Introduction to Bayesian Data Analysis

Lecture 1

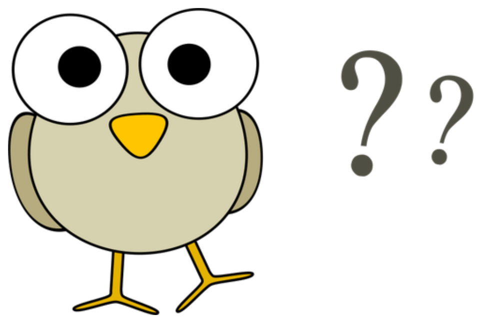
Julia Haaf

Summer 2025

Who are we?

- Dr. Nicole Cruz
    - Postdoc, University of Potsdam

- Dr. Julia Haaf
    - Professor, University of Potsdam



Who are you?

Overview over the workshop

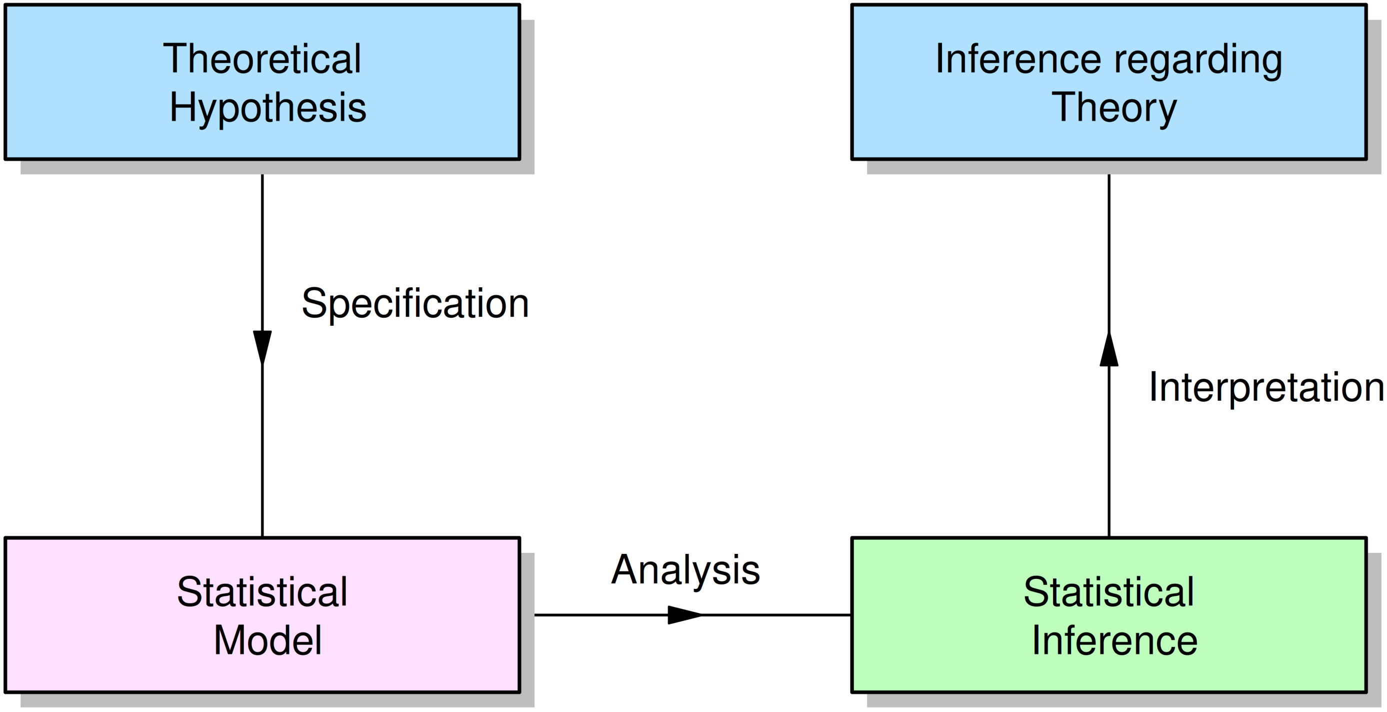
Sheet1

	Monday	Tuesday	Wednesday	Thursday	Friday Project Day
09:00:00 AM	**Welcome Session**	**Lecture 3**, Bayes factor	**Lecture 5**, Hierarchical modeling	**Lecture 6**, Nonlinear models	Project presentations
09:30:00 AM	**Lecture 1**, Introduction				
10:00:00 AM					
10:30:00 AM	*Coffee Break*				
11:00:00 AM					
11:30:00 AM	Excercises	Excercises	Excercises	Excercises	On your own
12:00:00 PM					
12:30:00 PM	*Lunch Break*				
01:00:00 PM					
01:30:00 PM					
02:00:00 PM	**Lecture 2**, Bayesian Workflow	**Lecture 4**, Linear regression	Q&A, reading time	Excercises	Small meetings on own project
02:30:00 PM					
03:00:00 PM	*Coffee Break*				
03:30:00 PM					On your own
04:00:00 PM	Excercises	Excercises	Excercises	Q&A, reading time	
04:30:00 PM					*Break*
05:00:00 PM					**Closing Remarks**
05:30:00 PM		**Keynote**: Ben Bolker		**Keynote**: Reinhold Kliegl	
06:00:00 PM					
06:30:00 PM					

Overview

1. Statistical Modeling
2. Bayesian Statistics
    - An Example
    - Prior
    - Bayes's Theorem
    - Posterior
    - Estimation in `R`

Inferential Statistics



Statistical Models

A statistical model is the mathematical representation of a series of statistical assumptions and relationships. Statistical models contain information about the generation of sample data from the population.

- Mathematical relationships between random variables and other non-random variables.
- Statistical model as “formal representation of a theory”.
- Statistical models represent the process of data generation.

Statistical Models

Statistical model as assumptions about the population

$\longrightarrow$ Statistical Model = Probability Distribution

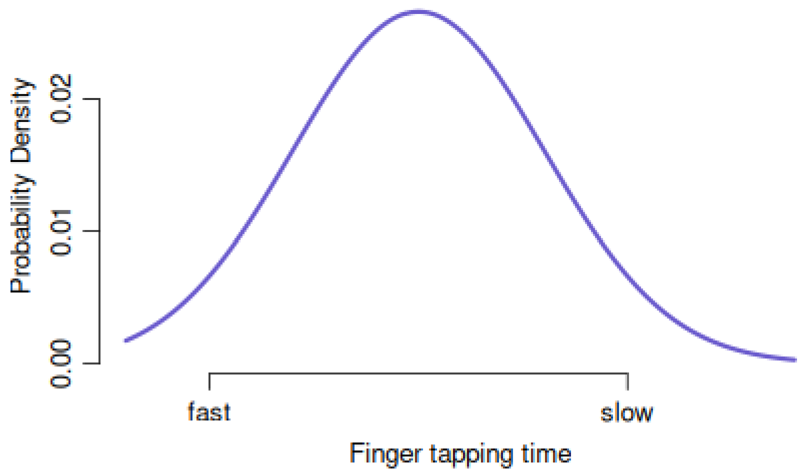
Statistical Models

Example: A single subject pressing a button repeatedly (Chapter 3.2.1)

- Finger tapping task (for a review, see Hubel et al. 2013)
- Procedure
  - blank screen (200 ms) + cross in the middle of a screen
  - as soon as they see the cross, they tap on the space bar as fast as they can until the experiment is over (361 trials).
- Dependent measure: Finger tapping times in milliseconds.
- Research question is: how long does it take for this particular subject to press a key?

Statistical Models

Example: A single subject pressing a button repeatedly (Chapter 3.2.1)

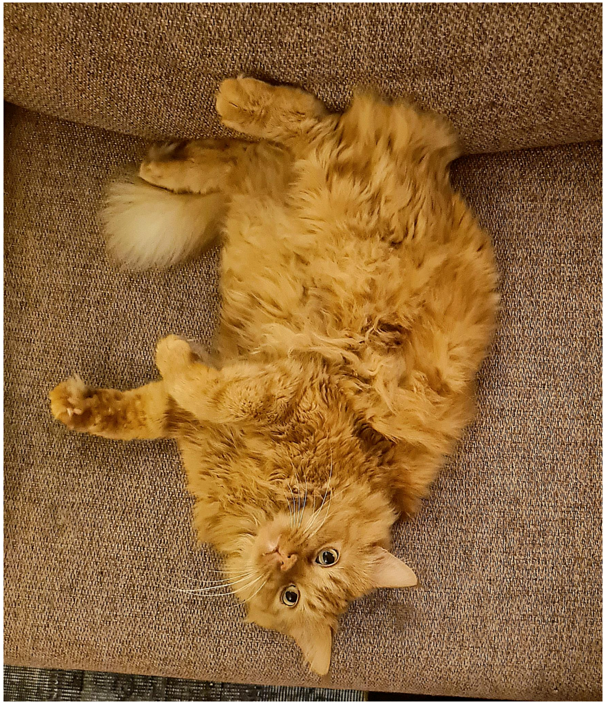


- **Modell:** $y \sim \text{Normal}(\mu, \sigma^2)$

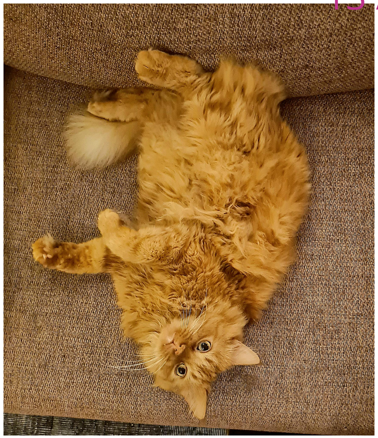
Bayesian Statistics

An Example

- This is Frank
- Frank enjoys eating, but he is somewhat picky
- We want to model the likelihood that Frank eats his food.



