Introduction to Bayesian statistics

Part 1 — Concepts

Jorge N. Tendeiro

Department of Psychometrics and Statistics
Faculty of Behavioral and Social Sciences
University of Groningen

j.n.tendeiro@rug.nl
https://github.com/jorgetendeiro/GSMS-2020

Bayes rule

- $\mathcal{D}$ = data
- θ = unknown parameter

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

In words,

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

The *evidence* does not depend on θ; let's hide it:

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

The symbol $\propto$ means "proportional to".

Bayes rule

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

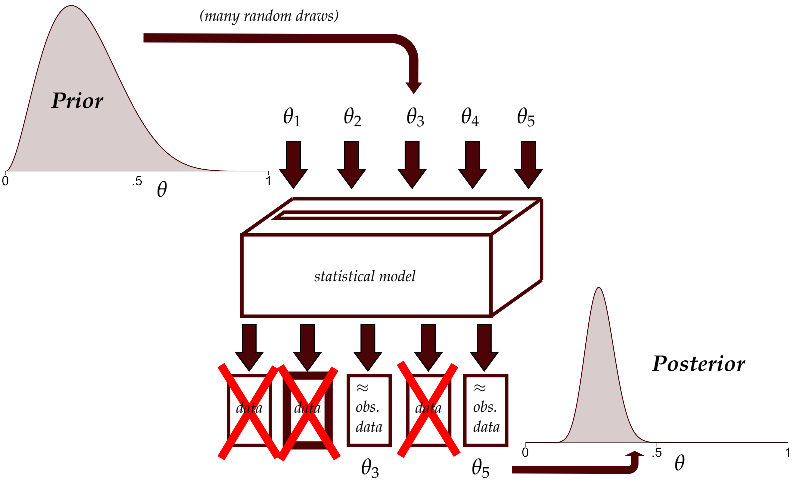
- *Prior*: Belief about the 'true' value of θ, *before looking at the data*.
- *Likelihood*: The statistical model, linking θ to data.
- *Posterior*: Updated knowledge about θ, in light of the observed data.

Bayes rule

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



Prior
0
.5
1
θ
(many random draws)
θ1 θ2 θ3 θ4 θ5
statistical model
data
data
≈ obs. data
data
≈ obs. data
θ3
θ5
Posterior
0
.5
1
θ

Bayes rule

The Bayes rule from the ABC perspective:

> *Find the values of θ that allow the model to predict data pretty much like our observed data.*

Humm. . .

MLE, anyone?

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

Bayes rule – Summary

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

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.

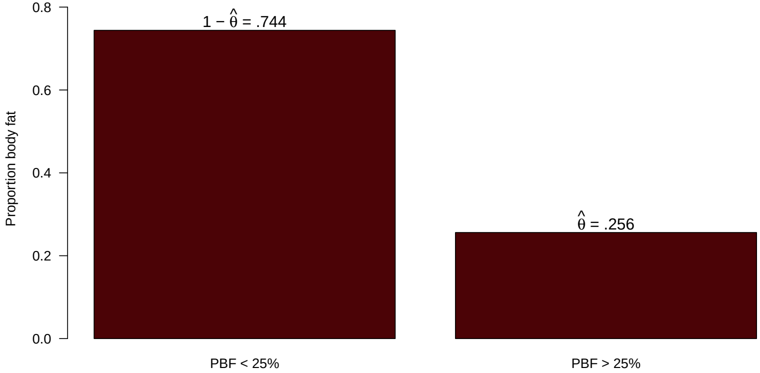
Bayes rule – Example

Data here:
https://dasl.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 infer the proportion of obese men in the population.

Let's denote the population proportion by θ.

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.

