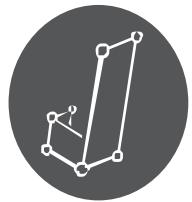
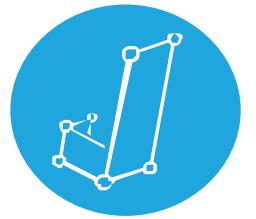


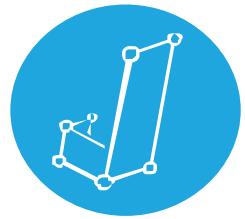
Theory and Practice of Bayesian Inference Using JASP



JASP

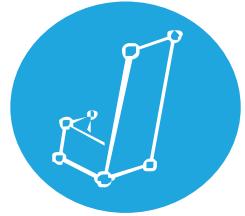
Johnny van Doorn

Practical Stuff

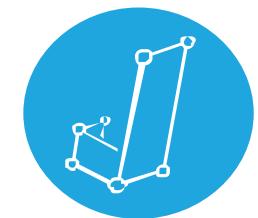


- Did everyone manage to install JASP 0.18.3?
- Can everyone access the workshop OSF page?
 - ▶ <https://osf.io/u2e9d/>
- Introduce myself

Outline



- Statistical Models
- How do Models Learn from Data?
- Comparing Statistical Models
- Extending to other Tests
- Some Practice



June 4, 2004

Magician-turned-mathematician uncovers bias in coin flipping

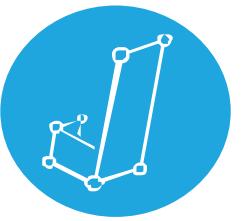
BY ESTHER LANDHUIS



Persi Diaconis has spent much of his life turning scams inside out. In 1962, the then 17-year-old sought to stymie a Caribbean casino that was allegedly using shaved dice to boost house odds in games of chance. In the mid-1970s, the upstart statistician exposed some key problems in ESP research and debunked a handful of famed psychics. Now a Stanford professor of mathematics and statistics, Diaconis has turned his attention toward simpler phenomena: determining whether coin flipping is random. Could a simple coin toss -- used routinely to decide which team gets the ball, for instance -- actually be rigged?

Source: <https://news.stanford.edu/pr/2004/diaconis-69.html>

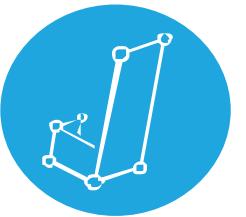
<https://osf.io/6a5hy/>



What is Bayesian Inference?

What are the tools we need?

What are Statistical Models?



A statistical model is a combination of a general statistical model (e.g., the binomial model) and a **statement** about a parameter value that describe a certain phenomenon

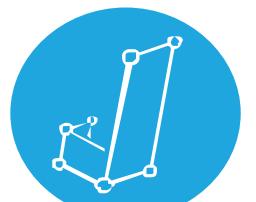
The general binomial model describes a series of chance-based events with a binary outcome, and is governed by a single parameter θ

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

For instance, for flipping a coin:
A binomial model with $\theta = 0.5$

What is the theoretical implication of this model?

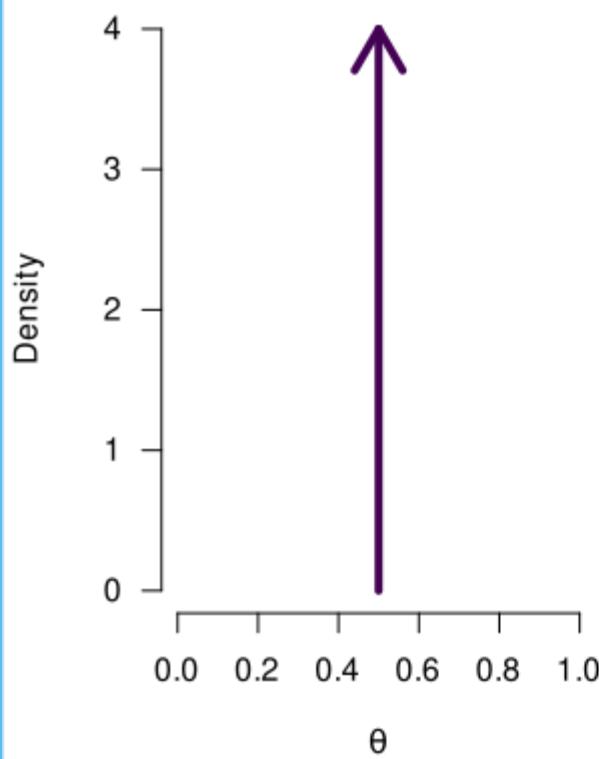
What are Statistical Models?



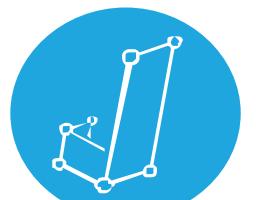
A binomial model with $\theta = 0.5$

We can reflect a model's statement by means of a probability distribution

Sarahs Model



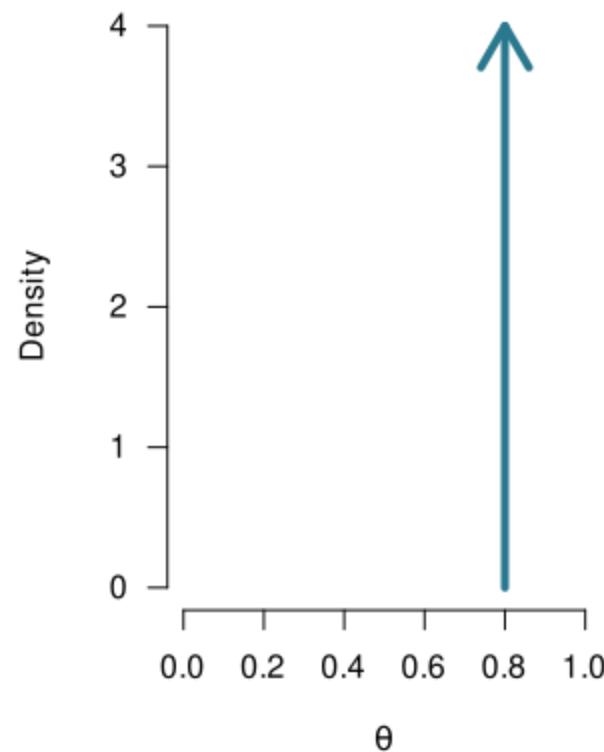
What are Statistical Models?



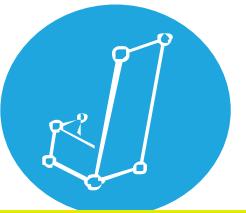
A binomial model with $\theta = 0.8$

We can reflect a model's statement by means of a probability distribution

Pauls Model



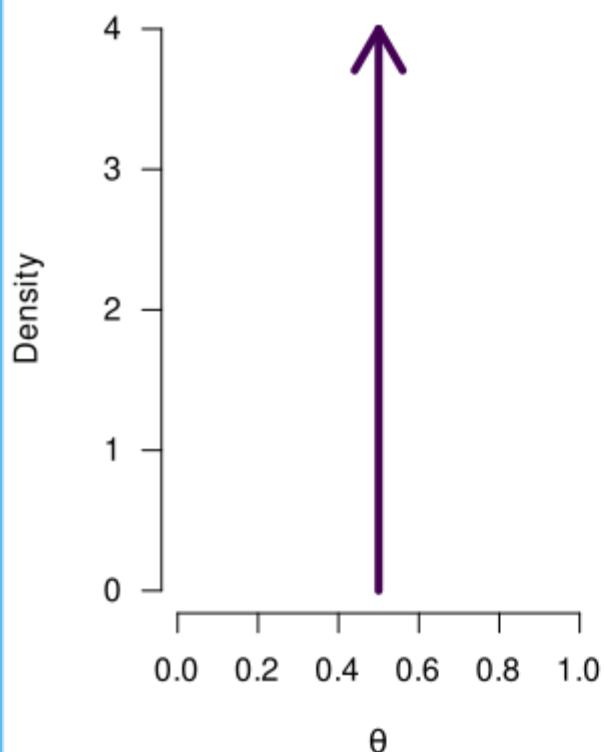
Statistical Models Make Predictions



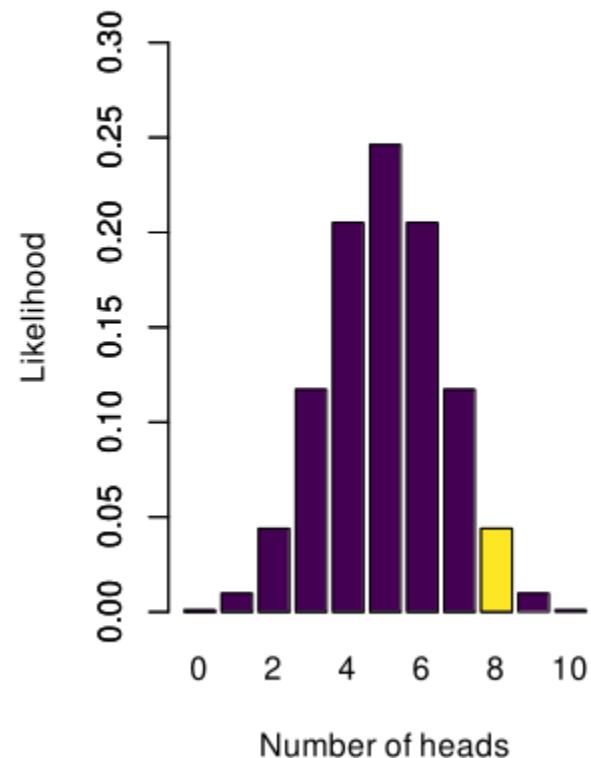
Based on what these models claim about θ , certain outcomes are more/less likely