An Example



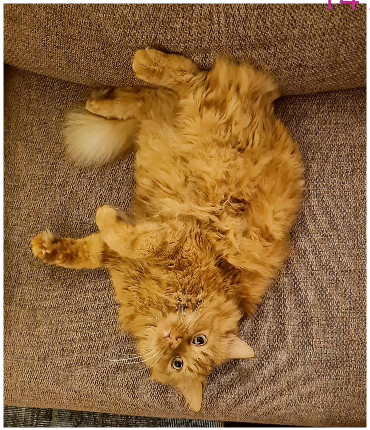
The Study

- For 20 days (i.e., 40 meals), we observe whether Frank eats his food
- Result: x of 40 meals were eaten.

The Statistical Model?

- $Y \sim \mathrm{Binomial}(N, \theta)$,
- $0 \leq \theta \leq 1, N = 40$

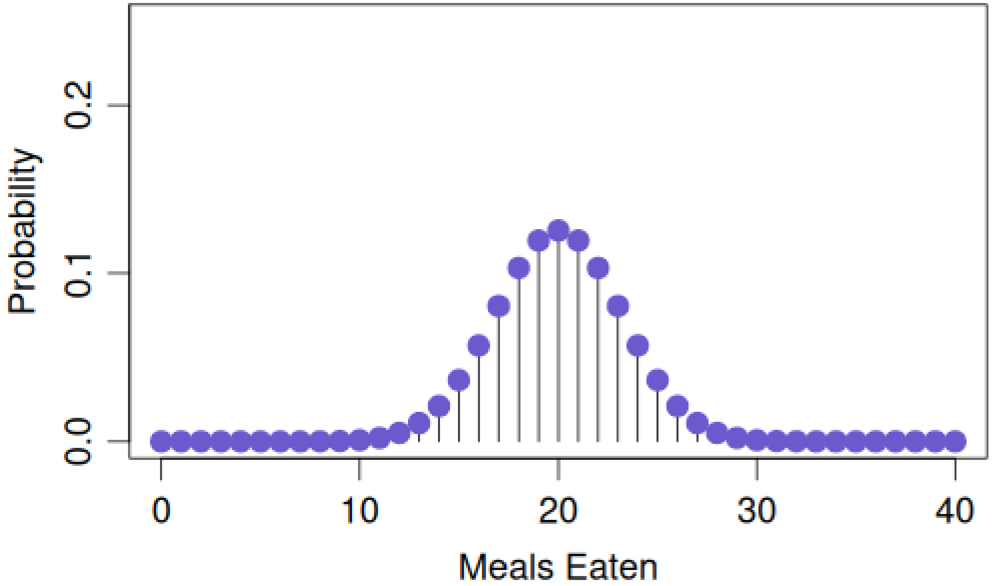
An Example



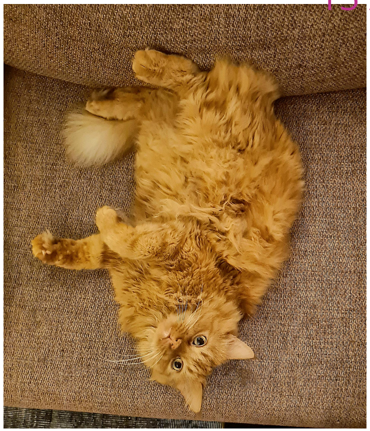
What is θ?

Before Knowing the Data

θ could be 0.5.



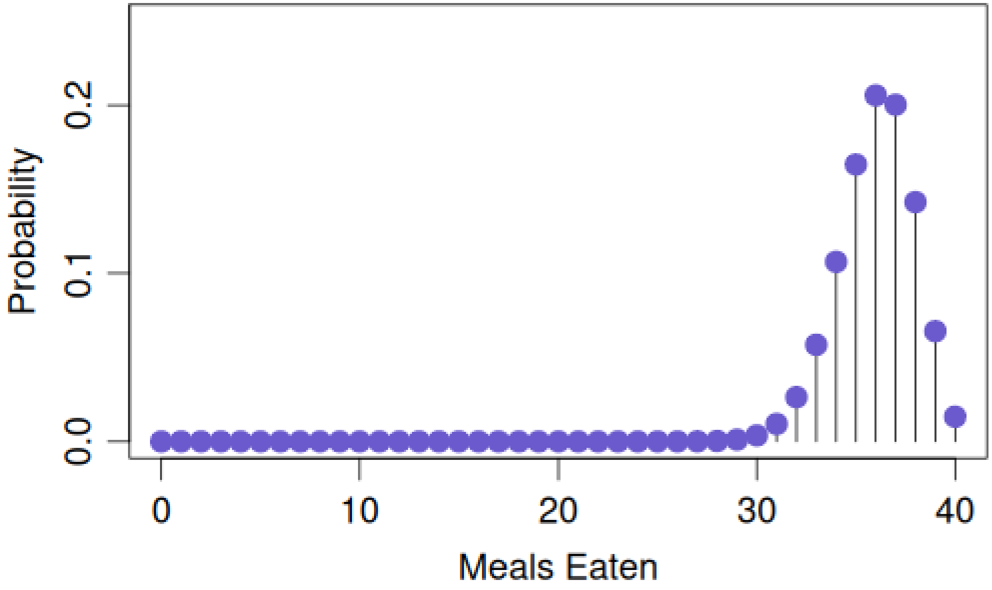
An Example



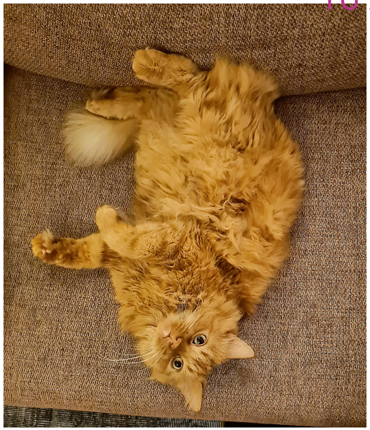
What is θ?

Before Knowing the Data

θ could be 0.9.



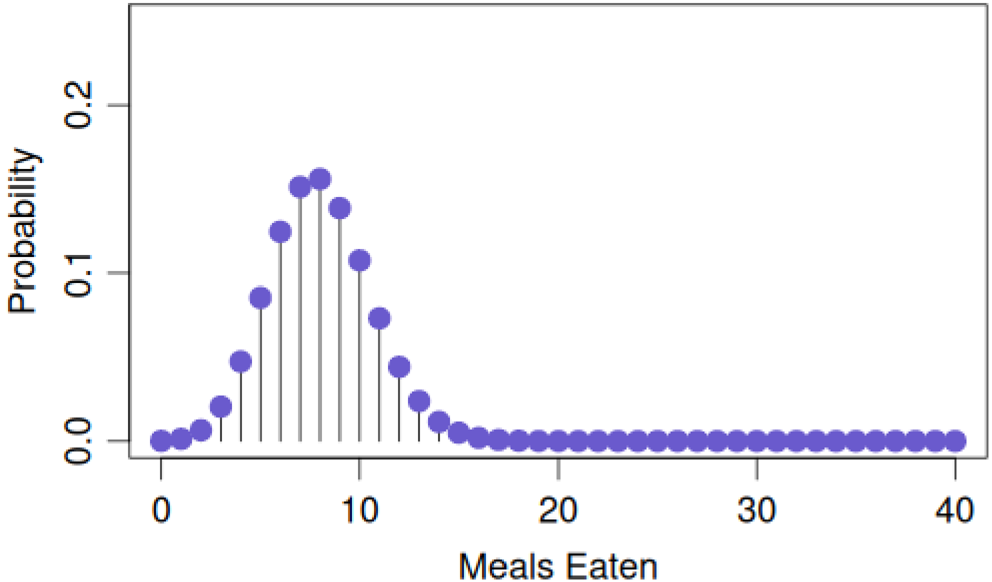
An Example



What is θ?

Before Knowing the Data

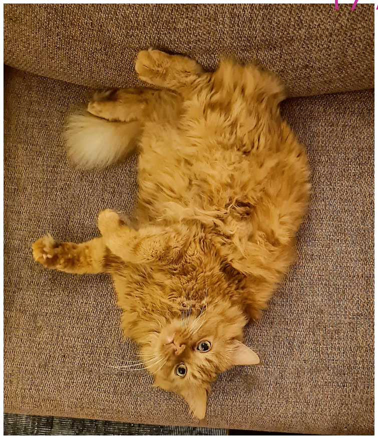
θ could be 0.2.