0	0	1	0	1	0	0	0	0	0

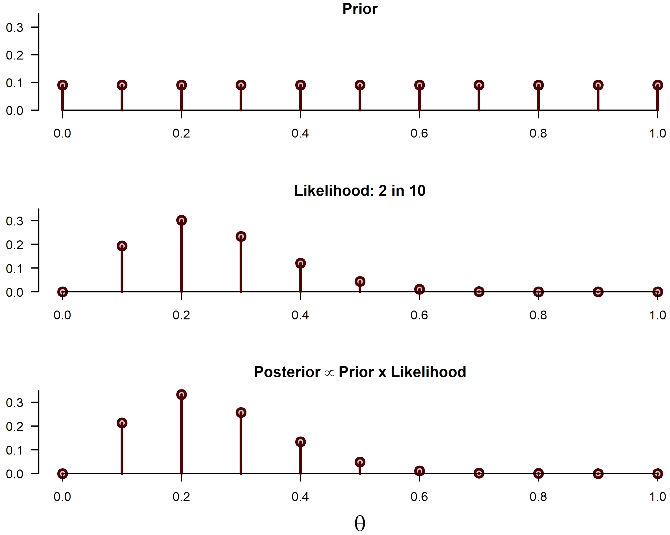
Bayes rule – Example

- *Model.* We'll use the binomial model. Assumptions:
  - ✓ Independence between measurements.
  - ✓ One population with underlying rate θ.
  - ✓ Random sample.

Bayes rule – Example

- *Prior.* We'll try several.

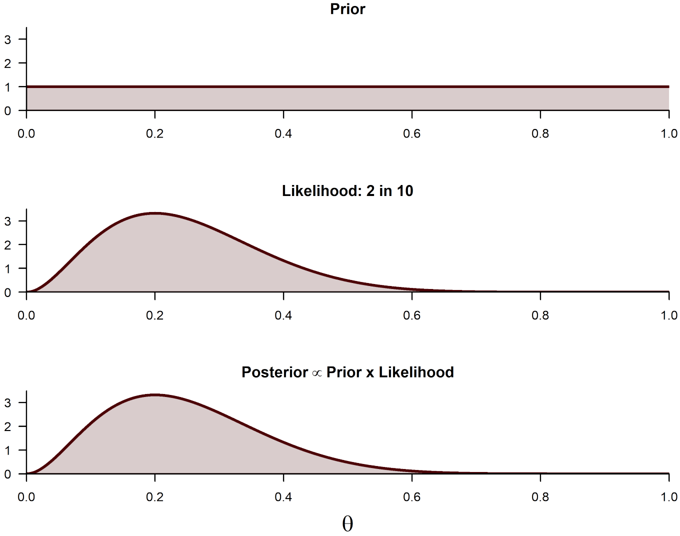
Bayes rule – Example



Bayes rule – Example

What happens if the prior is ‘uninformative’?

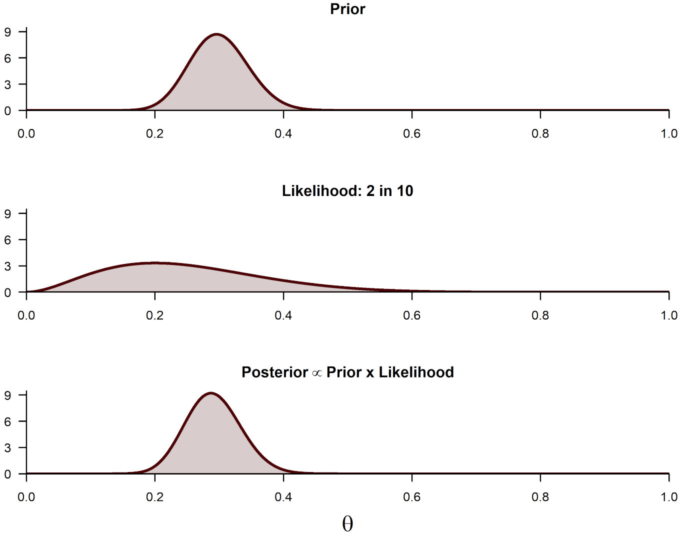
Bayes rule – Example



Bayes rule – Example

What happens if the prior is 'very informative'?

