* with the *data*, using the *statistical model* as the bridge between both.

## Bayes rule – Example

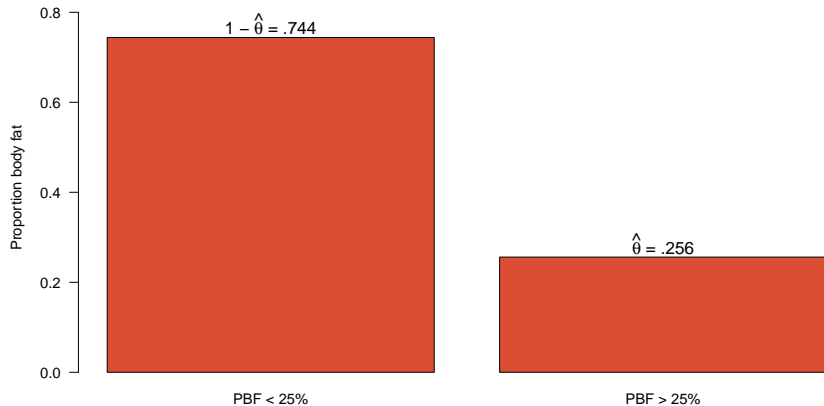
Data here:

<https://das1.datadescription.com/datafile/bodyfat/>.

- ▶ Various measurements of 250 men.
- ▶ To keep it simple: I dichotomize the percentage of body fat (PBF).
- ▶ 0 = PBF lower than 25%;  
1 = PBF larger than 25%.
- ▶ *Goal*: Infer the proportion of obese men in the population.

Let's denote the population proportion by  $\theta$ .

## Bayes rule – Example



## *Bayes rule – Example*

Let's use the Bayesian machinery.

Recall that we need three ingredients:

- ▶ Data.
- ▶ Prior.
- ▶ Model.

## Bayes rule – Example

*Data.* For now, let's only use the first 10 scores.

- ▶ Sample size: 10
- ▶ Number of men with PBF > 25%: 2
- ▶ Sample proportion:  $\hat{\theta} = \frac{2}{10} = .20$

0	0	1	0	1	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---

## Bayes rule – Example

*Model.* We'll use the binomial model. Assumptions:

- ▶ Independence between measurements.
- ▶ One population with underlying rate  $\theta$ .
- ▶ Random sample.

## *Bayes rule – Example*

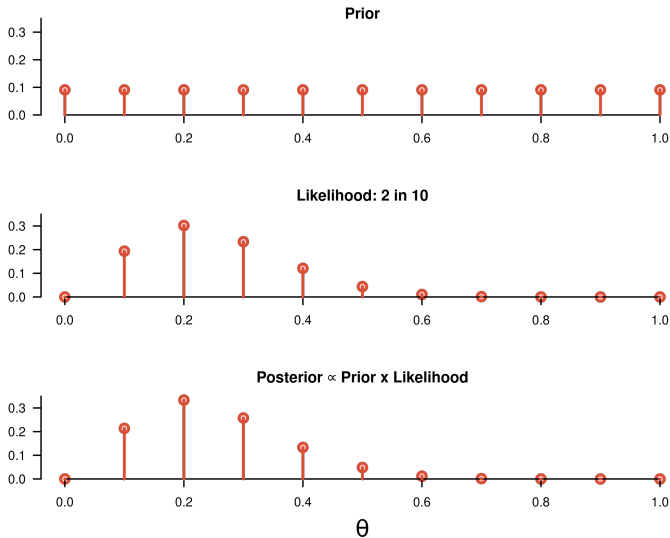
*Prior.* We'll try several.

