Introduction to Bayesian statistics[1]

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

[1] The template from these slides is inspired in that from Mark Andrews

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.

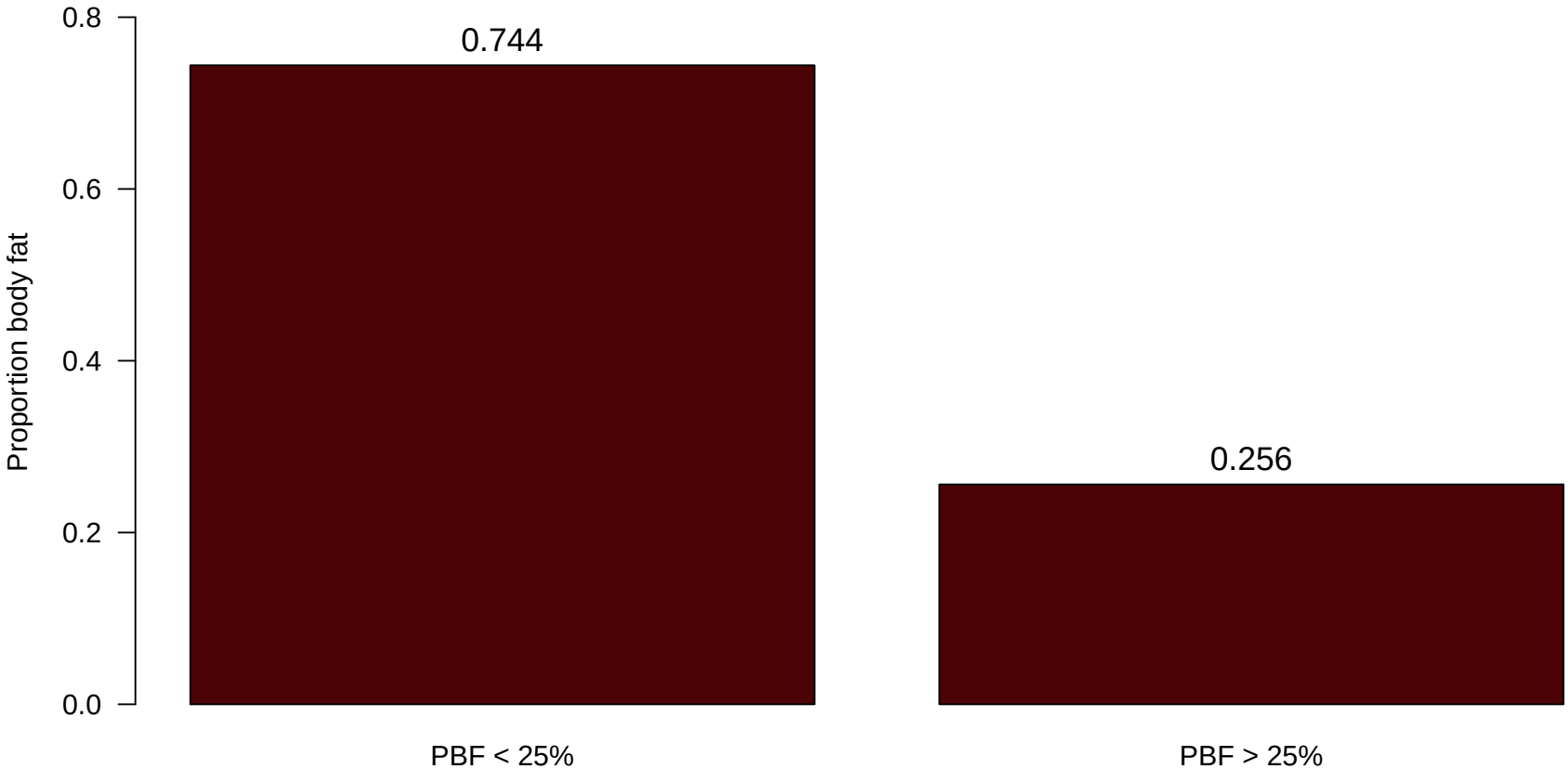
Bayes rule – Example

Data here:
https://dasl.datadescription.com/datafile/bodyfat/.

