



What's Your Prior?

Part I: Introduction

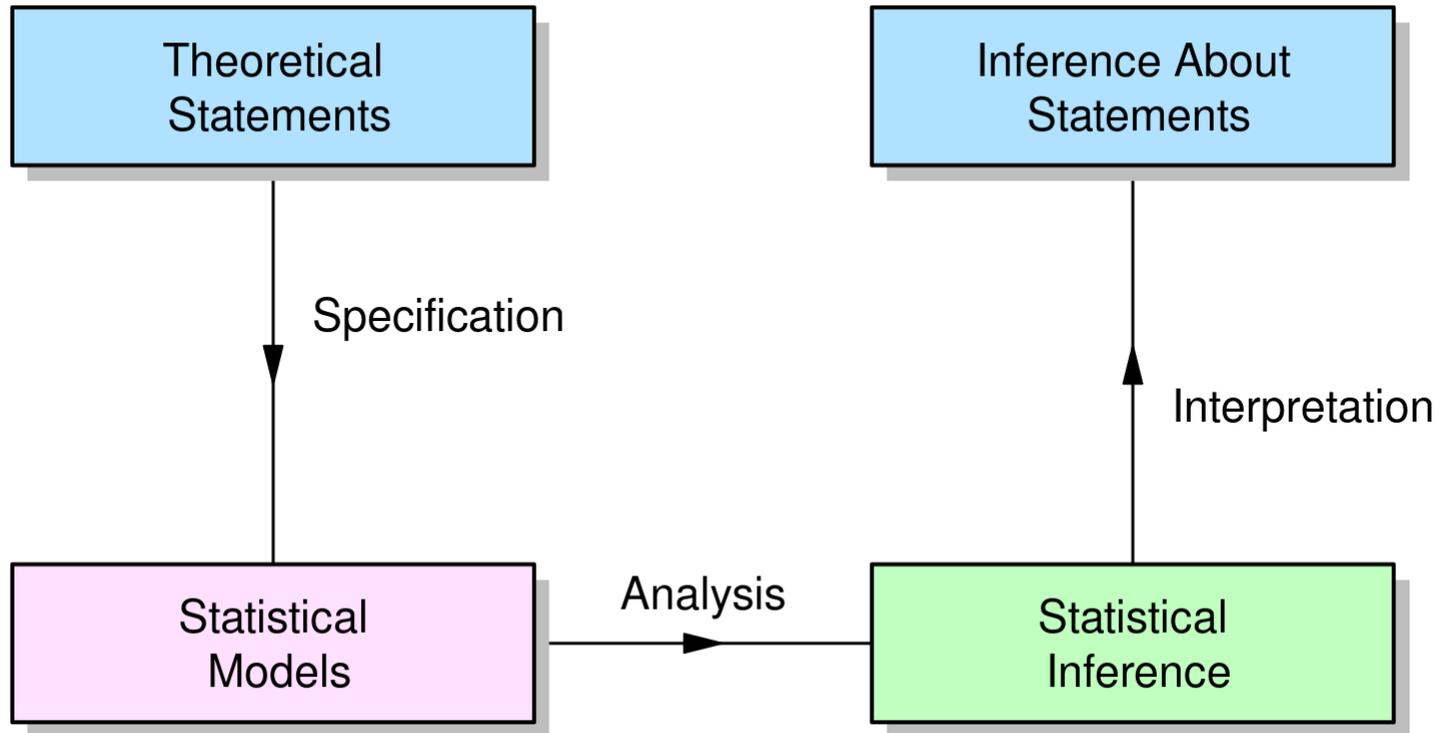
Julia Haaf

Outline

1. Posterior Updating (Or: Is my cat a picky eater?)
2. Priors for estimation
3. Priors for model comparison (Or: Is my cat fat?)
4. Prior prediction and prior sensitivity

Posterior Updating

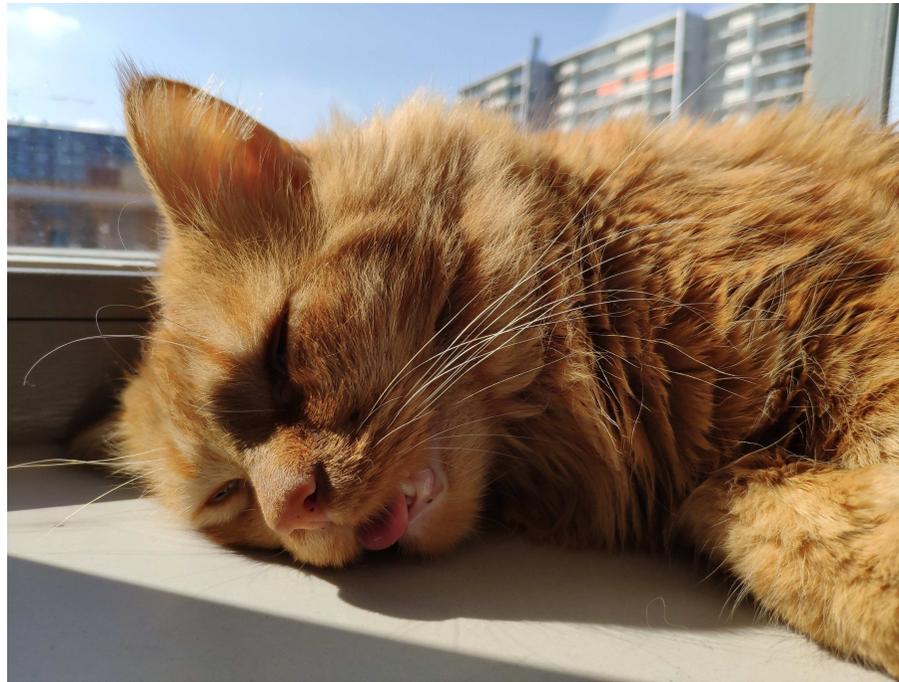
Overview of Modeling



An Example

- This is Frank.
- Frank likes to eat but he might be a tad picky.
- We may model how often Frank eats his food in a month:

$$Y \sim \text{Binomial}(\theta, 30)$$



An Example

Is Frank a picky eater?