## *Bayes rule – Example*

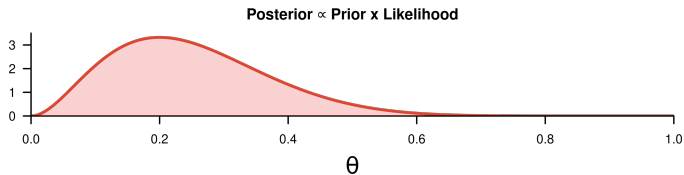
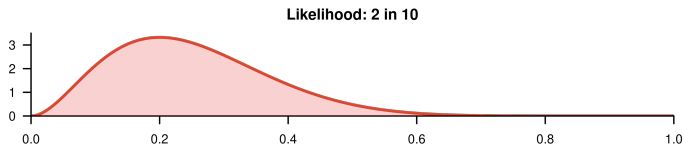
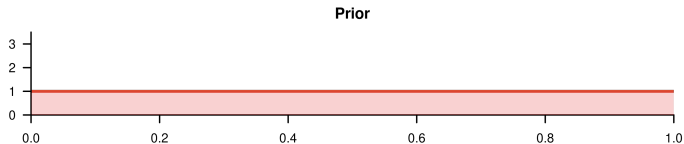
What happens if the prior is ‘uninformative’?



## Bayes rule – Example



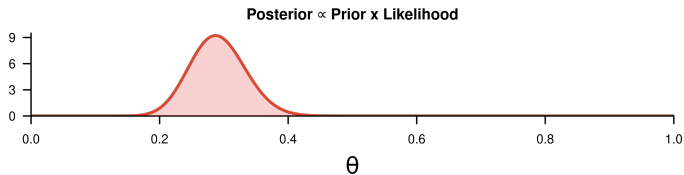
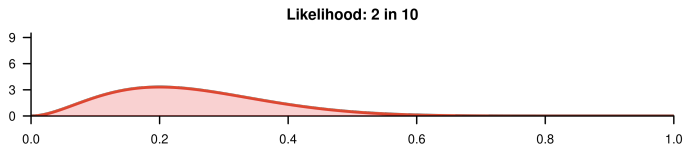
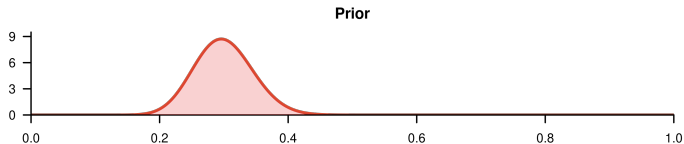
## Bayes rule – Example



## *Bayes rule – Example*

What happens if the prior is ‘very informative’?

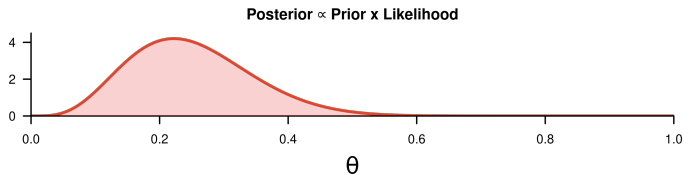
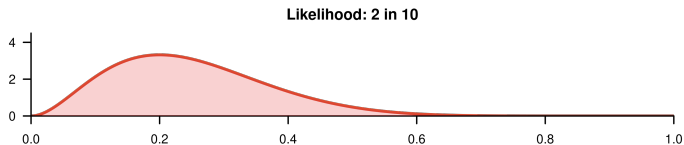
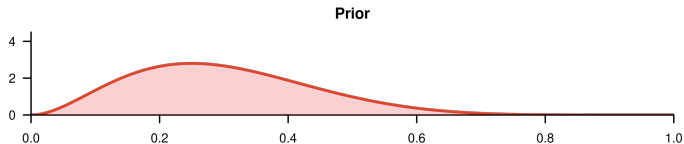
## Bayes rule – Example