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

Sarahs Model



Likely Outcomes under Sarahs Model

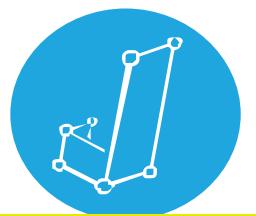


$$\frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k}$$

To create this figure, we take the binomial formula, and fill in $\theta = 0.5$

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

Statistical Models Make Predictions



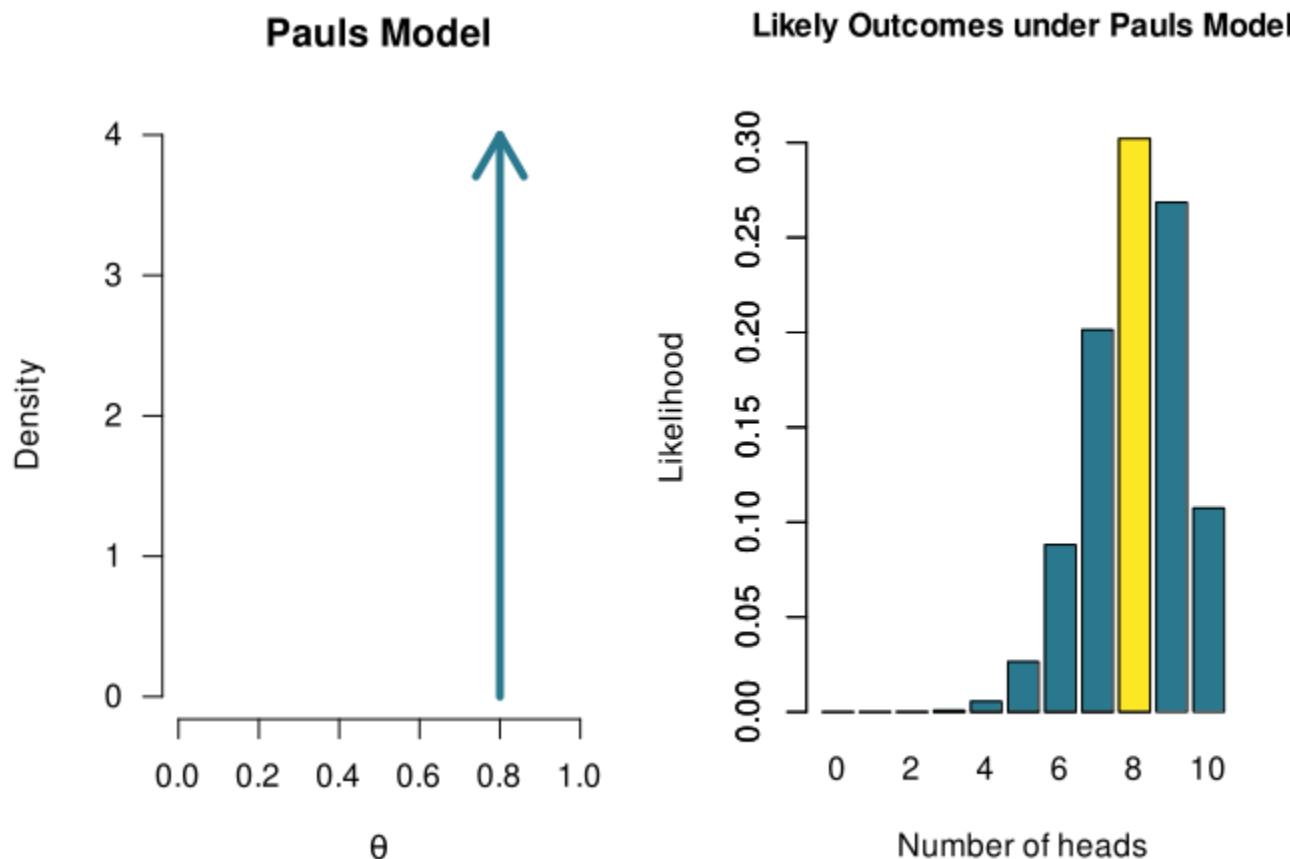
Based on what these models claim about θ , certain outcomes are more/less likely

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

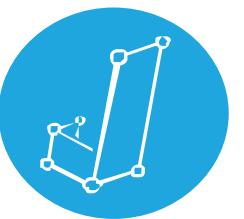
$$\frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k}$$

To create this figure, we take the binomial formula, and fill in $\theta = 0.8$

For instance, for an outcome of $k=8$ heads, the formula gives 0.302

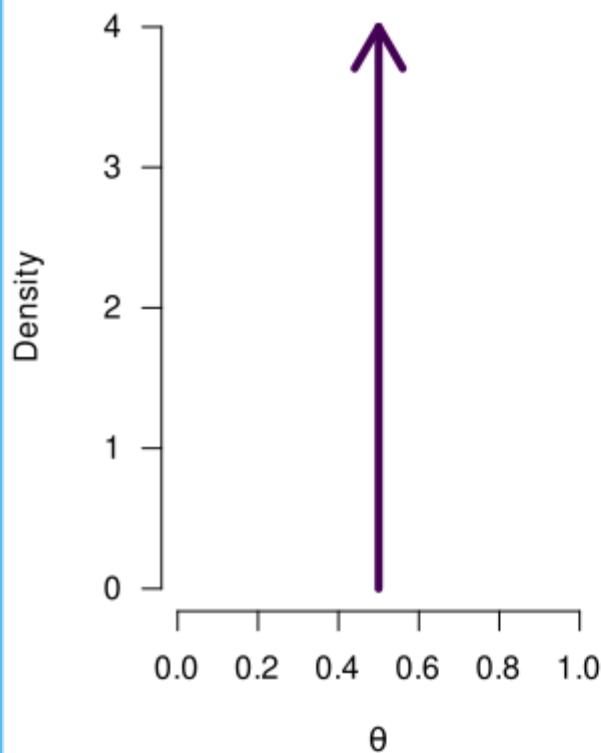


Statistical Models Make Predictions

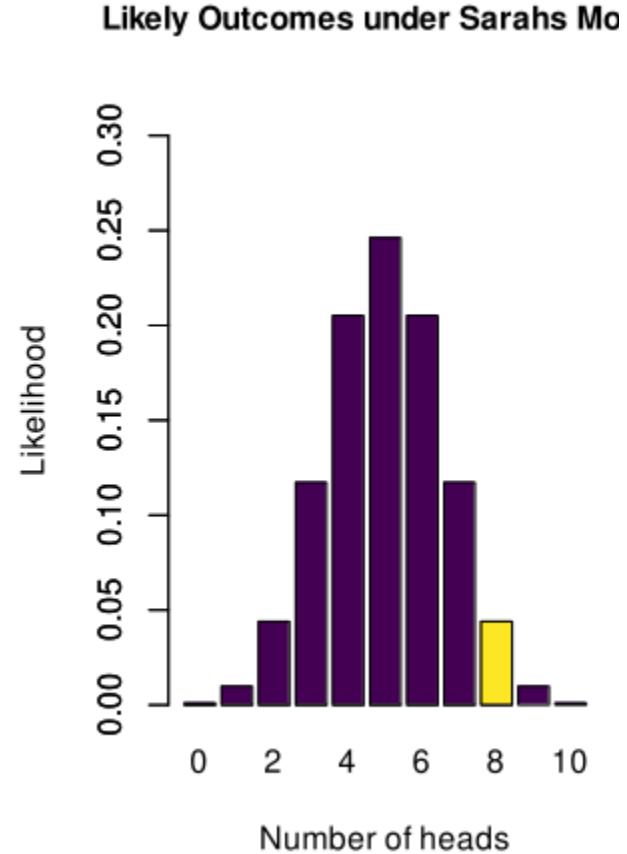


Based on what these models claim about θ , certain outcomes are more/less likely

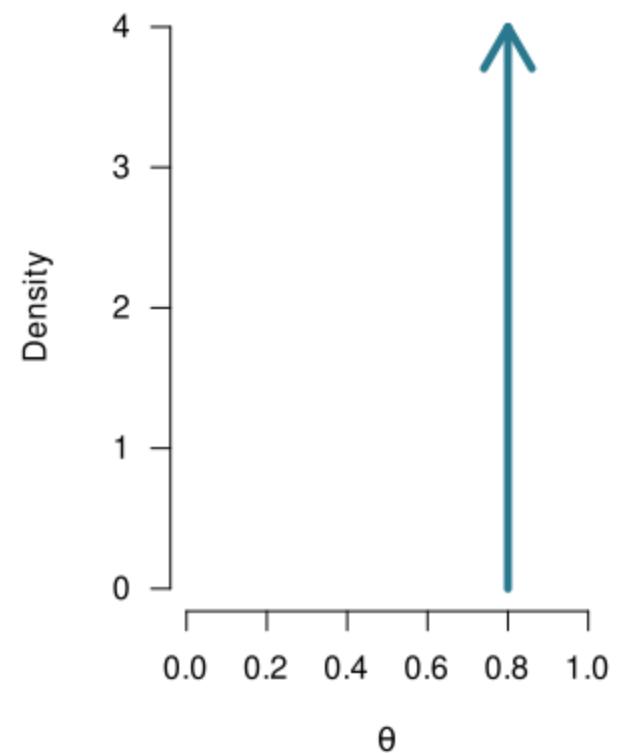
Sarahs Model



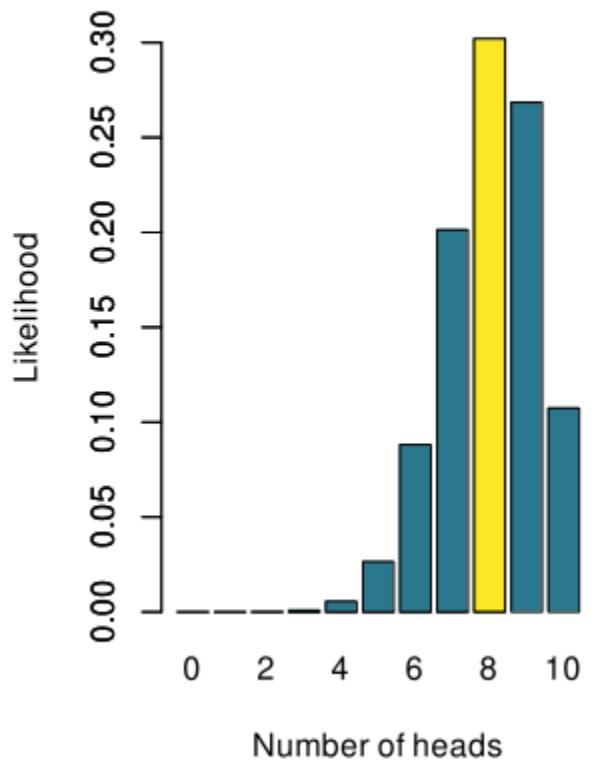
Likely Outcomes under Sarahs Mo



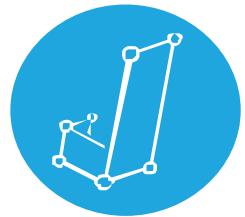
Pauls Model



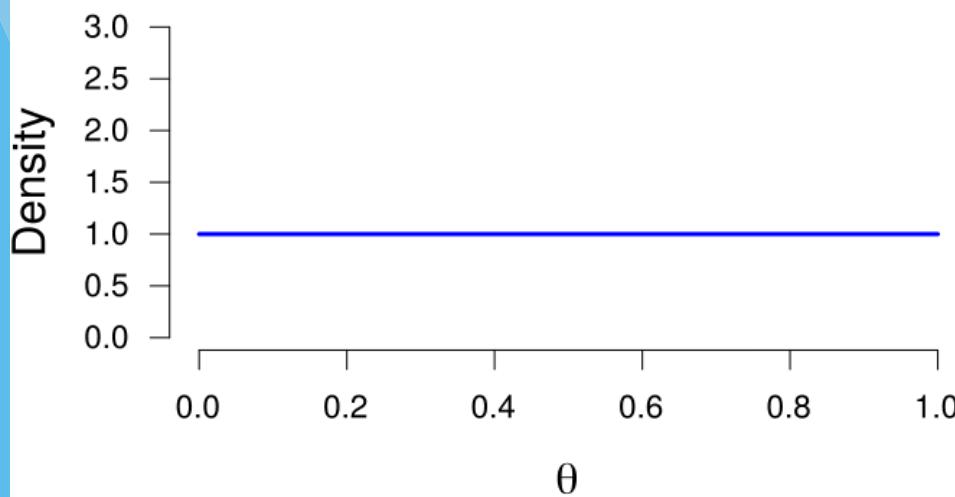
Likely Outcomes under Pauls Model



Models Can Also State a Range of Values



Beta Distribution ($a = 1, b = 1$)



We can reflect a model's statement by means of a probability distribution

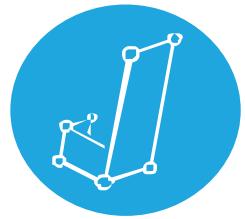
Introducing the beta distribution:

- It ranges from 0 to 1
- Its shape is determined by two values: a and b
 - If a and b equal 1, it is uniform

We use the beta distribution here because:

- A proportion is also between 0 and 1
- We can create many different shapes, which allows us to reflect many different prior ideas

Models Can Also State a Range of Values



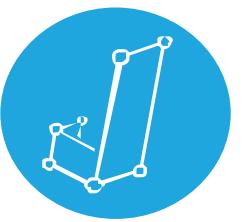
Introducing the beta distribution:

- It ranges from 0 to 1
- Its shape is determined by two values: a and b
 - If a and b equal 1, it is uniform

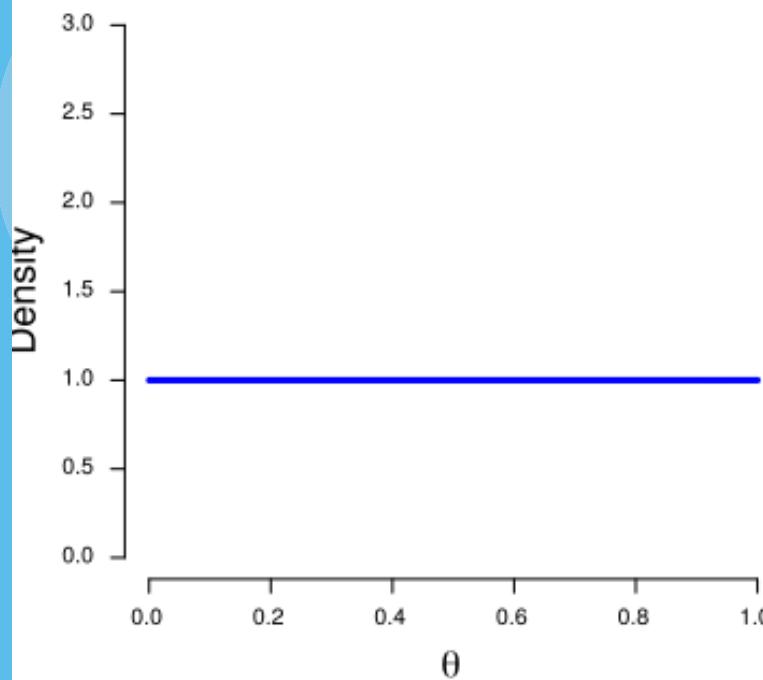
An applet that let you shape a beta distribution

<https://maglit.me/bayesian-intro-shiny>

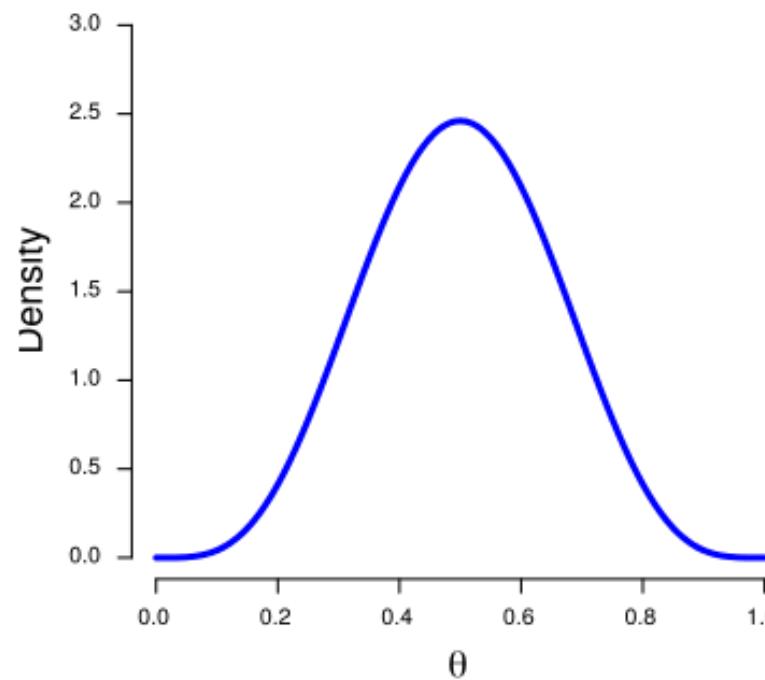
Models Can Also State a Range of Values



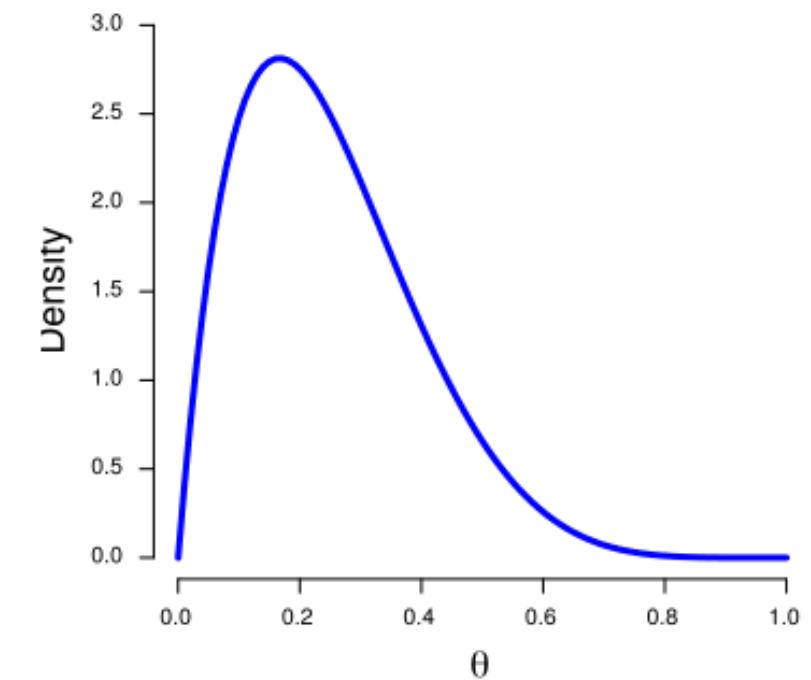
Beta Distribution ($a = 1, b = 1$)



Beta Distribution ($a = 5, b = 5$)



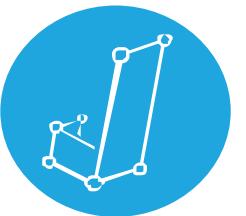
Beta Distribution ($a = 2, b = 6$)



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)



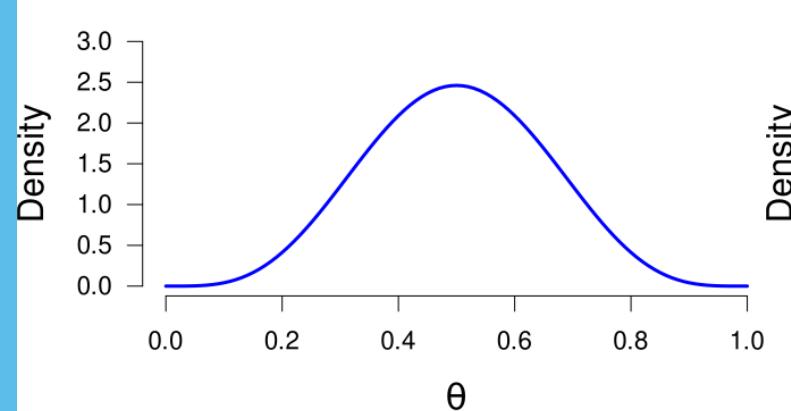
Beta Distribution Interpretation

In the context of a prior distribution for a proportion, the a and b can be interpreted as previously observed heads and tails.

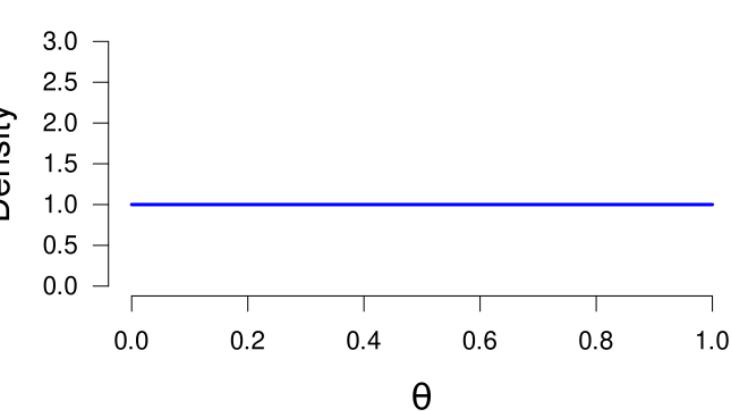
We cannot start with an empty distribution (a and b cannot be 0), no matter how clueless you are about something, there will always be starting point (e.g., $a = b = 1$)

Models that go all in on a single value have a very strong conviction:
Sarah believes as if she has seen infinitely many heads and tails already!

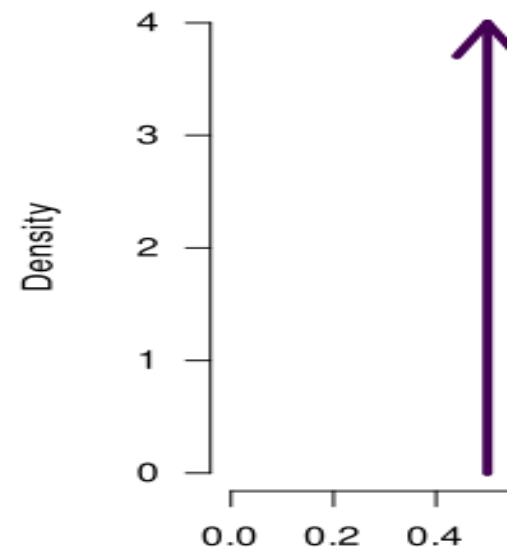
Beta Distribution ($a = 5, b = 5$)



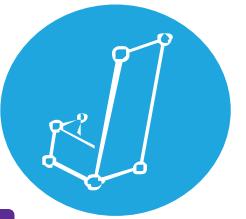
Beta Distribution ($a = 1, b = 1$)