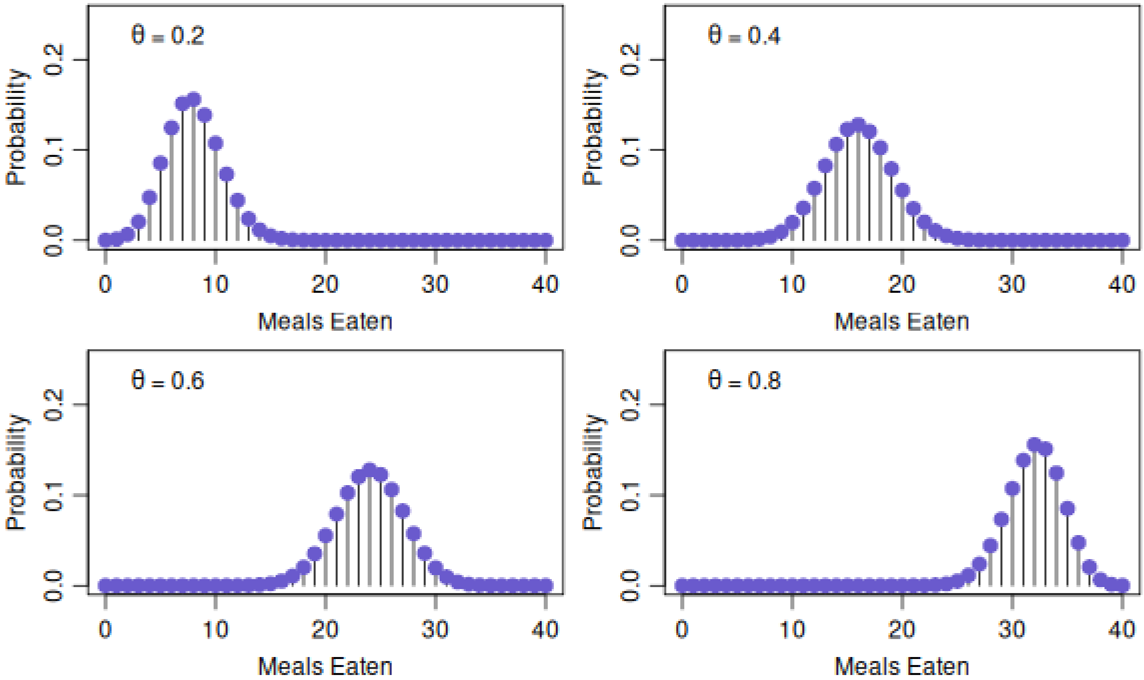


An Example

What is θ?

Before Knowing the Data





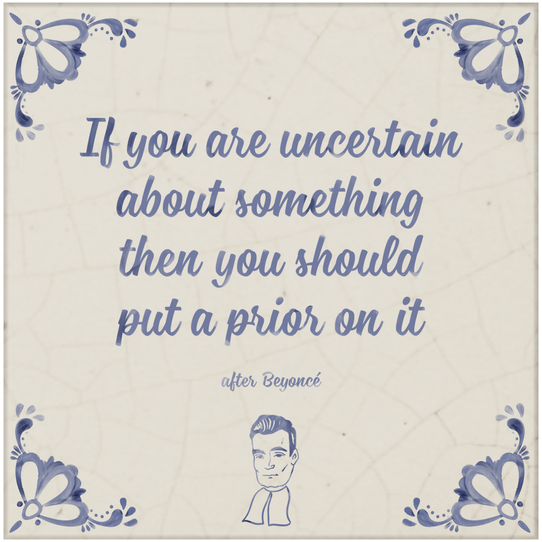
Parameters

- In Bayesian Statistics, parameters are also random variables (just like data).
- This means that parameters in a statistical model also receive a probability distribution.
- The probability distribution of the parameter changes when we observe data.
- Before we observe data, the probability distribution of a parameter is called the *prior.*

Prior

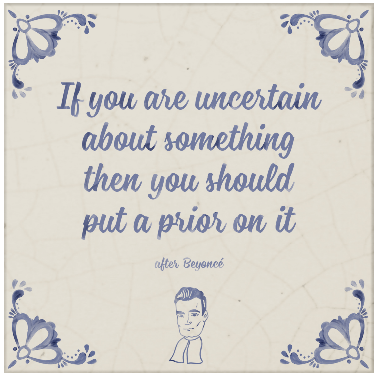
Prior

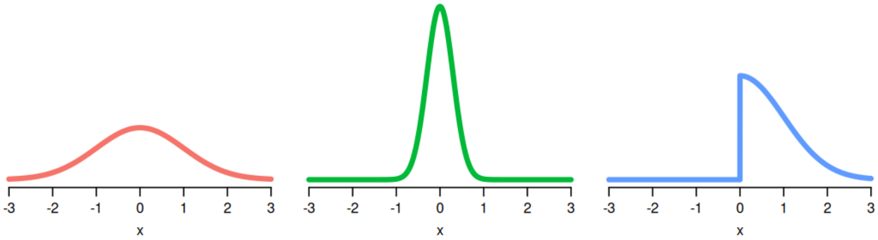
Definition: The prior distribution is an essential component of Bayesian inference. It represents the information about an uncertain parameter. This information is based on prior knowledge about plausible values of the parameter.



Prior

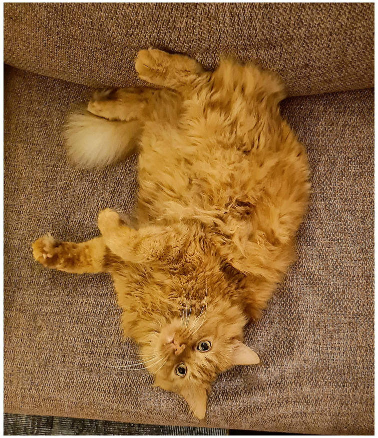
- Appropriate Distribution reflects the nature of the parameter
    - discrete vs. continuous
    - range: only positive values, only values between 0 and 1
- Knowledge and Uncertainty: The Prior reflects how much prior knowledge about a parameter is available.
    - The more prior knowledge, the more informed the distribution
    - i.e., the more possible parameter values are deemed implausible

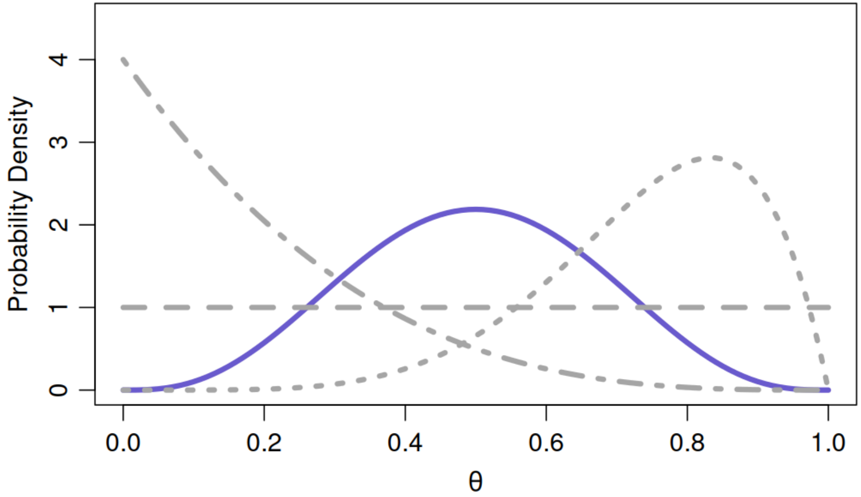




Prior

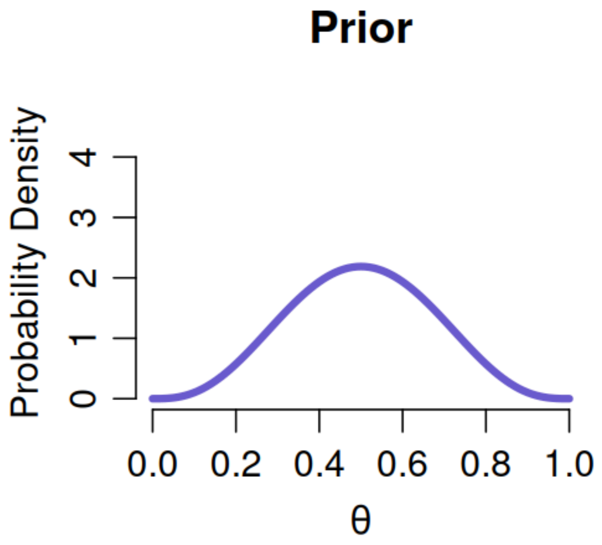
- $Y \sim \mathrm{Binomial}(N, \theta)$,
- $0 \leq \theta \leq 1, N = 40$
- $\theta \sim ???$
- Beta distribution: $\mathrm{Beta}(\alpha, \beta)$