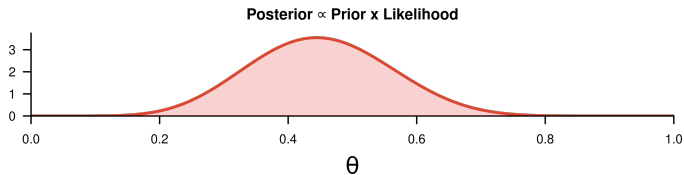
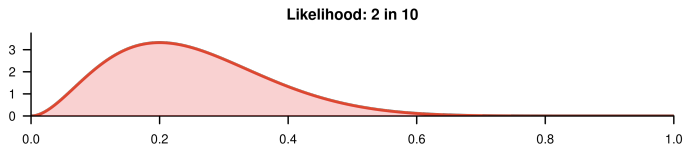
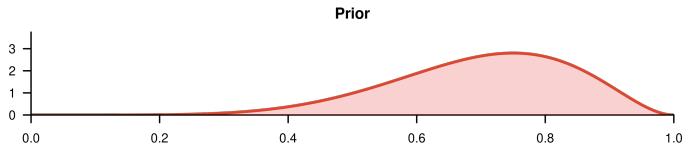
## *Bayes rule – Example*

What happens if neither the prior nor the likelihood dominates?

## Bayes rule – Example



## Bayes rule – Example



## Bayes rule – Example

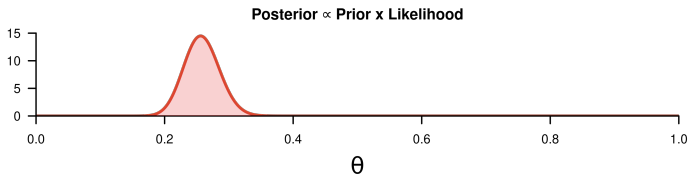
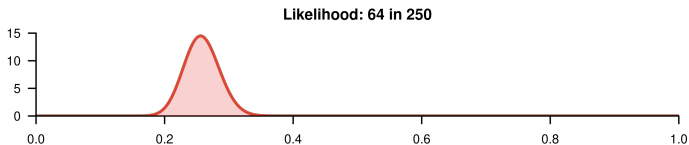
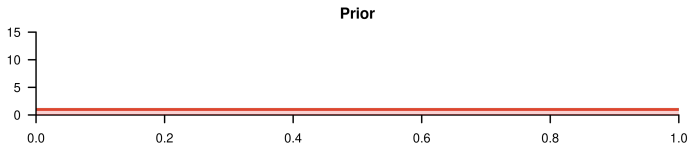
Let's now use all the data.

- ▶ Sample size: 250
- ▶ Number of men with PBF > 25%: 64
- ▶ Sample proportion:  $\hat{\theta} = \frac{64}{250} = .256$

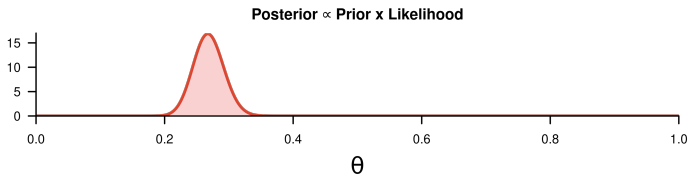
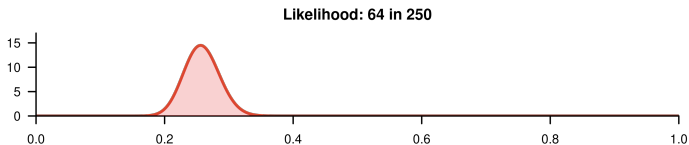
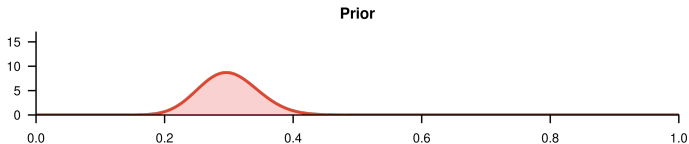
How does Bayes updating look like now, when the data dominate?