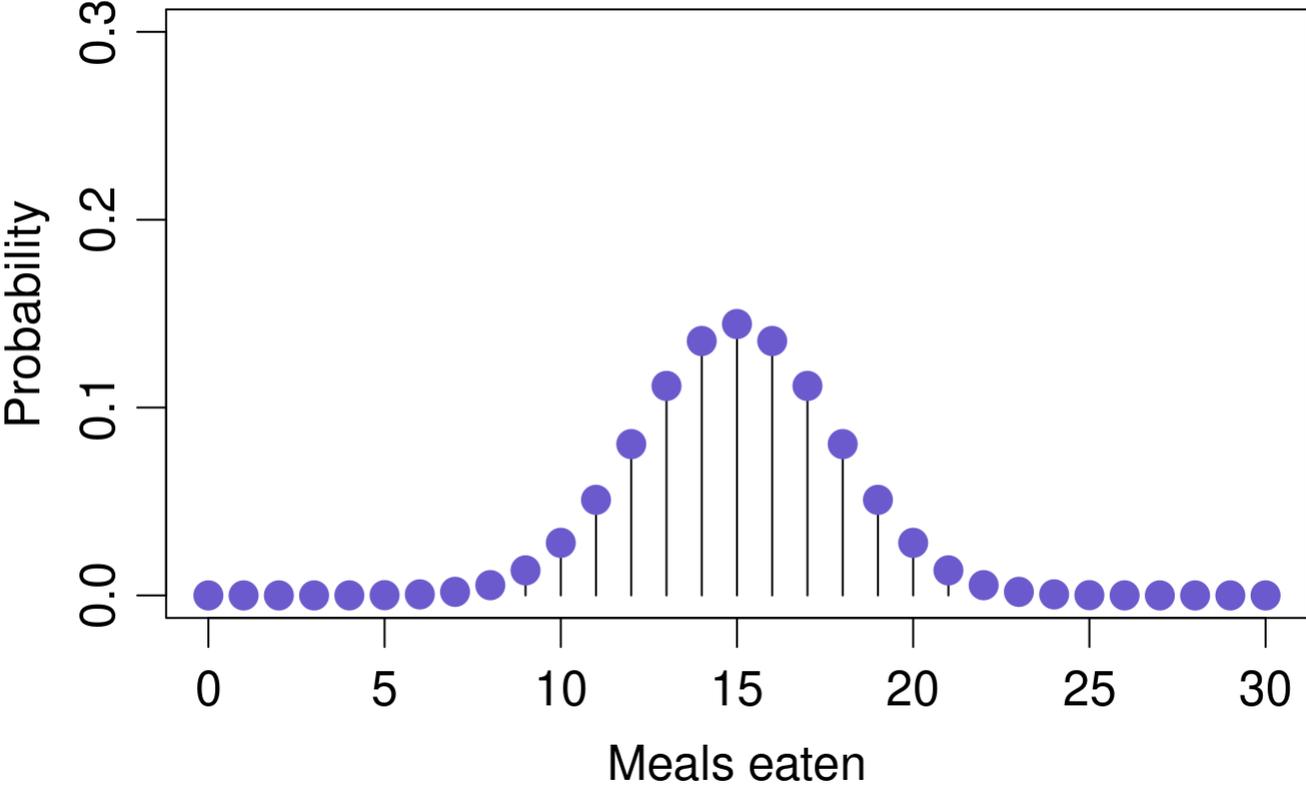
$Y \sim \text{Binomial}(\theta, 30),$

$\theta = .5.$



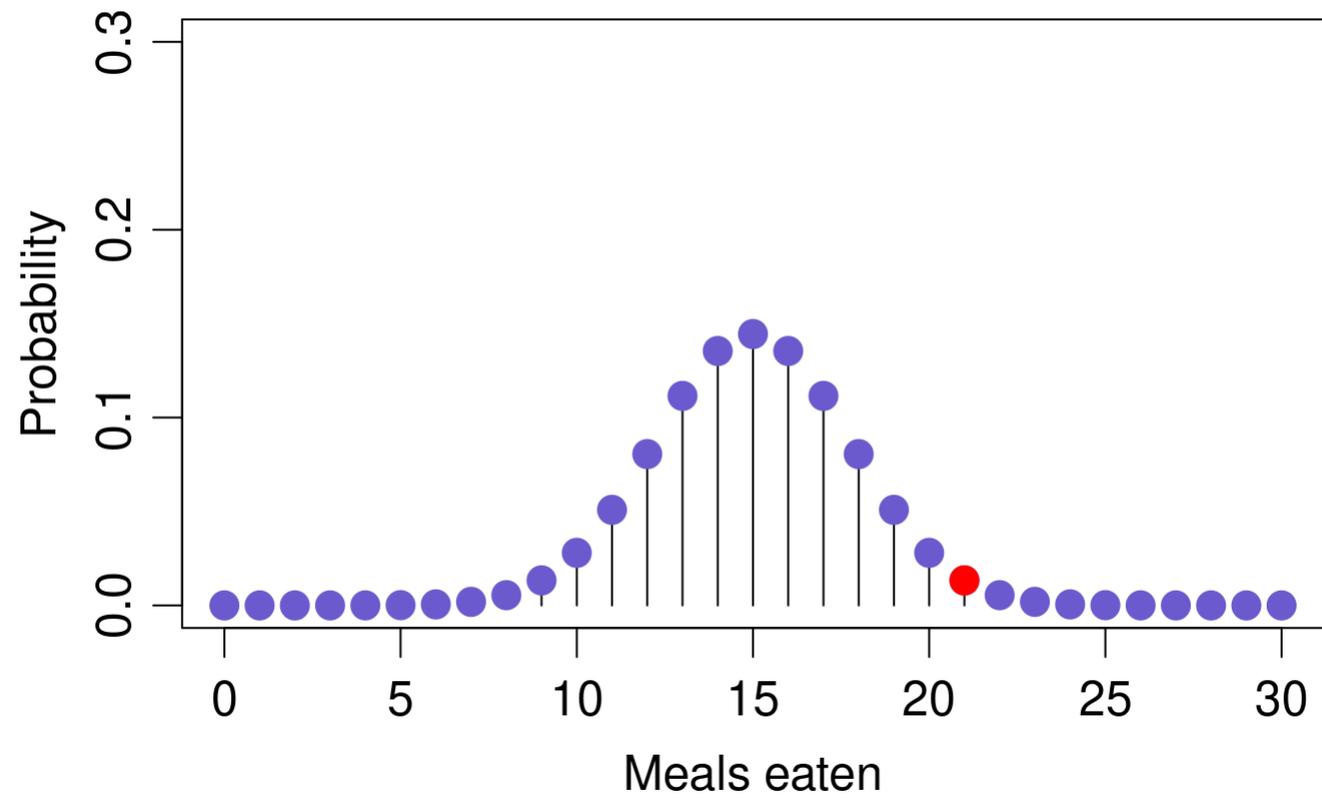
Models, an Example

$$Y \sim \text{Binomial}(\theta, 30), \quad \theta = .5.$$



Data

Let's say he ate 21 out of 30 meals. $Y = 21$.



Priors

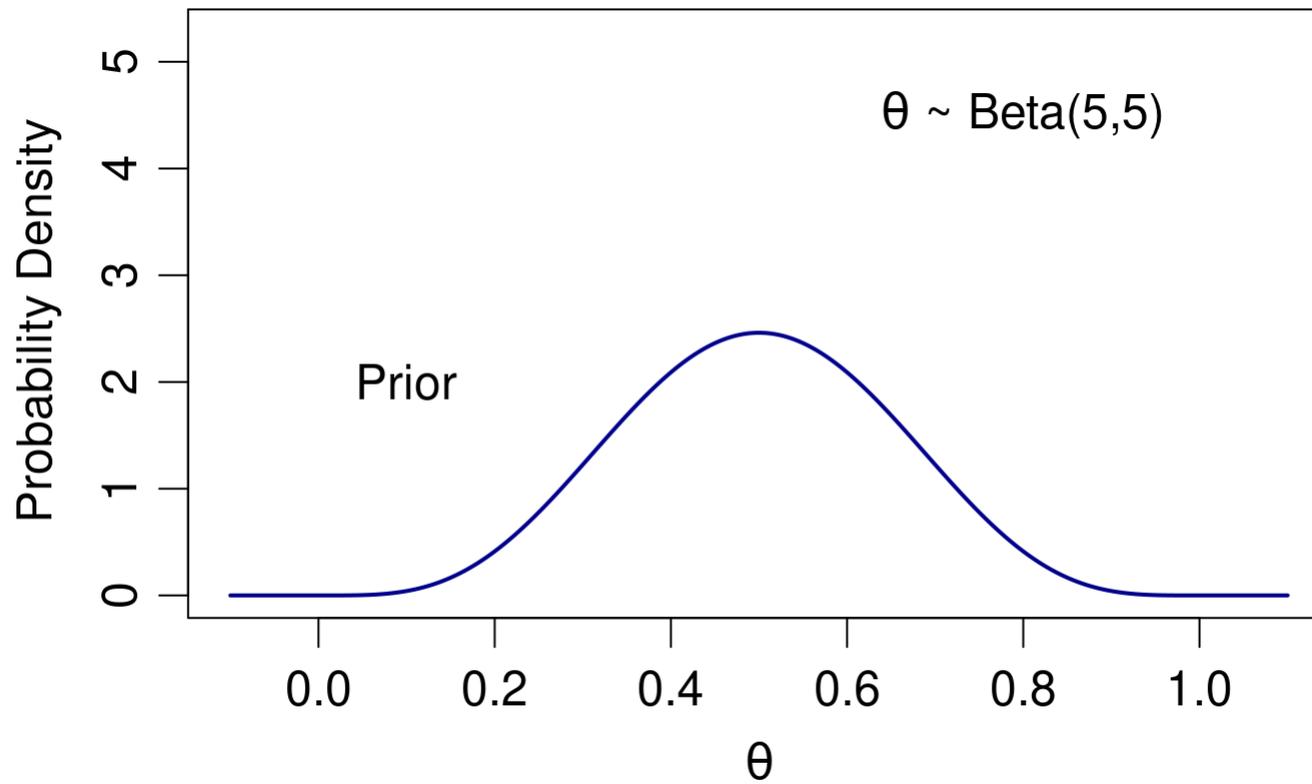
In Bayesian statistical analysis we typically would use a prior *distribution* for parameters.

$$Y|\theta \sim \text{Binomial}(\theta, N),$$
$$\theta \sim \text{Beta}(a, b).$$

If we assume Frank will most likely eat 5 out of 10 meals we may use $a = 5$ and $b = 5$.

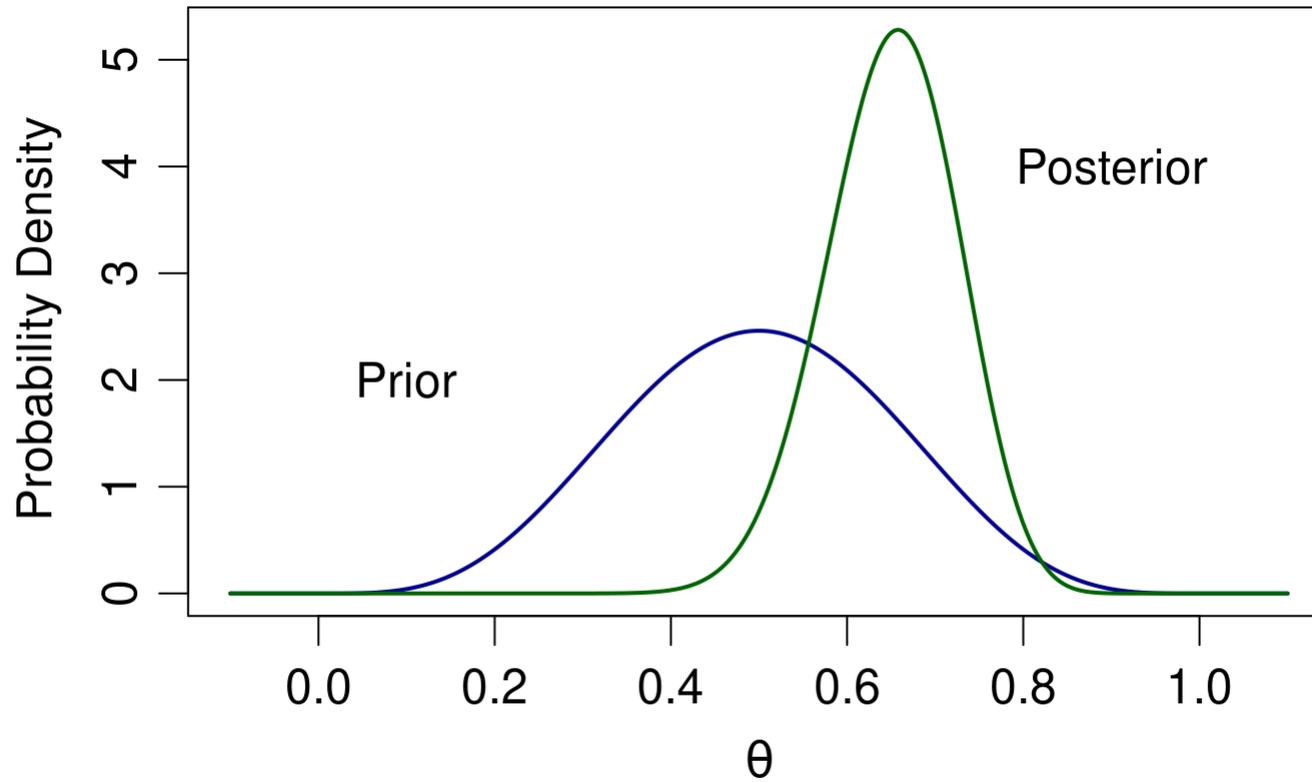
Posterior Updating

from $Pr(\theta)$...