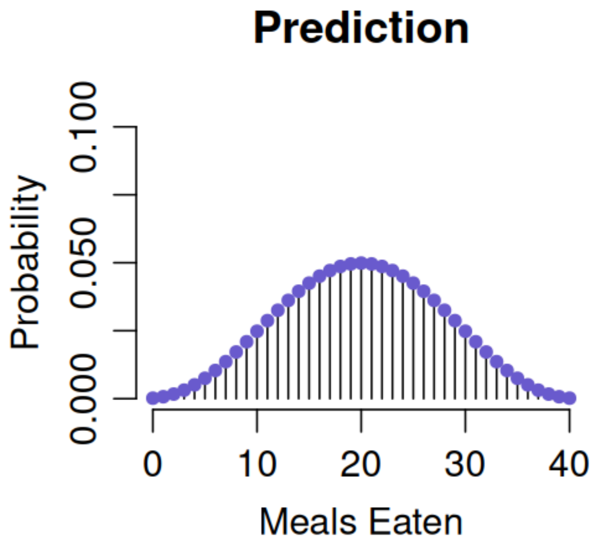




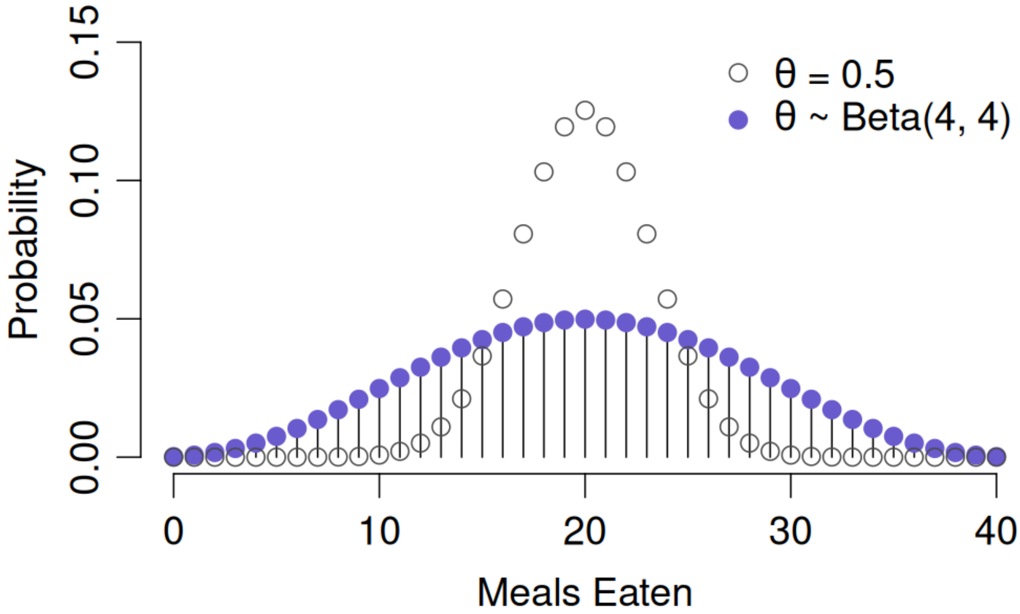
The Statistical Model

- $Y \sim \text{Binomial}(40, \theta)$,
- $\theta \sim \text{Beta}(4, 4)$



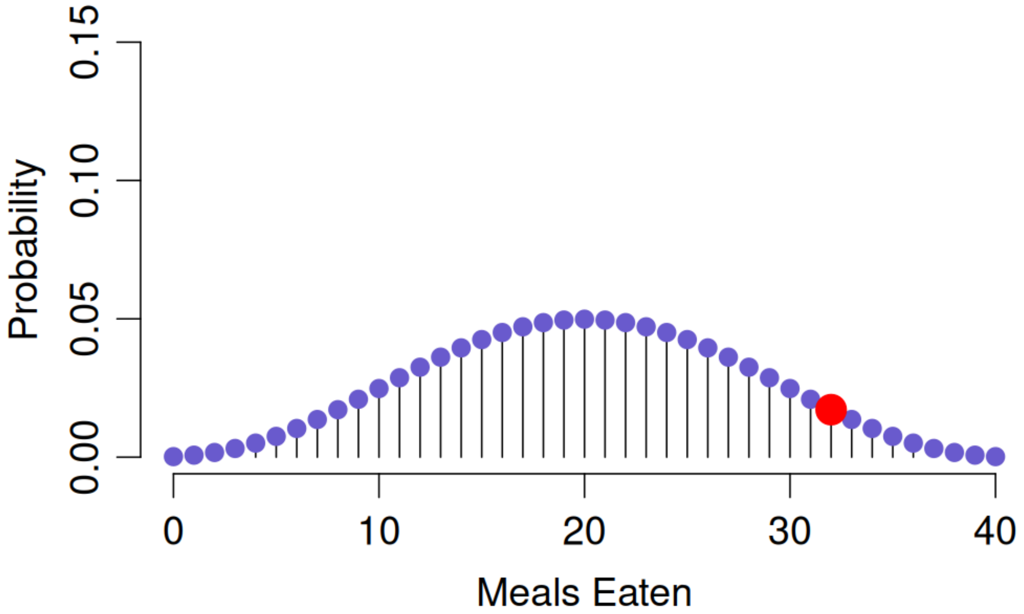


Prediction for Data



The prior distribution represents the uncertainty about the actual parameter value in the population.

Observation of Data

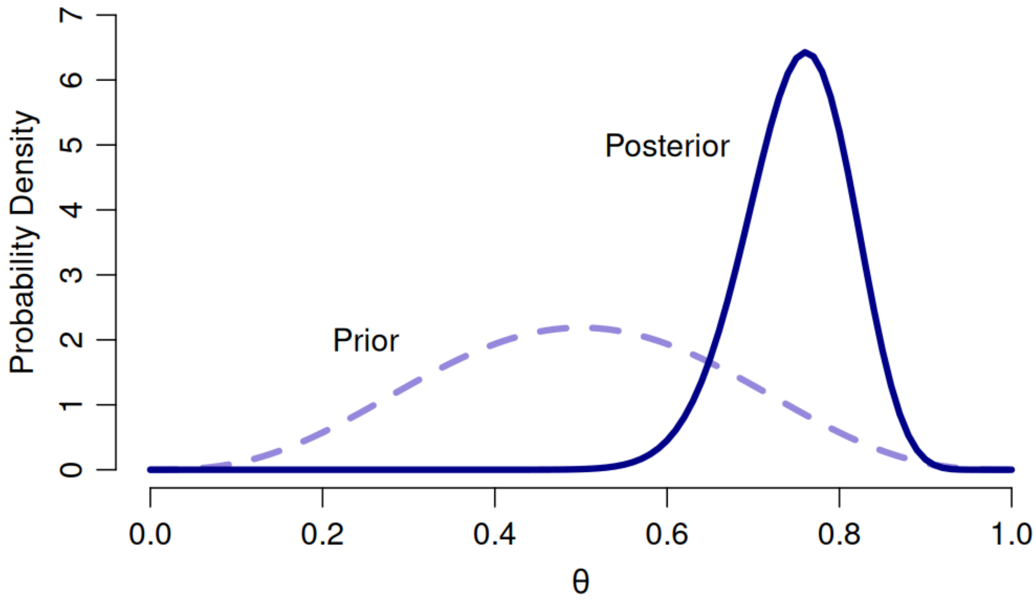


What happens when we observe data? $\rightarrow x = \mathbf{32}$

From Prior to Posterior

From Prior to Posterior

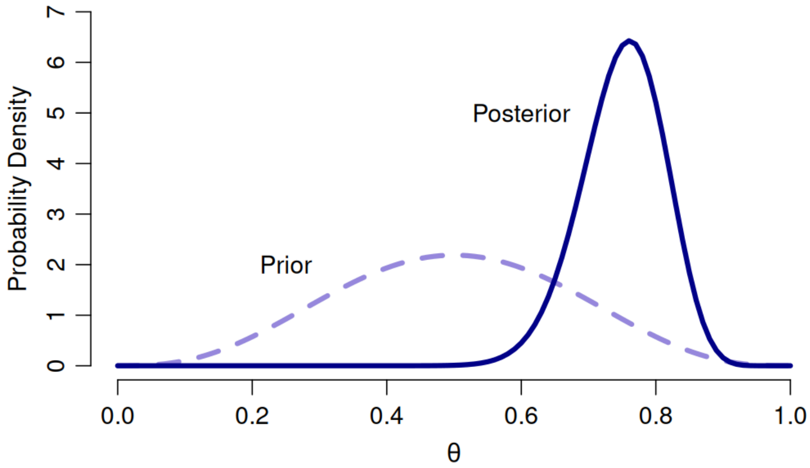
- When we observe data, we can update the distribution of the parameter
- The Prior distribution then becomes the Posterior distribution, where the new information from the data is integrated.



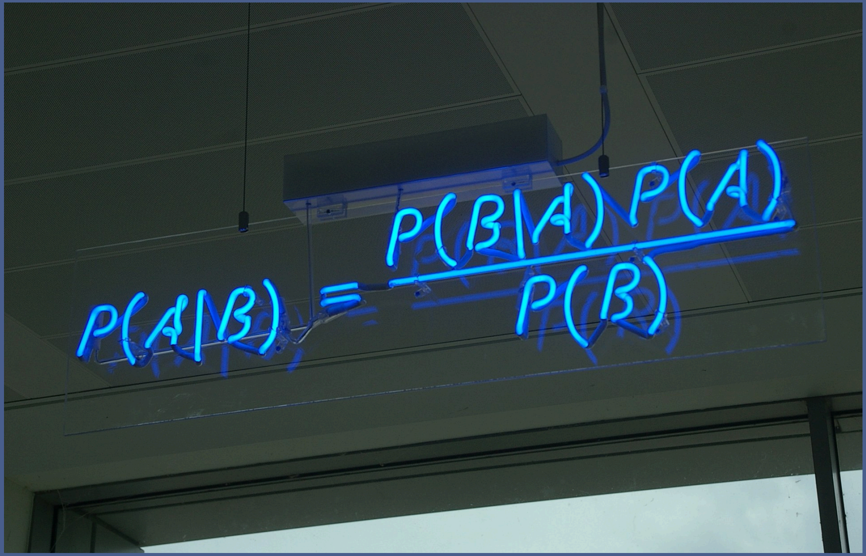
From Prior to Posterior

What changes?

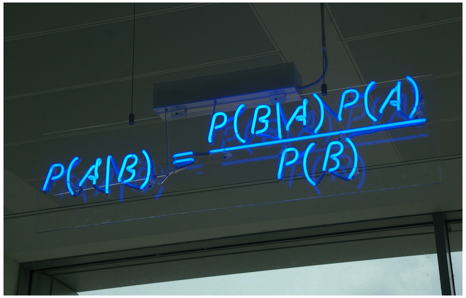
How do we arrive at the Posterior distribution?