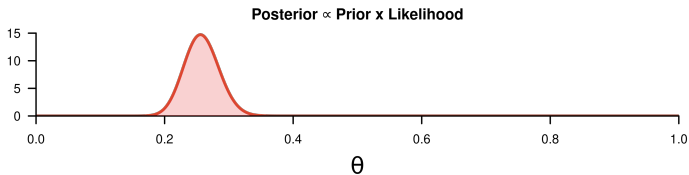
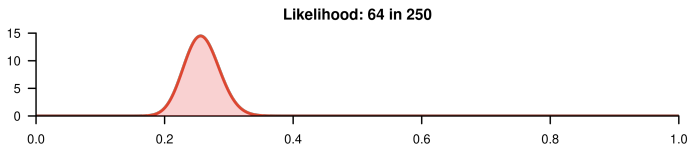
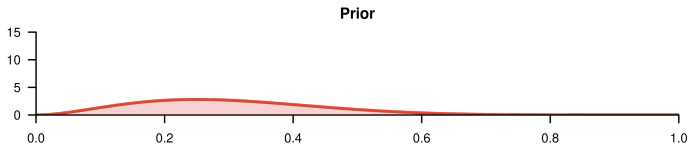
## Bayes rule – Example



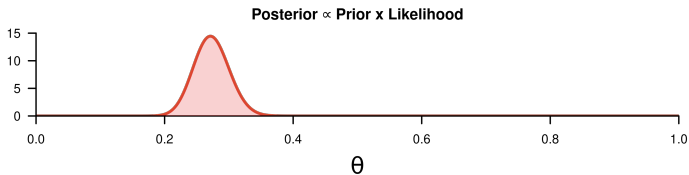
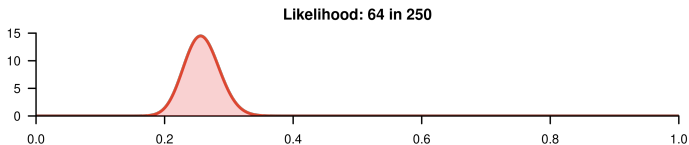
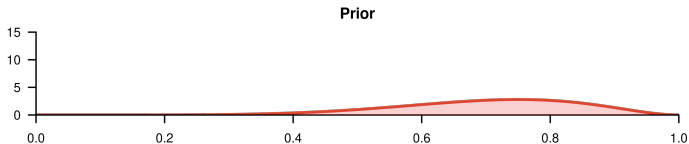
## Bayes rule – Example



## Bayes rule – Example



## Bayes rule – Example



## *Bayes rule – Some conclusions*

- ▶ Bayes rule highlights the parameter values that make the observed data look more plausible.
- ▶ The posterior distribution is a compromise between the information in the prior and the information in the data.

## *Bayes rule – Some conclusions*

How do priors typically affect posterior distributions?

- ▶ For 'uninformative' priors, posterior  $\approx$  likelihood.
- ▶ For 'very informative' priors, posterior  $\approx$  prior.

## *Bayes rule – Some conclusions*

How do data typically affect posterior distributions?

- ▶ For small sample sizes, posterior  $\approx$  prior.
- ▶ For large sample sizes, posterior  $\approx$  likelihood.

## Bayesian inference – Some criticism

I think I can hear some of you thinking right now...

*“Hey, but there are sooo many posterior distributions!”*

*“This seems all sooo subjective!”*

*“I sooo don’t like it!”*

:-)

Fair points.

Let me offer several counter-arguments.