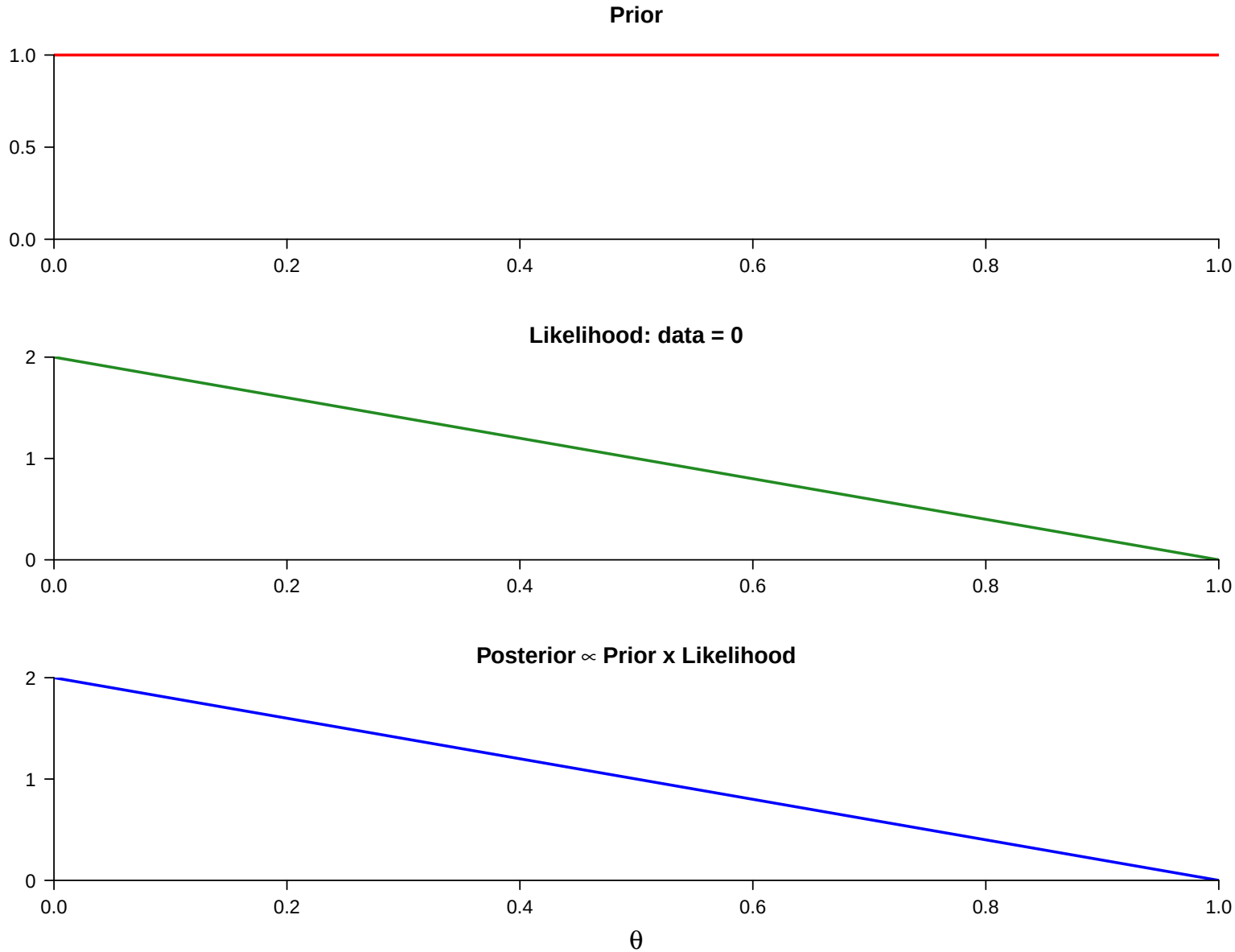
- Various measurements of 250 men.
- Keep it simple: Let's dichotomize the percentage of body fat (0 = PBF lower than 25%; 1 = PBF larger than 25%).
- *Goal*: Infer infer the proportion of obese men in the population.

```
url.data <- "https://dasl.datadescription.com/download/data/3079"
PBF.data <- read.csv(url(url.data), header = TRUE, sep = "\t")
PBF      <- ifelse(PBF.data$Pct.BF > 25, 1, 0)
prop.table(table(PBF))
```