Sarahs

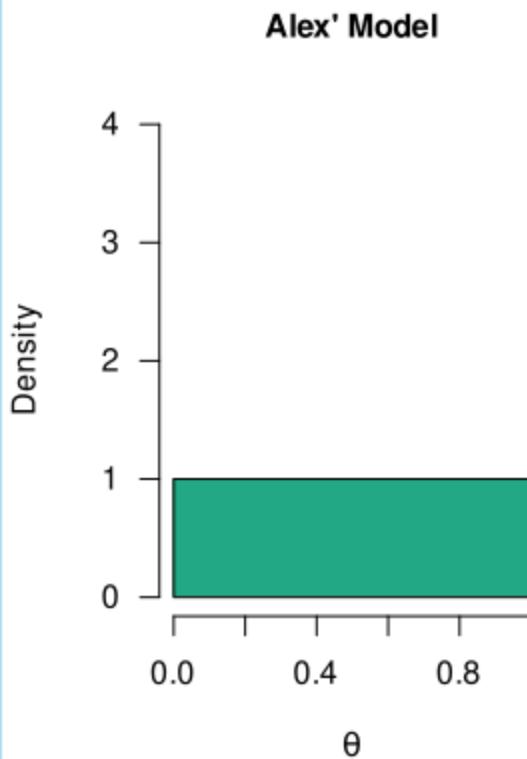


Models Can Also State a Range of Values

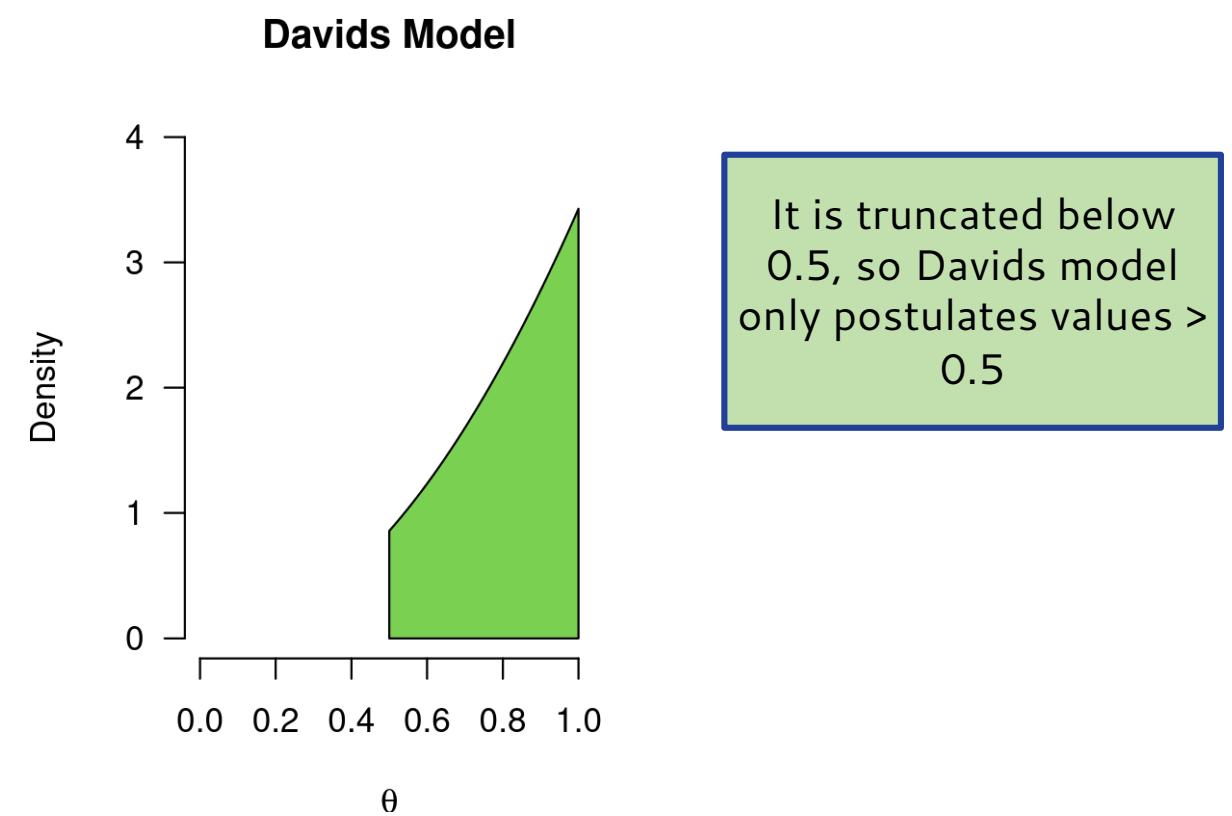


We can reflect a model's statement by means of a probability distribution

A binomial model with $\theta \sim \text{Beta}(1, 1)$

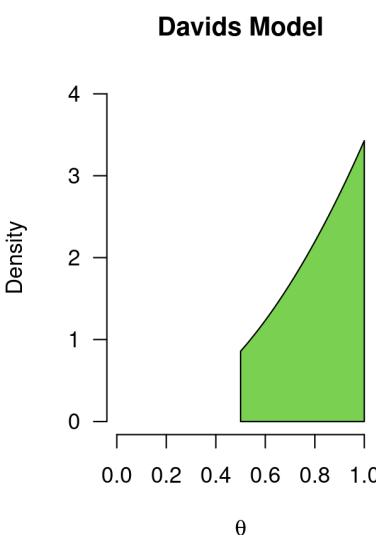
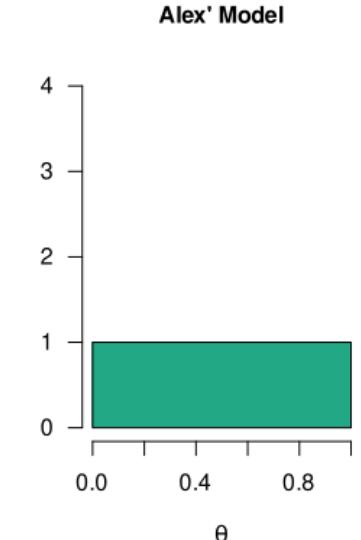
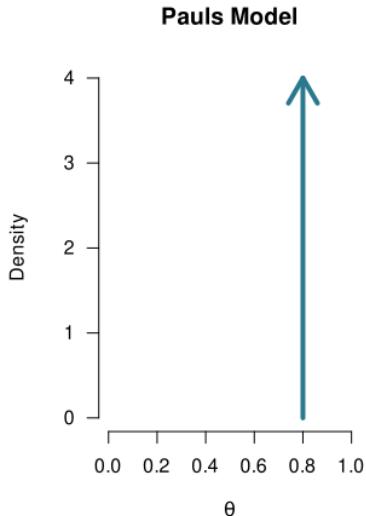
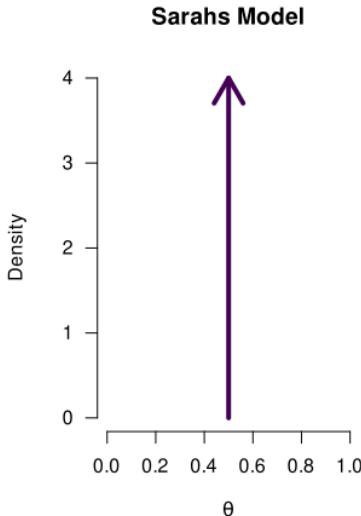
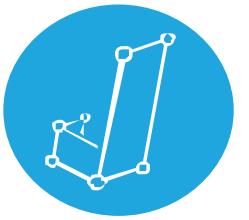


A binomial model with $\theta \sim \text{Beta}(3, 1)$



What are the theoretical implications of these models?

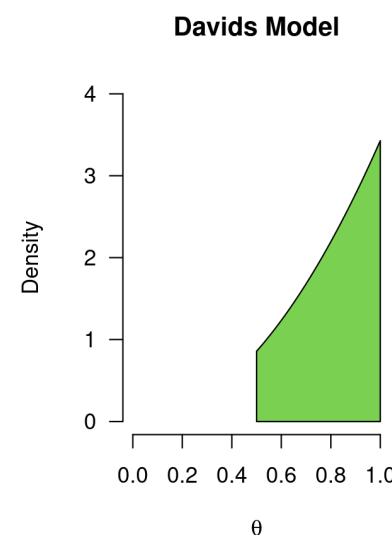
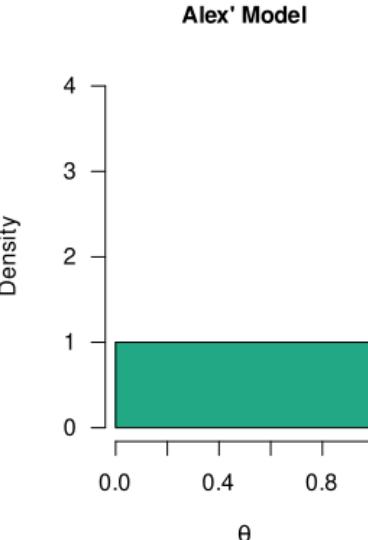
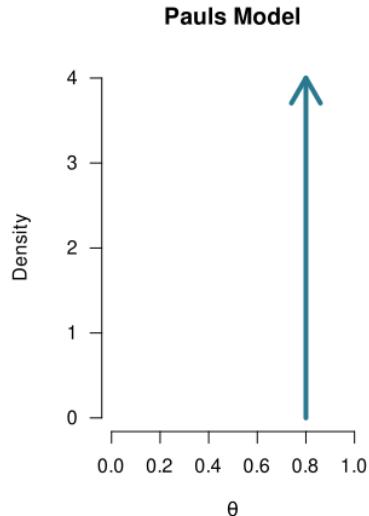
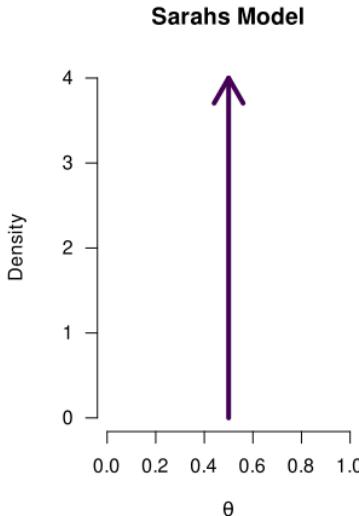
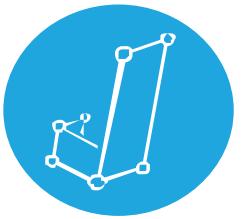
So, We Have All Sorts of Models...



Later we discuss
comparing these models
to each other

Now we focus on a single
model, and how they
learn from data

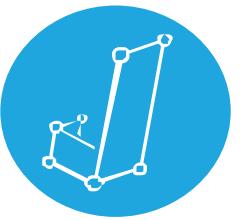
So, We Have All Sorts of Models...



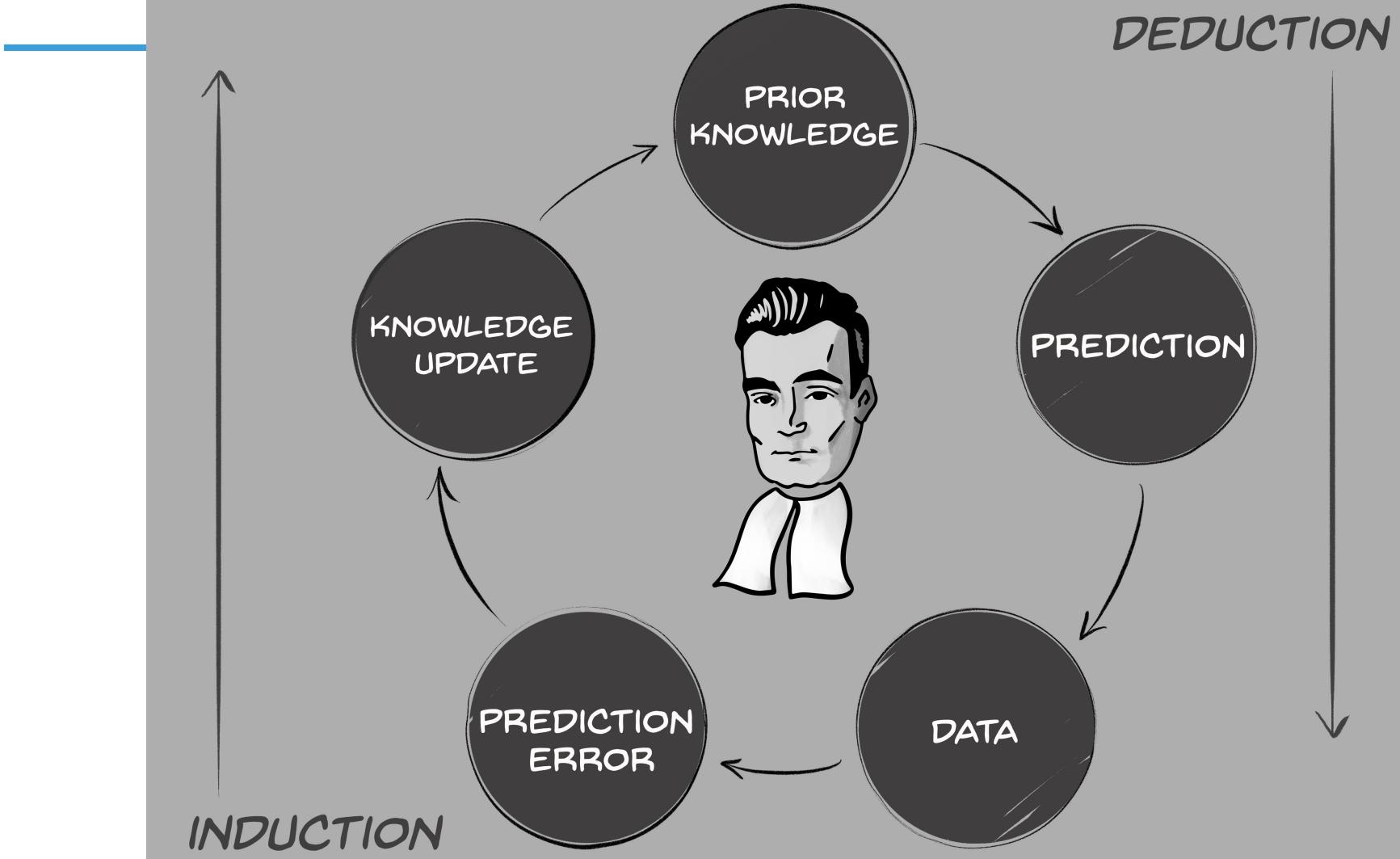
These models reflect prior knowledge/beliefs

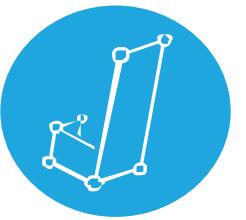
We will update this prior knowledge with data

To end up with posterior knowledge



BAYESIAN LEARNING CYCLE





The Scenario for Today

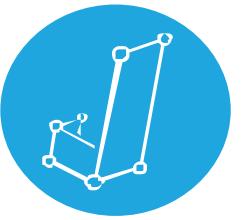


VS



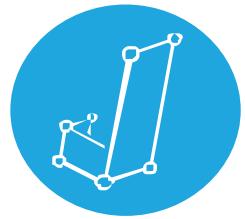
What is the proportion of Mac users in the population?