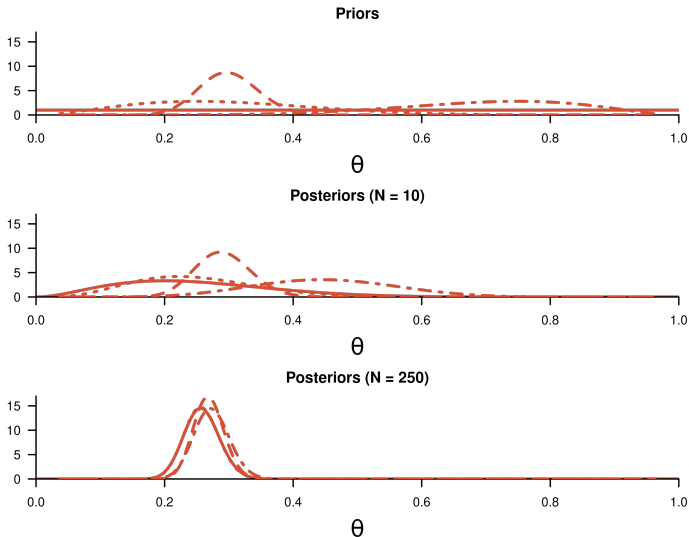


## *Bayesian inference – Counterarguments to criticism*

1. Posterior distributions are fairly stable across a wide range of reasonable priors, *for large data sets*.

More data  $\Rightarrow$  more information  $\Rightarrow$  more certainty.

# Bayesian inference – Counterarguments to criticism



## *Bayesian inference – Counterarguments to criticism*

2. *Illusion of certainty*: Pretending that results tell us more than is actually possible.

## *Bayesian inference – Counterarguments to criticism*

There is subjectivity in each step of the scientific way:

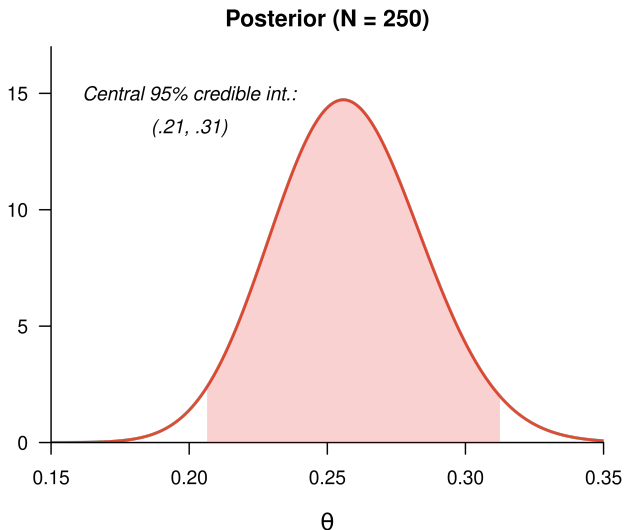
- ▶ Selection of participants.
- ▶ Number of assessments.
- ▶ Variables to measure.
- ▶ Variables to control.
- ▶ Variability across researchers / labs.
- ▶ Statistical model to use.
- ▶ Variables to (not) include in the model.
- ▶ ...

Then, try topping it up with few and noisy data...

It is *fair, logical, necessary*, that statistical inference reflects uncertainty.

Do embrace uncertainty!

## Bayesian inference – Counterarguments to criticism



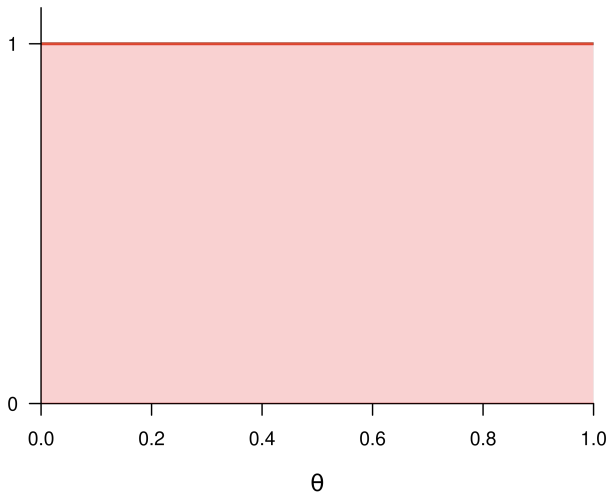
## *Bayesian inference – Counterarguments to criticism*

3. Priors allow incorporating useful information.

▶ What is known about the parameter?