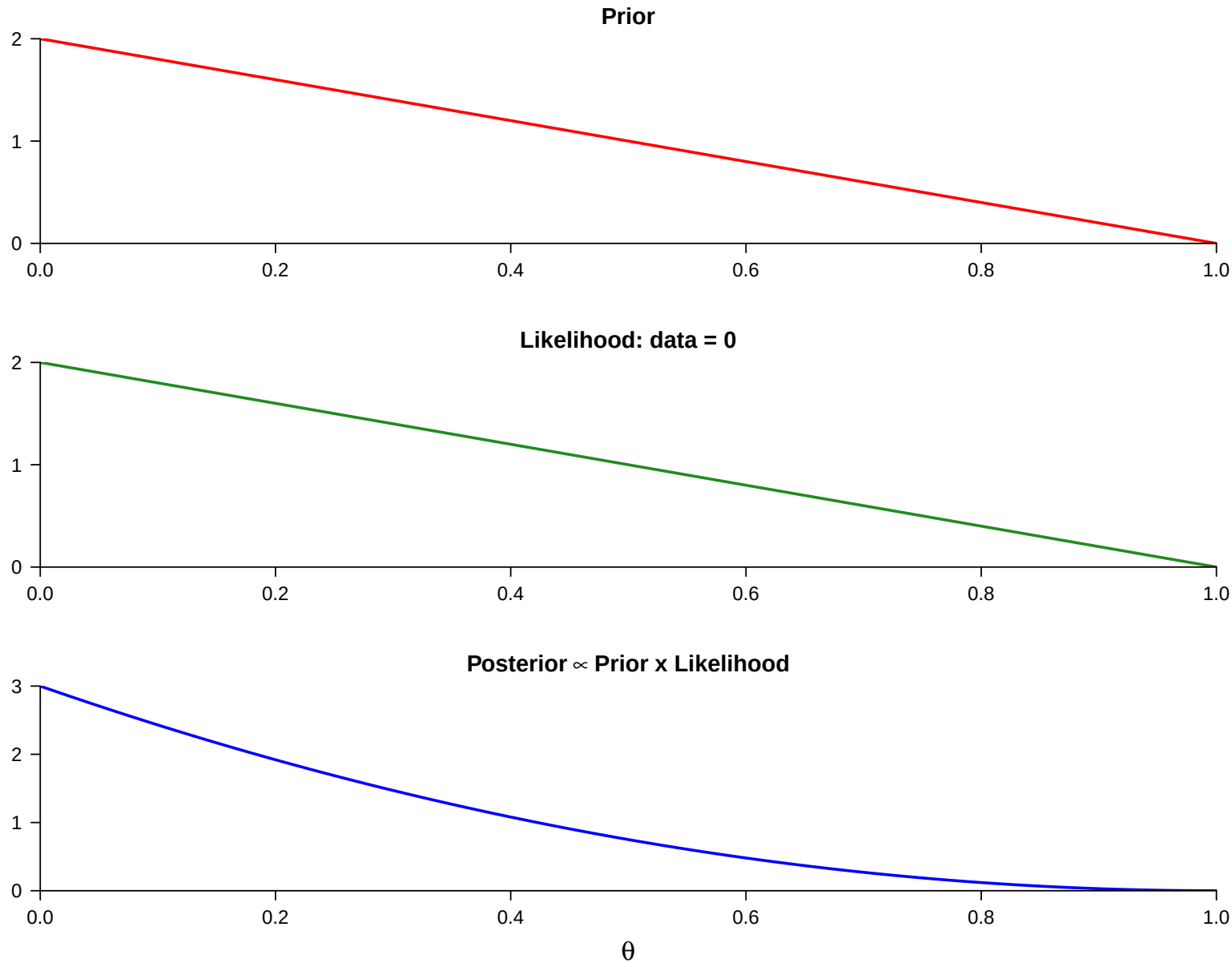
Bayes rule – Example



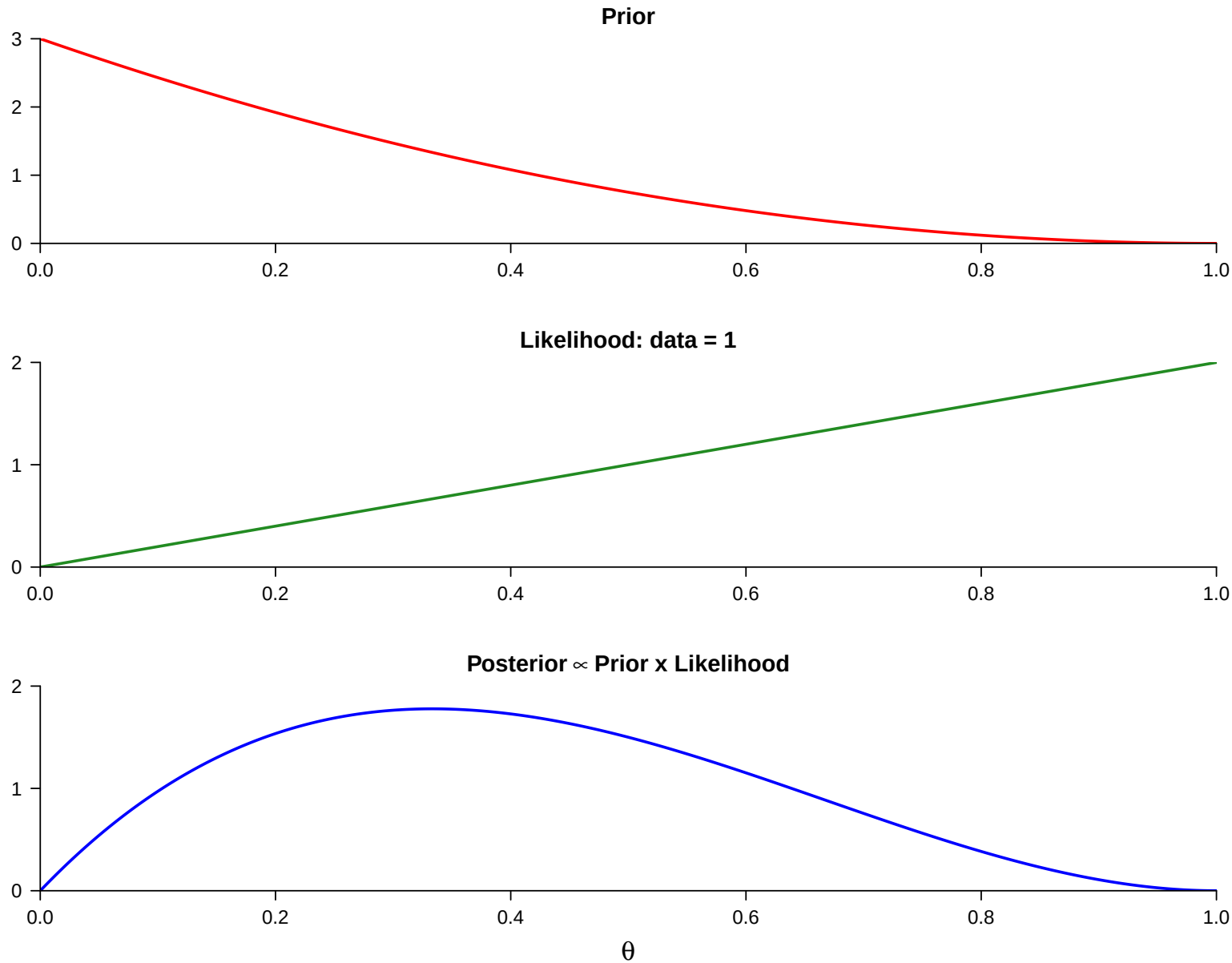
Bayes rule – Example (N = 1)



Bayes rule – Example (N = 2)



Bayes rule – Example (N = 3)



The rest

Therefore, the posterior distribution is basically a (rational, logically correct) means of merging together both our prior knowledged about some phenomenon with the information about the phenomenon that our data has to offer.

A small example

Let's make things concrete. I downloaded data from https://dasl.datadescription.com/datafile/bodyfat/, containing various measurements of 250 men. I focus on variable 'Pct.BF' (percentage of body fat) and dichotomize it (0 = PBF lower than 25%; 1 = PBF larger than 25%). I want to infer the proportion of obese men in the population.

```
PBF.data <- read.csv(url("https://dasl.datadescription.com/downl
                         header = TRUE, sep = "\t")
PBF      <- ifelse(PBF.data$Pct.BF > 25, 1, 0)
head(PBF)
```

```
## [1] 0 0 1 0 1 0
```

```
length(PBF)
```

```
## [1] 250
```

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