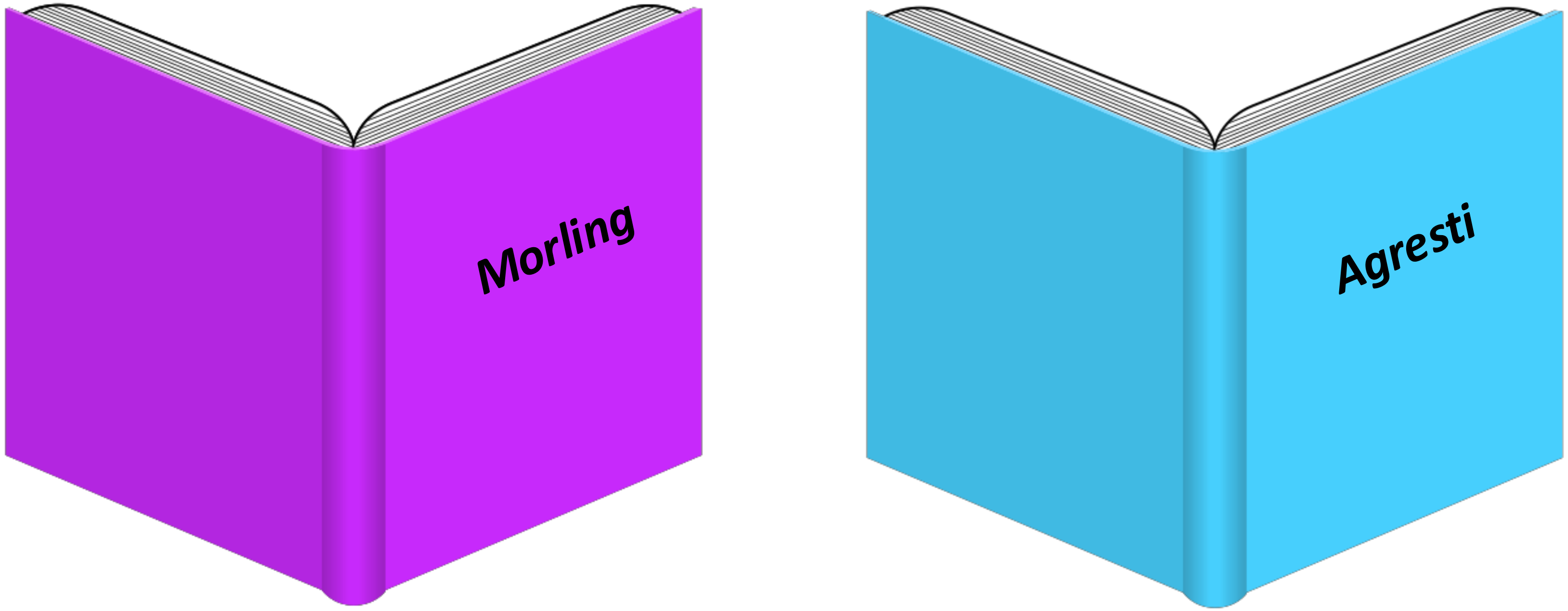


Research Methods and Statistics

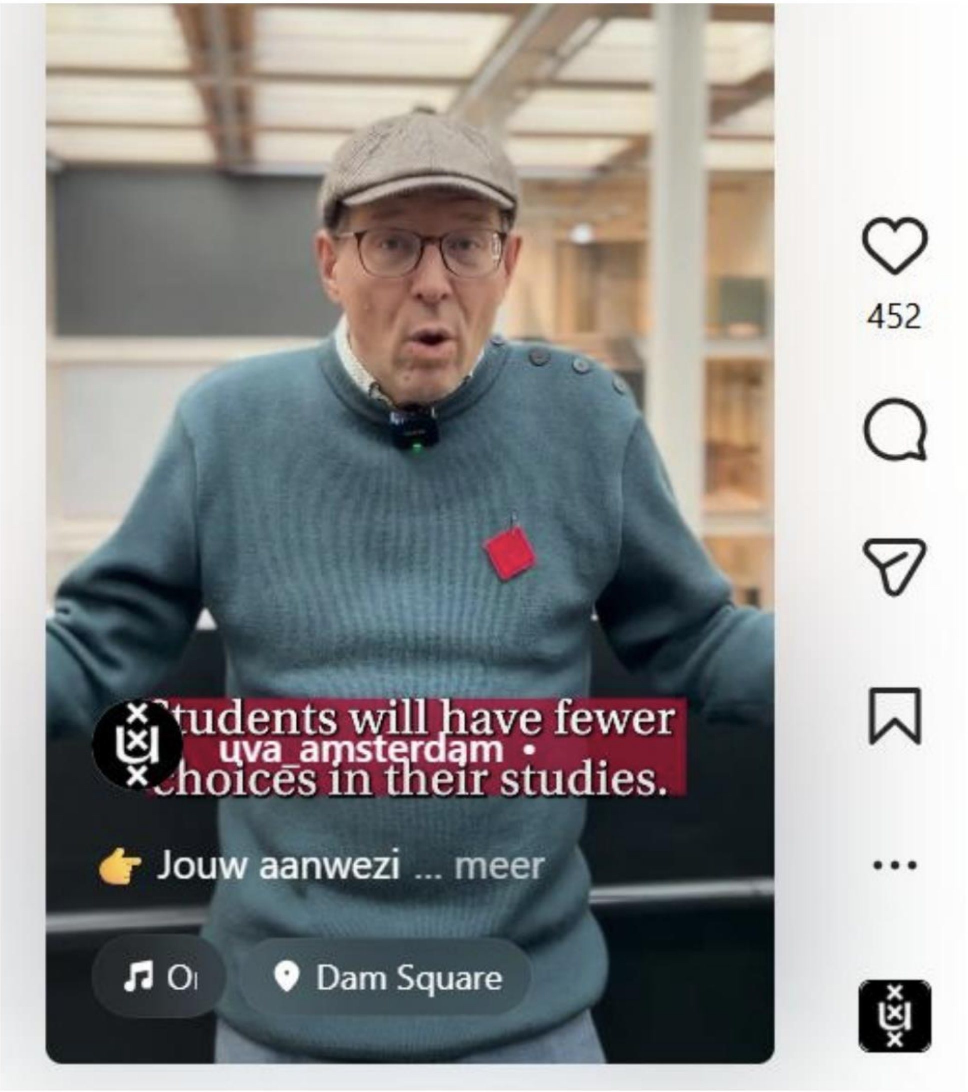
Lecture 23: Bayesian Hypothesis Testing

Johnny van Doorn

Pictures source: pixabay

Don’t forget:
Next week, 9 dec: protest 12AM at Dam square

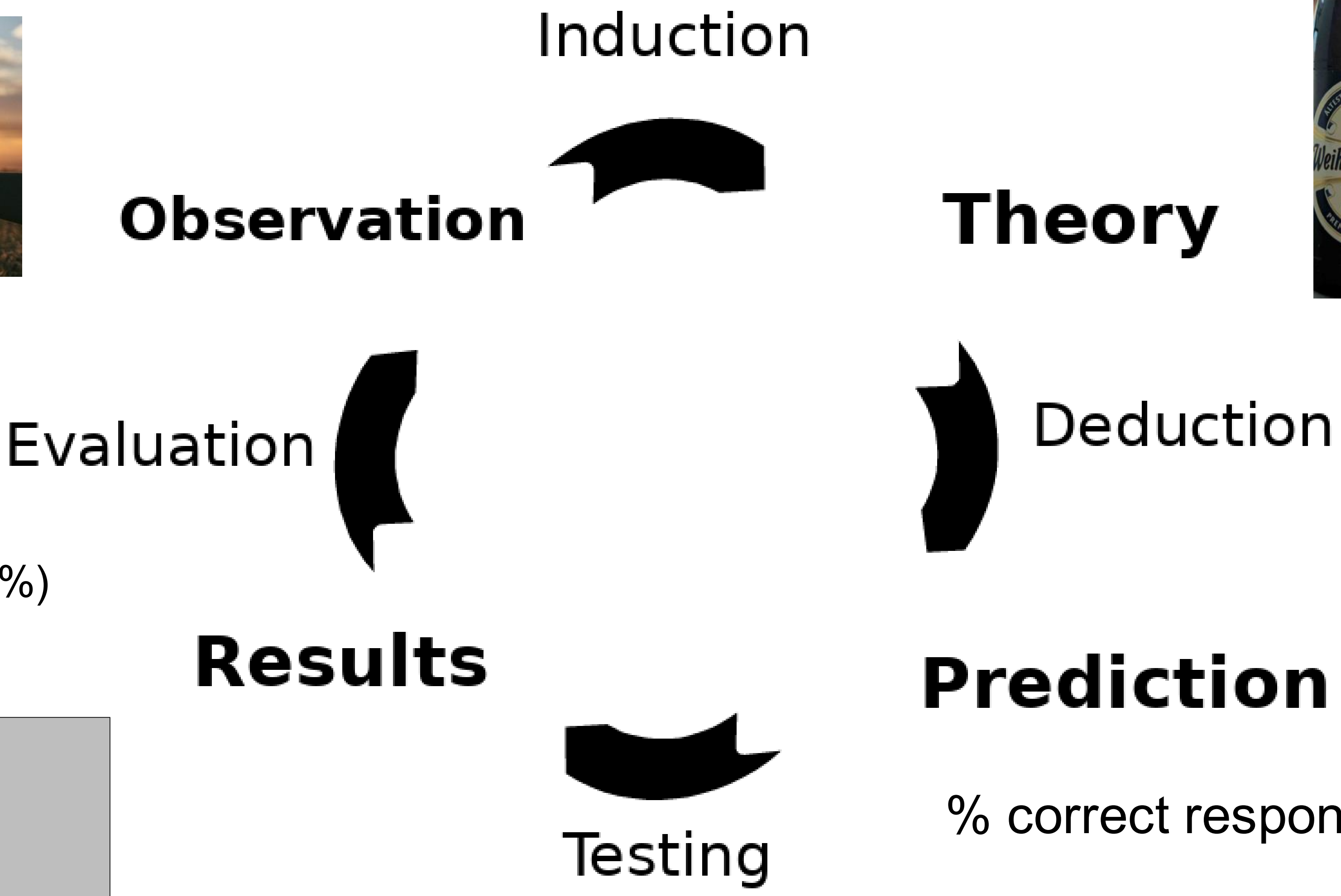
- No lecture!
- UvA professor Rens Bod (of WOinActie) explains why we should go to Dam square:
- https://www.instagram.com/reels/DRcAvEpjOlh/



The Next 3 Lectures

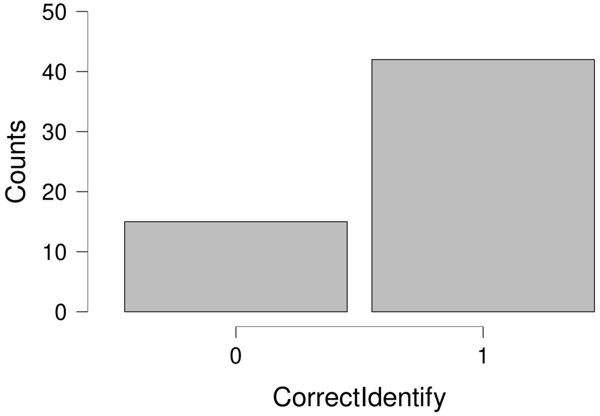
- ~~Today~~
  - ~~Introduction to Bayesian Estimation of a Proportion~~
- Tuesday
  - Bayesian Hypothesis Testing of a Proportion
- Thursday
  - Bayesian Inference for Correlation and T-Test
  - Statistics in the Wild

≠



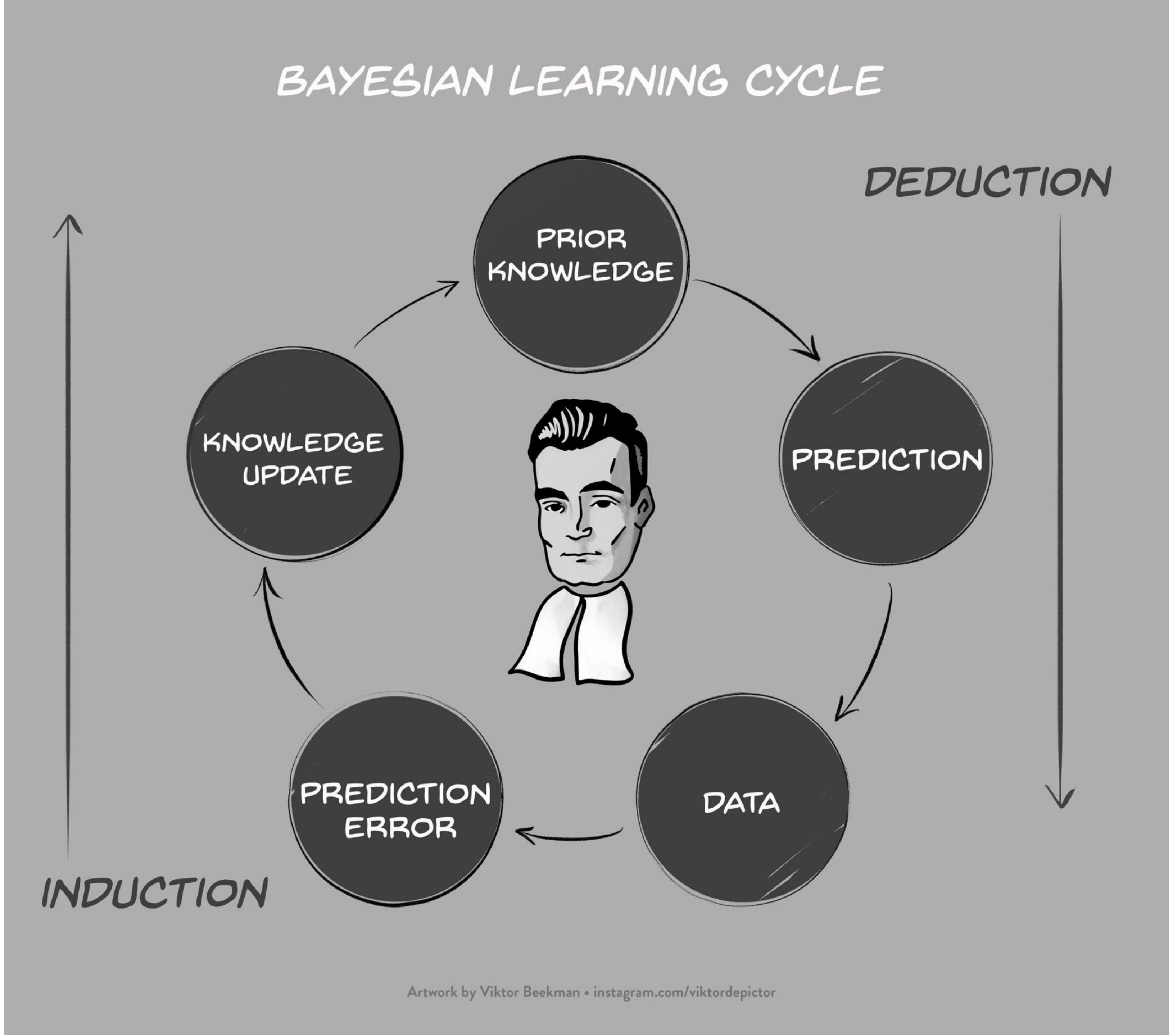
% correct responses > 50%

42 correct (73.7%)
15 incorrect



Today

- **Recap of last week**
- Bayesian Hypothesis Testing
    - Basic concepts
    - Testing a proportion
    - About the Bayes factor
- Recap
    - Practical stuff & next week
    - Example exam question



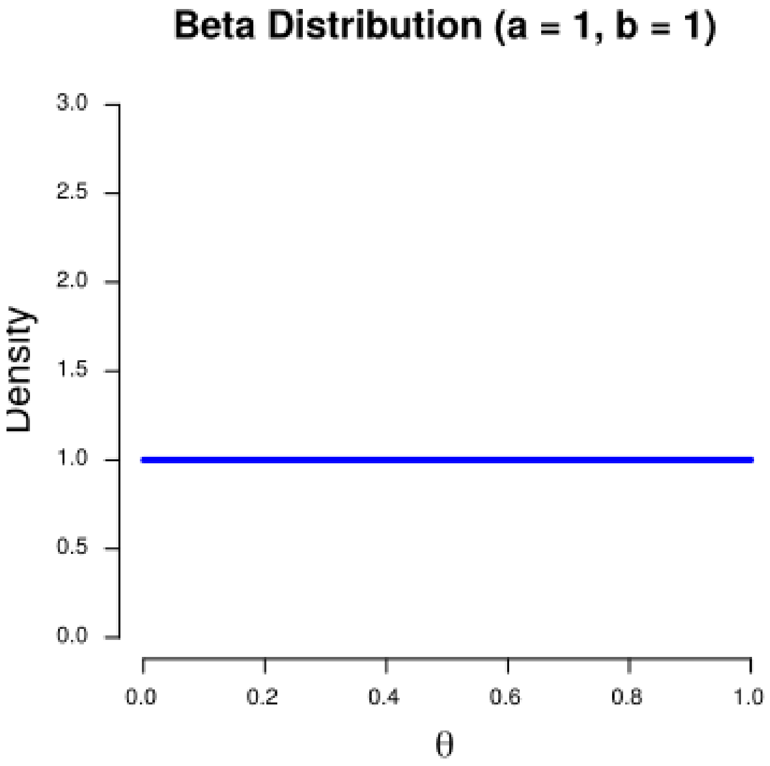
Source:
https://www.bayesianspectacles.org

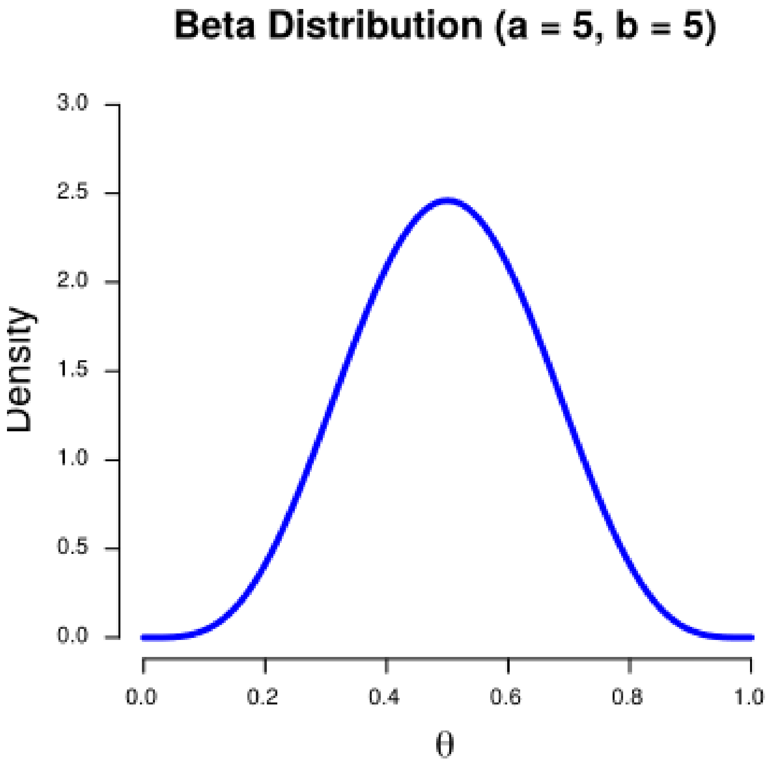
Bayesian Estimation

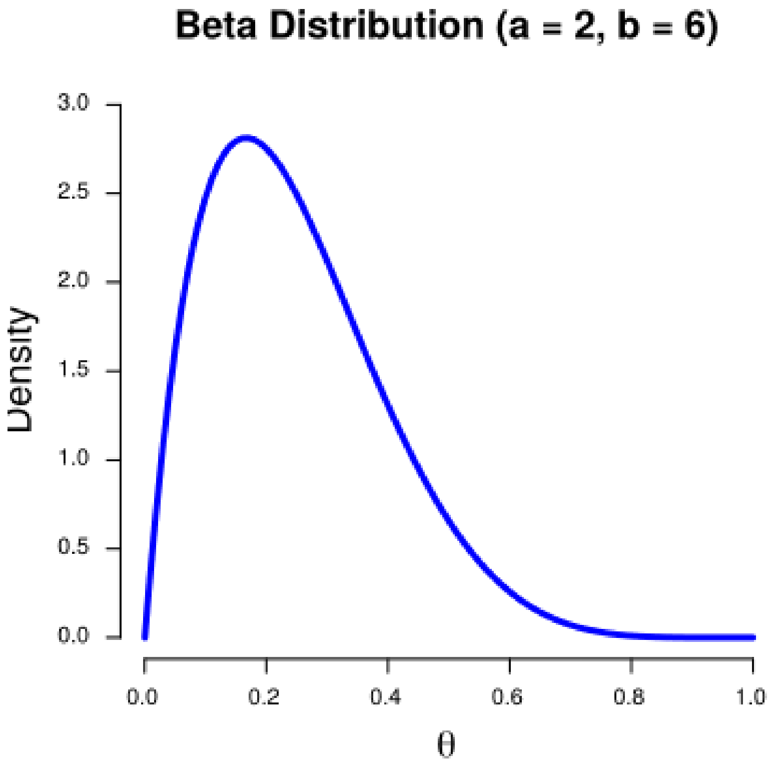
$$\underbrace{p(\theta \mid \text{data})}_{\substack{\text{Posterior beliefs} \\ \text{about the world}}} = \underbrace{p(\theta)}_{\substack{\text{Prior beliefs} \\ \text{about the world}}} \times \underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

This is on the level of the ***parameter***

Models Can Also State a Range of Values







A model that reflects the idea that all values of the proportion are equally plausible - we call this an *uninformative model*

A model that reflects the idea that values close to 0.5 are more plausible

A prior distribution that reflects the idea that values below 0.5 are more plausible (i.e., the coin is biased towards tails)

Estimating a Proportion: Predictive Updating Factor

The **likelihood** of the data, given a certain value of θ

This tells us something about how well a specific value of θ predicted the data (i.e., it is the quality of the prediction for this specific value)

$$\frac{P(\text{data} \mid \theta)}{P(\text{data})}$$

The **marginal likelihood**, across all values of θ

This tells us something how well θ predicted the data, **averaged** over all possible values of θ (i.e., it is the average quality of the prediction of the model)

Taken together, this ratio tells us how well each value of θ predicted the data, **relative** to all other values!