Posterior Updating

from $Pr(\theta)\dots$ to $Pr(\theta|Y)$.



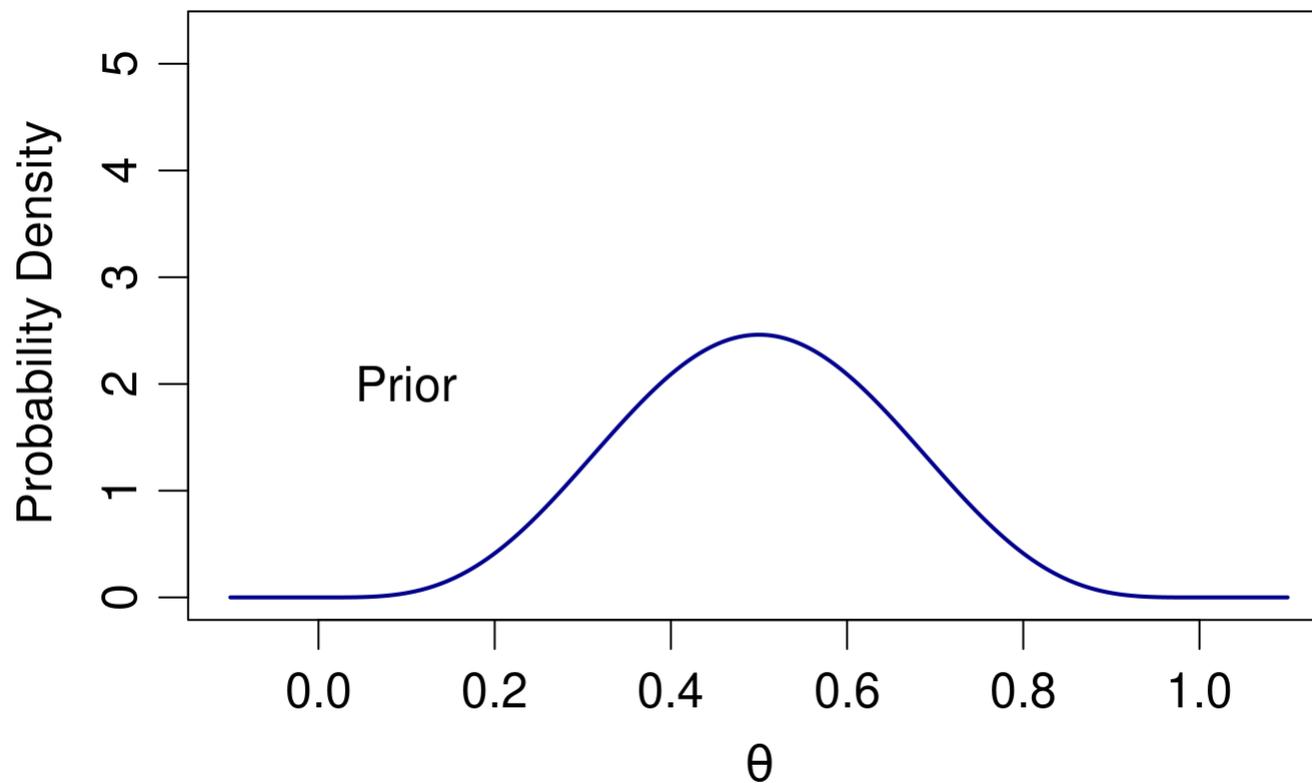
Bayes' Rule:

$$Pr(\theta|Y) = Pr(\theta) \frac{Pr(Y|\theta)}{Pr(Y)}$$

- $Pr(\theta|Y)$ is the *posterior distribution* of θ .
- $Pr(\theta)$ is the *prior distribution* of θ .
- $Pr(Y|\theta)$ is the likelihood of the data.
- $Pr(Y)$ is the marginal likelihood.

In search for a prior for Frank's eating habits.

What should it look like?



Opinions

- “Priors should represent your belief about the parameter.”
- “Priors should represent what experts in the field might expect.”
- “Priors should contain as little information as possible. Let the data speak.”
- “Priors should represent reasonable expectation about likely ranges of the parameter.”



Matching priors to goals of analysis

- There are priors that are most suitable for estimation.
- And there are priors most suitable for model comparison.
- And there are priors that are pretty good for both.
- Oh, and not everyone agrees on this classifications (or what “good means”).



Priors for Estimation

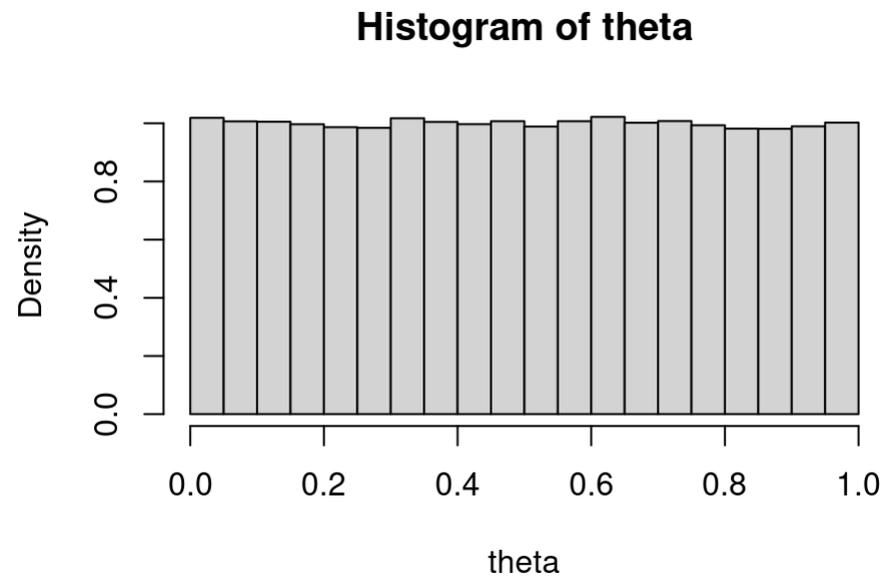
(Pretend) You Don't Know a Lot

- Flat priors.
- Jeffreys priors.
- Reference priors.
- (Default priors.)

Flat Priors

$$\theta \sim \text{Uniform}(0, 1).$$

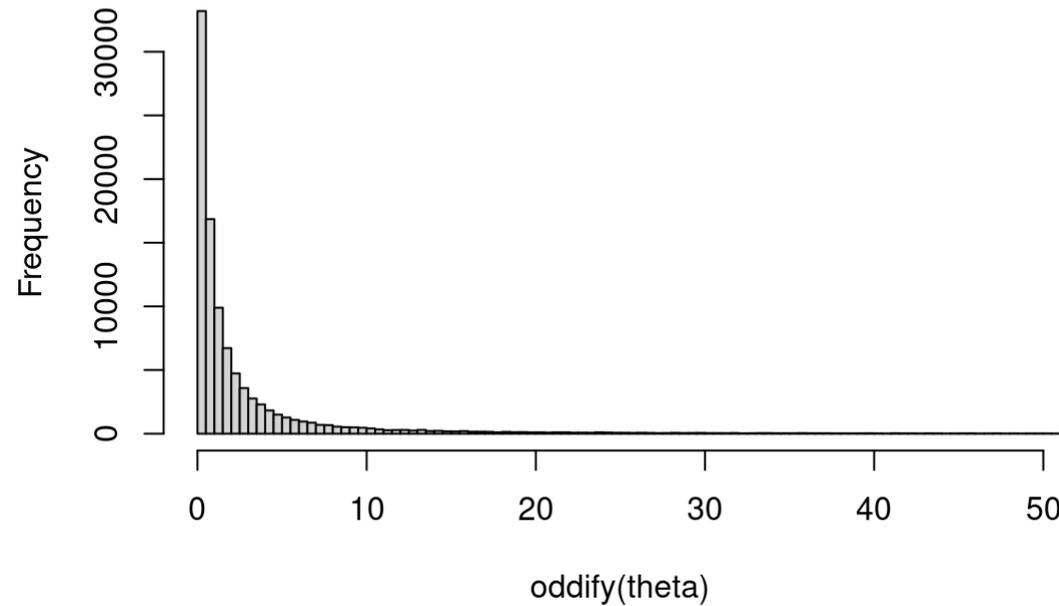
```
theta <- runif(100000)  
hist(theta, probability = T)
```



Non-informative = flat?