Let's not pretend we do not know anything.

## *Bayesian inference – Counterarguments to criticism*



## Bayesian inference – Counterarguments to criticism

About PBF:

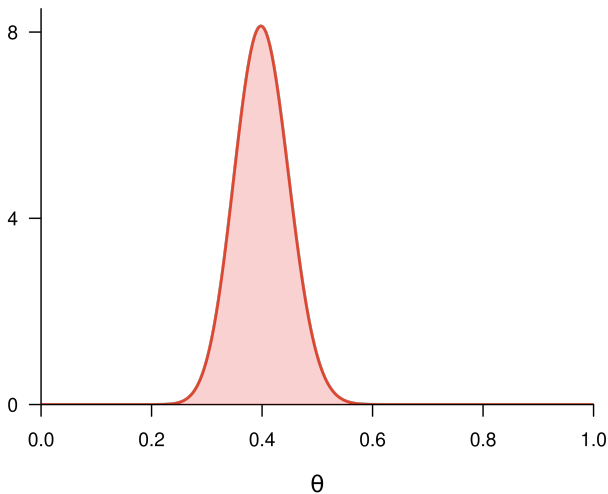
- ▶ It is actually known that the proportion of obese men in (some...) population is **about 40%**.
- ▶ We can (we *should*!) take this into account.
- ▶ And *that* is what the prior is for.

(Do you know how much variability to expect?  
Then include this in the prior too!!)

One of Bayes' advantages: Accummulation of evidence.  
Use it!



## *Bayesian inference – Counterarguments to criticism*



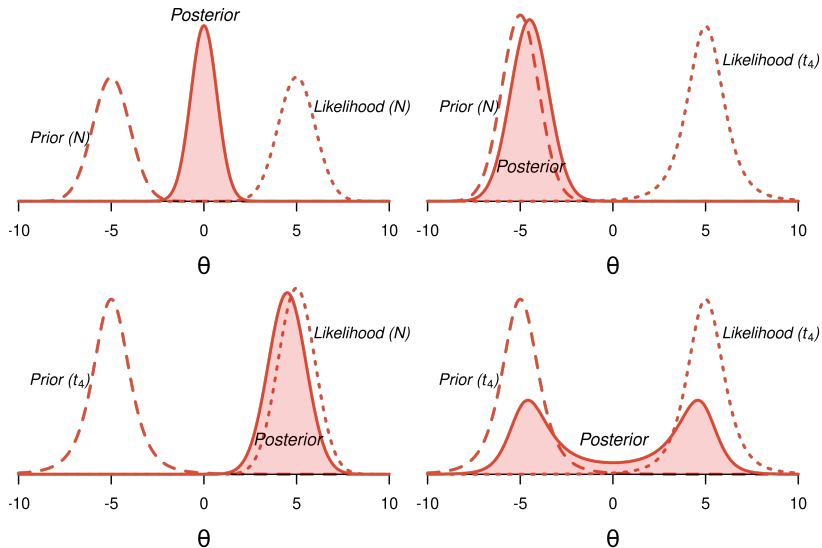
## *Bayesian inference – Counterarguments to criticism*

Of course, strange things can occur.

*“A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.”*  
(Stephen Senn)

(Inspired by *Aki Vehtari* and *John Kruschke*.)

# Bayesian inference – Counterarguments to criticism



## *Bayesian inference – Summarize results*

The posterior distribution is the *Holy Grail* in Bayesian statistics.