Estimating a Proportion: Predictive Updating Factor

The **likelihood** of the data, given a certain value of θ

This tells us something about how well a specific value of θ predicted the data (i.e., it is the quality of the prediction for this specific value)

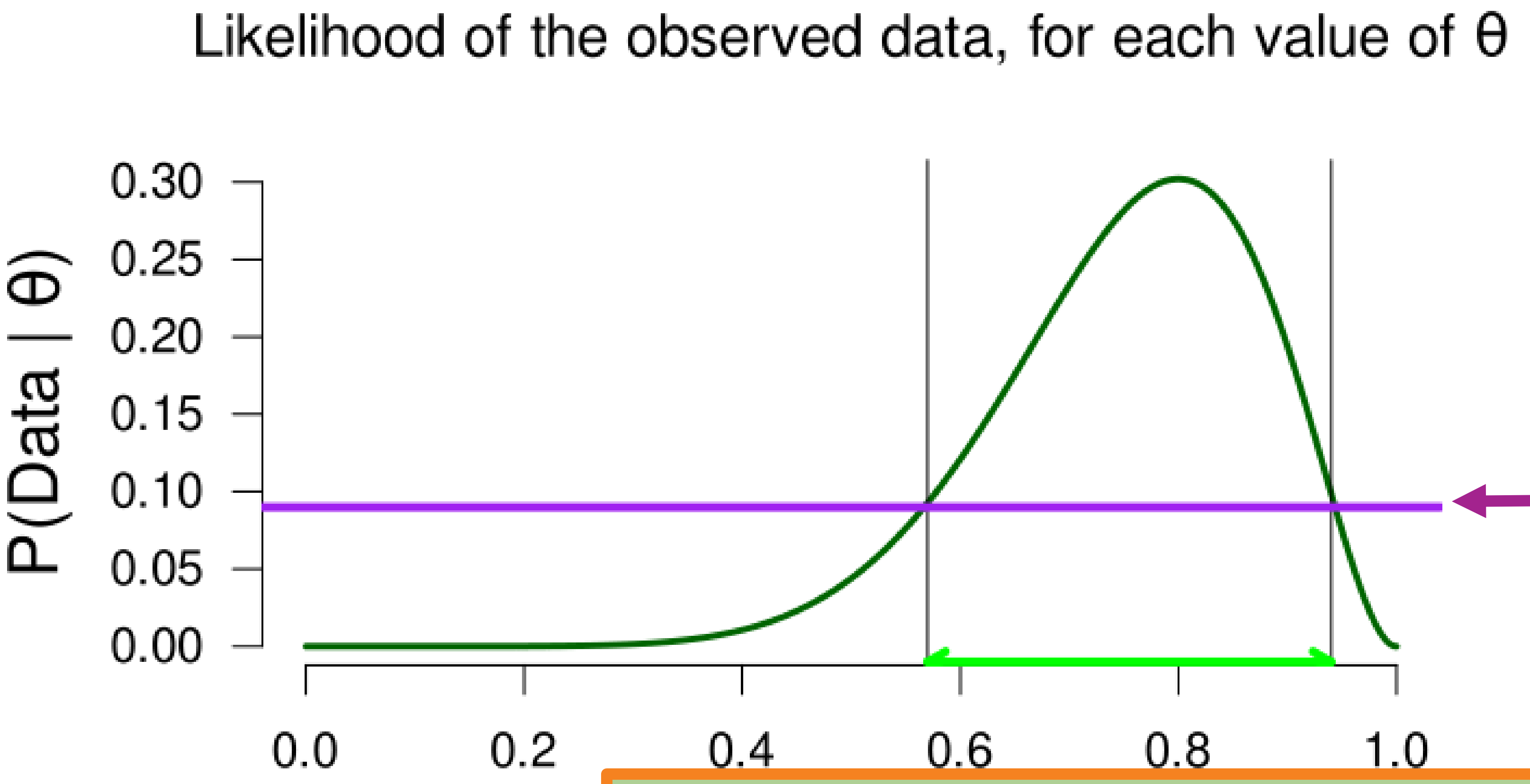
$$\frac{P(\text{data} \mid \theta)}{P(\text{data})}$$

The **marginal likelihood**, across all values of θ

This tells us something how well θ predicted the data, averaged over all possible values of θ (i.e., **it is the average quality of the prediction of the model)**

Taken together, this ratio tells us how well each value of θ predicted the data, **relative** to all other values!

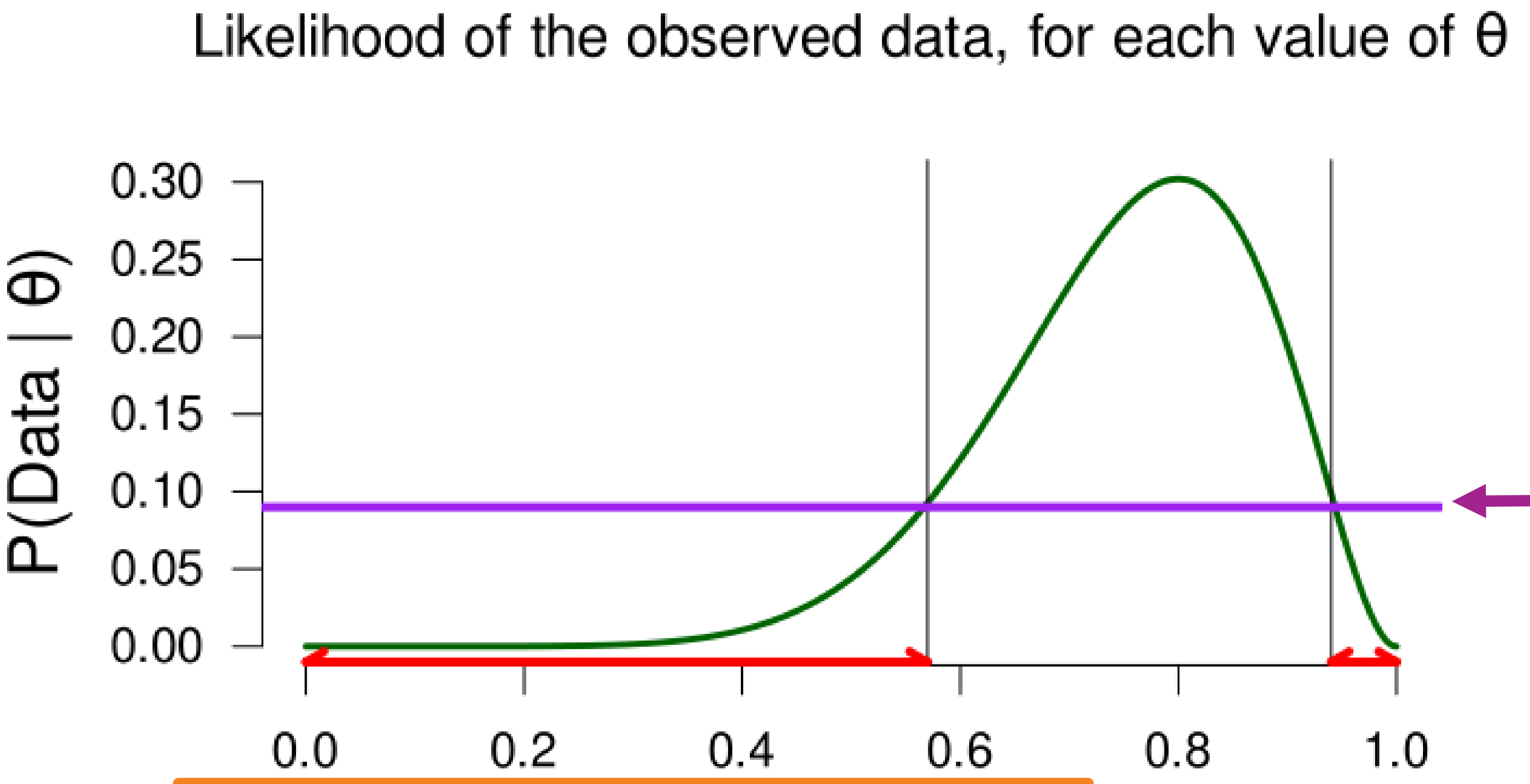
Estimating a Proportion: Predictive Updating Factor



The **marginal likelihood**, across all values of θ (= 0.091)

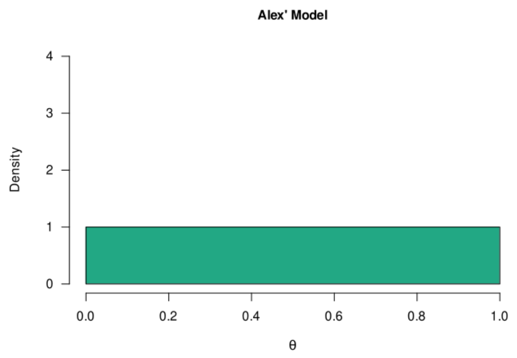
Values of θ that predicted the data better than average (i.e., their likelihoods are **greater** than the marginal likelihood)

Estimating a Proportion: Predictive Updating Factor



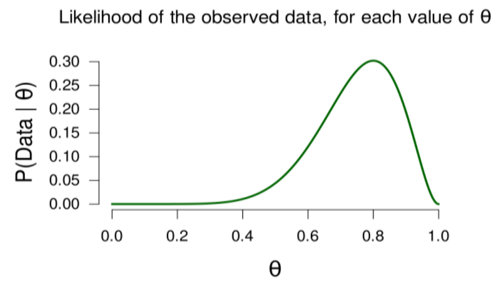
The **marginal likelihood**, across all values of θ (= 0.091)

Values of θ that predicted the data worse than average (i.e., their likelihoods are **less** than the marginal likelihood)



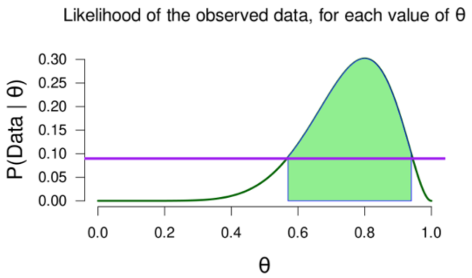
We start with our prior beliefs

We update those with data

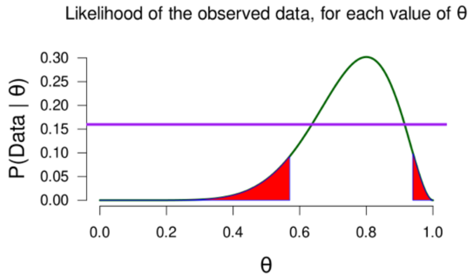


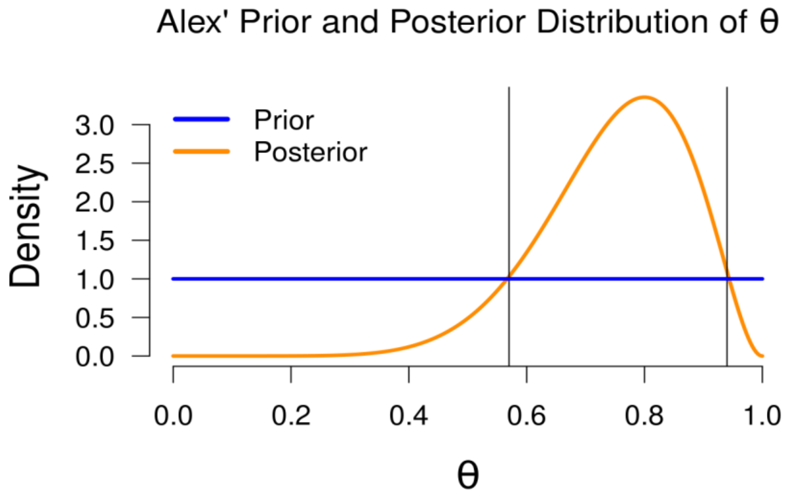
Values of θ that predicted the data better than average receive a boost in plausibility (i.e., their updating ratio >1

Values of θ that predicted the data worse than average receive a penalty in plausibility (i.e., their updating ratio < 1



We end with our posterior beliefs





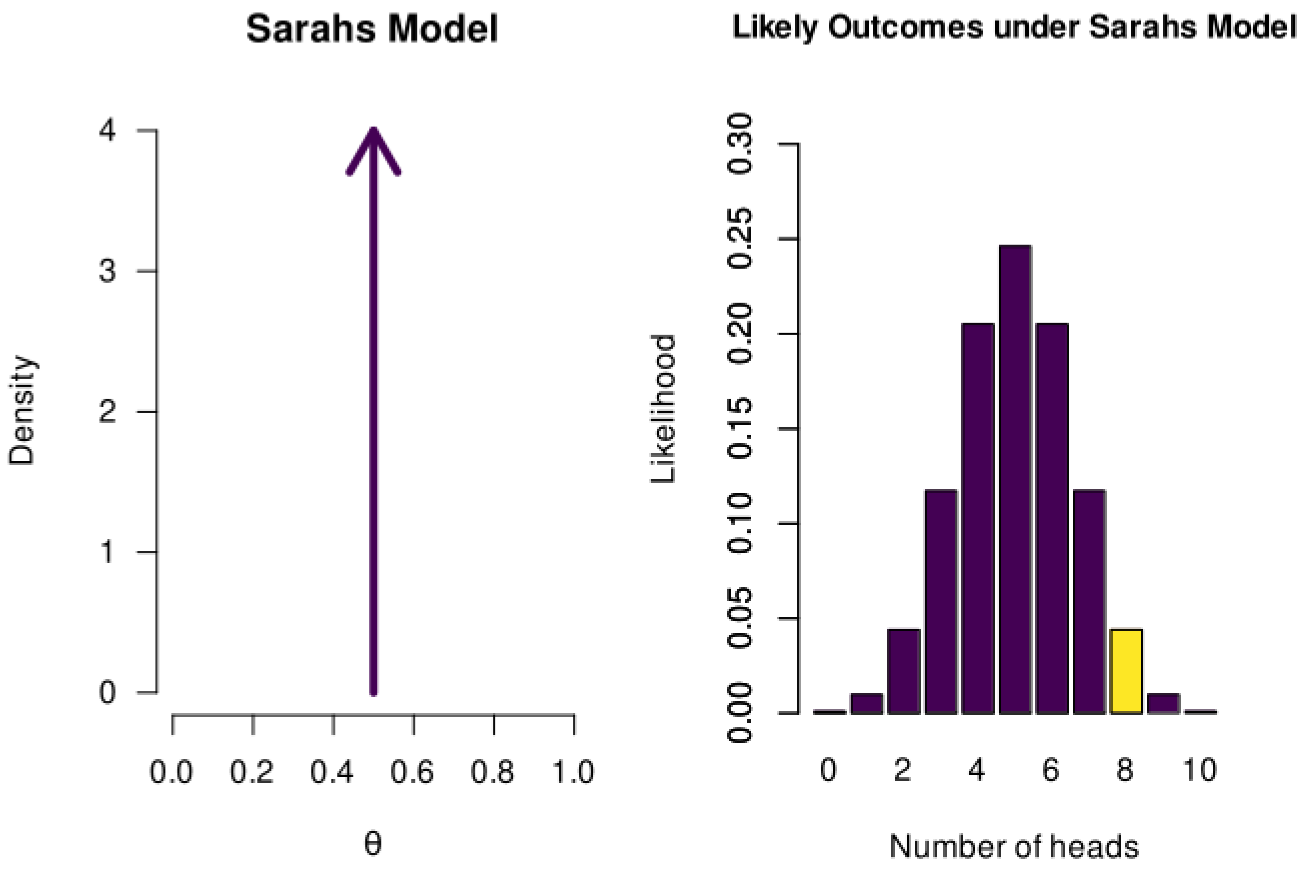
Today

- Recap of last week
- **Bayesian Hypothesis Testing**
    - Basic concepts
    - Testing a proportion
    - About the Bayes factor
- Recap
    - Practical stuff & next week
    - Example exam question

Statistical Models Make Predictions

Based on what these models claim about θ, certain outcomes are more/less likely

The yellow bar indicates how likely an outcome of 8/10 heads is, under Sarah's model



$$\frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

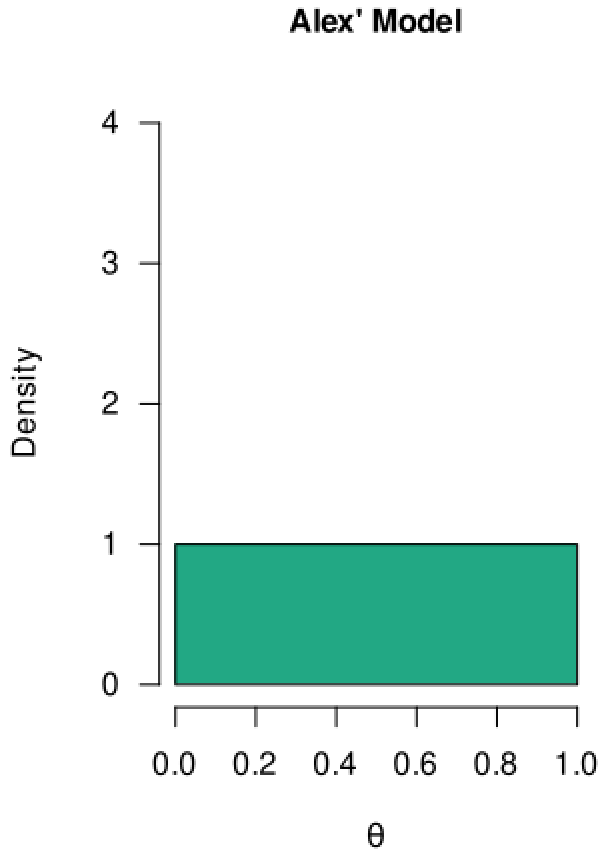
To create this figure, we take the binomial formula, and fill in θ = 0.5