What are the odds of Frank eating the food, $OR = \frac{\theta}{1-\theta}$?

Histogram of oddify(theta)

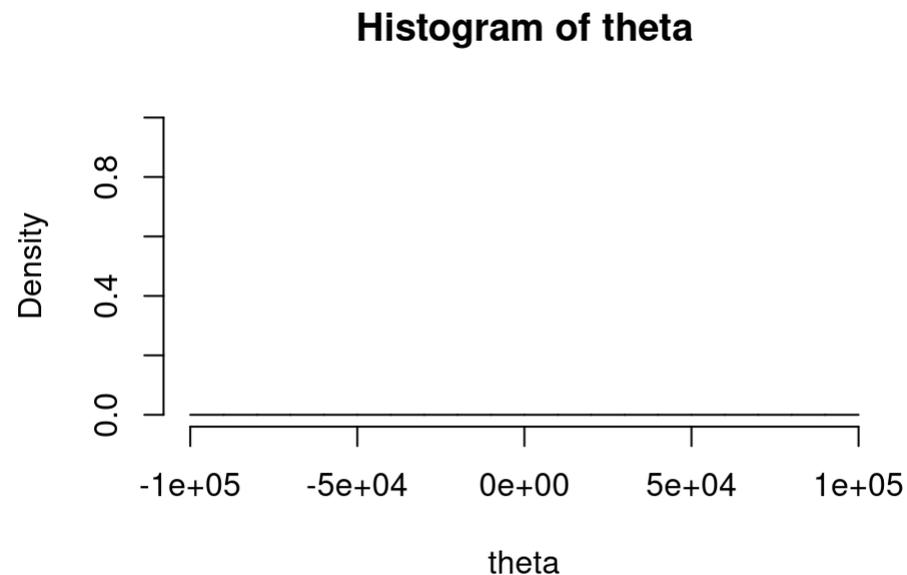


Flat Priors Can Be Improper

Suppose θ does not have constrained support (mean difference between two measures):

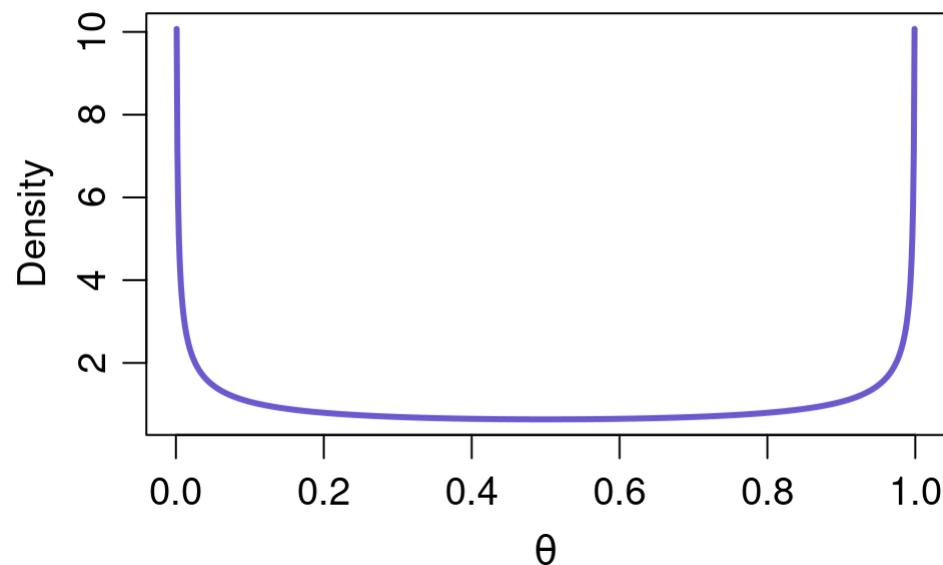
$$\theta \sim \text{Uniform}(-\infty, \infty).$$

```
theta <- runif(100000, min = -100000, max = 100000)
hist(theta, probability = T, ylim = c(0, 1))
```



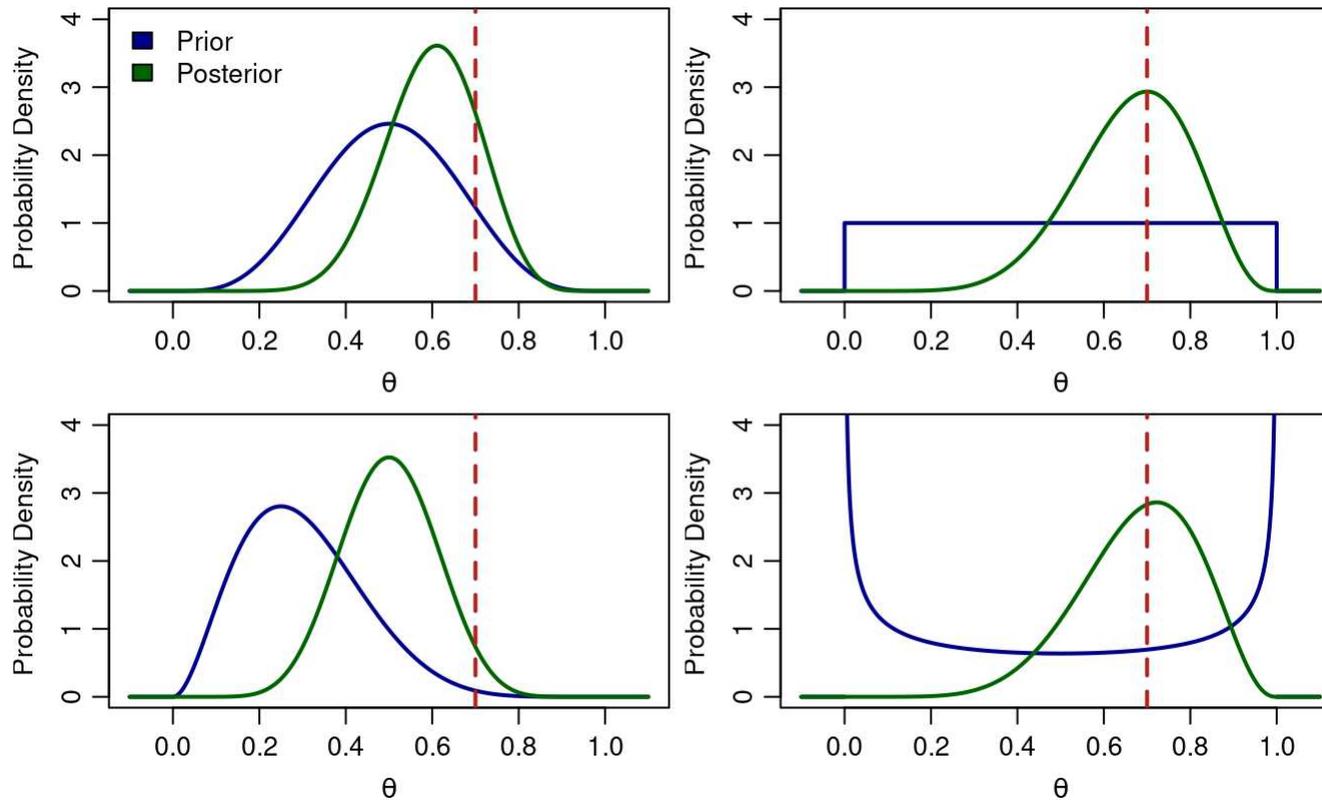
Jefferys Priors and Reference Priors

- Goal: Allow the data to have the maximum effect on the posterior estimates.
- Process: Maximize divergence from hypothetical data.
- The same for one-dimensional parameters.



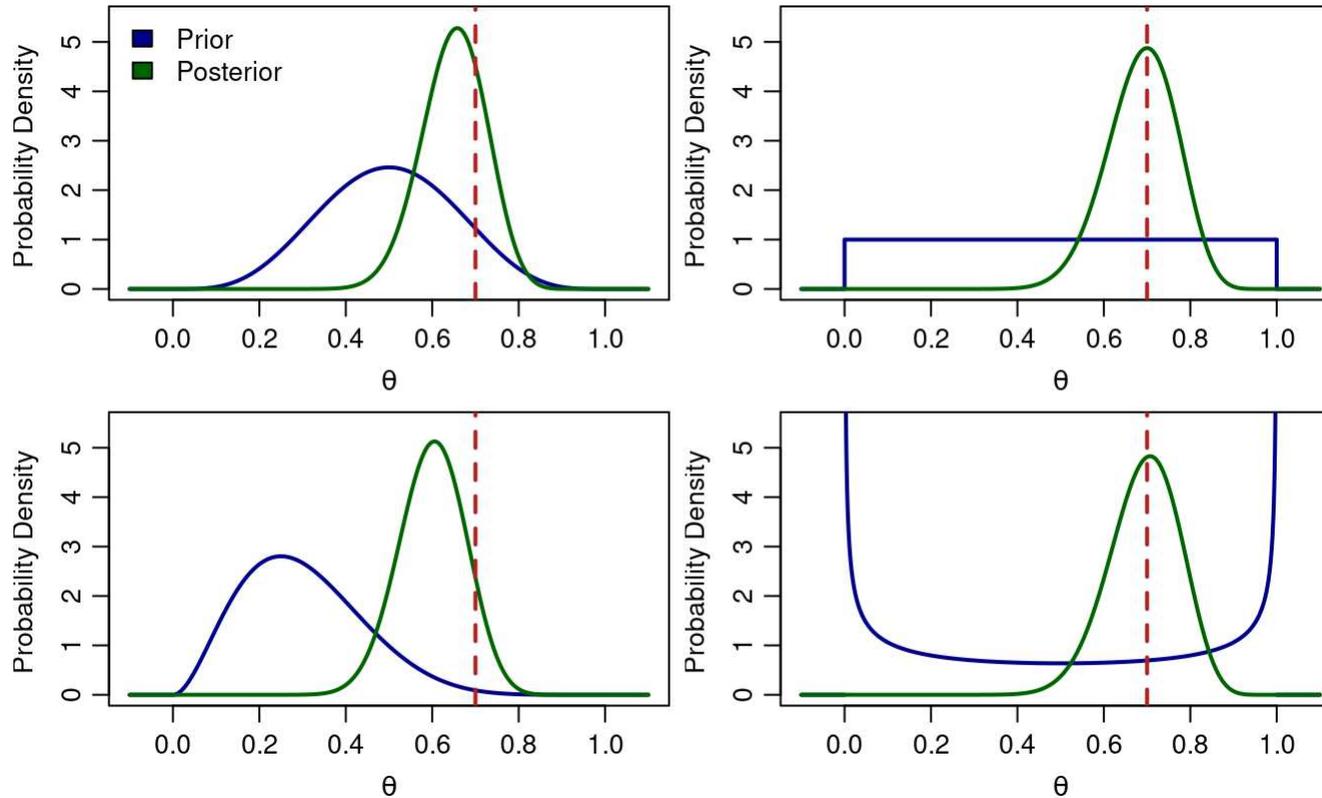
Why Priors for Estimation Really Don't Matter That Much