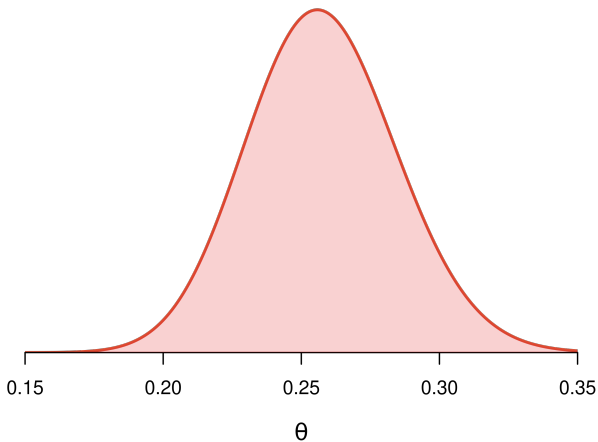
It reflects our current knowledge of the world, conditional on:

- ▶ The chosen model.
- ▶ The chosen prior(s).
- ▶ The observed data.

How can we summarize the information in the posterior distribution?

## *Bayesian inference – Summarize results*

**Posterior (N = 250)**



## *Bayesian inference – Summarize results*

*Point estimates.*

Commonly used:

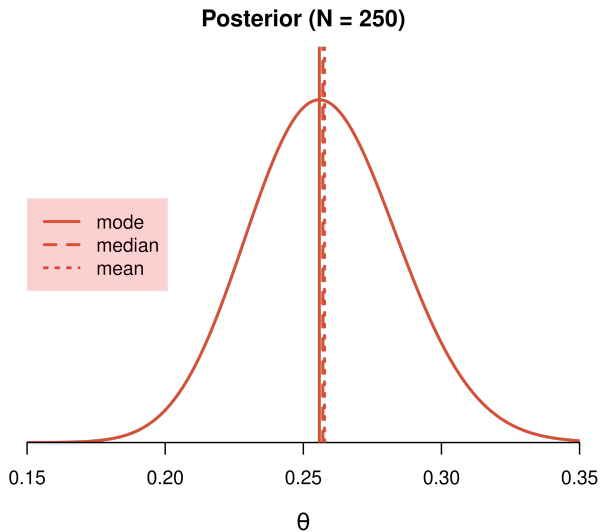
- ▶ posterior mean
- ▶ posterior mode
- ▶ posterior median.

For the PBF data based on 250 scores:

post. mean  $\approx$  post. mode  $\approx$  post. median  $\approx$  .26.

(Recall:  $\hat{\theta} = .256$ .)

## Bayesian inference – Summarize results



## Bayesian inference – Summarize results

*Interval estimates.*

I will focus on the 95% credible interval throughout.

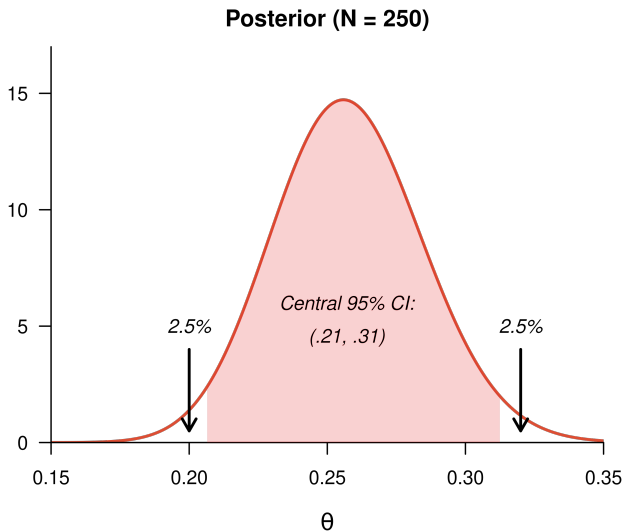
There are some variants (do not overly worry about these nuances):

- ▶ *Central 95% credible interval.*  
With 2.5% probability out on each tail.
- ▶ *95% HDI (highest density interval).*  
The shortest interval covering area .95.