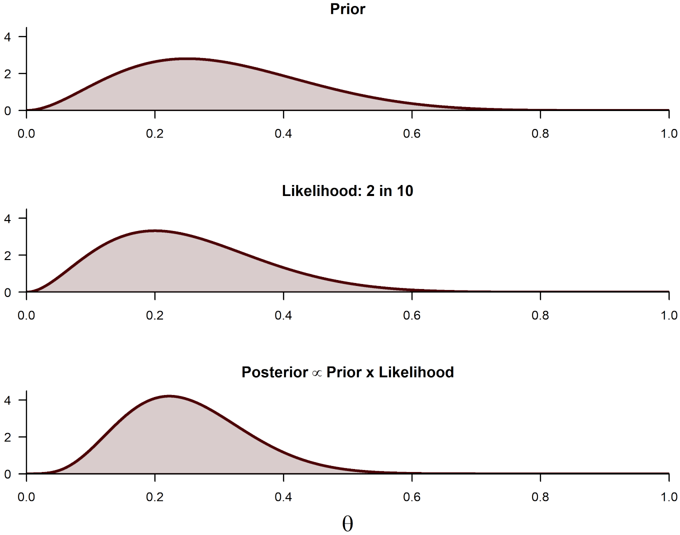
Bayes rule – Example



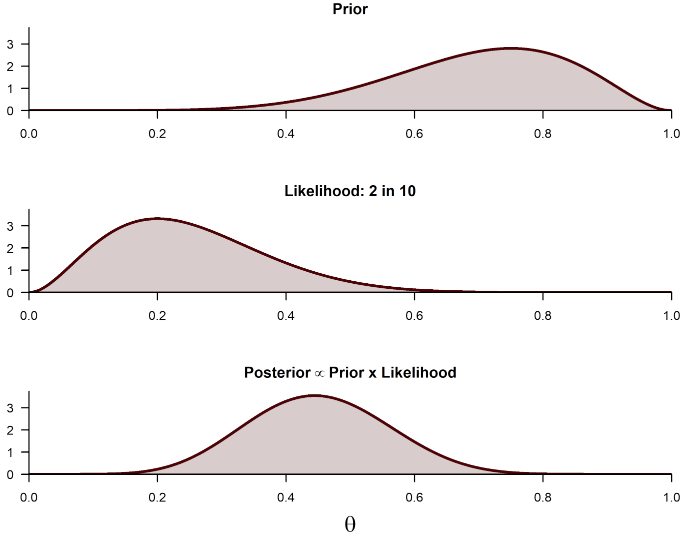
Bayes rule – Example

What happens if neither the prior nor the likelihood dominates?

Bayes rule – Example



Bayes rule – Example



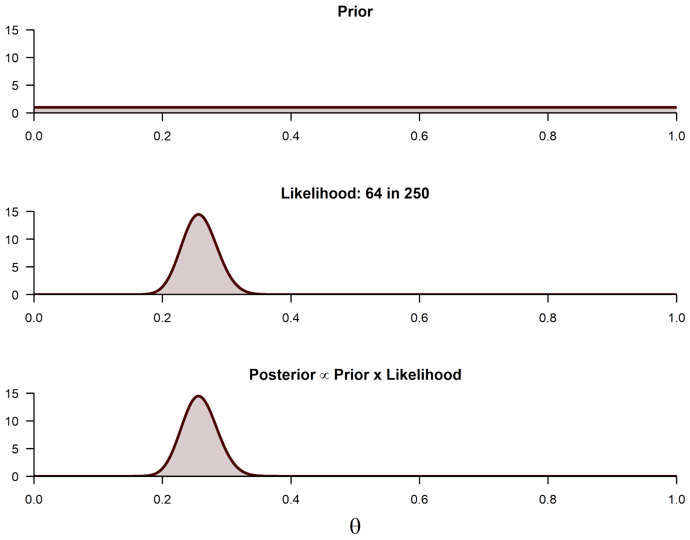
Bayes rule – Example

Let's now use all the data.