Frank eats his food 7 out of 10 times.



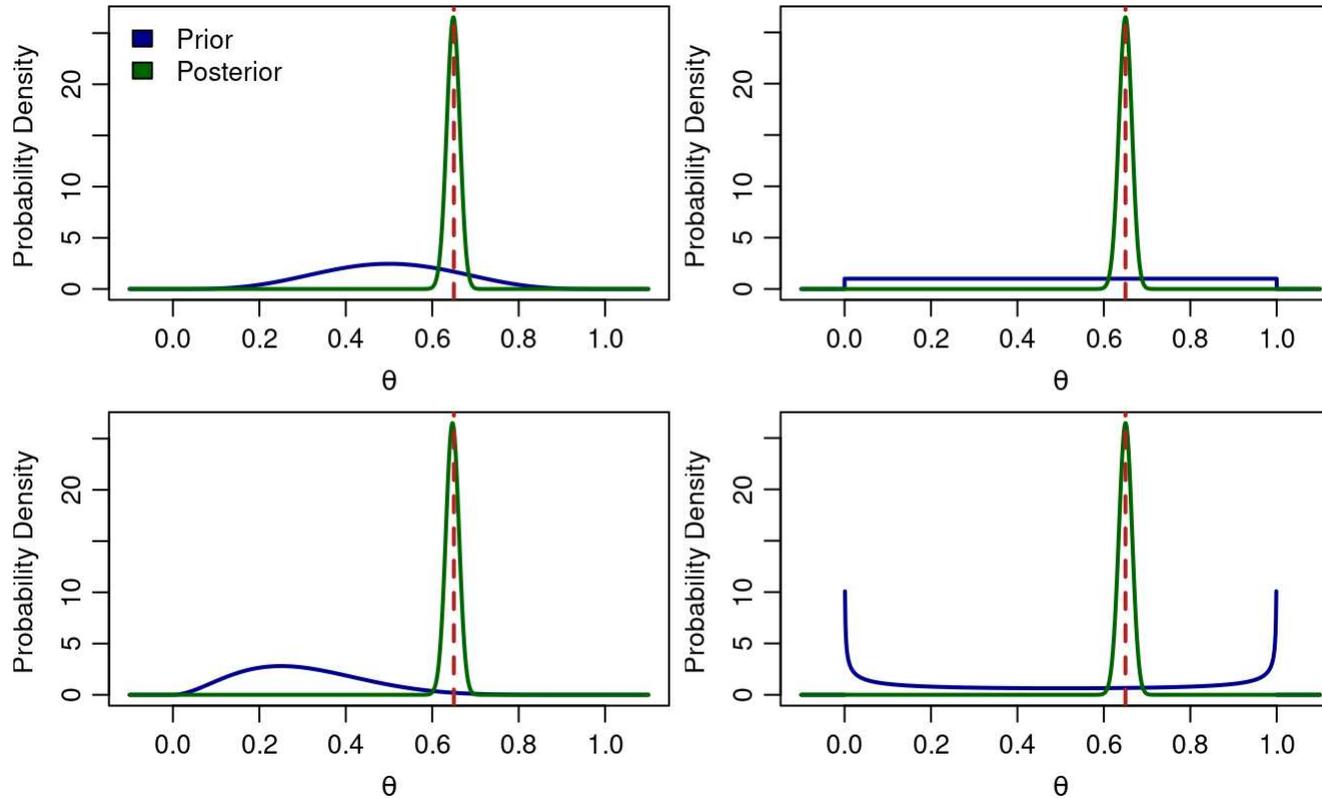
Why Priors for Estimation Really Don't Matter That Much

Frank eats his food 21 out of 30 times.



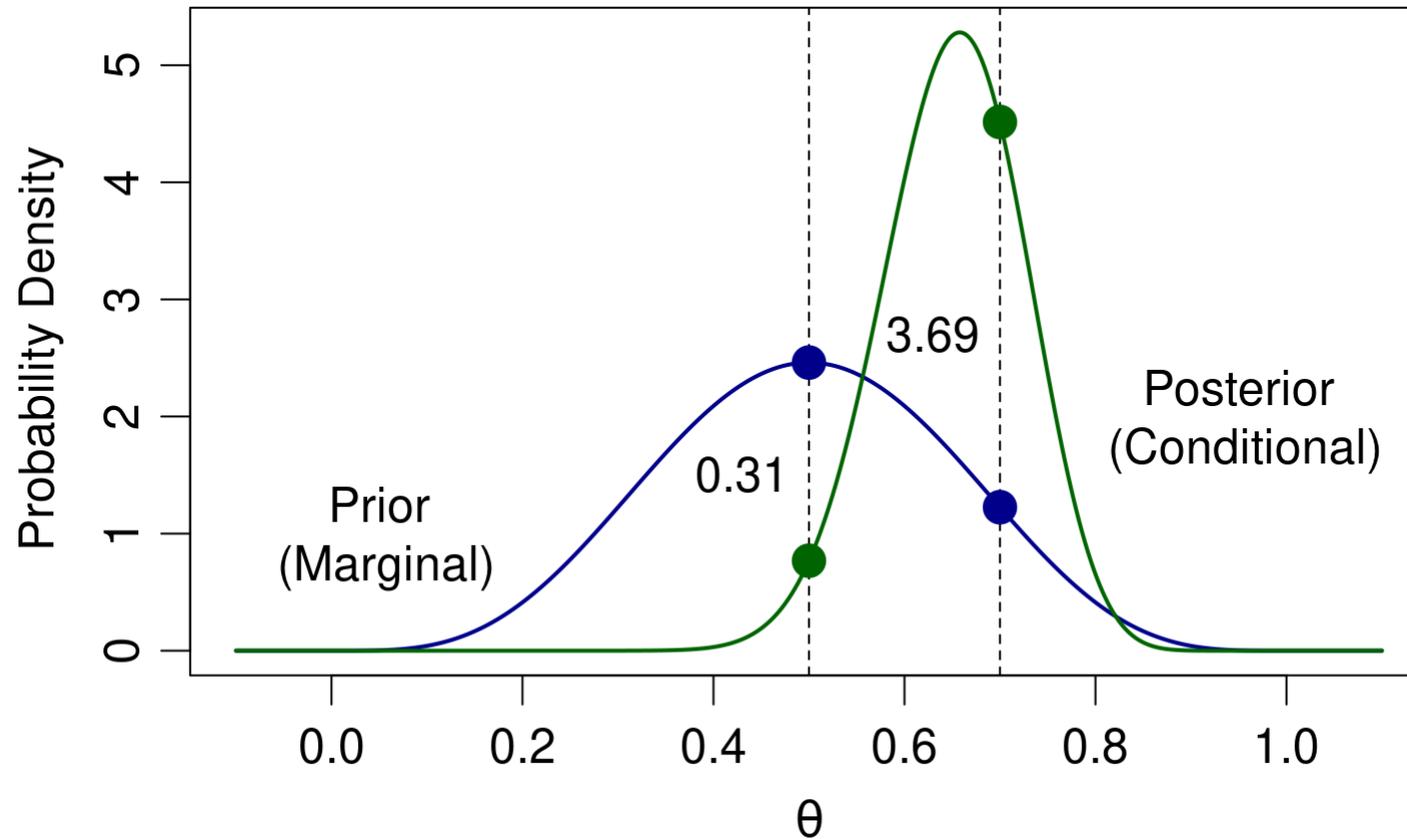
Why Priors for Estimation Really Don't Matter That Much

Frank eats his food 650 out of 1000 times.