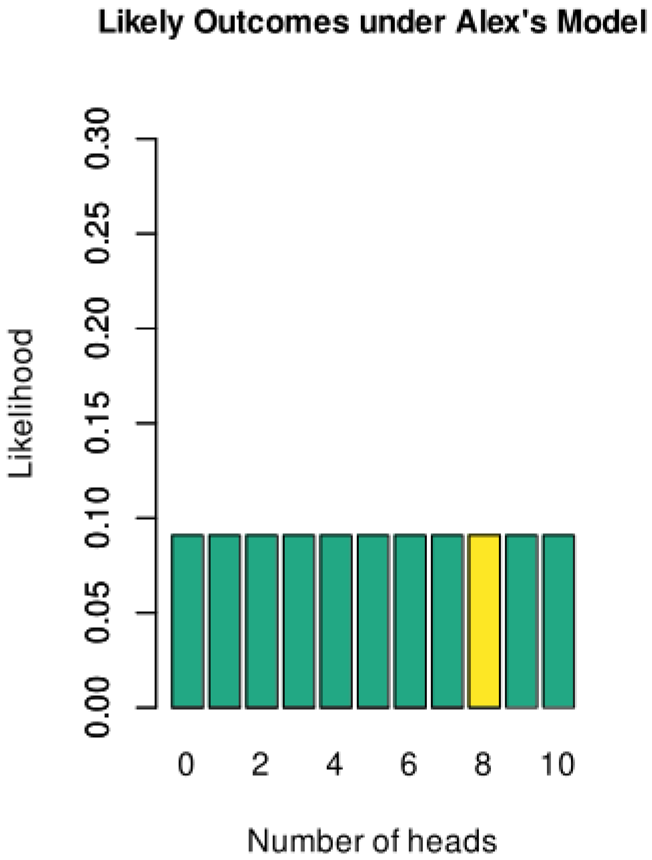
For instance, for an outcome of k=8 heads, the formula gives 0.0439

Statistical Models Make Predictions

Based on what these models claim about θ, certain outcomes are more/less likely

The yellow bar indicates how likely an outcome of 8/10 heads is, under Alex' model





$$\frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

Alex does not postulate a single value of θ, but a whole range of values

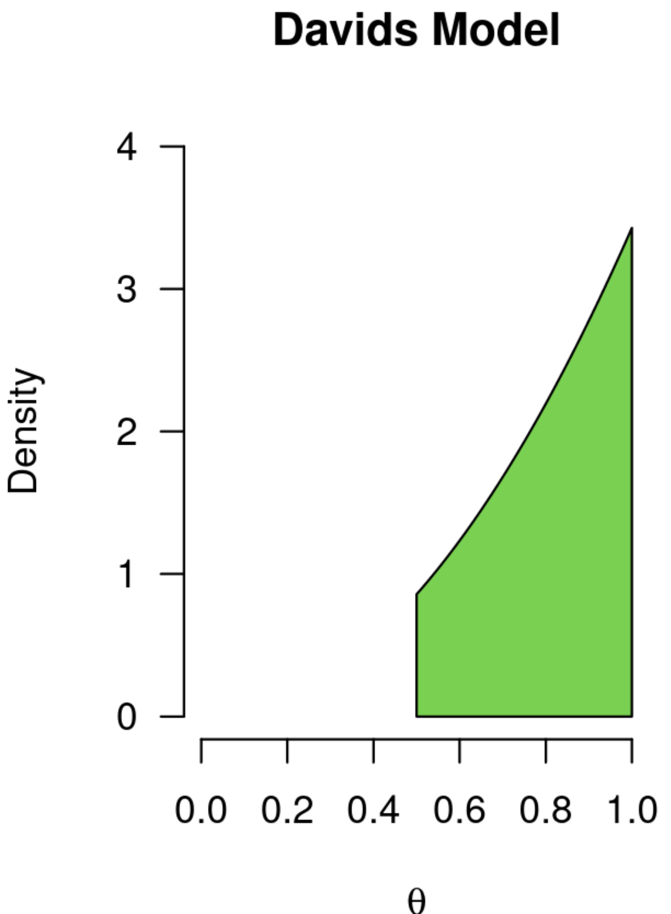
To obtain the likelihood of 8 heads for Alex' model, we consider every postulated value of θ between **0 and 1,** compute the likelihood of the data for each value, and average across all of these likelihoods, weighted by the prior density at each point

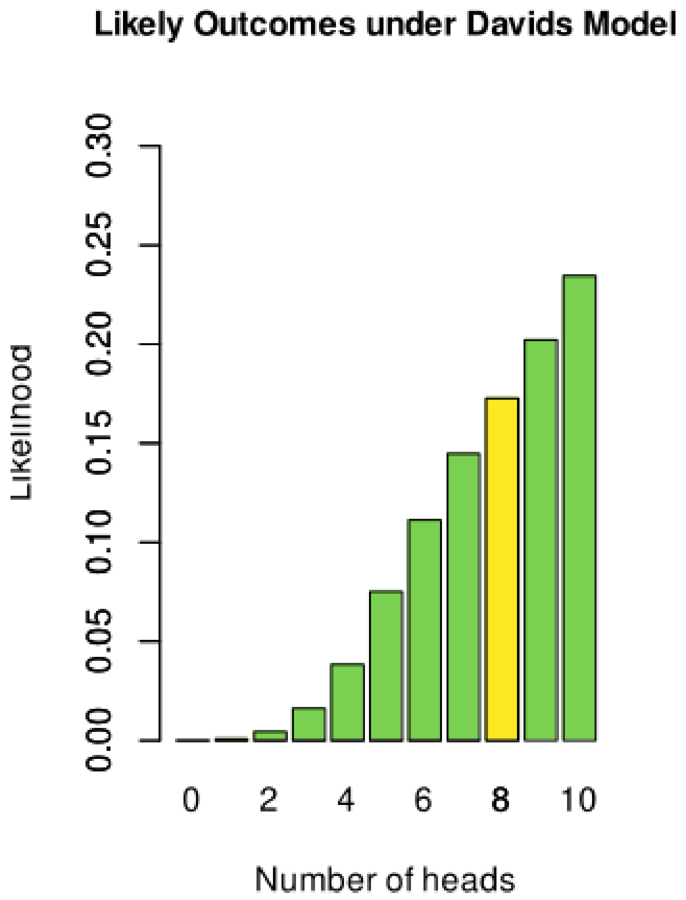
→ The **marginal likelihood**, across all values of θ

Statistical Models Make Predictions

Based on what these models claim about θ, certain outcomes are more/less likely

The yellow bar indicates how likely an outcome of 8/10 heads is, under David's model





$$\frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

David does not postulate a single value of θ, but a whole range of values

To obtain the likelihood of 8 heads for David's model, we consider every postulated value of θ between **0.5 and 1,** compute the likelihood of the data for each value, and average across all of these likelihoods, weighted by the density at each point

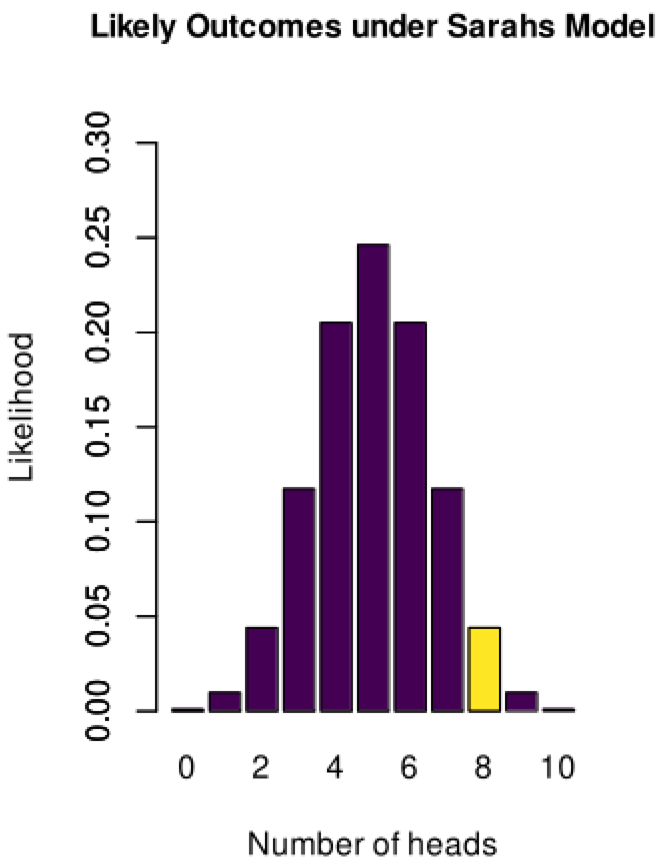
→ The **marginal likelihood**, across all values of θ

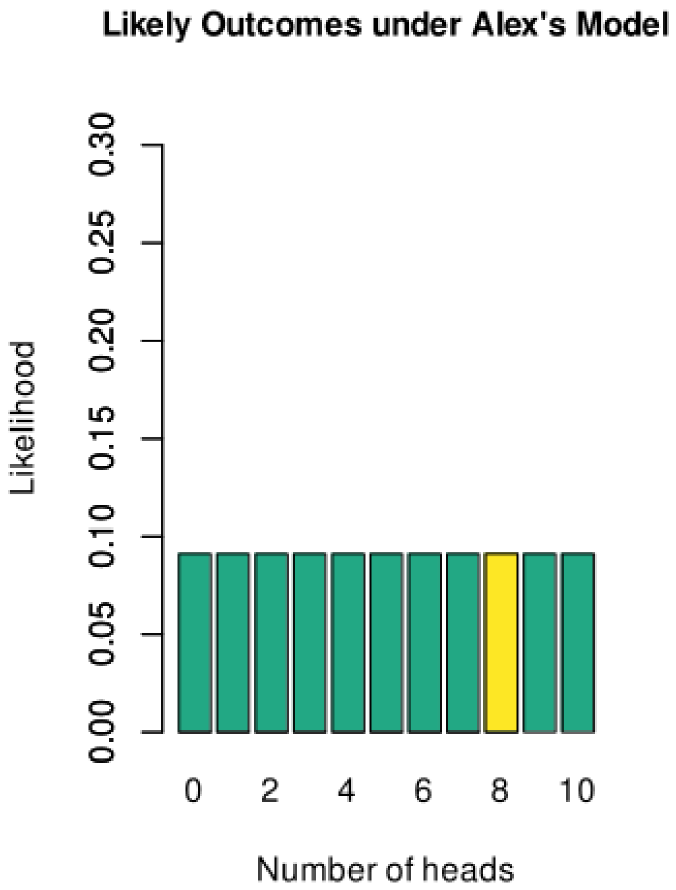
Bayesian Hypothesis Testing

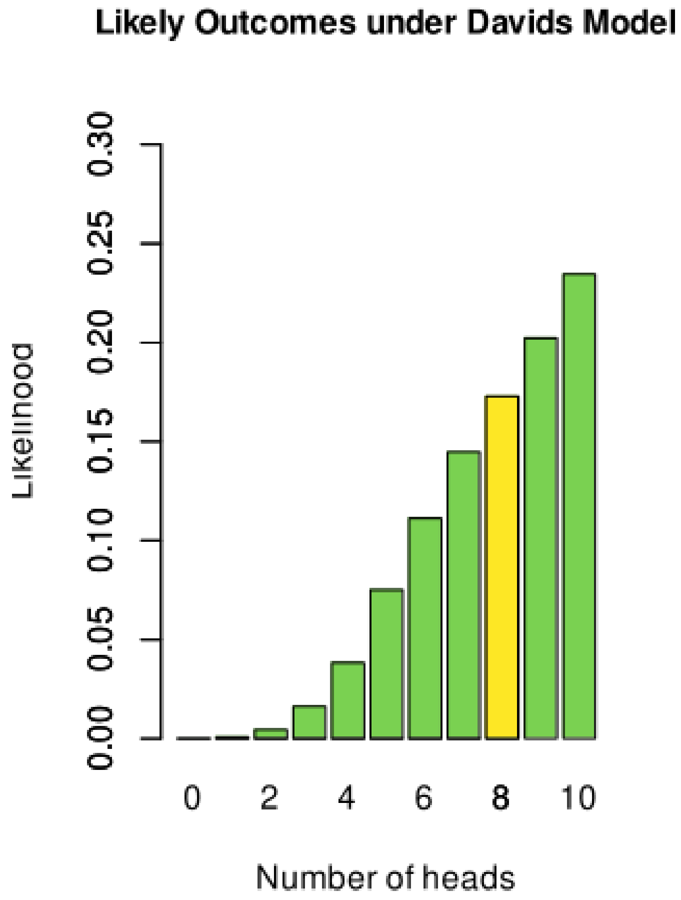
- We generally want to compare two models, H_0 and H_1
- Which model is better supported by the data?

Bayesian Hypothesis Testing

The yellow bar (marginal likelihood) indicates how likely the observed data (8/10 heads) are, under each model





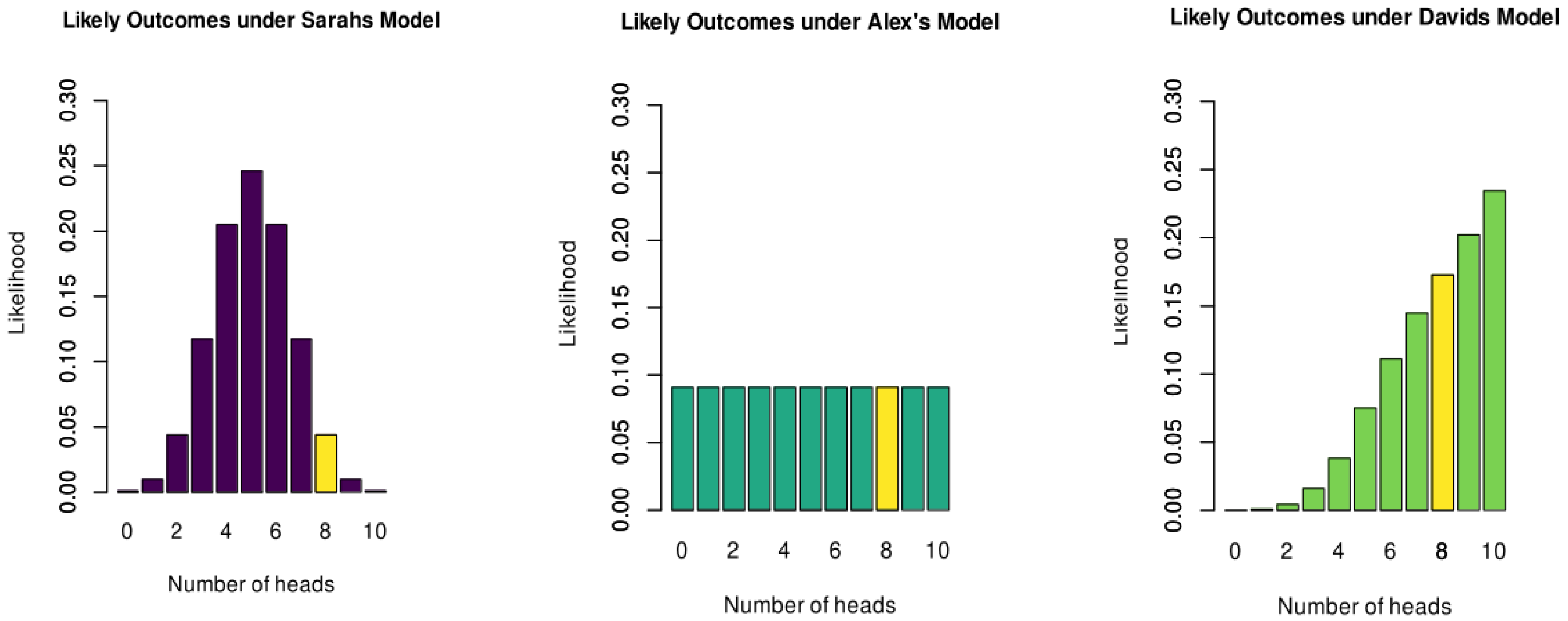


Bayesian Hypothesis Testing

We can compare these marginal likelihoods to do model comparison!

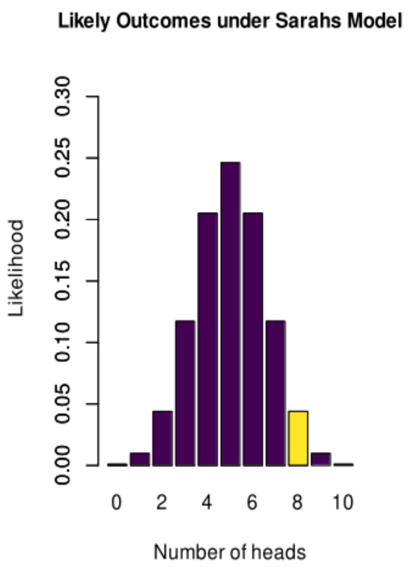
We look at how likely the observed data are under all these models

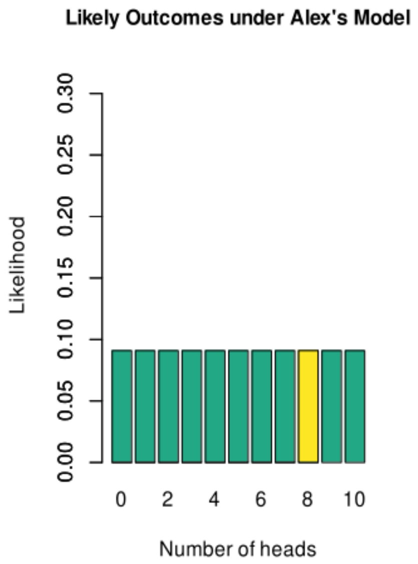
The ratio of these likelihoods tells us which of the two models predicted the best!

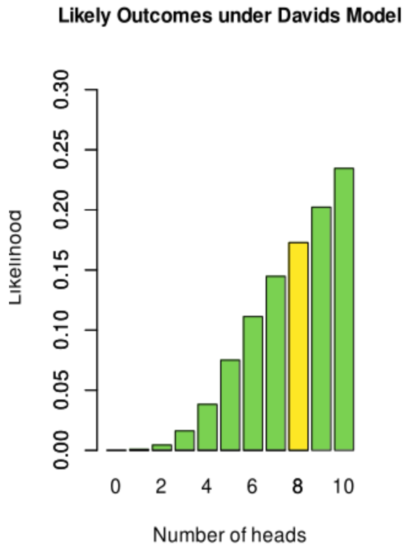


Bayesian Hypothesis Testing

We can compare these marginal likelihoods to do model comparison!