Bayes' Theorem



$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

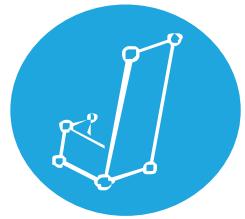
Applying Bayes' Theorem to Statistical Inference



We can replace "A" and "B", with parameter value " θ " and observed "data"

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

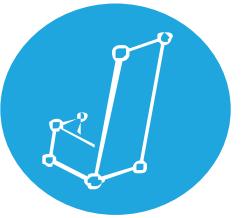
Applying Bayes' Theorem to Statistical Inference



We can rewrite the theorem slightly:

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

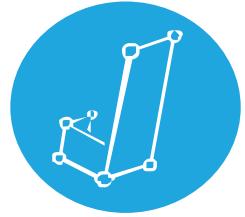
Bayesian Knowledge Updating



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

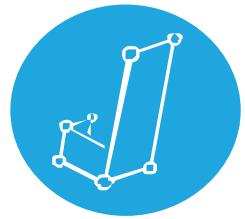
In the context of statistical analysis, the “world” refers to values of a parameter, such as the population proportion, or population correlation

Estimating a Proportion: Prior Distribution

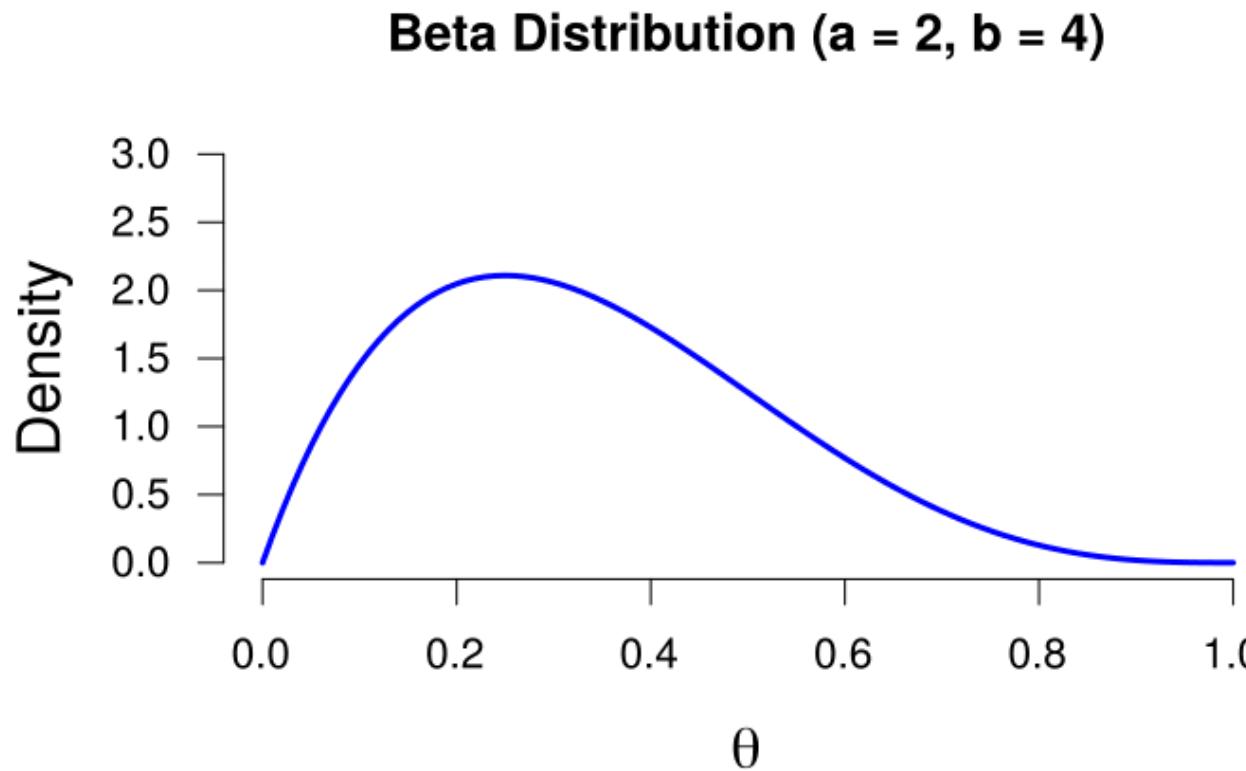


- The prior distribution reflects our beliefs about the parameter, **before** observing the data
- In other words, one of the models we have seen so far!

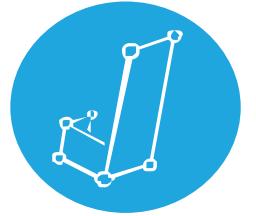
Estimating a Proportion: Prior Distribution



- Let's use the Beta prior distribution with $a = 2$, $b = 4$
 - (would you use a different prior?)



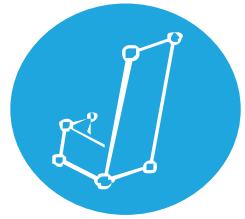
Bayesian Knowledge Updating



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

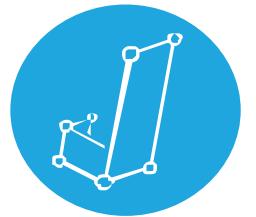
We update our prior knowledge by observing data

Estimating a Proportion: Some Fabricated Data



- We observe the following data ($n = 10$):
 - 3 Mac users
 - 7 Windows users
- Our *statistic* is the observed proportion: $3/10 = 0.3$

Estimating a Proportion: Predictive Updating Factor



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

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

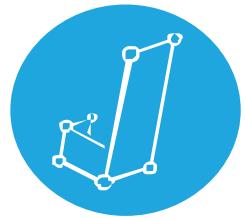
This tells us something about how well a specific value of θ predicted the data (i.e., it is the quality of the prediction for this specific value)

The **marginal likelihood**, across all values of θ in the model

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

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 by this model)

Estimating a Proportion: Predictive Updating Factor

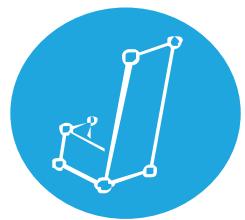


The **likelihood** of the data, given a certain value of θ in the model

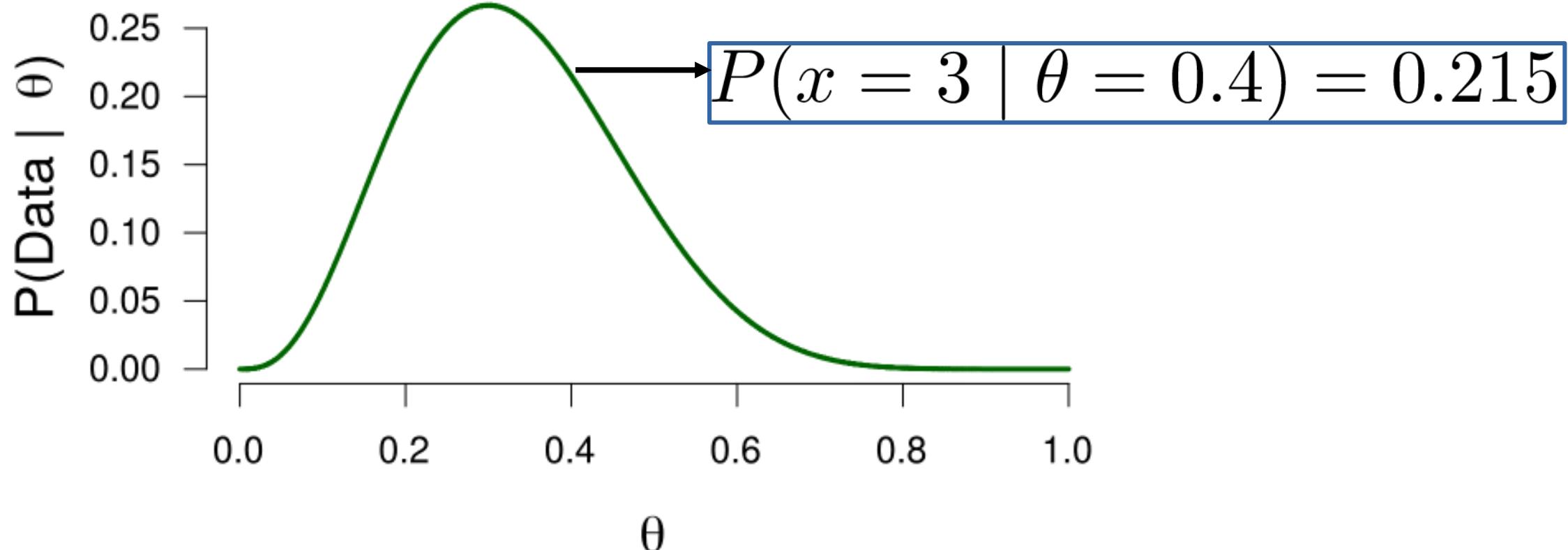
$$\underline{P(\text{data} \mid \theta)}$$

For instance, how likely is our observed data, if θ equals 0.4?