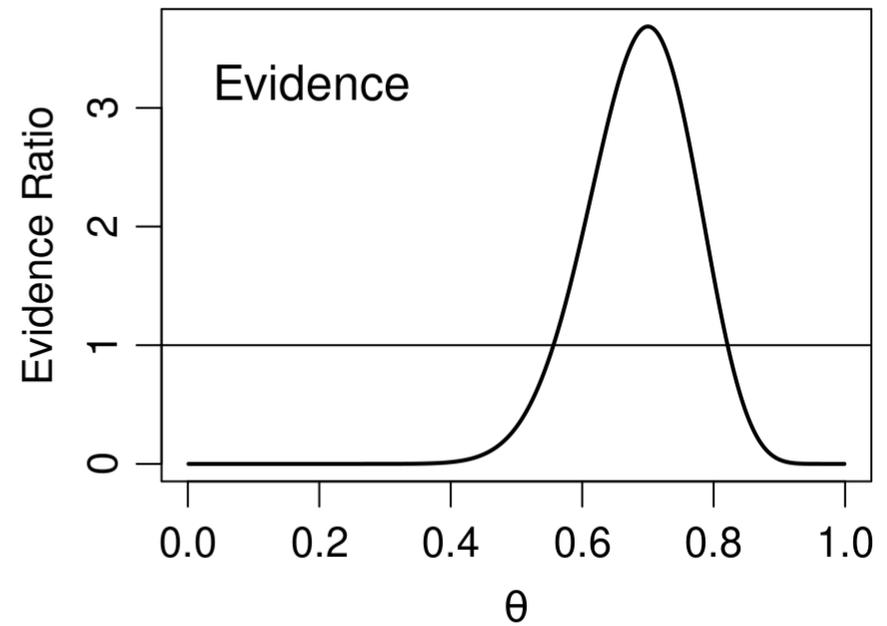
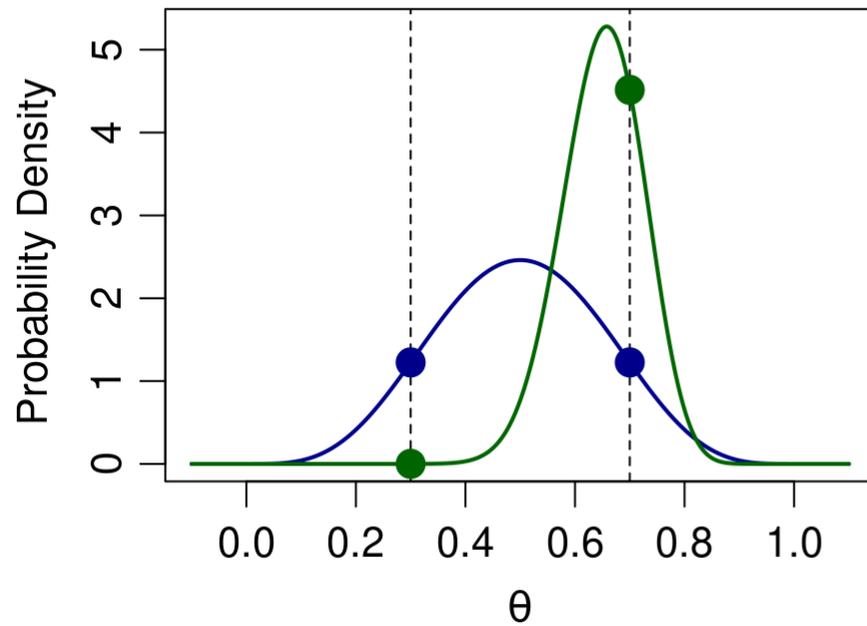
Priors for Model Comparison

Why Priors for Model Comparison Really Do Matter



Why Priors for Model Comparison Really Do Matter

Updating factor



Example: A Cat and a Diet

Frank might be picky, but once we figured out what he likes he is getting pretty fat. So we put him on a diet for 30 days.



Day

day 1

7.2

NA

day 2

7.12

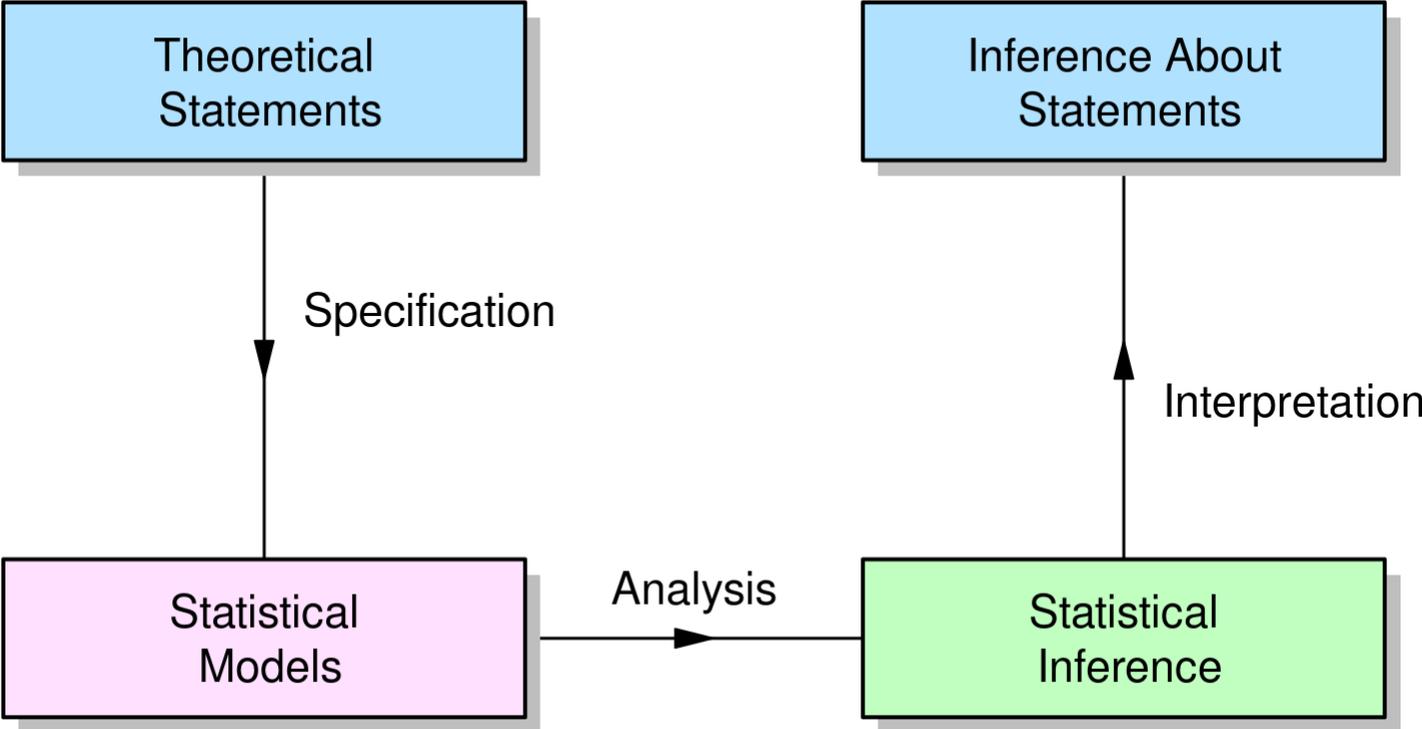
-0.08

day 3

7.07

0.05

Overview of Modeling



Models for Comparison

- \mathcal{M}_1 : Frank loses weight because of the diet.
- \mathcal{M}_0 : Frank does not lose weight because of the diet.
- Change in weight might be normally distributed: $Y_i \sim \text{Normal}(\mu, \sigma^2)$
- For model \mathcal{M}_1 we might want a truncated normal prior on μ : $\mu \sim \text{Normal}_-(a, b^2)$
- For model \mathcal{M}_0 $\mu = 0$.
- Prior on σ^2 ?
- Prior settings on a and b ?

Models for Comparison

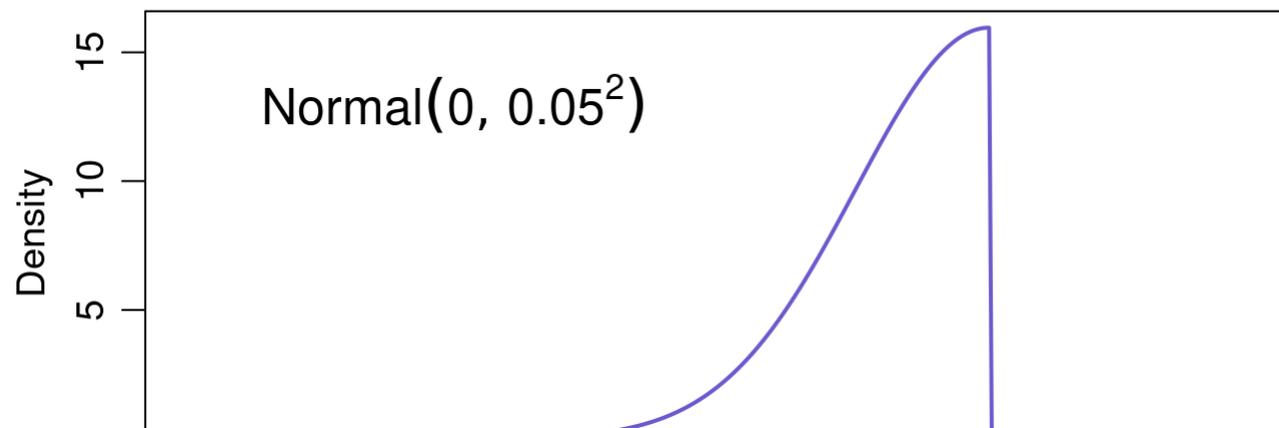
- \mathcal{M}_1 : Frank loses weight because of the diet.
- \mathcal{M}_0 : Frank does not lose weight because of the diet.
- Change in weight might be normally distributed: $Y_i \sim \text{Normal}(\mu, \sigma^2)$
- For model \mathcal{M}_1 we might want a truncated normal prior on μ : $\mu \sim \text{Normal}_-(a, b^2)$
- For model \mathcal{M}_0 $\mu = 0$.
- Prior on σ^2 ?
- Prior settings on a and b ?