Bayes's Theorem

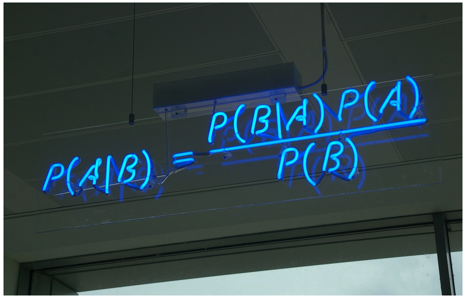


Bayes's Theorem



- The process by which the Prior distribution becomes the Posterior distribution is fundamental probability theory
- $P(\theta|x) = P(\theta)\frac{P(x|\theta)}{P(x)}$
- The Prior distribution is multiplied by another term to arrive at the Posterior distribution

Bayes's Theorem



$$P(\theta|x) = P(\theta)\frac{P(x|\theta)}{P(x)}$$

- $P(\theta|x)$: Posterior, after knowing the data
- $P(\theta)$: Prior, before knowing the data
- $P(x|\theta)$: Model of the Data, statistical model given the parameters
- $P(x)$: Probability of the Data, Prediction for the data across different parameter values.

The Posterior Distribution

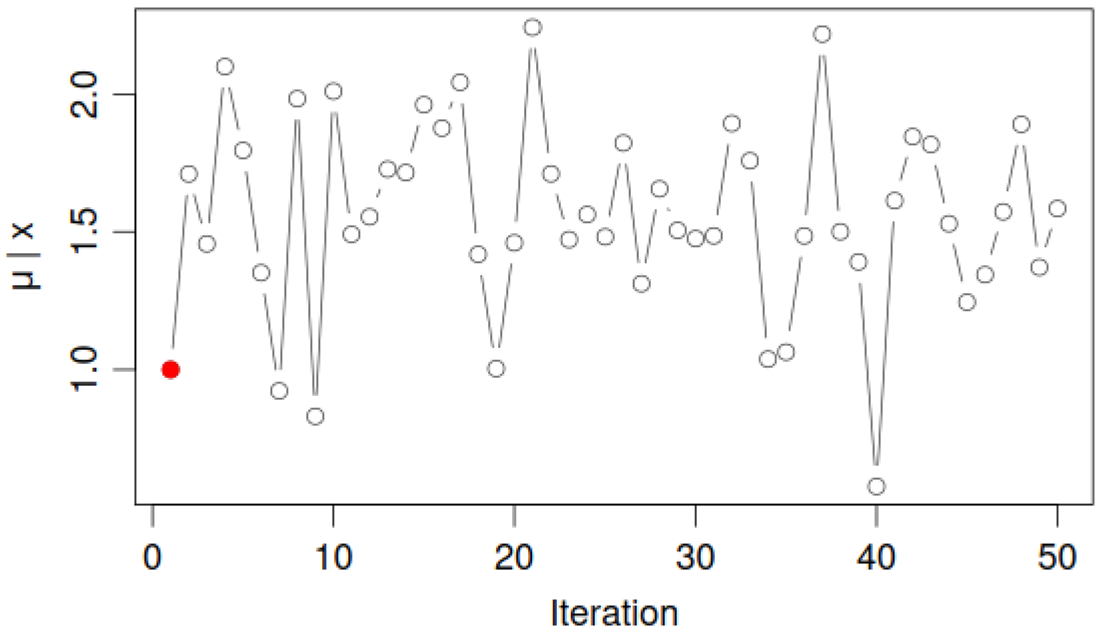
- The mathematical process to calculate the Posterior distribution is usually very complex
- An example: $P(x) = \int P(x|\theta)P(\theta)d\theta$
- Therefore, we often use estimation methods that estimate the Posterior distribution (e.g., in R) based on the Prior and Model of the Data

The Posterior Distribution

- Option 1: The Posterior distribution can be calculated mathematically
  - e.g., the Posterior distribution for the probability parameter that Frank eats his food
  - $\theta | x \sim \text{Beta}(4 + 32, 4 + 8)$
  - More generally: For the Binomial-Beta model, the Posterior is: $\theta | x \sim \text{Beta}(a + x, b + (N - x))$
- Option 2: The Posterior distribution can only be estimated
  - Estimation methods based on *Markov Chain Monte Carlo* estimators