Likelihood of data under Sarah’s model: 0.044

Likelihood of data under Alex’ model: 0.091

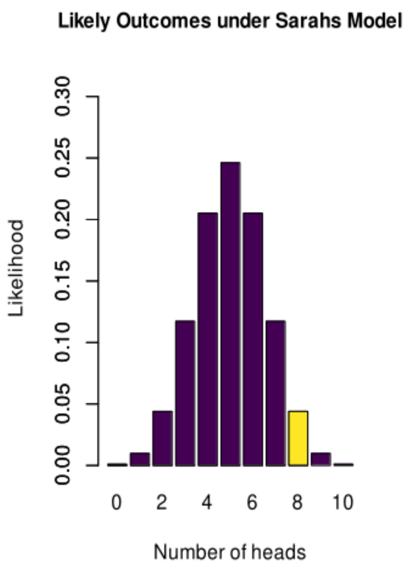
Likelihood of data under David’s model: 0.172

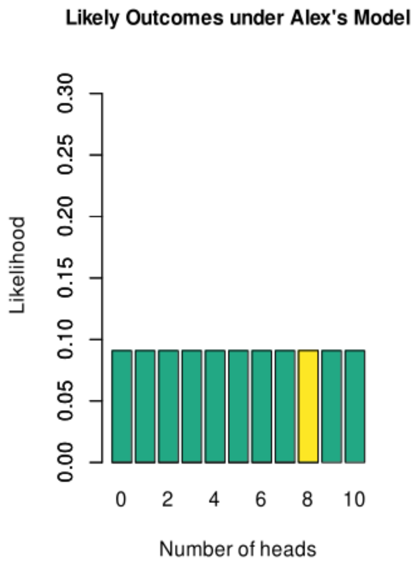
The data are 0.044 / 0.091 = 0.48 times as likely under Sarah’s model as under Alex’ model

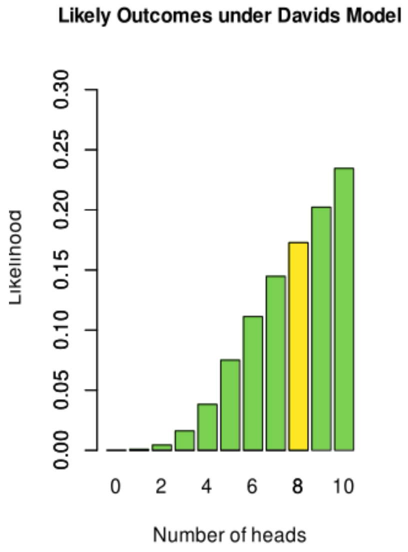
The data are 0.091 / 0.172 = 0.53 times as likely under Alex’ model as under David’s model

Bayesian Hypothesis Testing

We can compare these marginal likelihoods to do model comparison!







Likelihood of data under Sarah's model: 0.044

Likelihood of data under Alex' model: 0.091

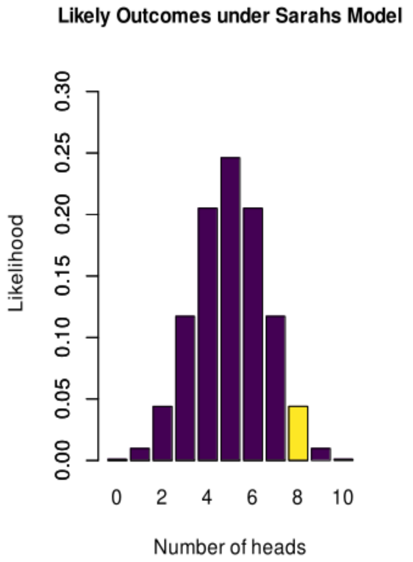
Likelihood of data under David's model: 0.172

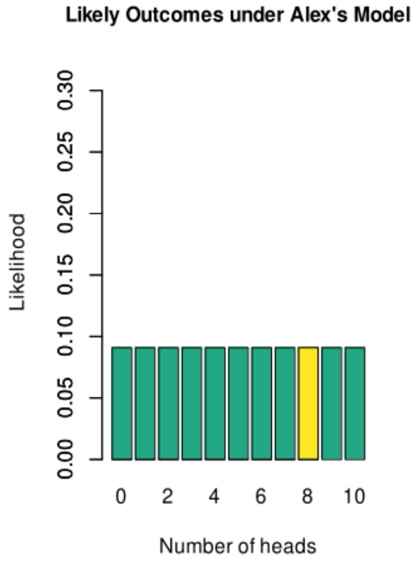
The data are 0.091 / 0.044 = 2.07 times as likely under Alex' model as under Sarah's model

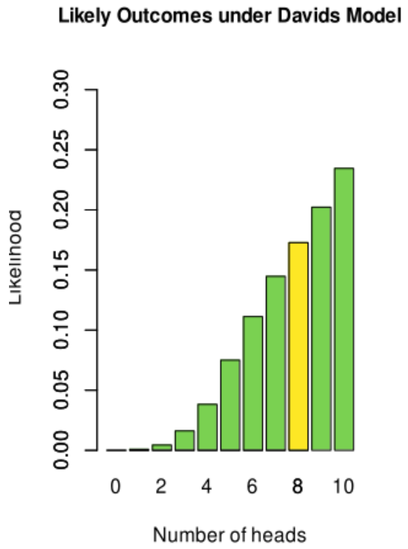
The data are 0.172 / 0.091 = 1.89 times as likely under David's model as under Alex' model

Bayesian Hypothesis Testing

We can compare these marginal likelihoods to do model comparison!







Likelihood of data under Sarah's model: 0.044

Likelihood of data under Alex' model: 0.091

Likelihood of data under David's model: 0.172

The data are 0.091 / 0.044 = 2.07 times as likely under Alex' model as under Sarah's model

The data are 0.172 / 0.091 = 1.89 times as likely under David's model as under Alex' model

This ratio is the Bayes factor: the holy grail of Bayesian model comparison!

Today

- Recap of last week
- Bayesian Hypothesis Testing
    - **Basic concepts**
    - Testing a proportion
    - About the Bayes factor
- Recap
    - Practical stuff & next week
    - Example exam question

Bayesian Hypothesis Testing

$$\underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}}$$

Bayesian Hypothesis Testing

$$\underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

Bayesian Hypothesis Testing

$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\substack{\text{Posterior beliefs}\\ \text{about hypotheses}}} = \underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs}\\ \text{about hypotheses}}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive}\\ \text{updating factor}}}$$

This is on the level of the ***hypothesis***

Bayesian Hypothesis Testing: Prior Odds

$$\underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}}$$

The prior odds specify how plausible a hypothesis is, **relative to another hypothesis**, before seeing the data

Bayesian Hypothesis Testing: Prior Odds

$$\underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}}$$

The prior odds specify how plausible a hypothesis is, relative to another hypothesis, **before seeing the data**

Bayesian Hypothesis Testing: Prior Odds

$$\underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}}$$

For instance, we can believe the *alternative* hypothesis is 5 times more likely, a priori. This means that the prior odds equal 5

Or, we can believe the *null* hypothesis is 5 times more likely, a priori. This means that the prior odds equal 0.2 (because the null hypothesis is in the denominator)

Or, we can believe both hypotheses to be equally likely, a priori. This means that the prior odds equal 1
→ usually done

"If you had 100€ to bet on the winning hypothesis, how would you divide it?"

Bayesian Hypothesis Testing: Predictive Updating Factor

How well did the alternative hypothesis predict the data?

$$\underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

How well did the null hypothesis predict the data?

What Do the Hypotheses Predict?

Theory

Deduction

Prediction

% correct responses > 50%

$$\mathcal{H}_0 : \theta = 0.5$$

The null hypothesis postulates that people **cannot** taste the difference, so their proportion correct is at chance level (50%)

It predicts that the observed proportion correct will be around 0.5

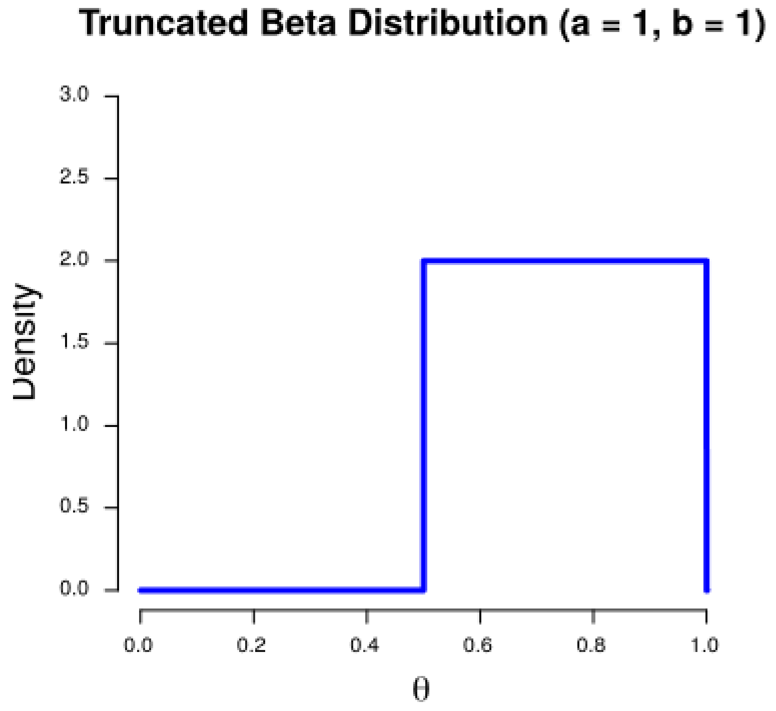
$$\mathcal{H}_1 : \theta > 0.5$$

The alternative hypothesis postulates that people **can** taste the difference, so their proportion correct is above chance level (>50%) → one-sided

It predicts that the observed proportion will be greater than 0.5

What Does the Alternative Hypothesis Predict?

The prior distribution reflects the statement made by a model. If it is one-sided, the prior is truncated: all values not predicted by the H_1 have a density of 0

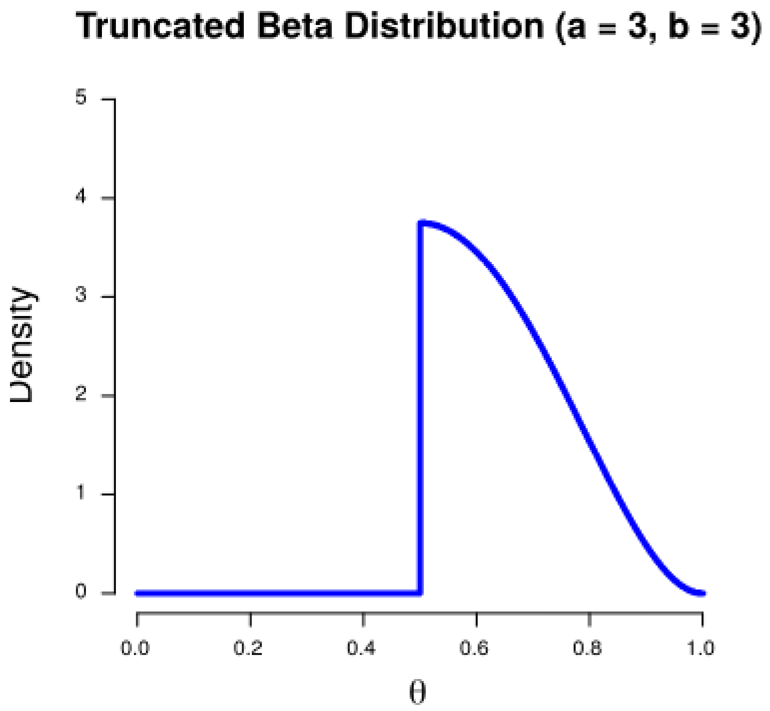


$$\mathcal{H}_1 : \theta > 0.5$$

The prior formalizes the predictions of the alternative hypothesis: Here, it predicts that all values of θ between 0.5 and 1 are equally likely.

"If you had 100€ to bet on likely values of θ, how would you divide it?"

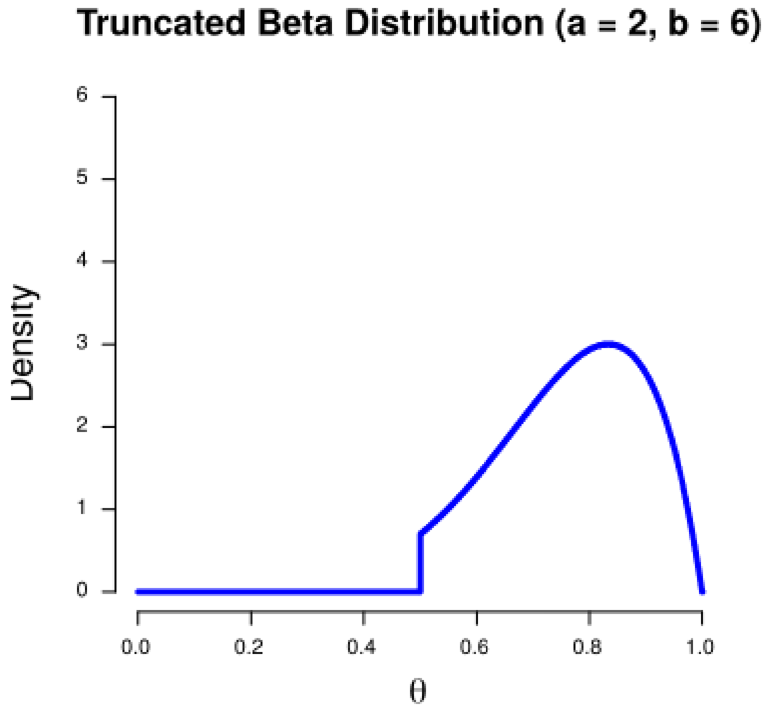
What Does the Alternative Hypothesis Predict?



$$\mathcal{H}_1 : \theta > 0.5$$

The prior formalizes the predictions of the alternative hypothesis: Here, it predicts that only values of θ between 0.5 and 1 are possible, and that values of θ closer to 0.5 are more likely than values close to 1.

What Does the Alternative Hypothesis Predict?

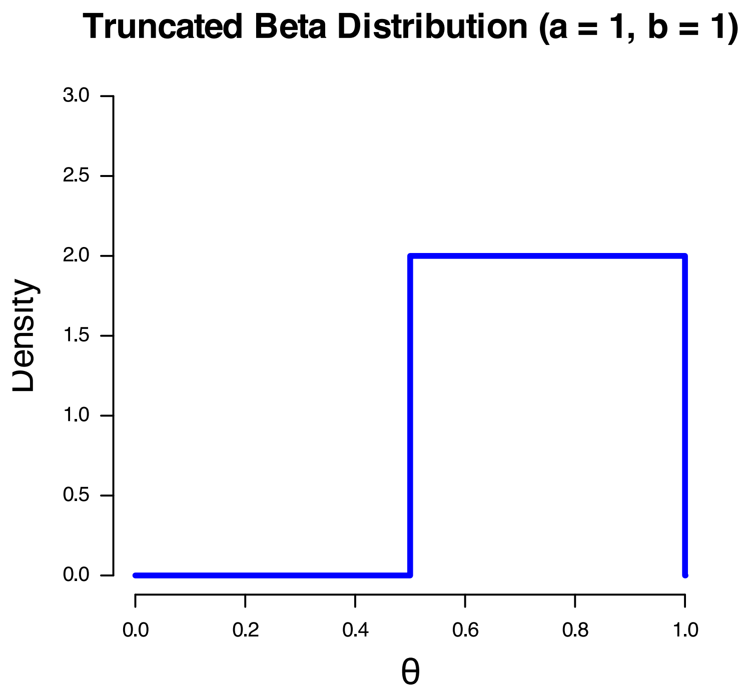


$$\mathcal{H}_1 : \theta > 0.5$$

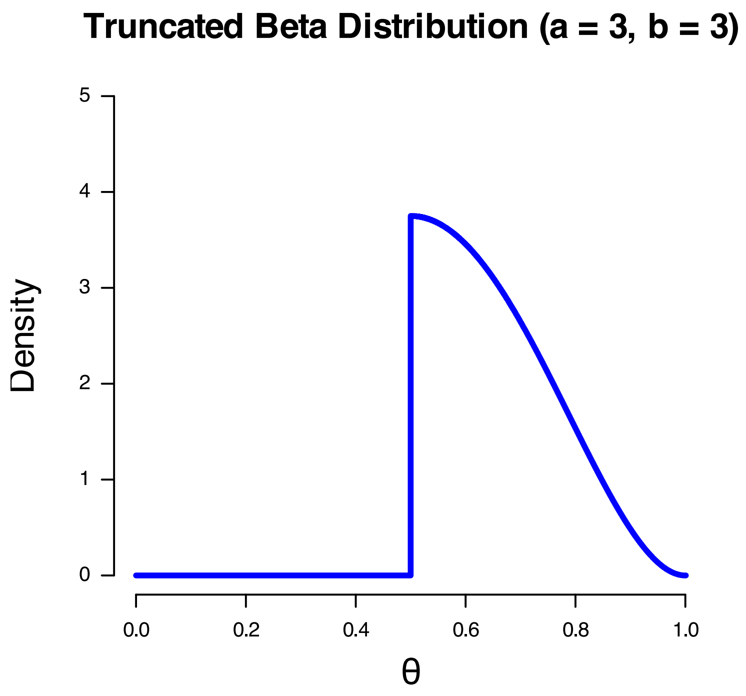
The prior formalizes the predictions of the alternative hypothesis: Here, it predicts that only values of θ between 0.5 and 1 are possible, and that values of θ closer to 1 are more likely than values close to 0.5.

What Does the Alternative Hypothesis Predict?

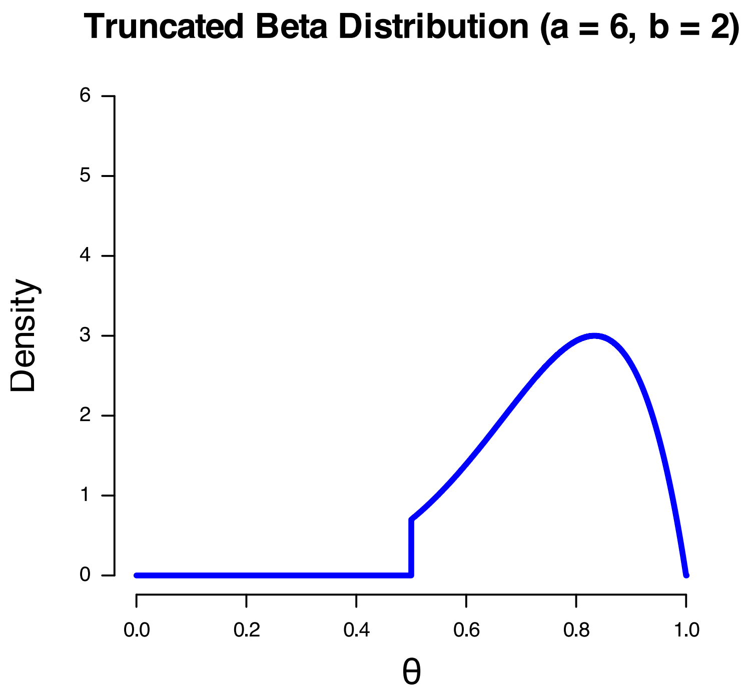
"If you had 100€ to bet on likely values of θ, how would you divide it?"