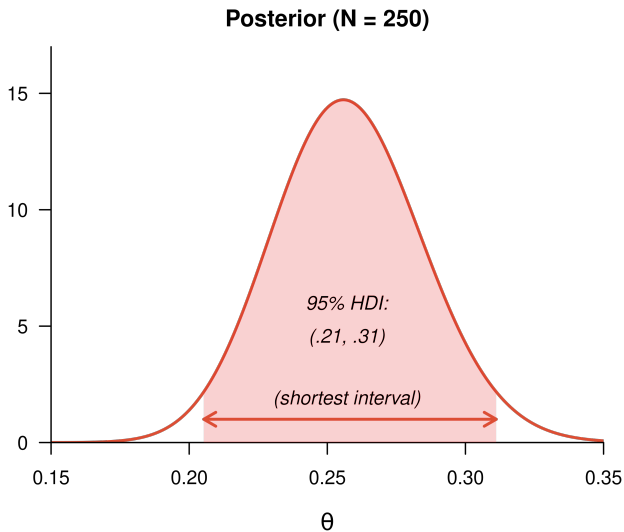
For the PBF data they practically coincide.



## Bayesian inference – Summarize results



## Bayesian inference – Summarize results



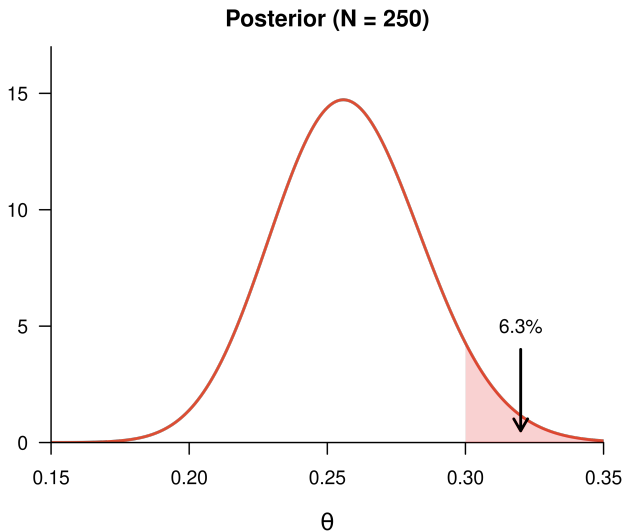
## *Bayesian inference – Summarize results*

The posterior distribution allows computing probabilities for any events involving parameters.

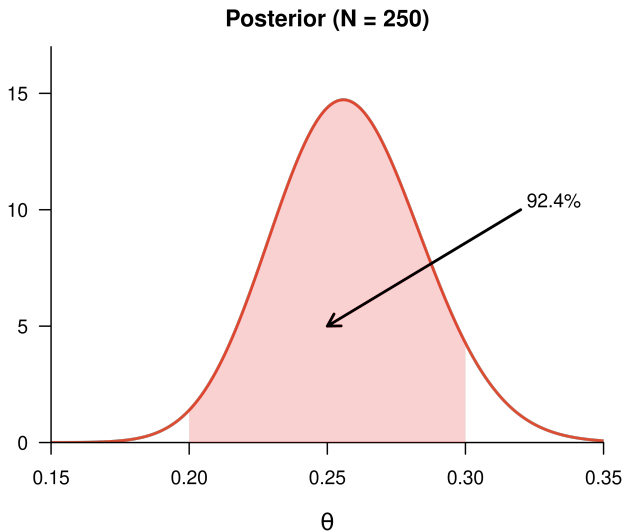
For instance:

- ▶ What is the (posterior) probability that the population proportion of obese men is larger than 30%?
- ▶ What is the (posterior) probability that the population proportion of obese men is between 20% and 30%?

## *Bayesian inference – Summarize results*



## *Bayesian inference – Summarize results*



## *Next*

Specific examples will be dealt with in Part 2.

More concepts will be introduced as we proceed.

Now where's that cup of coffee?