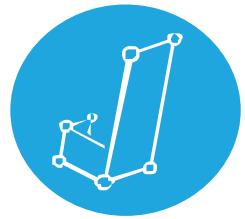
Estimating a Proportion: Predictive Updating Factor



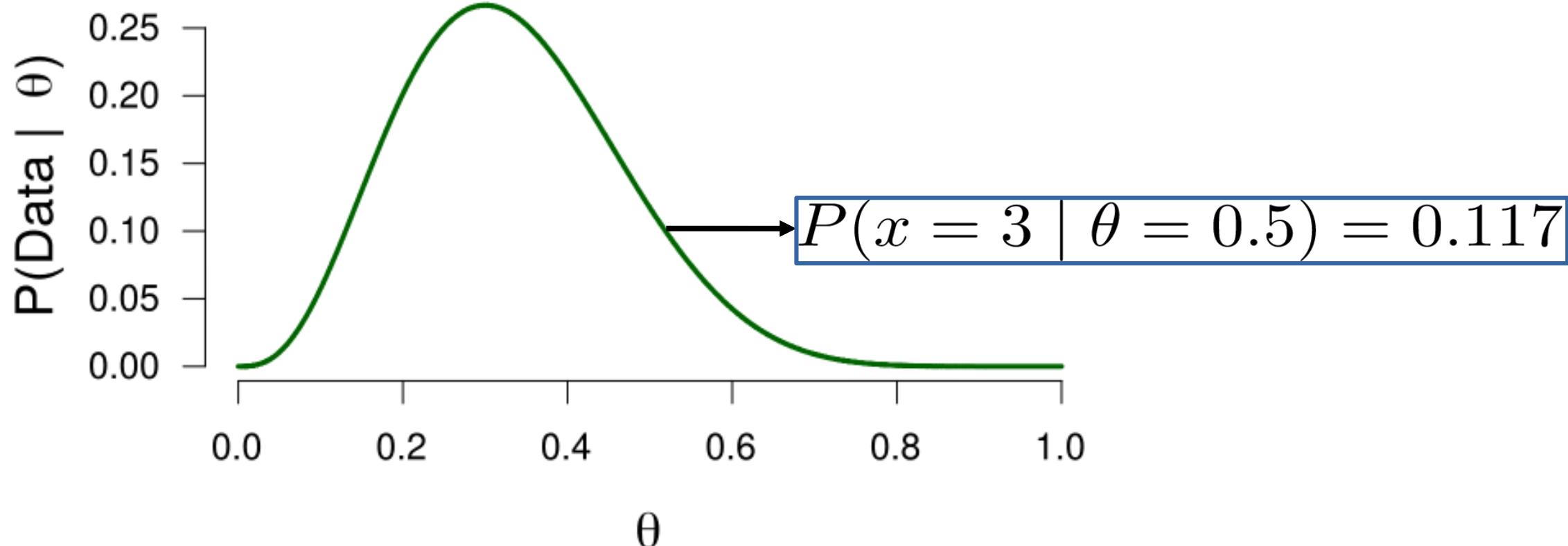
Likelihood of the observed data, for each value of θ



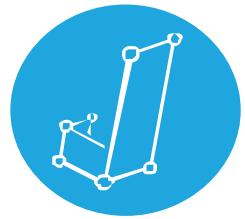
Estimating a Proportion: Predictive Updating Factor



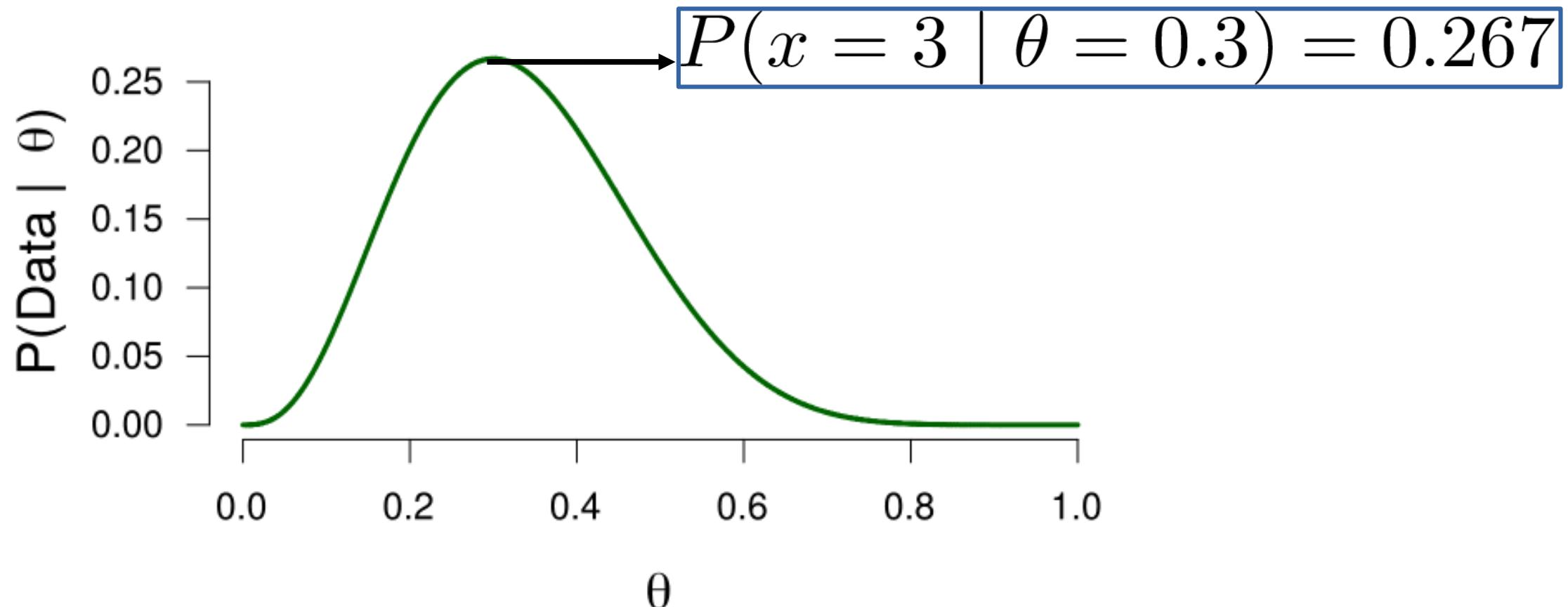
Likelihood of the observed data, for each value of θ



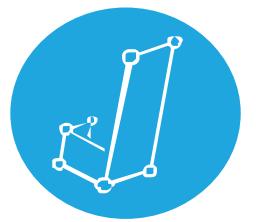
Estimating a Proportion: Predictive Updating Factor



Likelihood of the observed data, for each value of θ



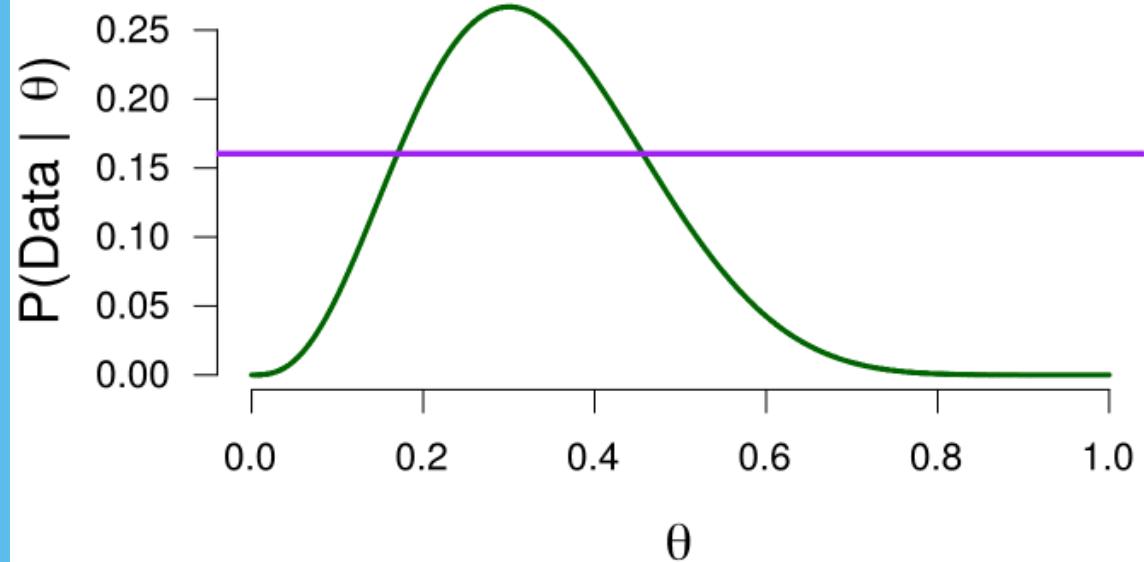
Estimating a Proportion: Predictive Updating Factor



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

The marginal likelihood, across all values of θ

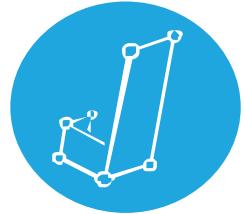
Likelihood of the observed data, for each value of θ



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 by this model)

Usually this is very difficult and we need computers to compute this. In our case, the marginal likelihood equals 0.16

Estimating a Proportion: Predictive Updating Factor



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

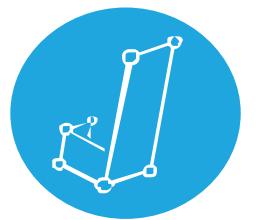
$$\frac{P(\text{data} \mid \theta)}{P(\text{data})}$$

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

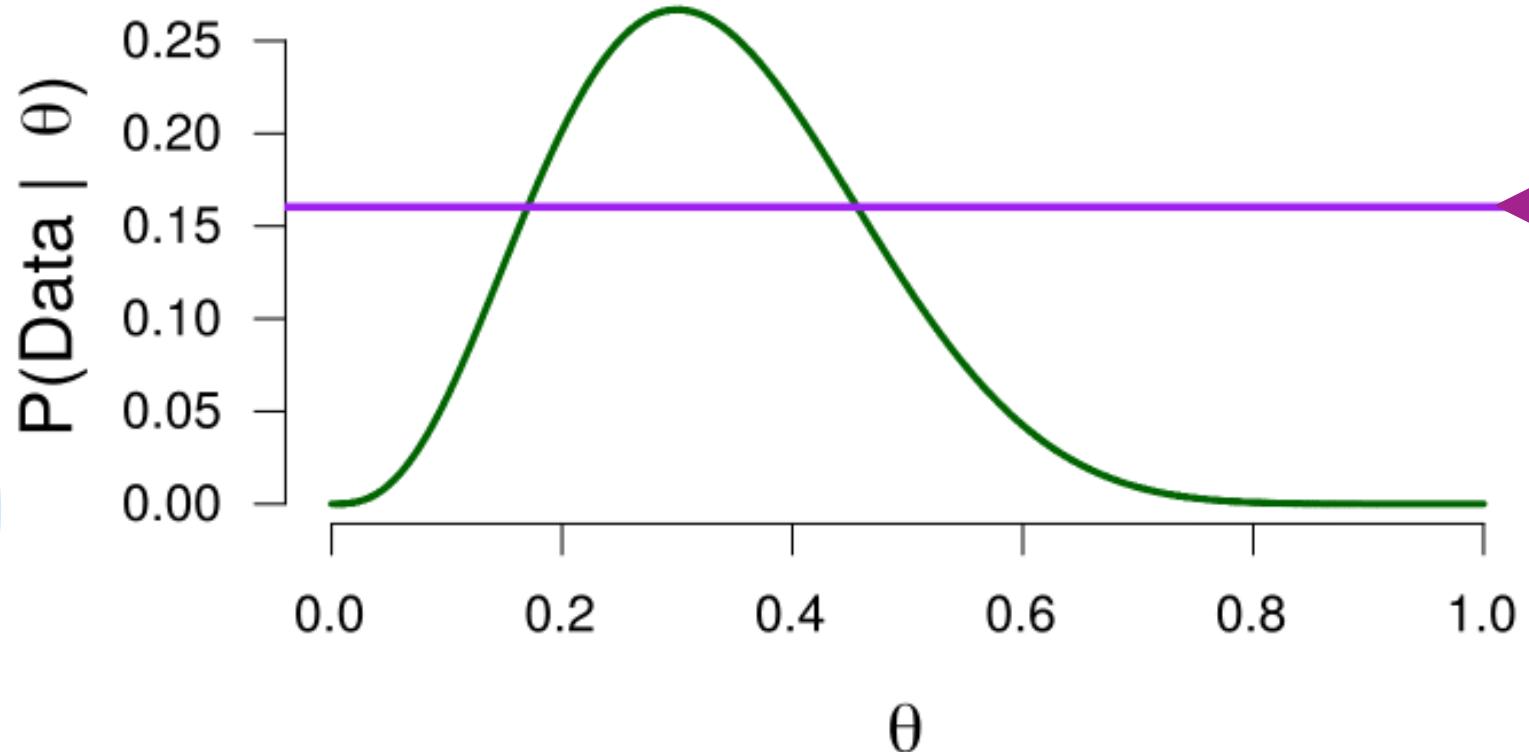
Taken together, this ratio tells us how well each value of θ predicted the data, **relative** to all other values!