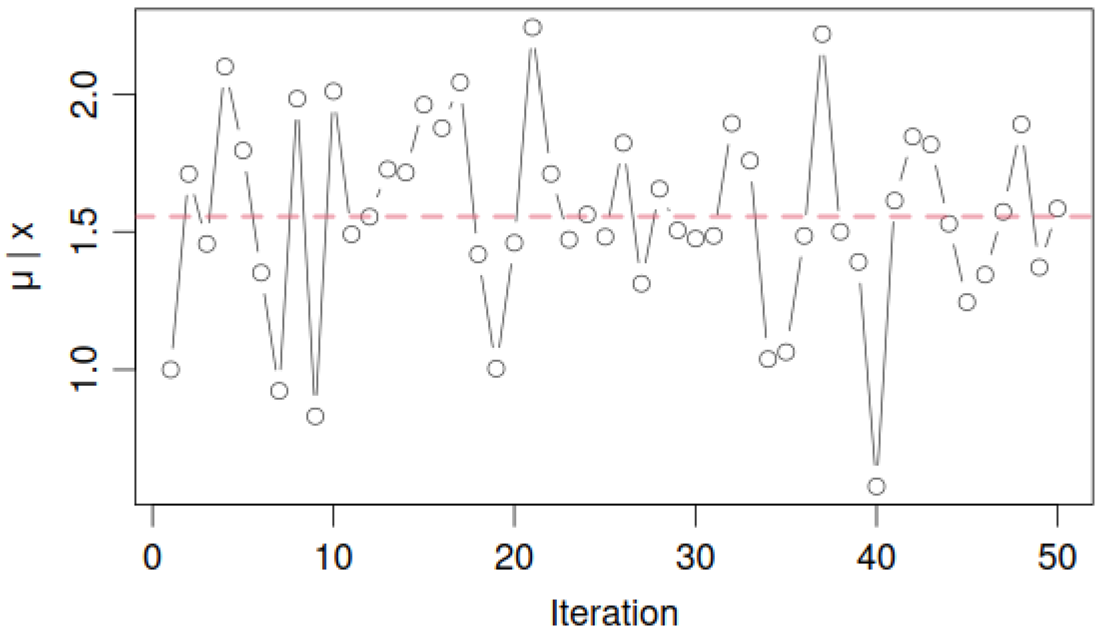
The Posterior Distribution

The MCMC process



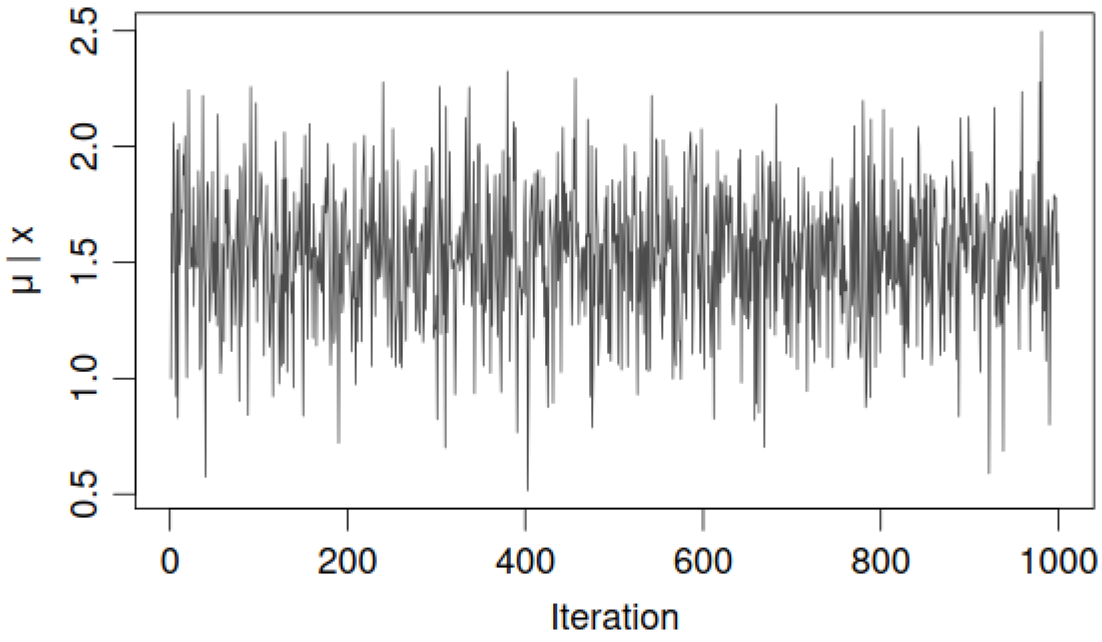
The Posterior Distribution

The MCMC process



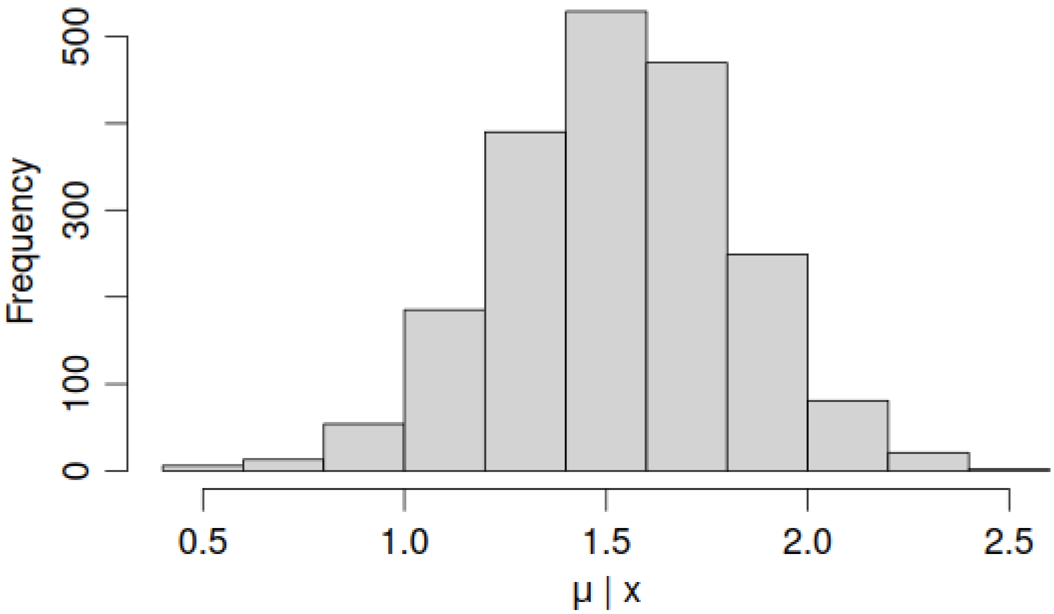
The Posterior Distribution

The MCMC process

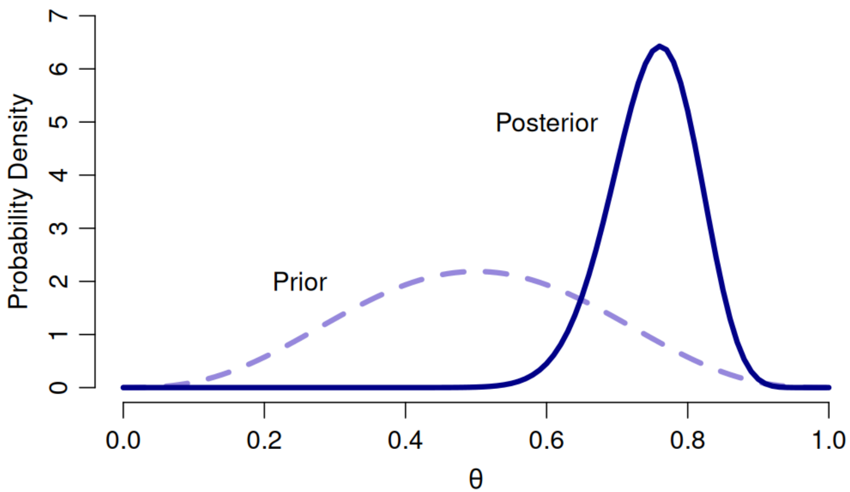


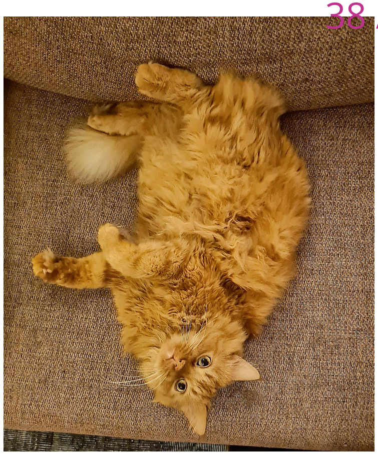
The Posterior Distribution

The Result



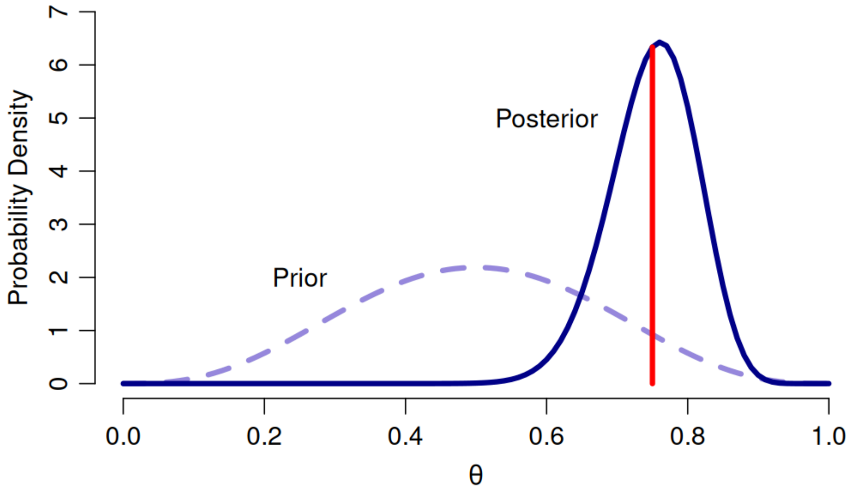
Back to Frank

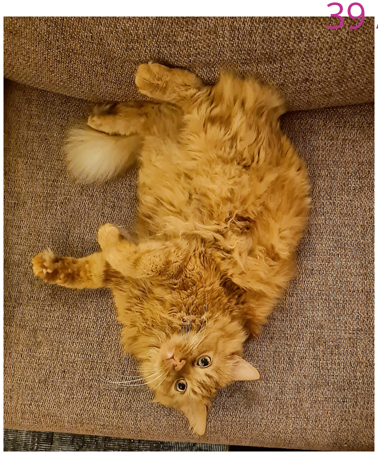




- What can we now learn about the probability that Frank eats his food?

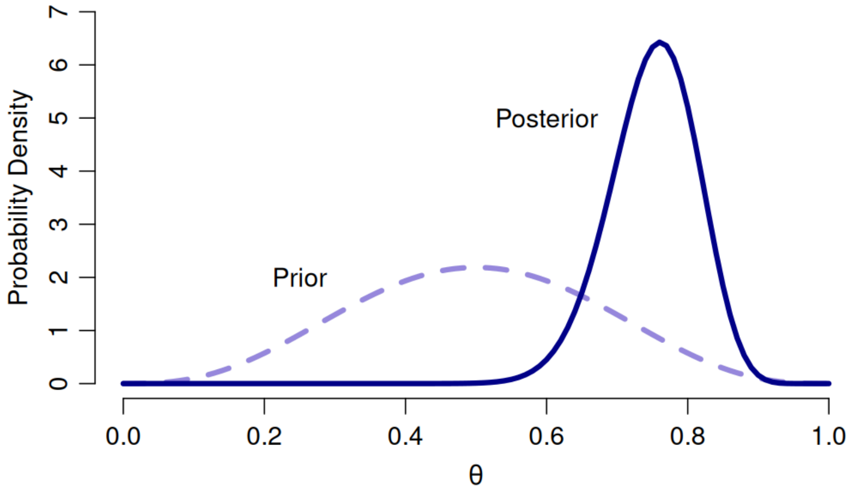
Back to Frank

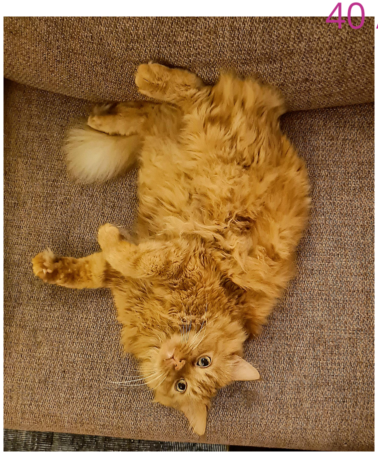




- What is the mean? $\hat{\theta} = 0.75$

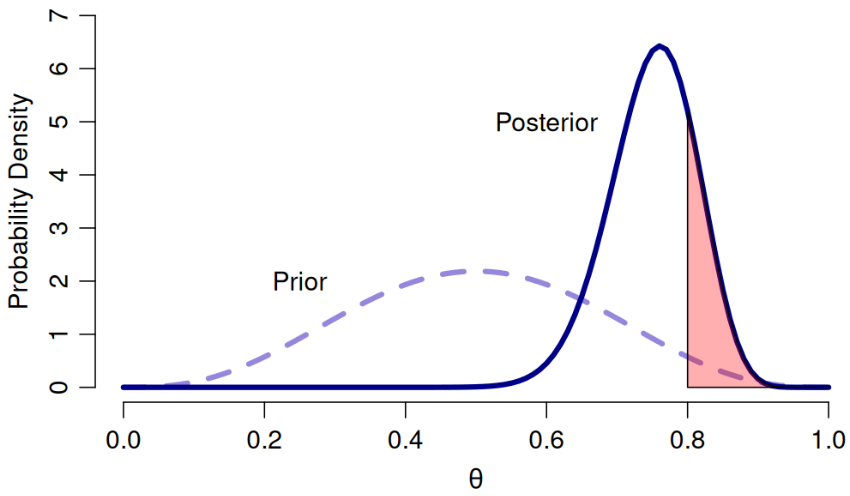
Back to Frank

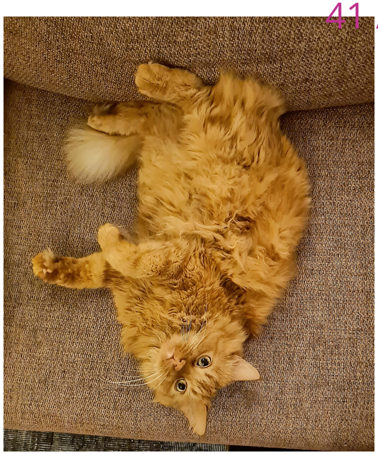




- What is the probability that Frank eats more than 80% of the meals?