If the observed proportion (i.e., the data) equals 0.55, it will win **as much** money as if the observed proportion equals 0.8



If the observed proportion (i.e., the data) equals 0.55, it will win **more** money than if the observed proportion equals 0.8

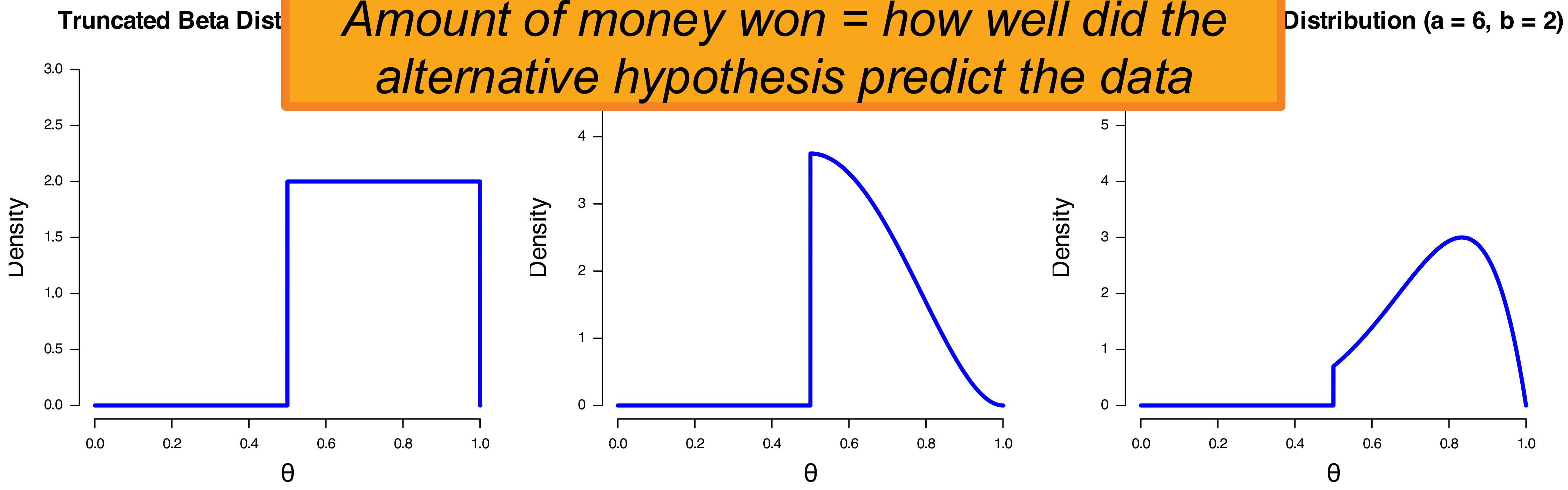


If the observed proportion (i.e., the data) equals 0.55, it will win **less** money than if the observed proportion equals 0.8

What Does the Alternative Hypothesis Predict?

"If you had 100€ to bet on likely values of θ, how would you divide it?"

Amount of money won = how well did the alternative hypothesis predict the data



If the observed proportion (i.e., the data) equals 0.55, it will win as much money as if the observed proportion equals 0.8

If the observed proportion (i.e., the data) equals 0.55, it will win more money than if the observed proportion equals 0.8

If the observed proportion (i.e., the data) equals 0.55, it will win less money than if the observed proportion equals 0.8

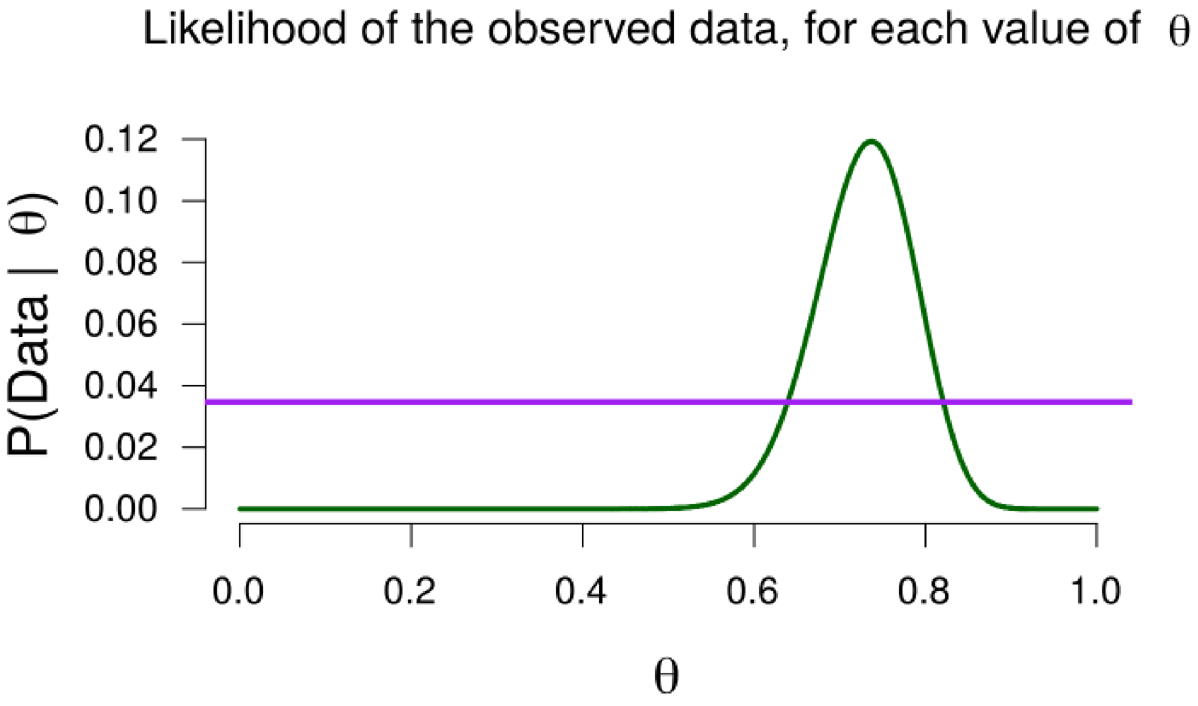
The Marginal likelihood tells us how well the model predicted the data

$$P(\text{data})$$

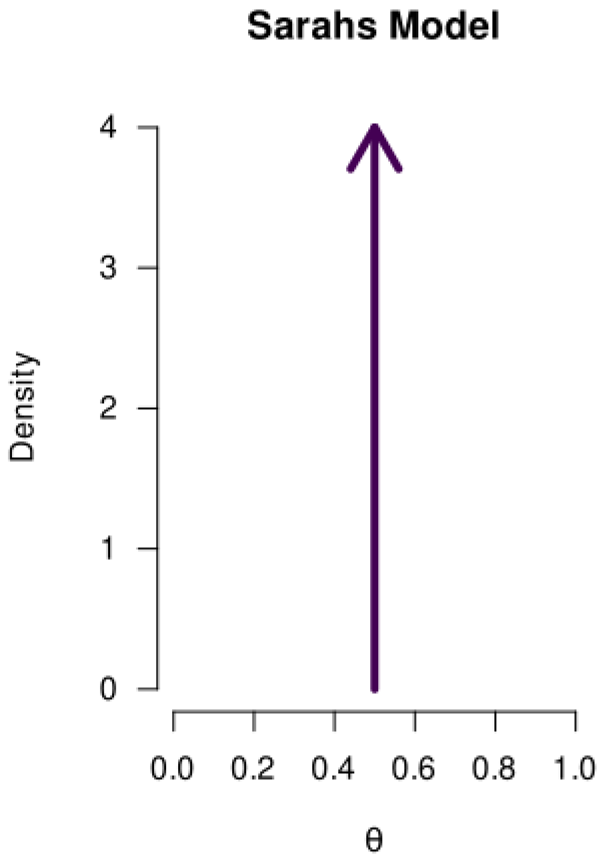
The **marginal likelihood**, across all values of θ

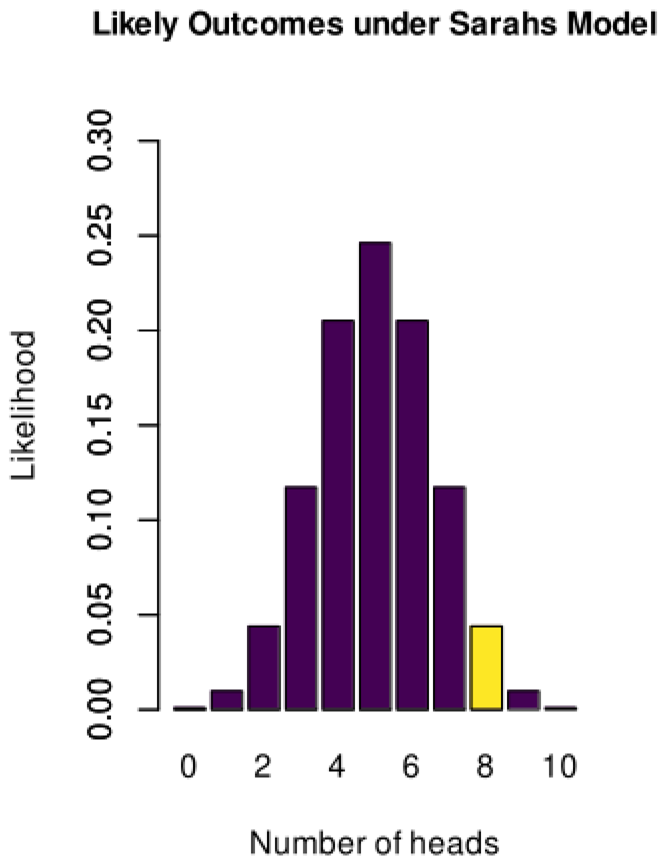
This tells us something how well θ predicted the data, **averaged** over all possible values of θ (i.e., it is the average quality of the prediction)

When averaging, the likelihood for each value of θ is weighted by the prior density of that point (i.e., by how much "money" H_1 bet on that point)



What Does the Null Hypothesis Predict?





"If you had 100€ to bet on likely values of θ, how would you divide it?"

Bayesian Hypothesis Testing: Predictive Updating Factor

The average likelihood across all values predicted by H_1
(i.e., the marginal likelihood)

$$\underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

The average likelihood across all values predicted by H_0 (i.e., the likelihood of the data, given the test value – in this case, $\theta = 0.5$)

A computational shortcut for calculating this factor is the Savage-Dickey density ratio:
- Take the prior density at the point of testing ($\theta = 0.5$)
- Take the posterior density at the point of testing ($\theta = 0.5$)

The predictive updating factor equals the ratio of those numbers!

Bayesian Hypothesis Testing: Predictive Updating Factor

The average likelihood across all values predicted by H_1
(i.e., the marginal likelihood)

$$\underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

The average likelihood across all values predicted by H_0 (i.e., the likelihood of the data, given the test value – in this case, $\theta = 0.5$)

A computational shortcut for calculating this factor is the Savage-Dickey density ratio:
- Take the prior density at the point of testing ($\theta = 0.5$)
- Take the posterior density at the point of testing ($\theta = 0.5$)

The predictive updating factor equals the ratio of those numbers!
This means that (when we look at point of testing):
- If prior density > posterior density, we obtain evidence for the alternative hypothesis
- if prior density < posterior density, we obtain evidence for the null hypothesis

Bayesian Hypothesis Testing: Posterior Odds

$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\substack{\text{Posterior beliefs} \\ \text{about hypotheses}}}$$

The posterior odds specify how plausible a hypothesis is, relative to another hypothesis, **after** seeing the data

Bayesian Hypothesis Testing

$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\substack{\text{Posterior beliefs} \\ \text{about hypotheses}}} = \underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

Bayesian Hypothesis Testing

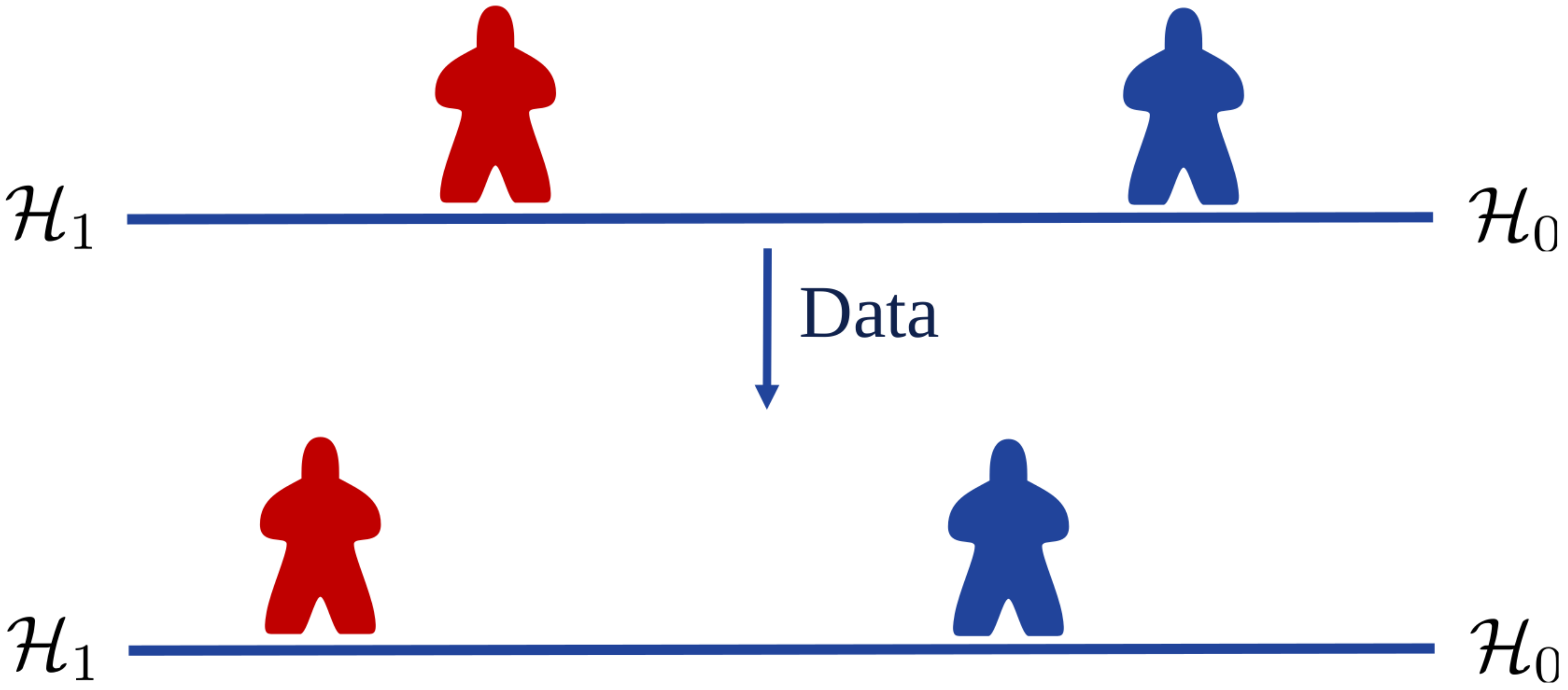
$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\substack{\text{Posterior beliefs} \\ \text{about hypotheses}}} = \underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}} \times \boxed{\underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}}$$

Bayes factor BF_{10}

Bayesian Hypothesis Testing

Different starting points (i.e., different prior beliefs)

Bayesian Hypothesis Testing



Bayesian Hypothesis Testing

Different updated locations (i.e., different posterior beliefs)

Bayesian Hypothesis Testing

But they take a step of equal size in the same direction:
Change in beliefs, quantified by the Bayes factor, is the same for both

Bayes Factor Interpretation

- **Example 1: BF_{10} = 20**

The data are twenty times more likely under H_1 than under H_0

- **Example 2: BF_{10} = 1/5 → BF_{01} = 5**

The data are five times more likely under H_0 than under H_1

- **Example 3: BF_{10} = 1**

The data are equally likely under H_1 and under H_0

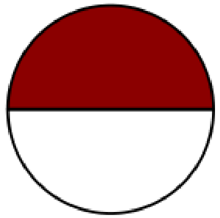
Bayes Factor Interpretation

BF	Evidence
1 – 3	Anecdotal
3 – 10	Moderate
10 – 30	Strong
30 – 100	Very strong
>100	Extreme

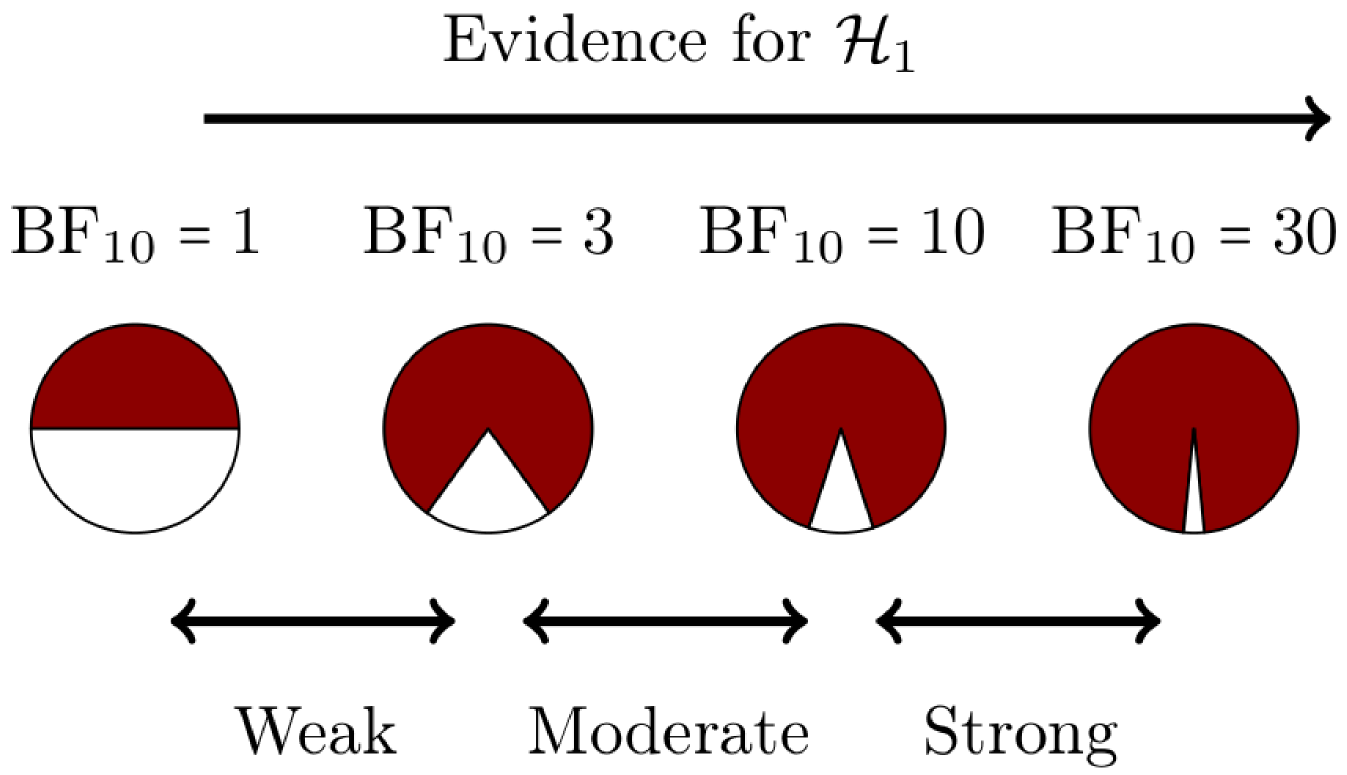
The Bayes factor is a **continuous** metric for relative predictive quality, these classifications are merely for providing intuition!