1/30

```
## [1] 0.03333333
```

Models for Comparison

- \mathcal{M}_1 : Frank loses weight because of the diet.
- \mathcal{M}_0 : Frank does not lose weight because of the diet.
- Change in weight might be normally distributed: $Y_i \sim \text{Normal}(\mu, \sigma^2)$
- For model \mathcal{M}_1 we might want a truncated normal prior on μ : $\mu \sim \text{Normal}_-(a, b^2)$
- For model \mathcal{M}_0 $\mu = 0$.
- Prior on σ^2 ?
- Prior settings on a and b ?



The Blessing of Default Priors

- BayesFactor package in R and JASP use default priors for the t-test called JZS prior.
- Prior structure is constant.
- Prior setting is on the scale of the effect size (think Cohen's d).

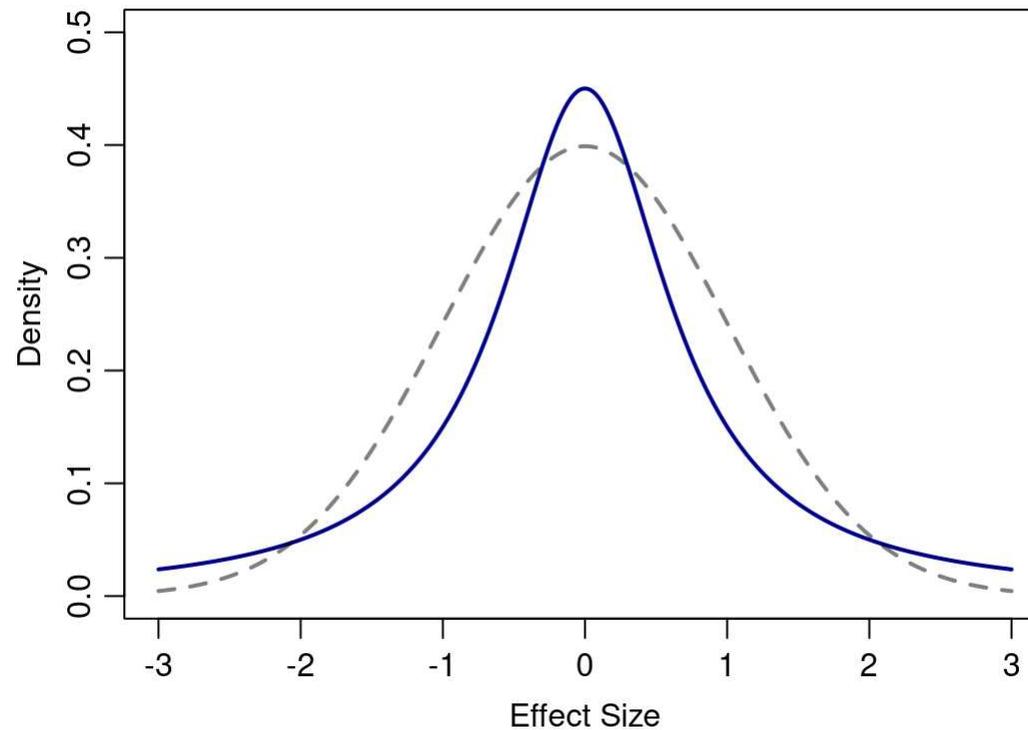
The Blessing of Default Priors

- BayesFactor package in R and JASP use default priors for the t-test called JZS prior.
- Prior structure is constant.
- Prior setting is on the scale of the effect size (think Cohen's d).

```
BayesFactor::ttestBF(x = Change[-1]  
  , mu = 0  
  , nullInterval = c(-Inf, 0)  
  , rscale = 1/sqrt(2))
```

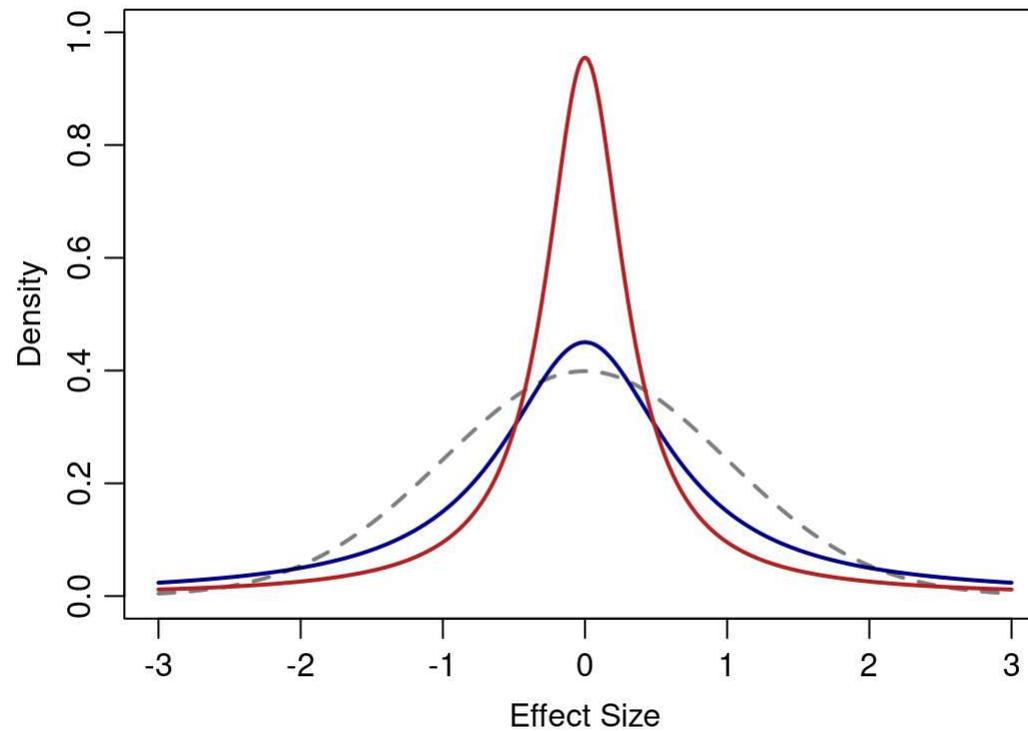
The Blessing of Default Priors

```
BayesFactor::ttestBF(x = Change[-1]  
  , mu = 0  
  , nullInterval = c(-Inf, 0)  
  , rscale = 1/sqrt(2))
```



The Blessing of Default Priors

```
BayesFactor::ttestBF(x = Change[-1]  
  , mu = 0  
  , nullInterval = c(-Inf, 0)  
  , rscale = 1/3)
```



The Blessing of Default Priors

```
BayesFactor::ttestBF(x = Change[-1]
                      , mu = 0
                      , nullInterval = c(-Inf, 0)
                      , rscale = 1/3)

## Bayes factor analysis
## -----
## [1] Alt., r=0.333 -Inf<d<0      : 3.095844 ±0%
## [2] Alt., r=0.333 !(-Inf<d<0) : 0.1448153 ±0%
##
## Against denominator:
##   Null, mu = 0
## ---
## Bayes factor type: BFoneSample, JZS
```