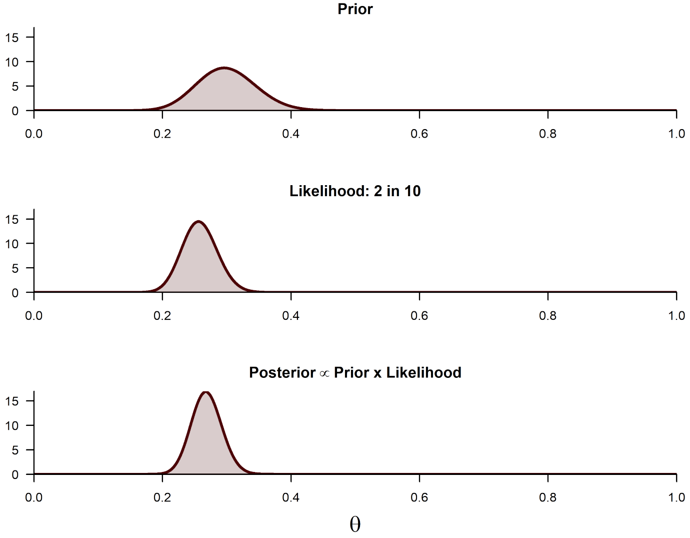
- Sample size: 250
- Number of men with PBF $> 25\%$: 64
- Sample proportion: $\widehat{\theta} = \frac{64}{250} = .256$

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

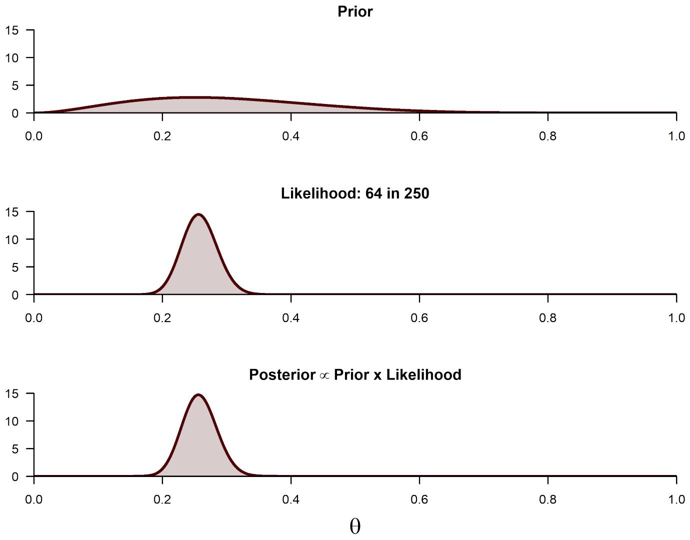
Bayes rule – Example



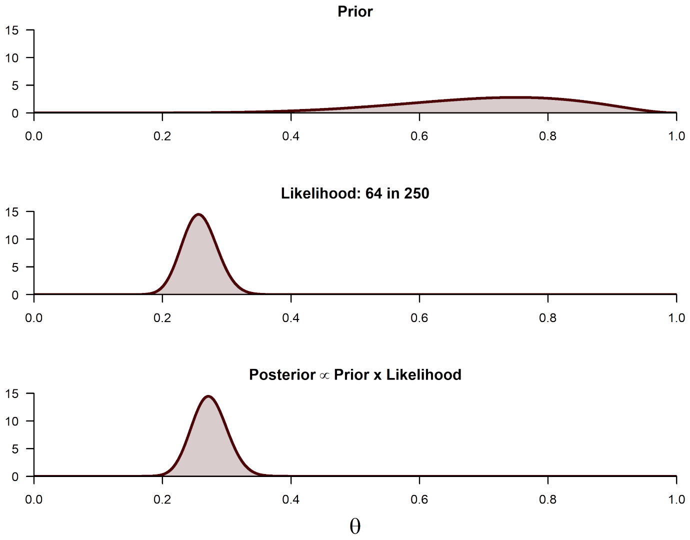
Bayes rule – Example



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.

References

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*(7), 1105. doi: 10.1037/xge0000560