Bayes Factor Interpretation

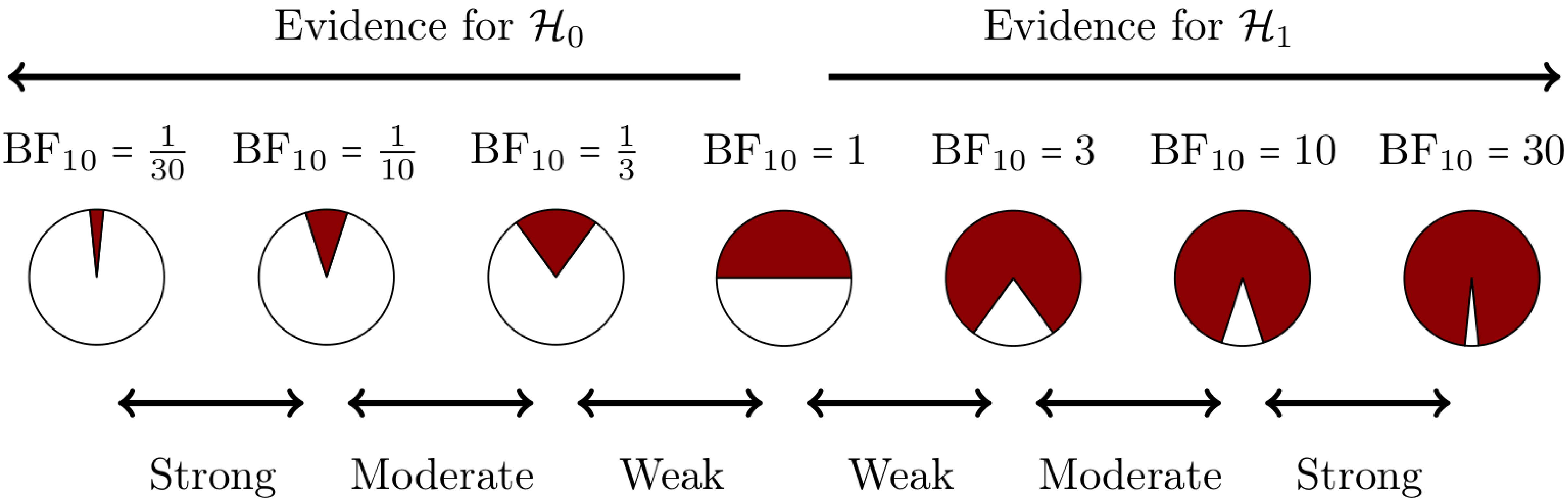
$\mathrm{BF}_{10} = 1$



Bayes Factor Interpretation



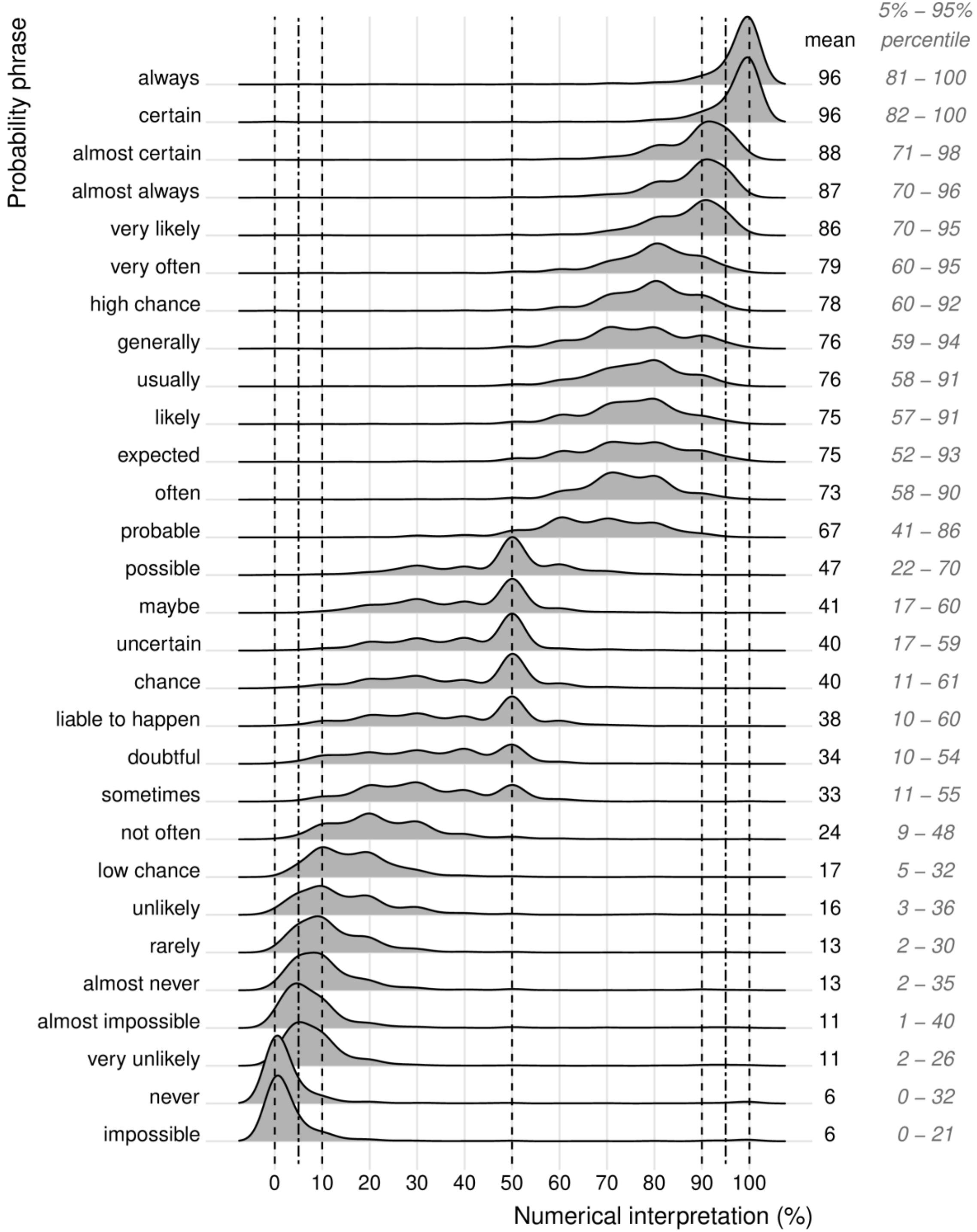
Bayes Factor Interpretation



For a comic that gives more intuition about interpretation, click here

The general challenge of translating between words and numbers:

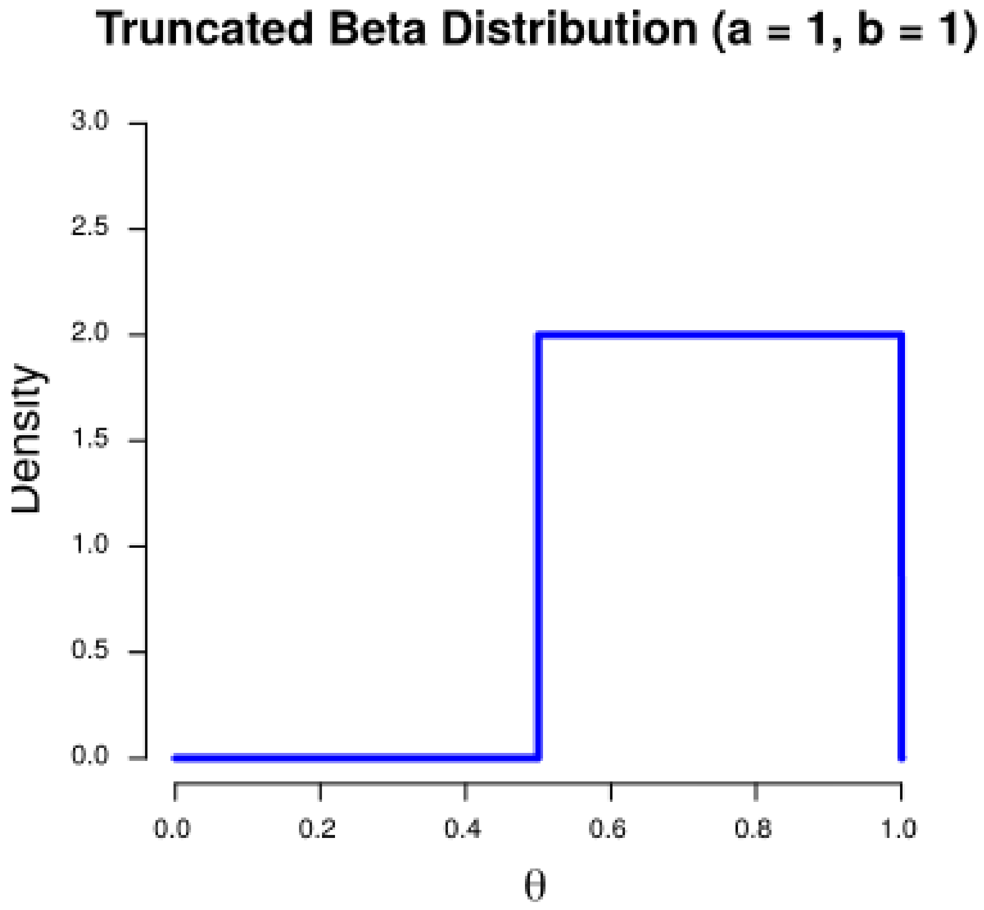
People have different interpretations of probabilistic words/phrases



Source:
[Willems, S., Albers, C., & Smeets, I. (2020). Variability in the interpretation of probability phrases used in Dutch news articles — a risk for miscommunication. Journal of Science Communication, 19.]

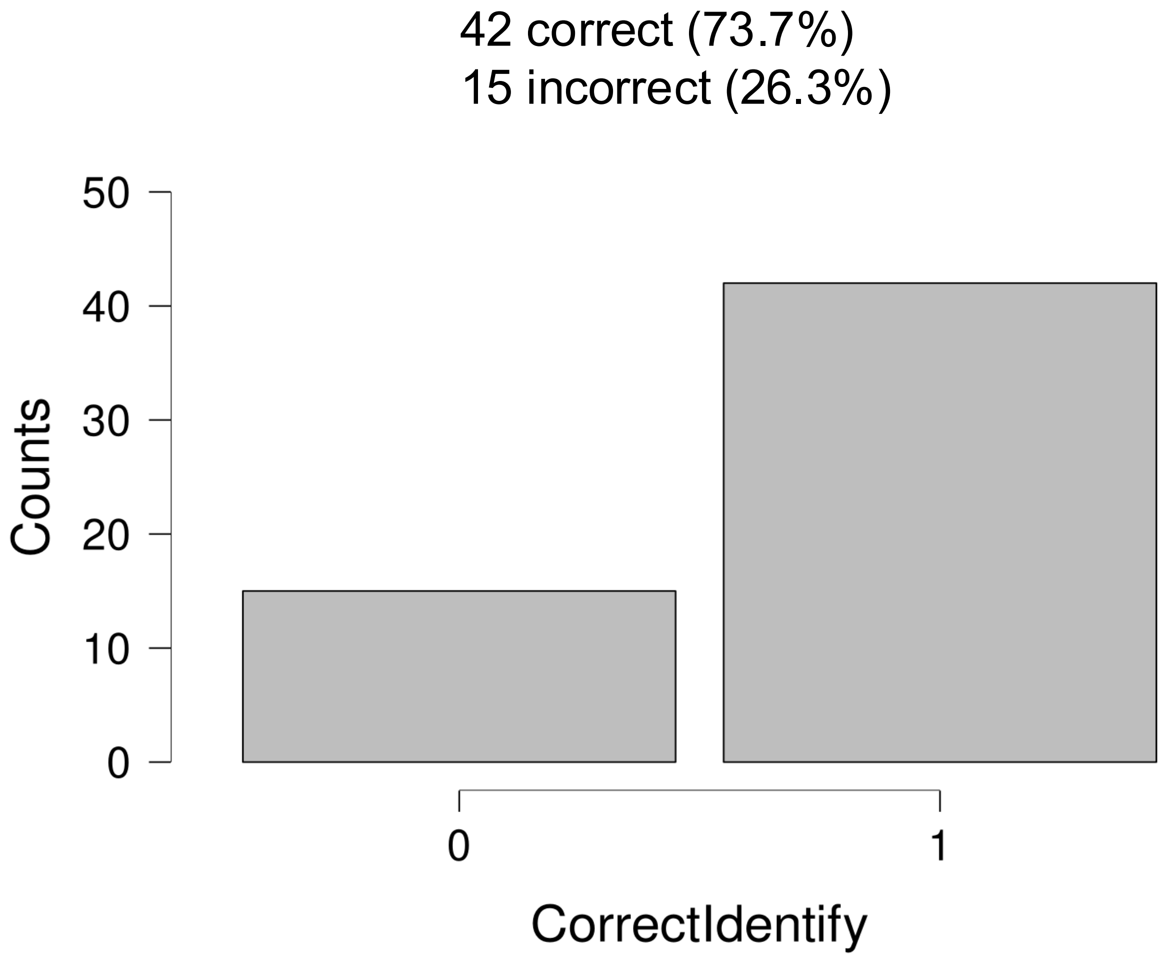
Example: Beer Tasting Experiment - Setup

$$\mathcal{H}_1 : \theta > 0.5$$

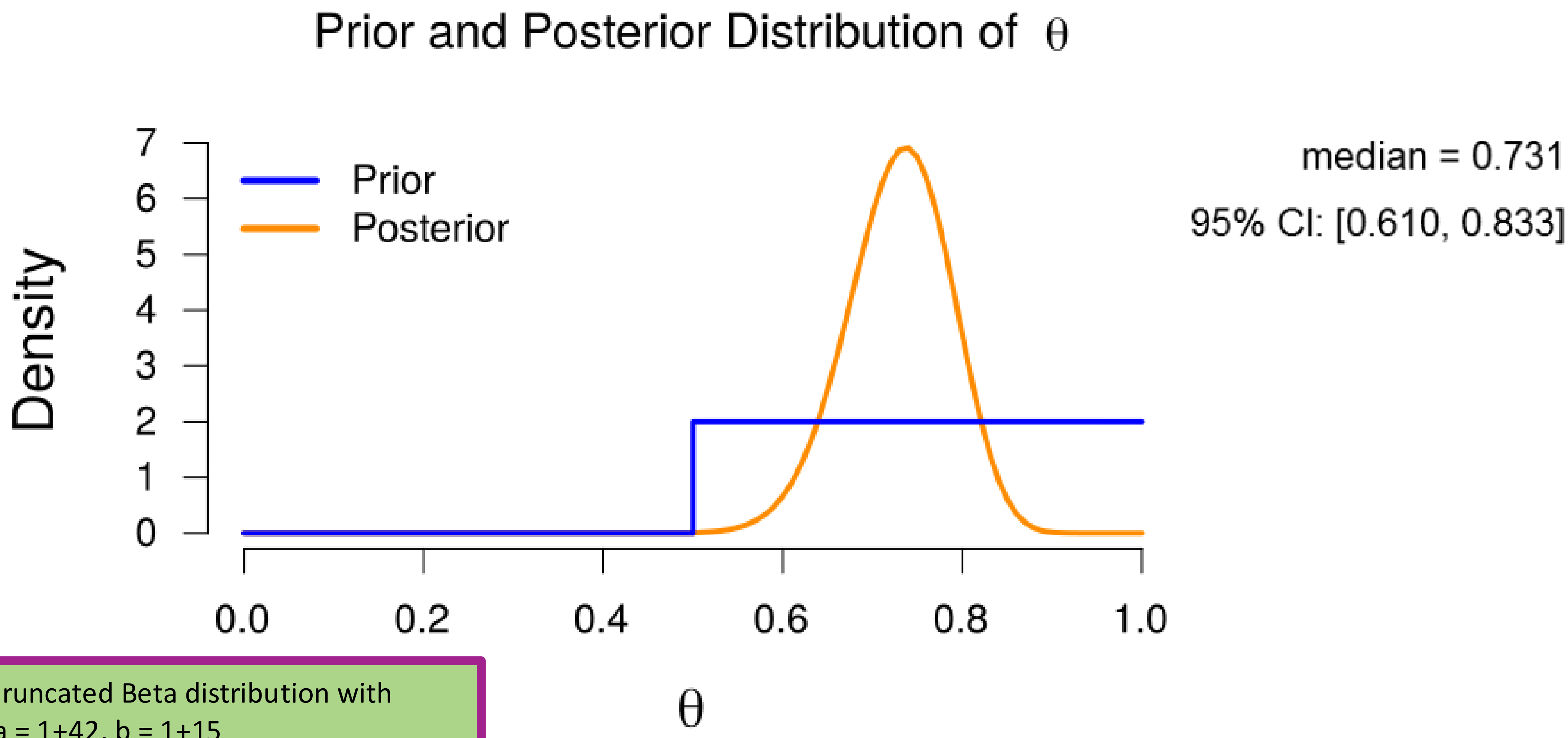


https://researchmethodsuva.shinyapps.io/test/

Example: Beer Tasting Experiment - Data

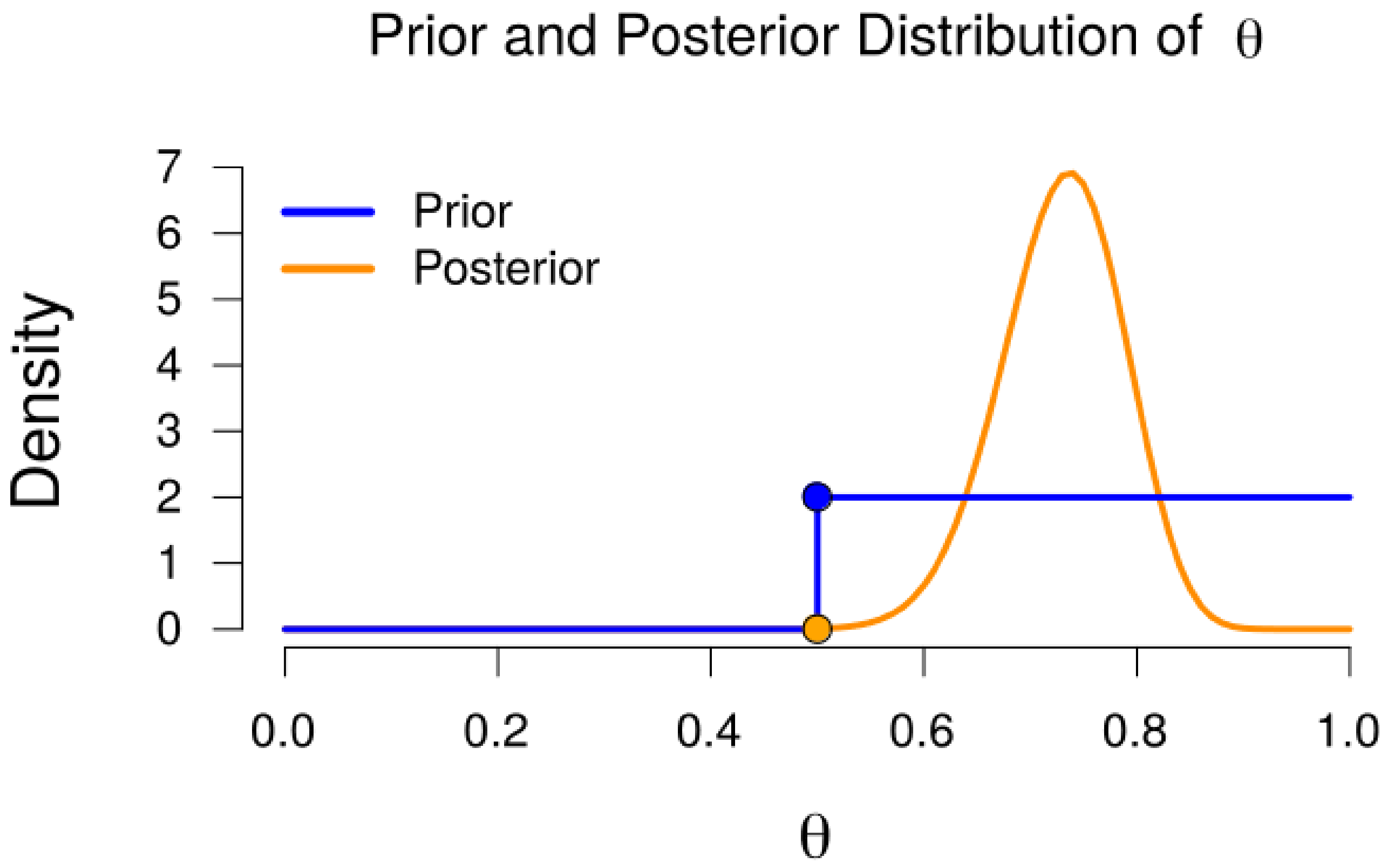


Example: Beer Tasting Experiment - Posterior



Posterior is a truncated Beta distribution with
a = 1+42, b = 1+15
(see computational trick last lecture, slide 55)