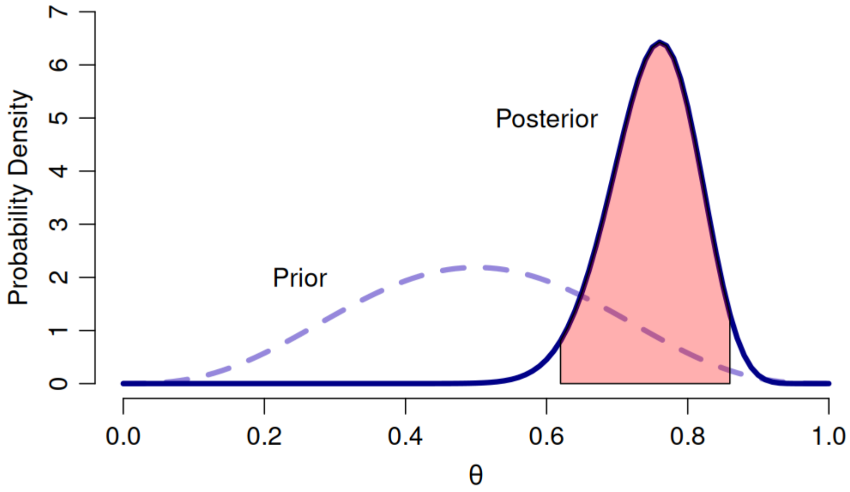
Back to Frank

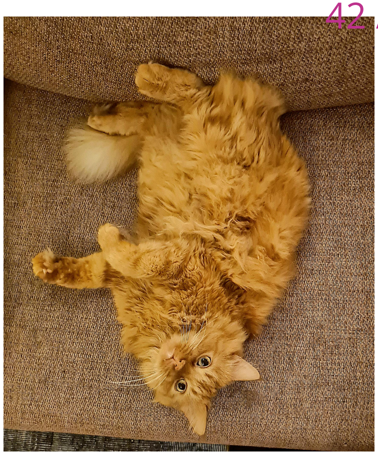




- What is the probability that Frank eats more than 80% of the meals?
  - $P(\theta > 0.8|x) = 0.2$

Back to Frank





- Which parameter values can we deem implausible?
- Credible Interval: 95% of the posterior distribution is between 0.62 and 0.86

Bayesian Parameter Estimation in R

Bayesian Parameter Estimation in R

In this course, we use the `R`-package `brms`.

- `brms` = Bayesian Regression Models using Stan
- Created by Paul-Christian Bürkner
- An interface to the Stan probabilistic programming language + sampling routines
- Allows fitting of a huge range of models using `R`'s formula syntax
- These slides aim to introduce the main functions

Bayesian Parameter Estimation in R

R formula syntax

term	meaning
`y ~ 1 + x`	`y` predicted by `x` (1 = intercept, included by default)
`y ~ 1 + x + z + x:z`	two variables and interaction (equivalent to `y ~ x*z`)
`(1 | g)`	allow the intercept to vary by levels of `g`
`(1 + x | g)`	allow intercept and slope to vary and model correlation between these 'random effects'
`(0 + x | g)`	fixed intercept, random slope
`(1 + x || g)`	model random effects but not correlation (fix ρ to 0)

Bayesian Parameter Estimation in R

`brm()`

```
brm(formula, data, family = gaussian(), prior = NULL,
  sample_prior = c("no", "yes", "only"),
  chains = 4, iter = 2000, warmup = floor(iter/2),
  cores = getOption("mc.cores", 1L))
```

see `?brm` for more

Bayesian Parameter Estimation in R

Binomial Model in `brms`

- $y_i = 0, 1, 1, 1, 0, \ldots$ for the ith meal
- $Y_i \sim \text{Bernoulli}(\pi)$,
  - where $\pi = \frac{\exp(\theta)}{\exp(\theta)+1}$ (inverse logit function).
  - Prior is placed on θ instead of π: $\theta \sim \text{Normal}(\mu_\theta, \sigma_\theta^2)$

Prior for θ?

$\theta = \log \frac{\pi}{1-\pi}$

(logit function)

```
p_sim <- rbeta(100000, 4, 4)
theta_sim <- log(p_sim / (1 - p_sim))
c(mean(theta_sim), sd(theta_sim))
```

```
## [1] -0.0002329545  0.7542324909
```

Bayesian Parameter Estimation in R

```r
library(brms)
frankdata <- data.frame(y = c(rep(1, 32), rep(0, 8)))
fit <- brm(data = frankdata
           , family = bernoulli(link = "logit")
           , y  ~ 0 + Intercept
           , prior = c(prior(normal(0, 0.75)
                             , coef = Intercept))
           , iter = 2000
           , warmup = 700)
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
##
## SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 4e-06 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition
would take 0.04 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
```

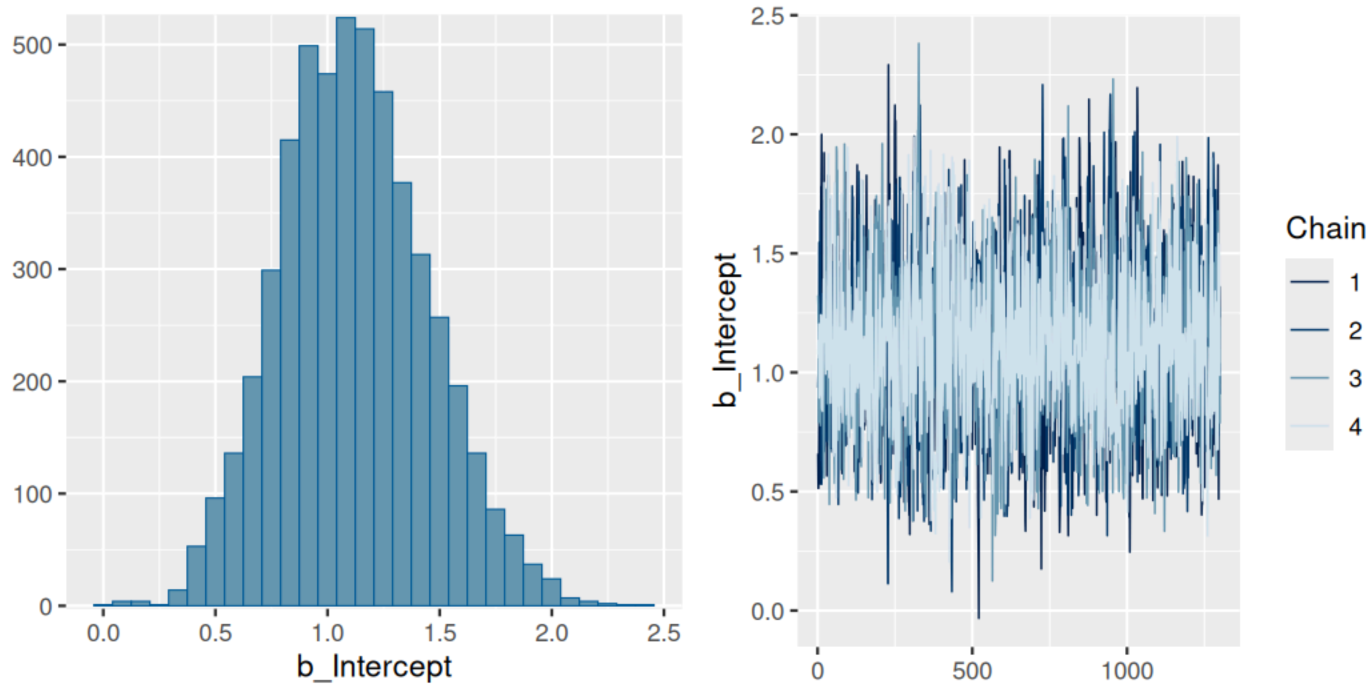
Bayesian Parameter Estimation in R

```
fit
```

```
##  Family: bernoulli
##   Links: mu = logit
## Formula: y ~ 0 + Intercept
##    Data: frankdata (Number of observations: 40)
##   Draws: 4 chains, each with iter = 2000; warmup = 700; thin = 1;
##          total post-warmup draws = 5200
##
## Regression Coefficients:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     1.12      0.33     0.50     1.80 1.00     1860     2335
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the
potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Bayesian Parameter Estimation in R

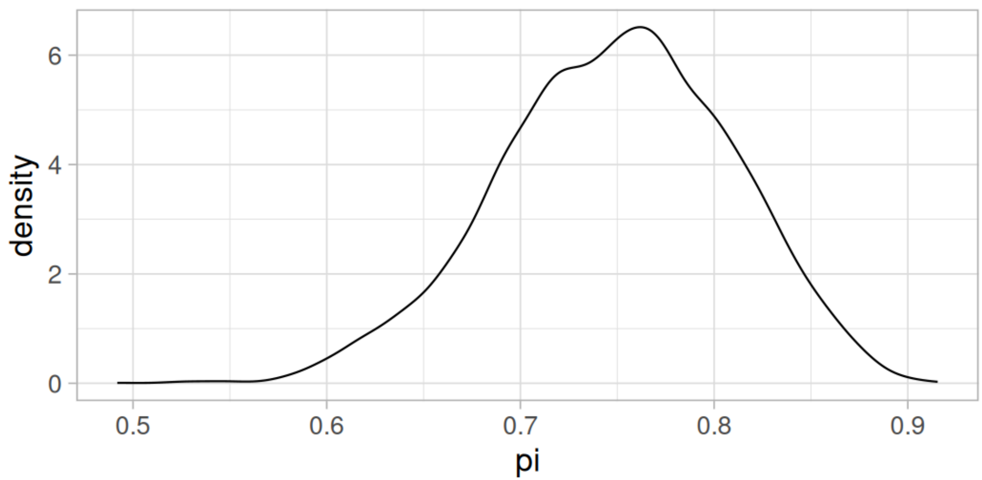
```
plot(fit)
```



Bayesian Parameter Estimation in R

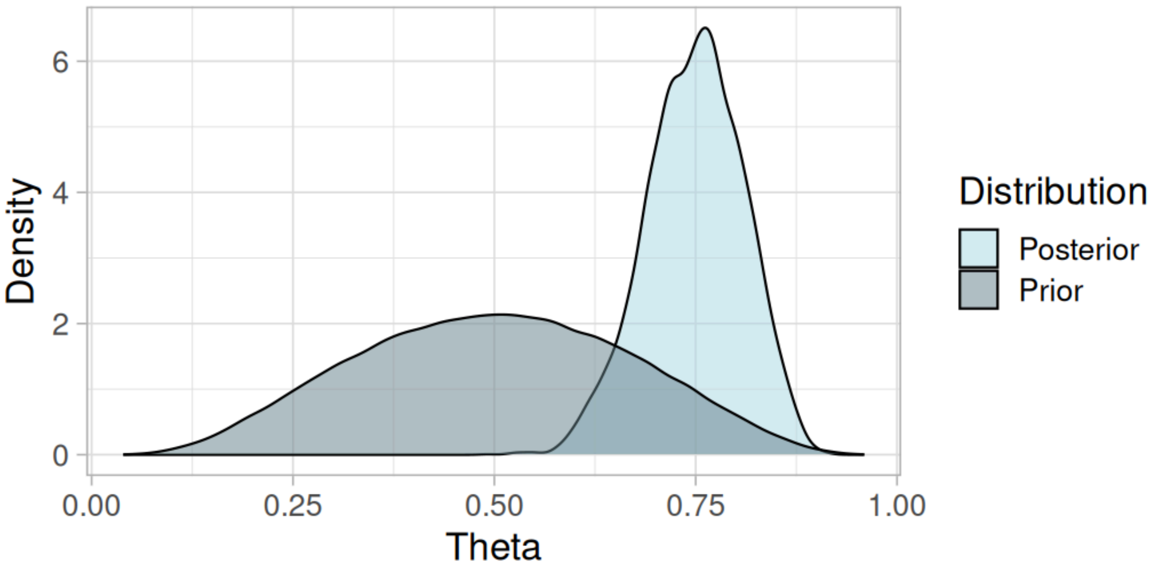
```
post <- as_draws_df(fit)
post %>%
  mutate(pi = exp(b_Intercept) / (1 + exp(b_Intercept))) -> post