Prior Prediction and Prior Sensitivity

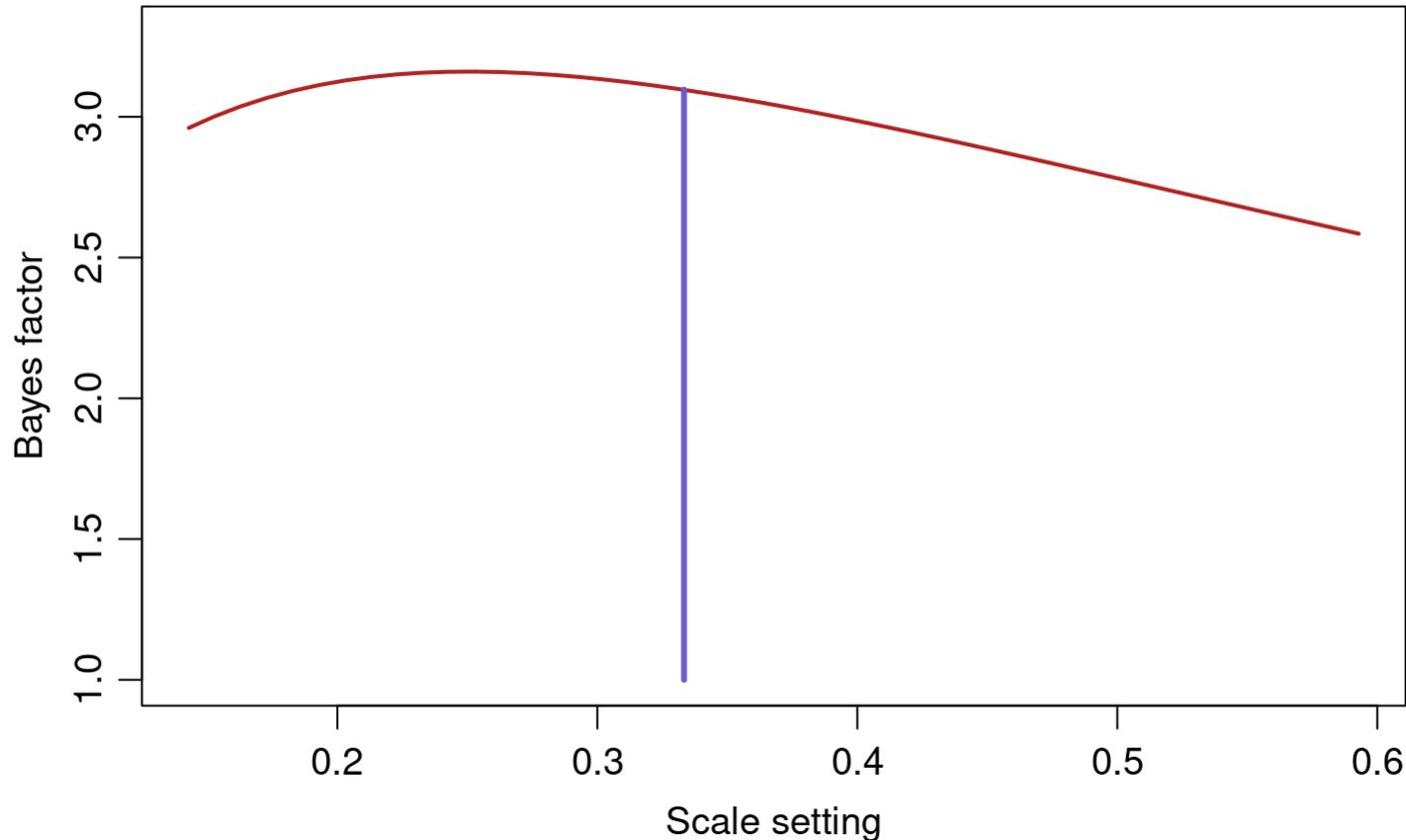
Ensuring you understand your priors and their influence on the analysis?

Prior Sensitivity

- How does your prior influence the results of the analysis?
- Redo the analysis for a range of priors
- If the results are (relatively) stable then we may trust them more.

Prior Sensitivity

- How does your prior influence the results of the analysis?
- Redo the analysis for a **reasonable** range of priors.
- If the results are (relatively) stable then we may trust them more.



Prior Sensitivity

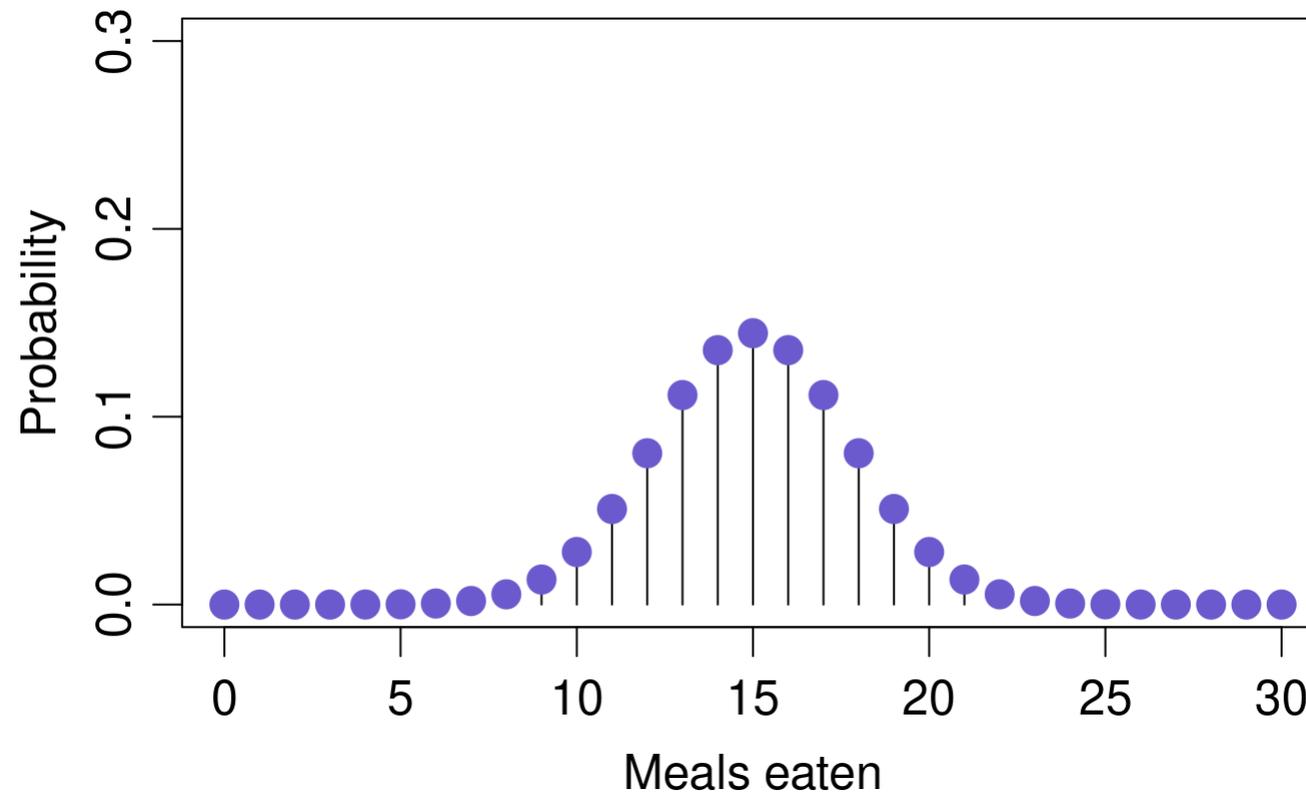
For two parameters

a	b
0.3	1.1
0.5	1.1
0.7	1.1
0.3	1.5
0.5	1.5
0.7	1.5
0.3	1.9
0.5	1.9
0.7	41/46 1.9

Prior Prediction

- What data are predicted by your prior?
- Are these predictions plausible?

$$Y \sim \text{Binomial}(\theta, 30), \quad \theta = .5.$$



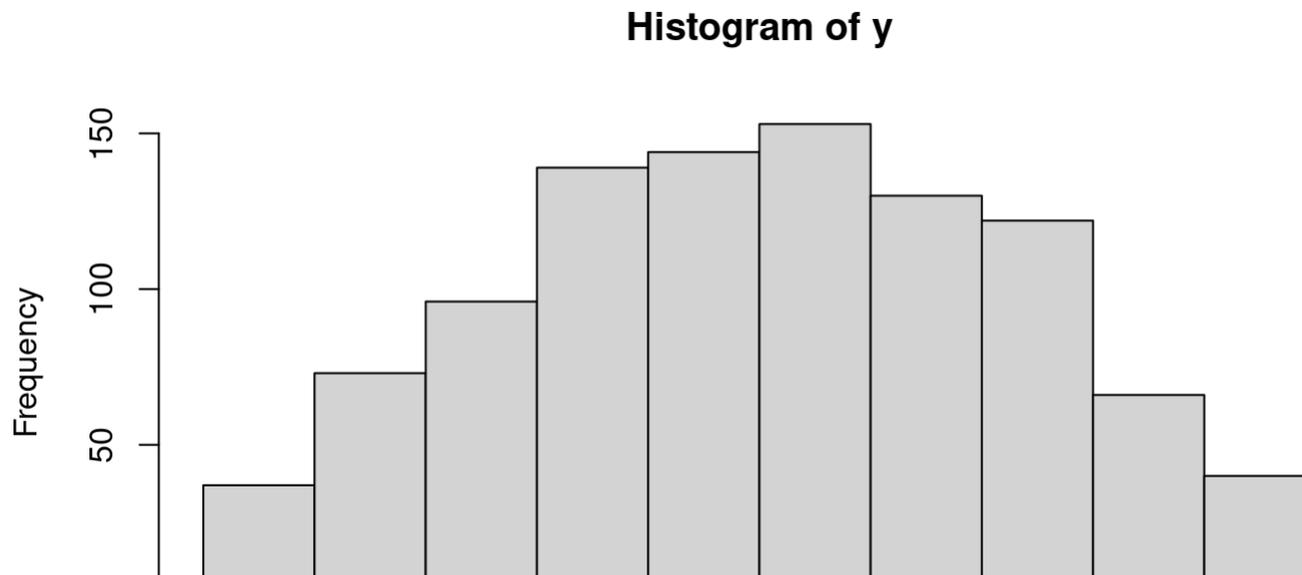
Prior Prediction

- Example: Memory performance in a two-alternative forced choice recognition task ($K = 100$).
- Proposed prior for the probability of a correct response: $\theta \sim \text{Beta}(2, 2)$.
- Procedure: Sample from the prior, simulate data based on the sample.

Prior Prediction

- Example: Memory performance in a two-alternative forced choice recognition task ($K = 100$).
- Proposed Prior: $\theta \sim \text{Beta}(2, 2)$
- Procedure: Sample from the prior, simulate data based on the sample.

```
M <- 1000 # simulation runs  
p.theta <- rbeta(M, 2, 2) # theta sampled from the prior  
y <- rbinom(1000, 100, p.theta)  
hist(y)
```



Prior Prediction

When it goes really wrong

- Using very wide normal priors (e.g., $\text{Normal}(0, 100)$ for likert-scale data).
- When priors are placed on transformed parameters (e.g., logistic regression models).



Thank you!