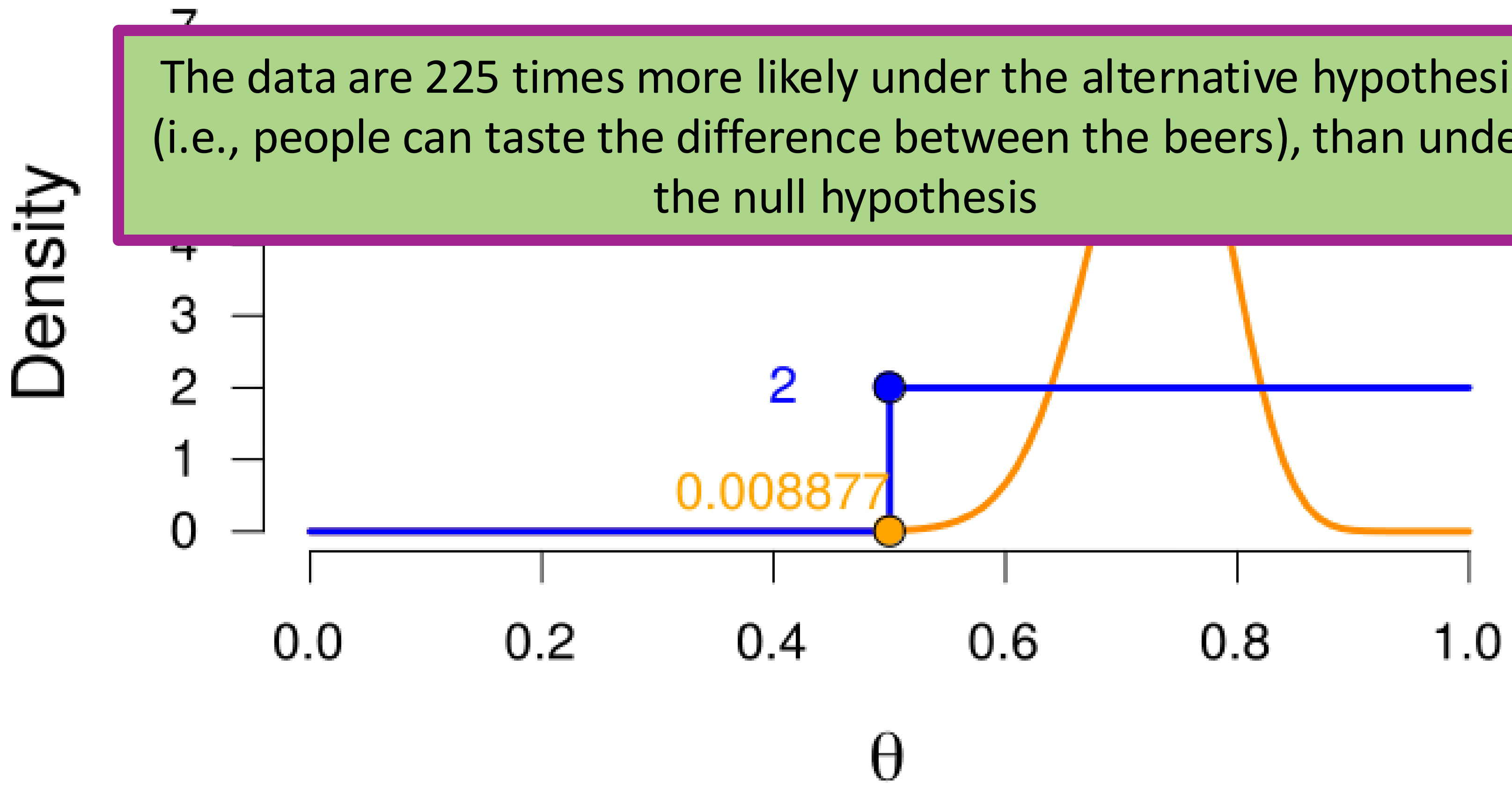
Example: Beer Tasting Experiment - Posterior



Example: Beer Tasting Experiment - Posterior

$$\text{Savage-Dickey: Bayes factor } \mathrm{BF}_{10} = \frac{2}{0.008877} = 225.3$$

The data are 225 times more likely under the alternative hypothesis (i.e., people can taste the difference between the beers), than under the null hypothesis



Example: Beer Tasting Experiment – Bayes Factor

The average likelihood across all values predicted by H_1
(i.e., the marginal likelihood)

Marginal likelihood = 0.03448

$$\underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

The average likelihood across all values predicted by H_0 (i.e., the likelihood of the data, given the test value (in this case, θ = 0.5)

$$\boxed{P(x = 42 \mid \theta = 0.5) = \frac{57!}{42!(57-15)!} 0.5^3 (1-0.5)^{57-15} = 0.000153}$$

$$\mathrm{BF}_{10} = \frac{0.03448}{0.000153} \approx 225.3$$

What Can We Do Now?

- We can **estimate** the population parameter
  - Credible interval & Posterior median

- We can **test** whether the parameter is equal to a certain value
  - Bayes factor

Both are based on the posterior distribution of the parameter!

Today

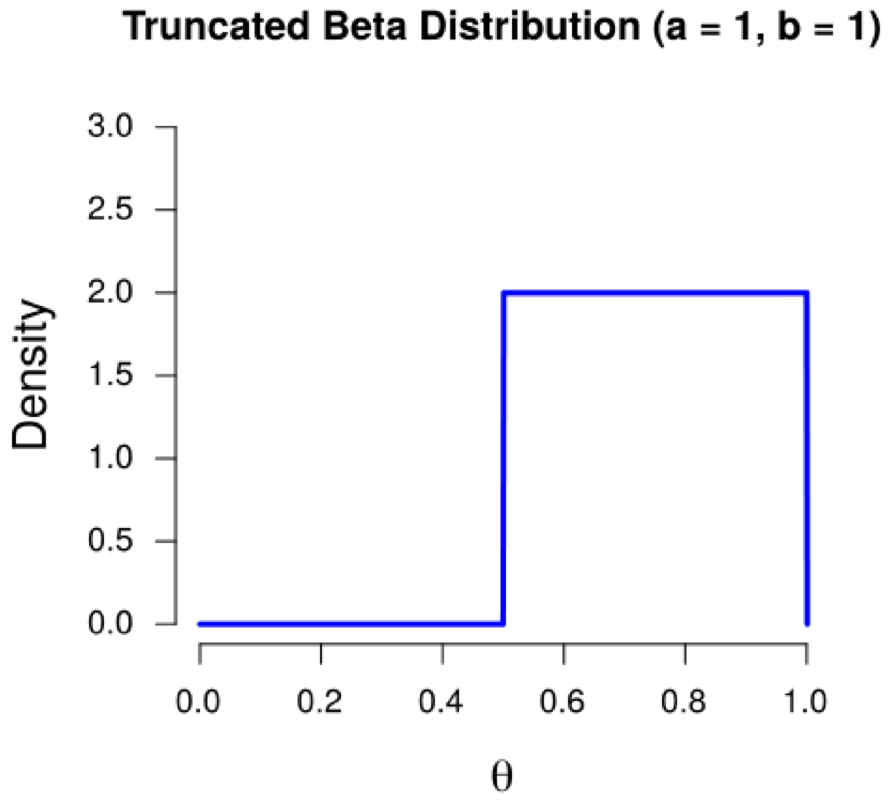
- Recap of last week
- Bayesian Hypothesis Testing
  - Basic concepts
  - Testing a proportion
  - **About the Bayes factor**
- Recap
  - Practical stuff & next week
  - Example exam question

Two-Sided Testing Vs. One-Sided Testing

- We just used the one-sided (positive) alternative hypothesis
- What would happen if we use a two-sided test?
- And what about a one-sided (negative) alternative hypothesis?

Two-Sided Testing Vs. One-Sided Testing

The alternative hypothesis governs the prior distribution. If it is one-sided, the prior becomes truncated: all values not predicted by the H_1 have a density of 0

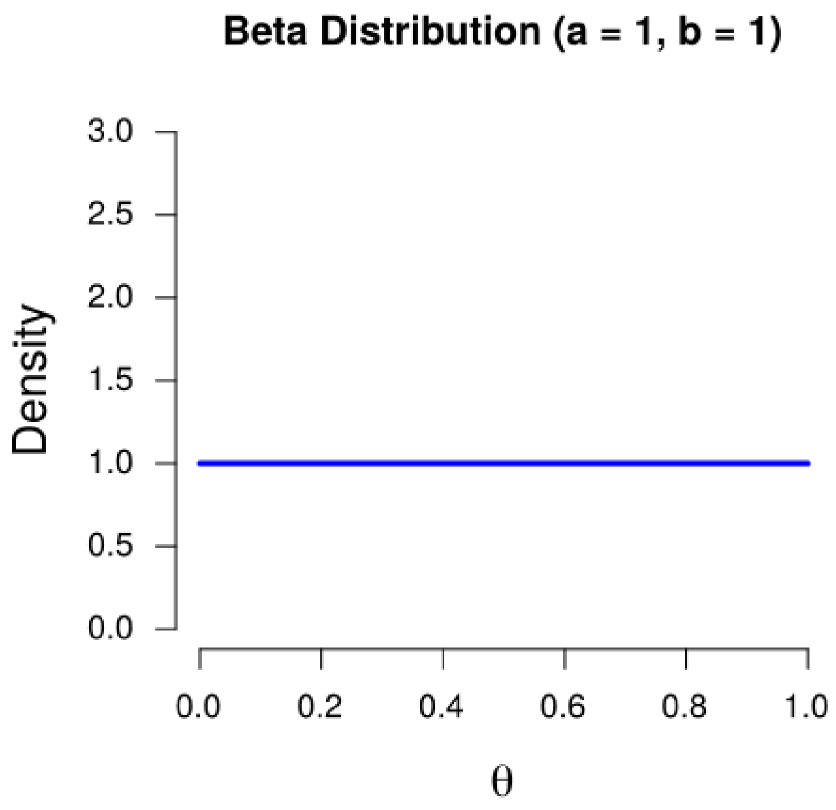


$$\mathcal{H}_1 : \theta > 0.5$$

The prior formalizes the predictions of the alternative hypothesis: Here, it predicts that all values of θ between 0.5 and 1 are equally likely.

“If you had 100€ to bet on likely values of θ, how would you divide it?”

Two-Sided Testing Vs. One-Sided Testing



$$\mathcal{H}_1 : \theta \neq 0.5$$

The prior formalizes the predictions of the alternative hypothesis: Here, it predicts that all values of θ between 0 and 1 are equally likely.

This means we bet on more values, so our bets are a lot more spread out!

"If you had 100€ to bet on likely values of θ*, how would you divide it?"*

Two-Sided Testing Vs. One-Sided Testing

The **marginal likelihood**, across all values of θ tells us something how well each θ predicted the data, **averaged** over all possible values of θ (i.e., it is the average quality of the model)

When averaging, the likelihood for each value of θ is weighted by the prior density of that point (i.e., by how much "money" H_1 bet on that point)

Because the two-sided H_1 spread out its bets a lot more, it received less winnings for the values of theta that predicted the data well (values close to the observed proportion of 0.737)
→ *Its marginal likelihood will be lower*

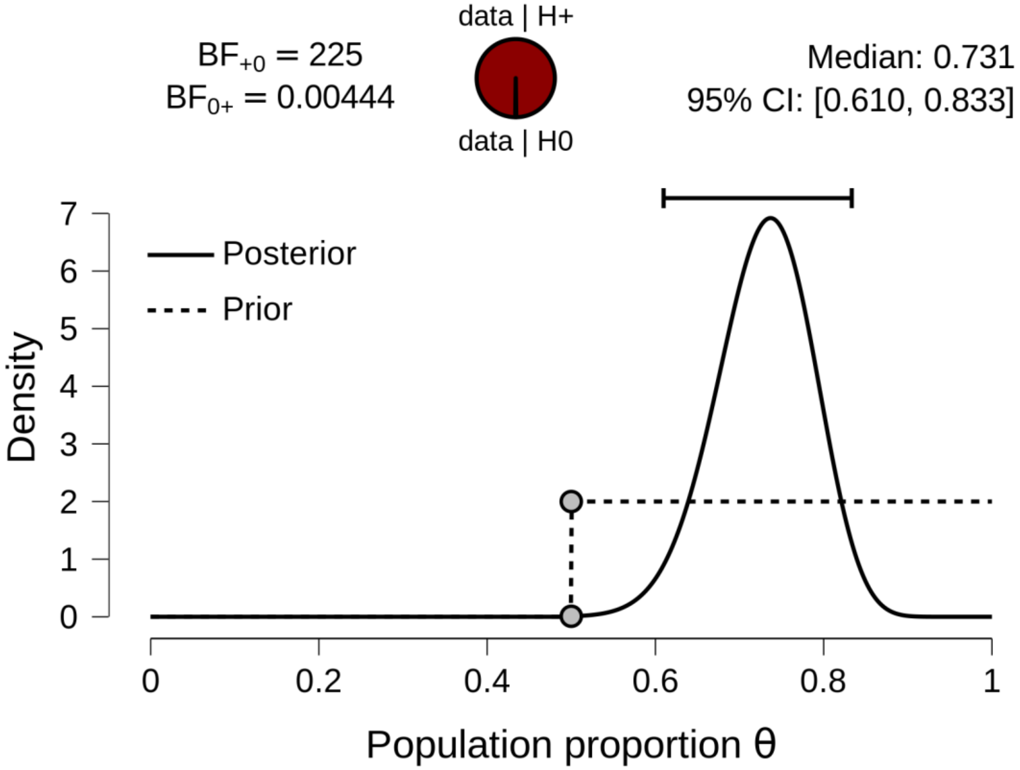
$$\mathcal{H}_1 : \theta \neq 0.5$$

The prior formalizes the predictions of the alternative hypothesis: Here, it predicts that all values of θ between 0 and 1 are equally likely.

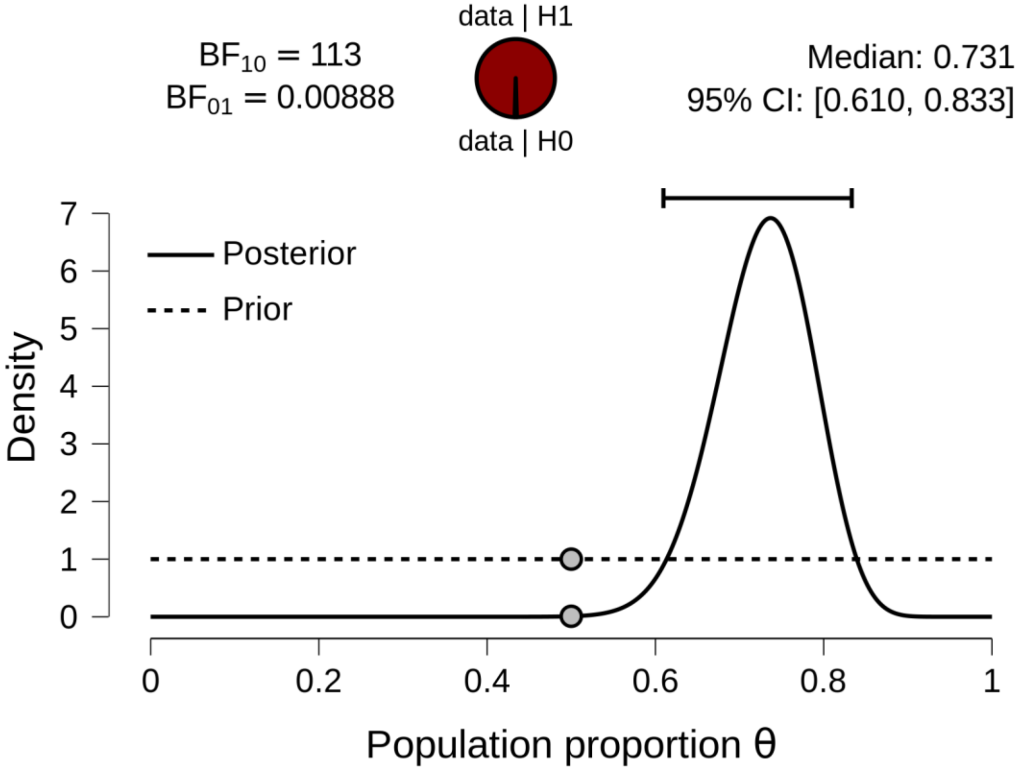
This means we bet on more values, so our bets are a lot more spread out!

"If you had 100€ to bet on likely values of θ*, how would you divide it?"*

Two-Sided Testing Vs. One-Sided Testing

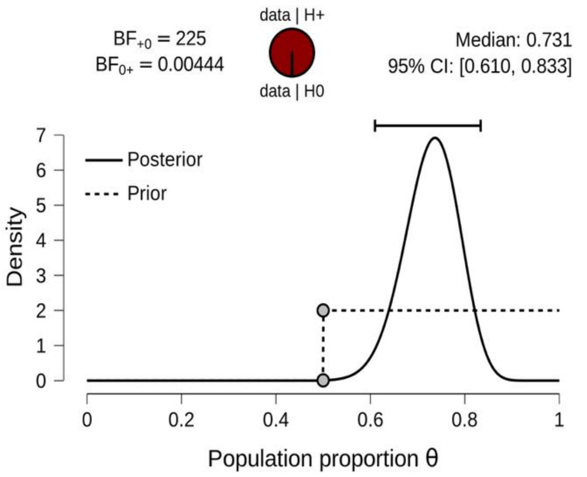


$$\mathcal{H}_1 : \theta > 0.5$$



$$\mathcal{H}_1 : \theta \neq 0.5$$

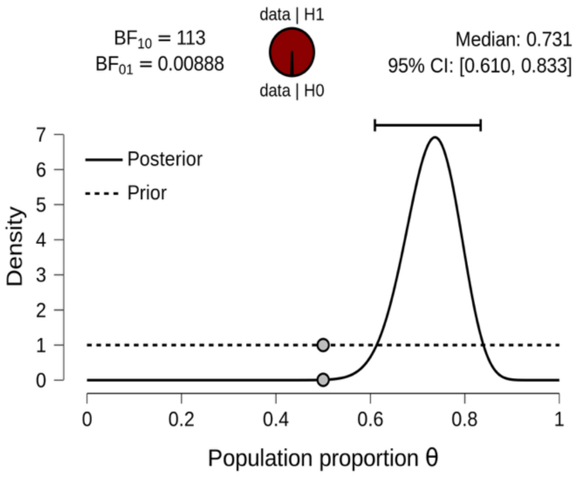
Two-Sided Testing Vs. One-Sided Testing



$$\mathcal{H}_1 : \theta > 0.5$$

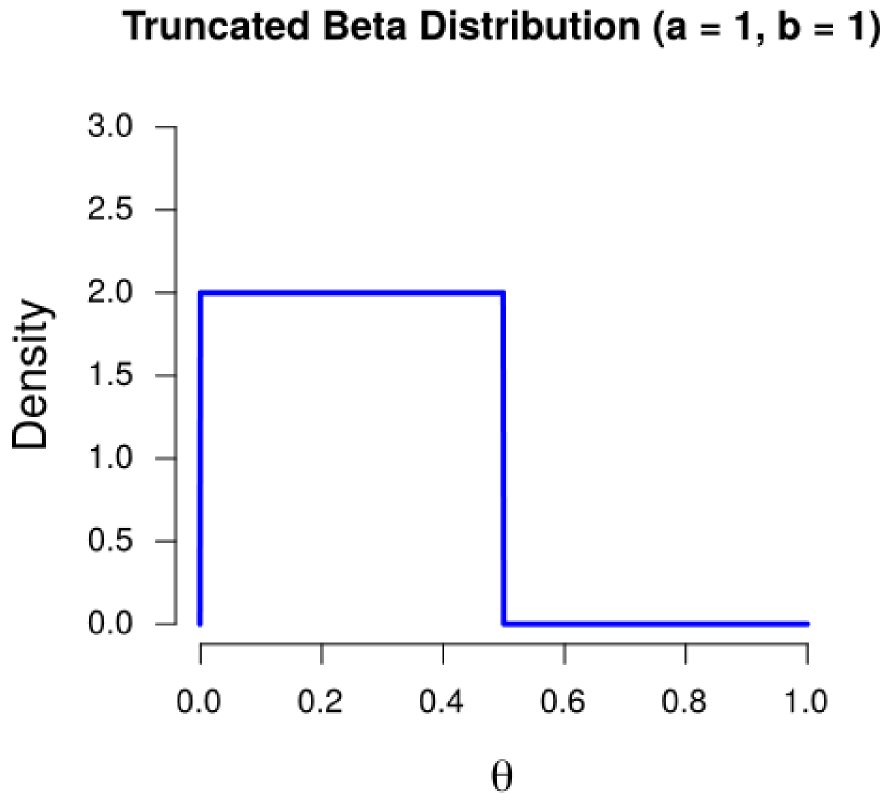
The positive one-sided hypothesis made a more specific prediction than the two-sided hypothesis. The values it **did** predict, predicted the data very well.