ggplot(post, aes(pi)) +
  geom_density() +
  theme_light(base_size = 16)
```



Bayesian Parameter Estimation in R

How much did we learn?



Bayesian Parameter Estimation in R

```
library(ggplot2)

bayes_binomial <- function(successes, failures, prior_alpha,
prior_beta){
  # Parameter of the Posterior
  aprime <- prior_alpha + successes
  bprime <- prior_beta + failures

  # Estimator for theta
  schaetzer <- aprime / (aprime + bprime)
  ci <- qbeta(c(0.025, 0.975), aprime, bprime)

  # Plot
  cols <- hcl(h = seq(15, 375
                      , length = 3)
              , l = 65, c = 100)[1:2]
  p <- ggplot(data.frame(x = 1), aes(x = x)) +
    xlim(c(0, 1)) +
    stat_function(fun = dbeta
                  , args = list(prior_alpha, prior_beta)
                  , geom = "area", alpha = 0.35, aes(fill = 'Prior'))
+
    stat_function(fun = dbeta
                  , args = list(aprime, bprime)
```

Bayesian Parameter Estimation in R

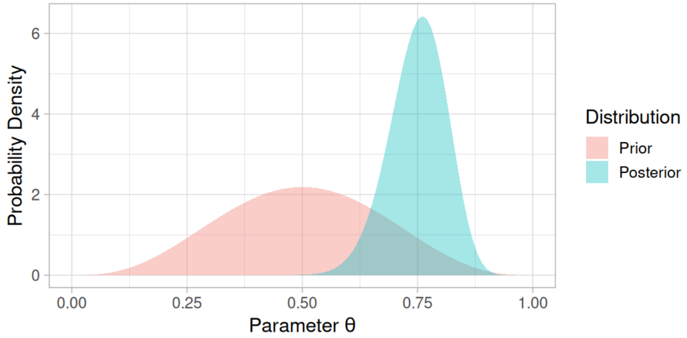
```r
prior1 <-
  bayes_binomial(successes = 32
                 , failures = 8
                 , prior_alpha = 4
                 , prior_beta = 4)

prior1$estimate; prior1$ci
```

```
## [1] 0.75
```

```
## [1] 0.6197427 0.8605509
```

```r
prior1$p
```



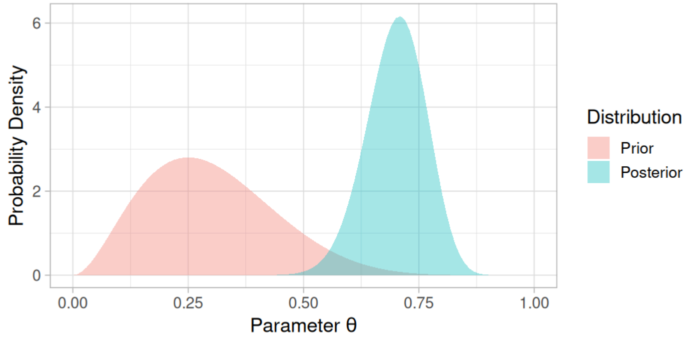

```r
prior2 <-
  bayes_binomial(successes = 32
                 , failures = 8
                 , prior_alpha = 3
                 , prior_beta = 7)

prior2$estimate; prior2$ci
```

```
## [1] 0.7
```

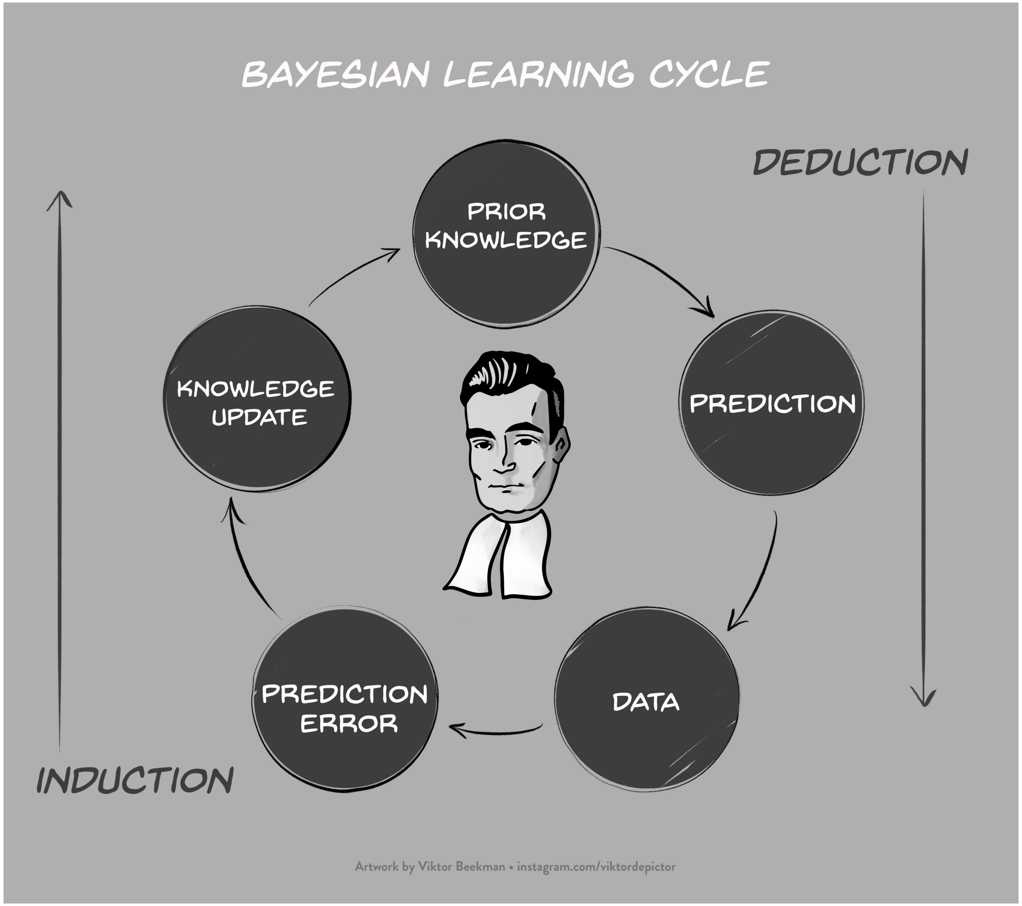
```
## [1] 0.5673703 0.8174806
```

```r
prior2$p
```



Wrap up

Wrap up


BAYESIAN LEARNING CYCLE
DEDUCTION
PRIOR KNOWLEDGE
PREDICTION
DATA
PREDICTION ERROR
KNOWLEDGE UPDATE
INDUCTION
Artwork by Viktor Beekman • instagram.com/viktordepictor

Assignments

Assignment 1

Literature

- Nicenboim, Schad, Vasishth, Introduction to Bayesian Data Analysis for Cognitive Science, Chapter 2
- Navarro, Chapter 17 Introduction and 17.1

:)