Now we know that the marginal (i.e., average) likelihood is 0.16. This means that when the likelihood of the data for a specific value of θ is greater than 0.16, that value of θ has predicted the data above average. If that is the case, the predictive updating factor is > 1 . If a specific value of θ predicted the data worse than average, that ratio is < 1 .

Estimating a Proportion: Predictive Updating Factor



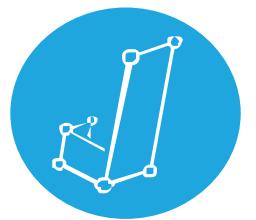
Likelihood of the observed data, for each value of θ



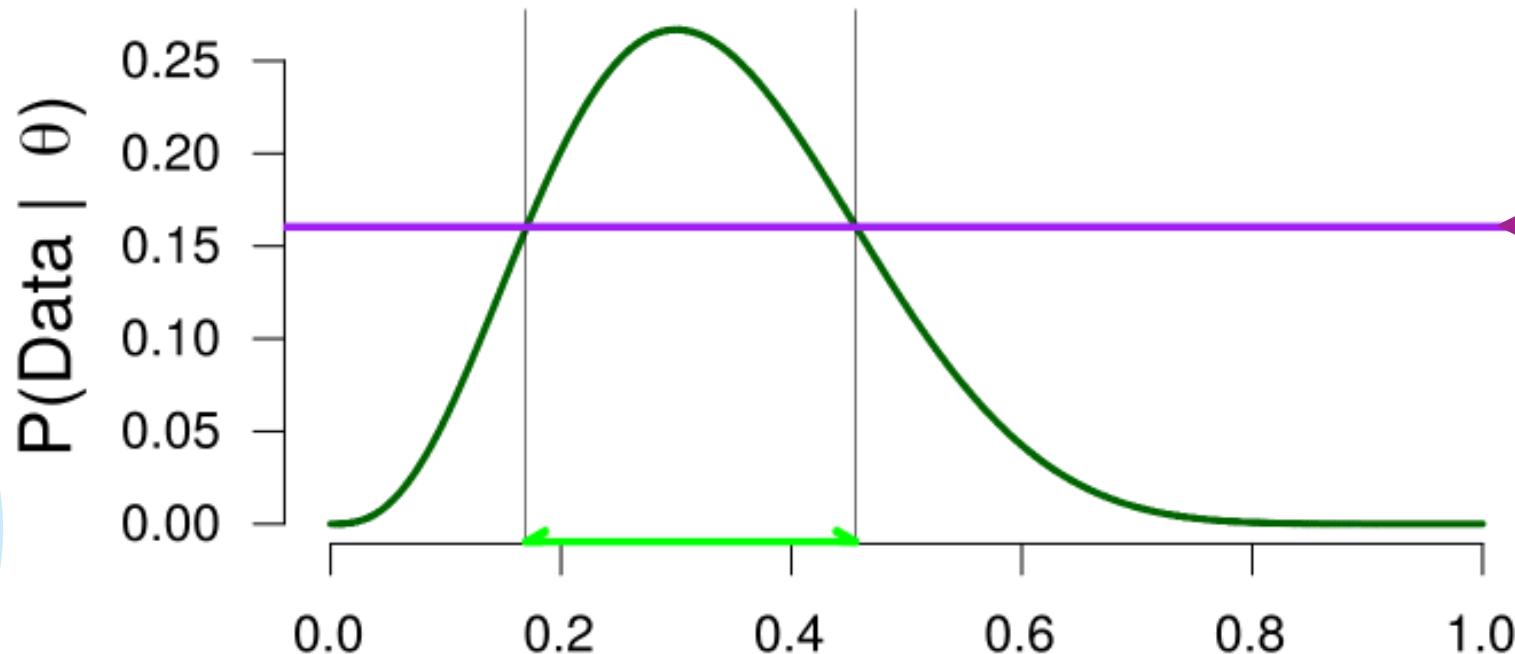
The marginal likelihood, across all values of θ ($= 0.16$)

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 by this model)

Estimating a Proportion: Predictive Updating Factor



Likelihood of the observed data, for each value of θ

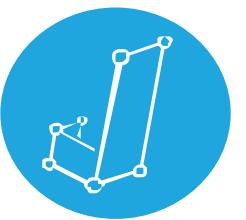


Values of θ that predicted the data better than average (i.e., their likelihoods are greater than the marginal likelihood)

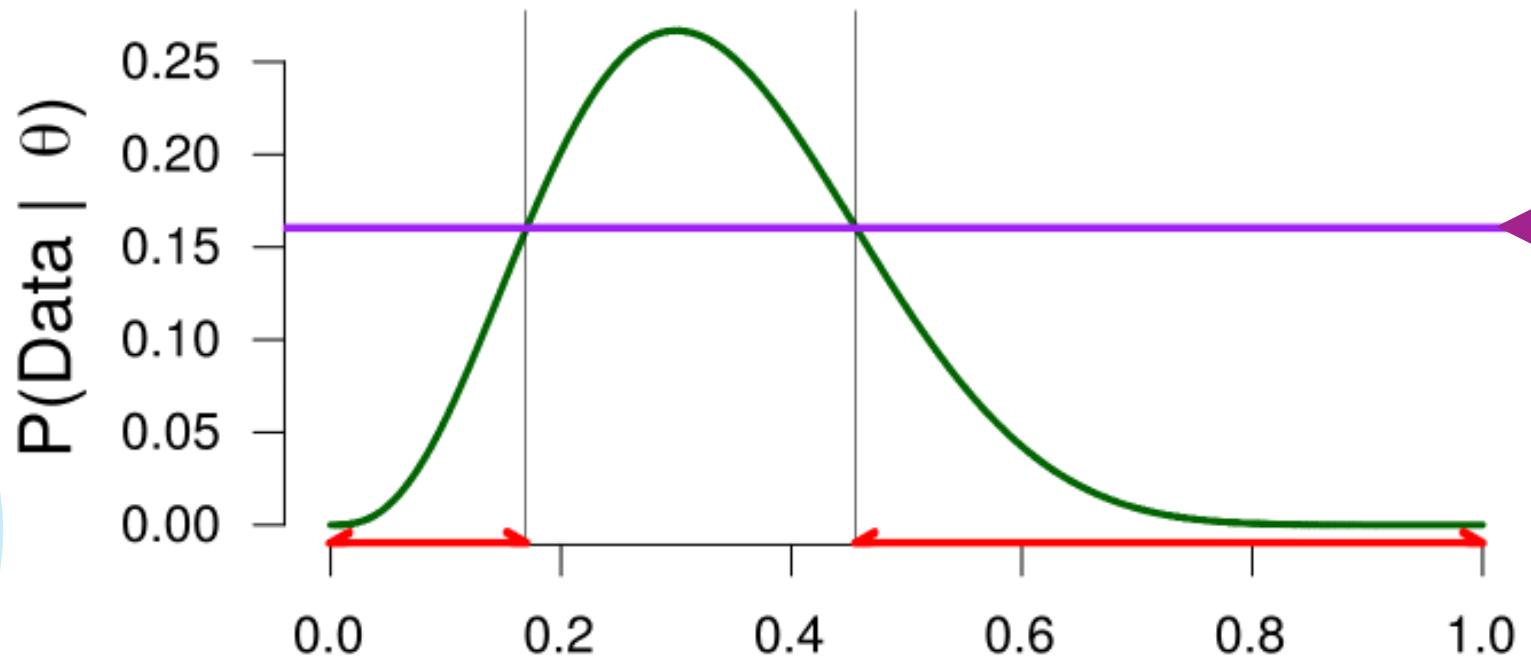
The **marginal likelihood**, across all values of θ ($= 0.16$)

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 by this model)

Estimating a Proportion: Predictive Updating Factor



Likelihood of the observed data, for each value of θ

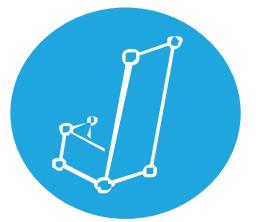


Values of θ that predicted the data worse than average (i.e., their likelihoods are less than the marginal likelihood)

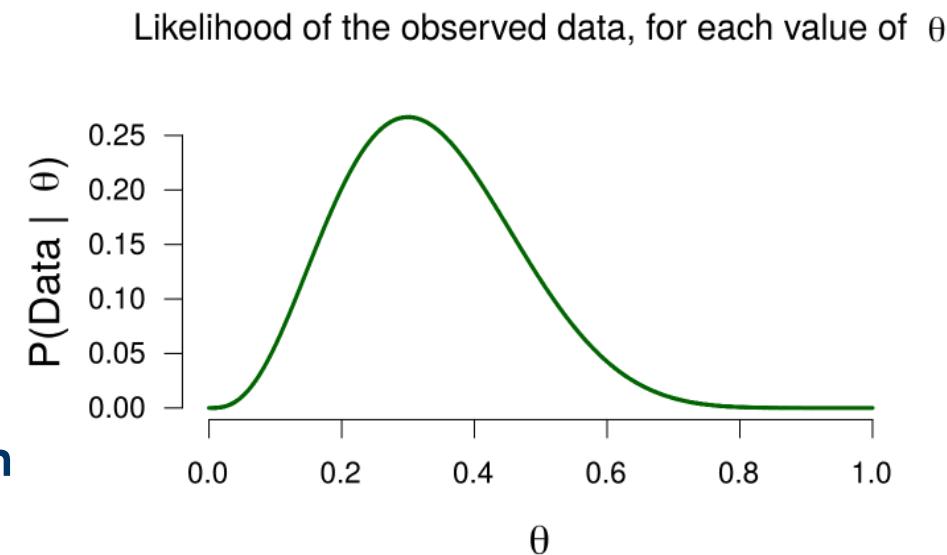
The **marginal likelihood**, across all values of θ ($= 0.16$)

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 by this model)

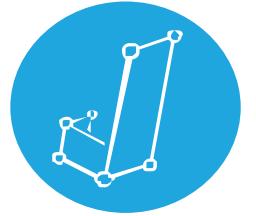
Estimating a Proportion: Some Notes on the Likelihood



- The likelihood is the same for different prior distributions/models
- However, the marginal likelihood is the average of the likelihoods, weighted by how the model divided its probability mass
- The marginal likelihood is therefore different for different prior distributions
- Since the marginal likelihood reflects the average quality of the predictions made by the model, we will use this later to perform **model comparison**