The concept of rewarding a more specific model, while it predicted the data equally well, is known as **parsimony**



$$\mathcal{H}_1 : \theta \neq 0.5$$

And what about a one-sided negative alternative hypothesis?

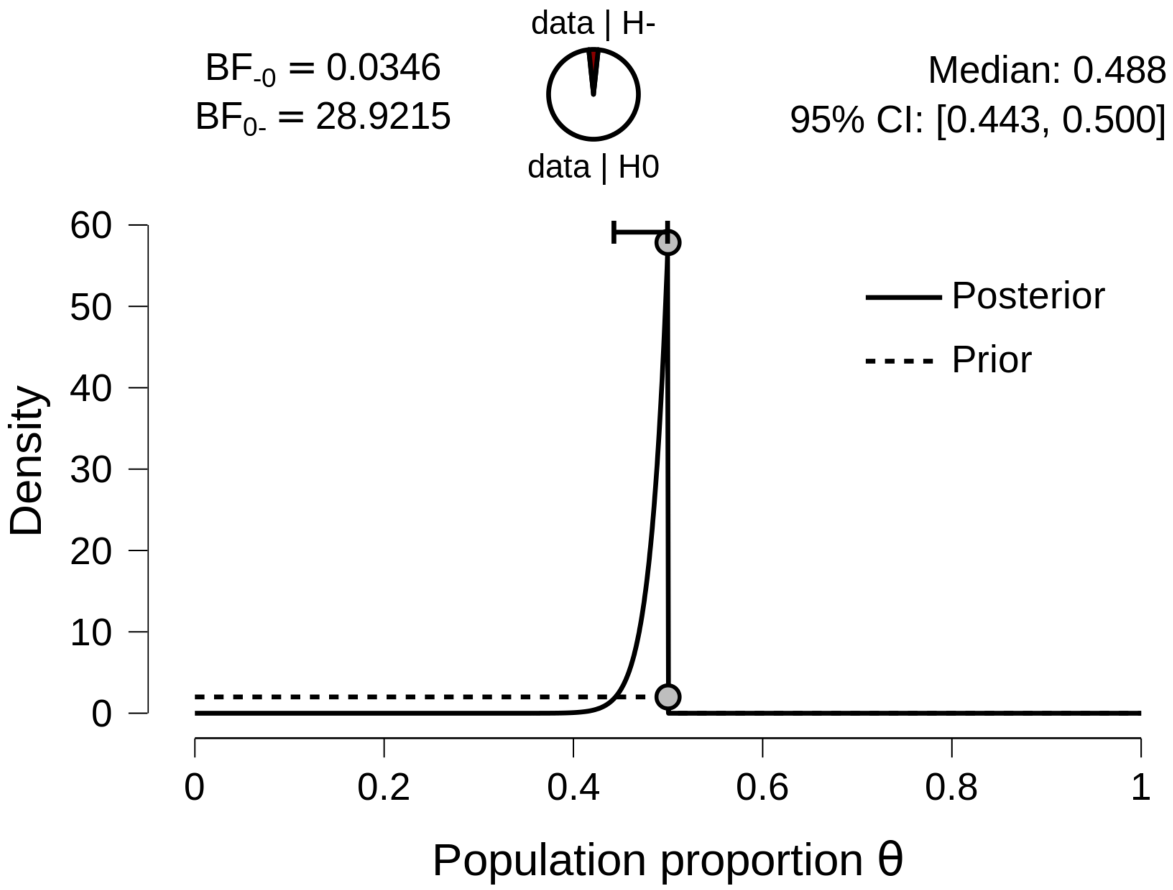


$$\mathcal{H}_1 : \theta < 0.5$$

The prior also formalizes the predictions of the alternative hypothesis: Here, it predicts that all values of θ between 0 and 0.5 are equally likely.

"If you had 100€ to bet on likely values of θ*, how would you divide it?"*

And what about a one-sided negative alternative hypothesis?



$$\mathcal{H}_1 : \theta < 0.5$$

This negative one-sided hypothesis predicts the data very poorly: it only bet money on values of θ that lead to a low likelihood. Values of θ close to 0.5 predicted the data well, **relative to those other values**, so the posterior is very peaked there.

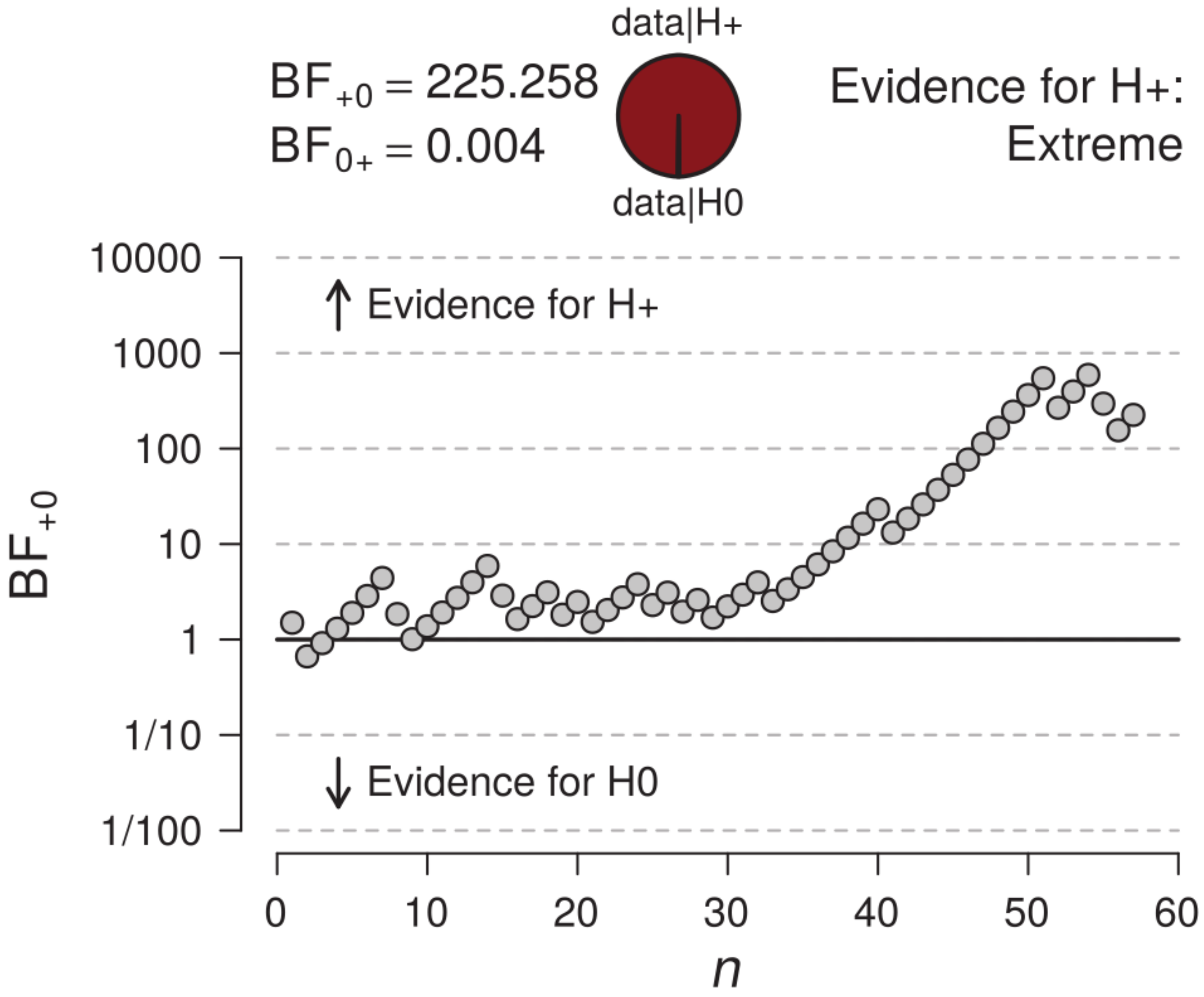
Here, H_0 outperforms H_1! This illustrates the **relative nature of the Bayes factor:** a high Bayes factor does not mean the hypothesis is true, it just means it predicted the data better (or less poorly) than the other hypothesis!

Sequential Analysis

- In the Bayesian framework we keep updating our beliefs
- It does not matter if we update it all at once, or one data point at a time (“Today’s posterior is tomorrow’s prior”)

Sequential Analysis

This means we can look at a plot of how our Bayes factor evolves as we accumulate knowledge!



Live Demonstration

www.jasp-stats.org

Today

- Recap of last week
- Bayesian Hypothesis Testing
  - Basic concepts
  - Testing a proportion
  - Live demonstration
- **Recap**
  - Practical stuff & next week
  - Example exam question

Recap

- Bayesian Hypothesis testing is another form of updating beliefs: we compare the predictions made by 2 different **hypotheses (or, models)** to update our beliefs about which hypothesis is better

- The Bayes factor is central: it is the predictive updating factor of our beliefs about hypotheses

- It is the ratio of each hypothesis' "predictive quality", measured by their marginal likelihoods: the average likelihood of all values of the parameter predicted by each respective hypothesis.

Recap

- The Bayes factor is a relative metric! Both hypothesis can predict very poorly: the Bayes factor tells you which did the least poorly

- The Bayes factor can be monitored as evidence accumulates

- We can investigate the effects of the prior distribution (more next lecture)

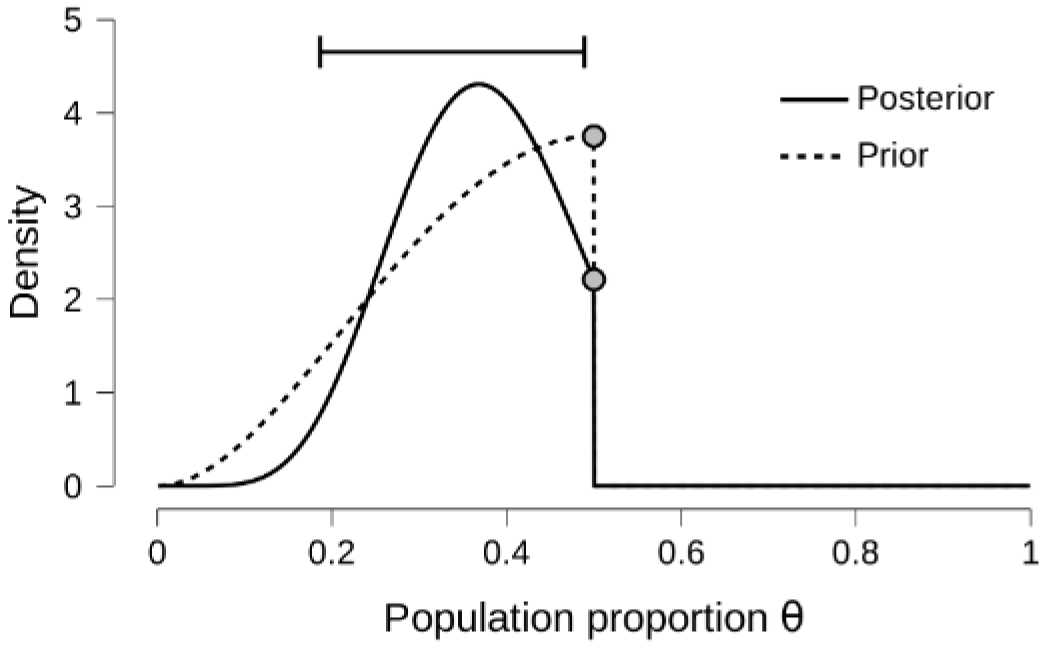
Source: https://www.bayesianspectacles.org/library/

Practical Stuff

- No formula sheet for Bayes (Bayesian questions more conceptual)
  - This week Bayesian WA, next additional WA with old Bayesian exam questions
- JASP
  - On the exam you only need to be able to work with the "Summary Statistics" Module (binomial test, t-test, correlation)
  - For example, Questions in WA, Question 6.4 in book
  - Other questions will provide JASP output and ask you to interpret

Example Exam Question

Someone tests whether a proportion is less than 0.5 and computes the following posterior distribution. What can we say about the Bayes factor?



a) Bayes factor $BF_{-0} > 1$ (i.e., will favor the negative one-sided hypothesis)

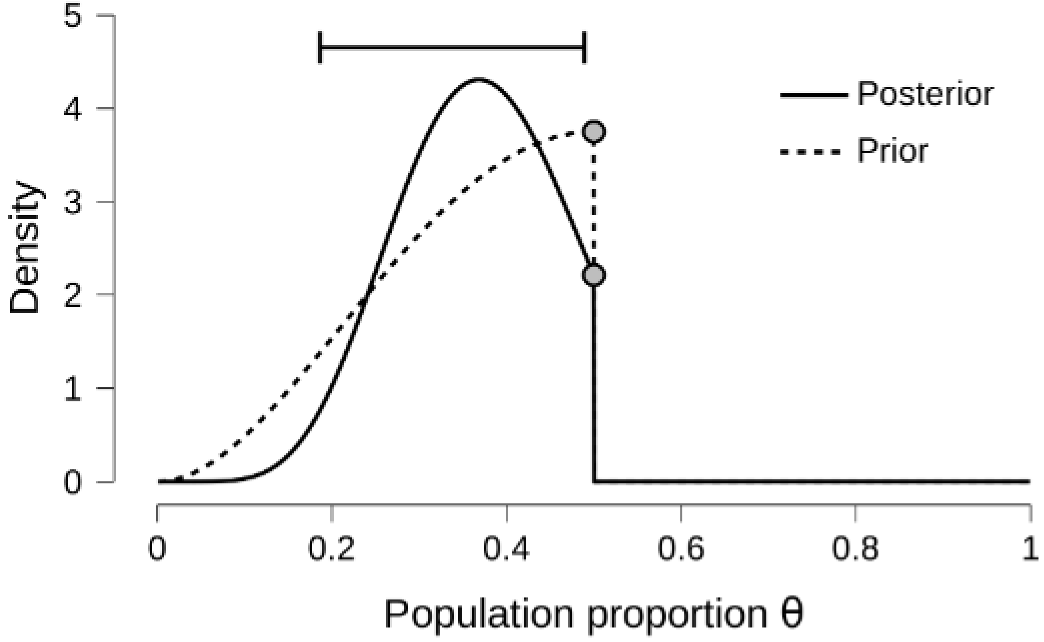
b) Bayes factor $BF_{-0} < 1$ (i.e., will favor the null hypothesis)

c) The Bayes factor BF_{-0} will equal 1

Example Exam Question

Someone tests whether a proportion is less than 0.5 and computes the following posterior distribution. What can we say about the Bayes factor?

Correct!
(posterior density for θ = 0.5) < (prior density for θ = 0.5)



a) **Bayes factor $BF_{-0} > 1$ (i.e., will favor the negative one-sided hypothesis)**

b) Bayes factor $BF_{-0} < 1$ (i.e., will favor the null hypothesis)

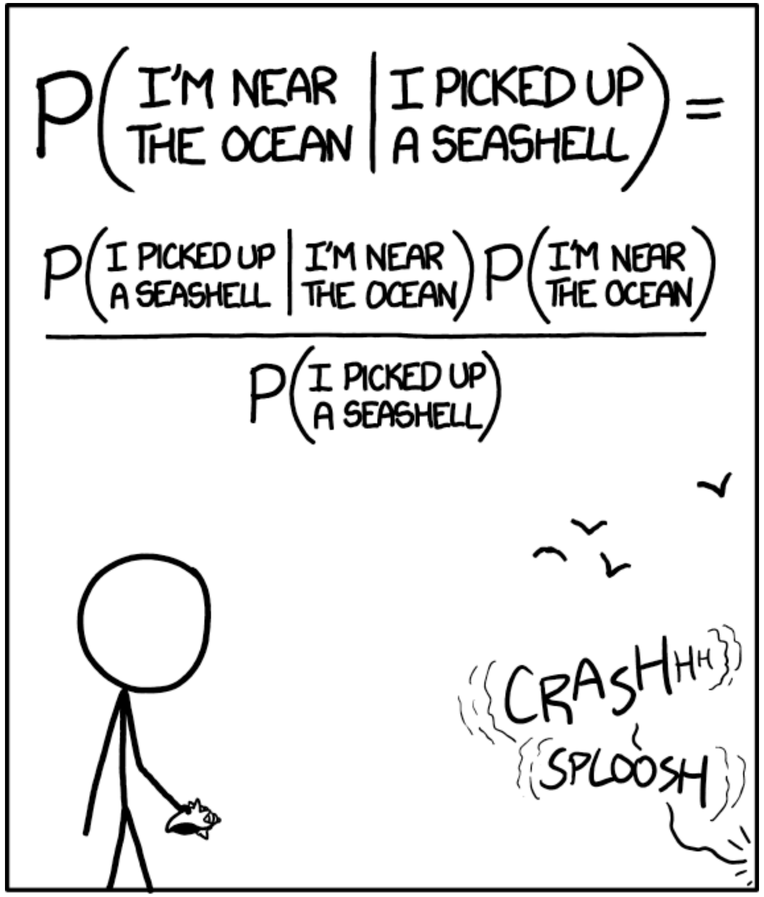
c) The Bayes factor BF_{-0} will equal 1

Thursday

- Prior distribution: curse or blessing?
- Look at Bayesian versions of t-test and correlation
- Statistics in the wild (frequentist and Bayesian)

Questions?

Thank you for your attention



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

https://xkcd.com/1236/

Bonus Article

For those who want to read a more elaborate article about how to apply Bayesian analysis in the real world – me and my labmates wrote this paper about applied guidelines for conducting and reporting a Bayesian analysis. It covers the full empirical cycle (from planning to reporting) in a general way, and provides some more context to the stuff we have talked about in class.