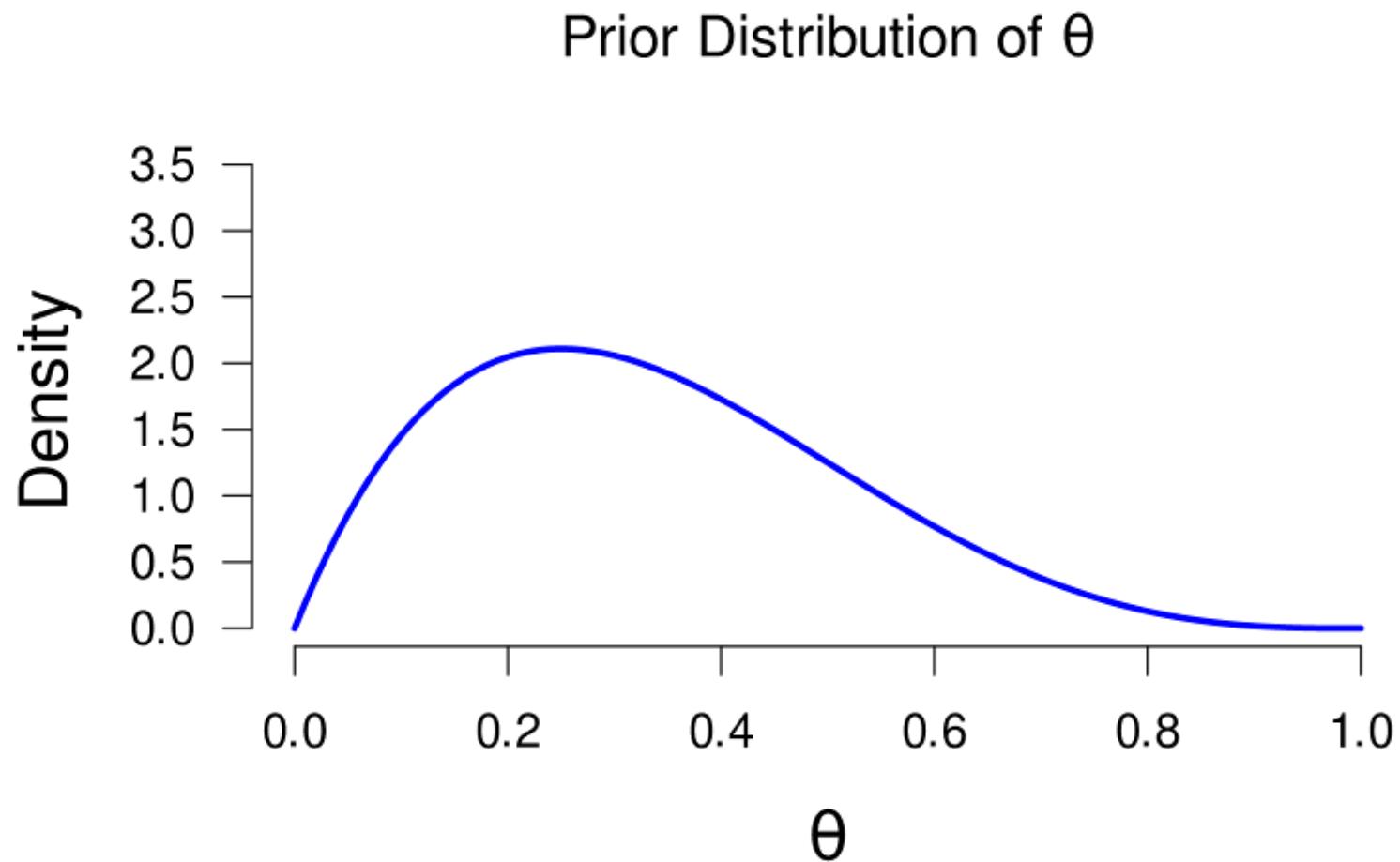
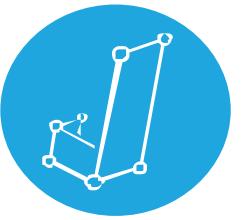


Bayesian Knowledge Updating

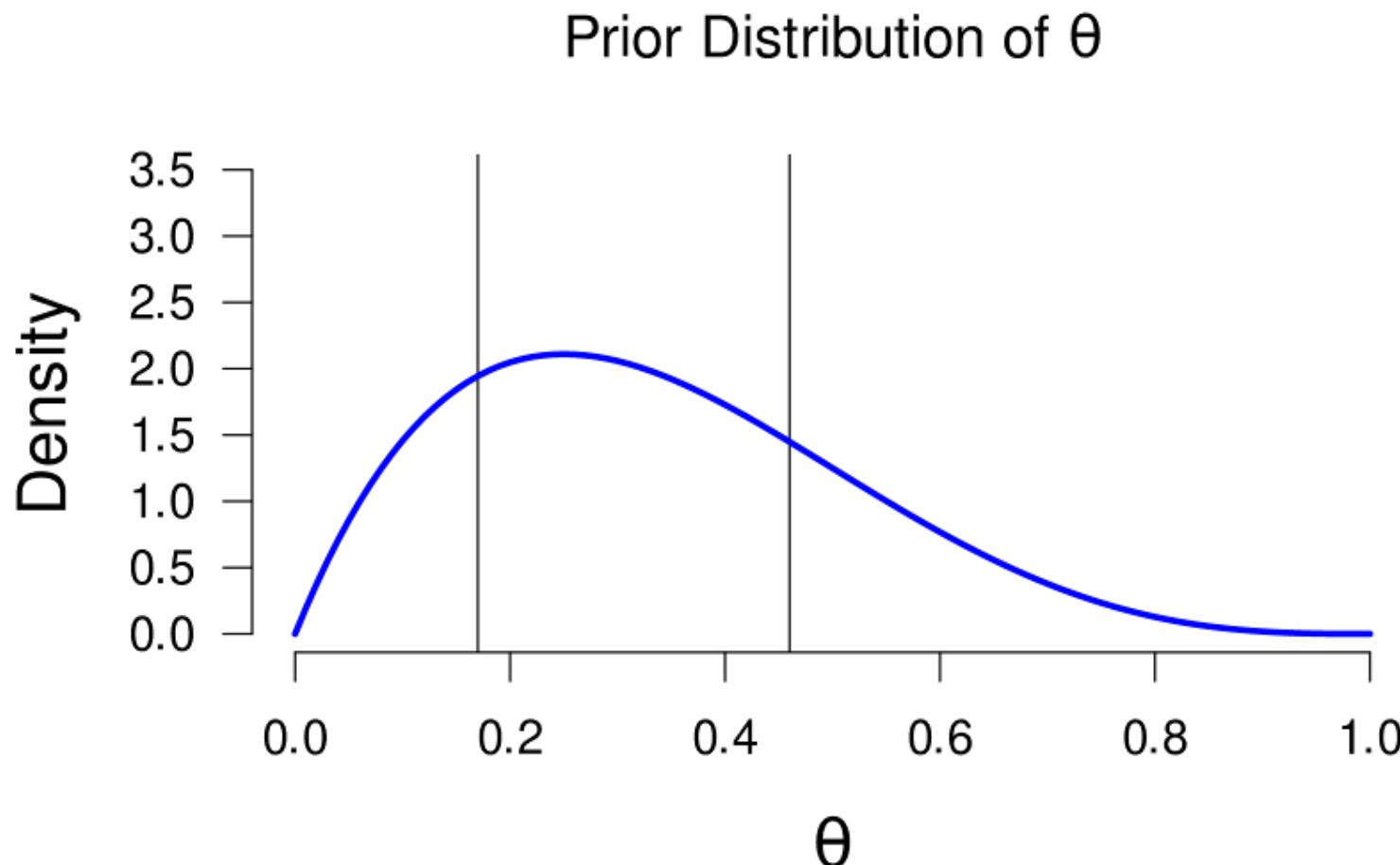
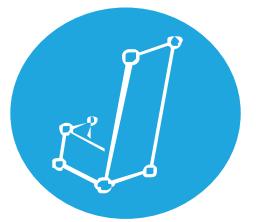


$$\underbrace{p(\theta \mid \text{data})}_{\text{Posterior beliefs about the world}} = \underbrace{p(\theta)}_{\text{Prior beliefs about the world}} \times \underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\text{Predictive updating factor}}$$

Estimating a Proportion: Posterior Distribution

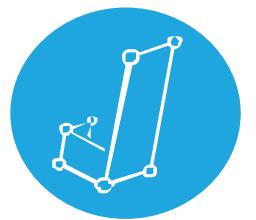


Estimating a Proportion: Posterior Distribution

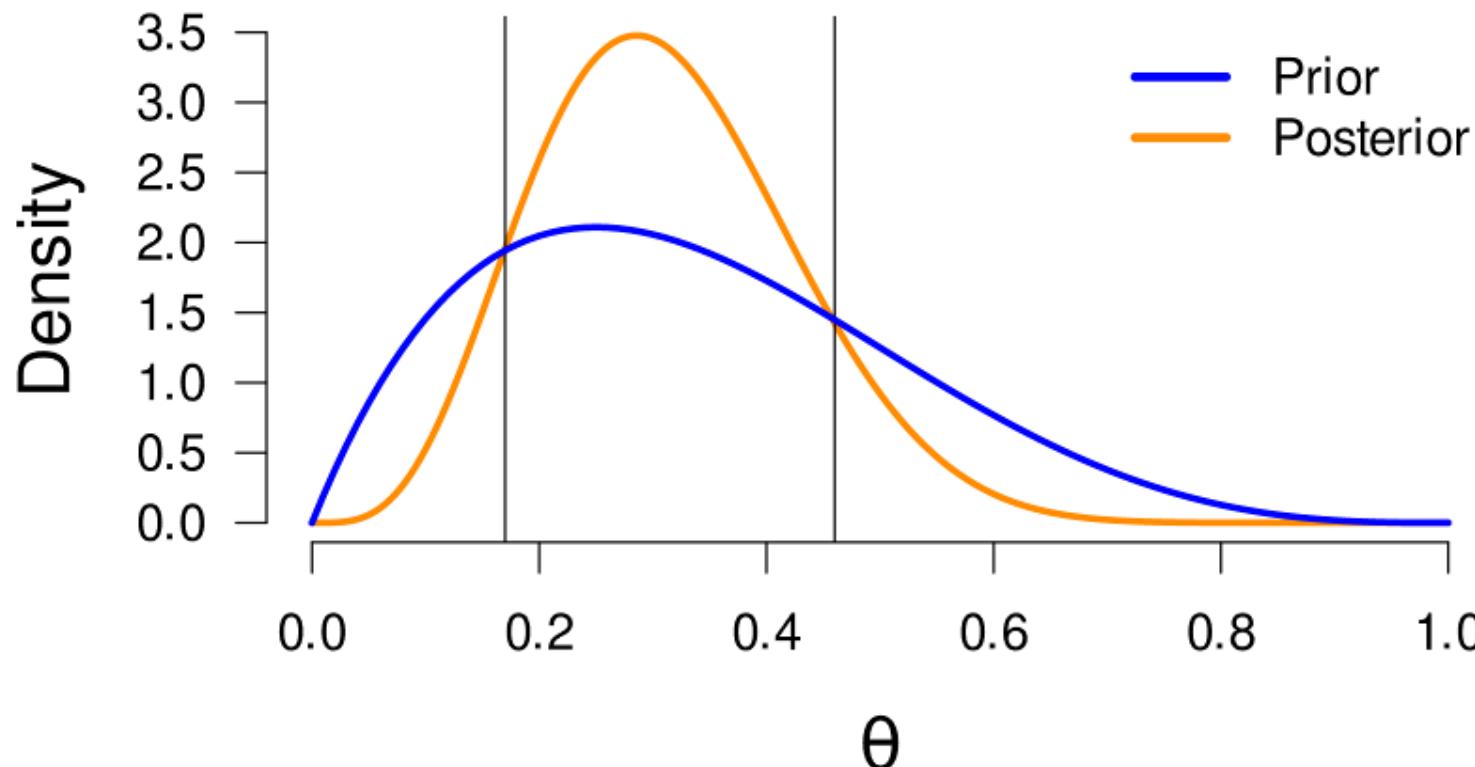


Same lines as on slide 34/35! It shows which values of theta received a decrease/increase in plausibility, because of the data

Estimating a Proportion: Posterior Distribution

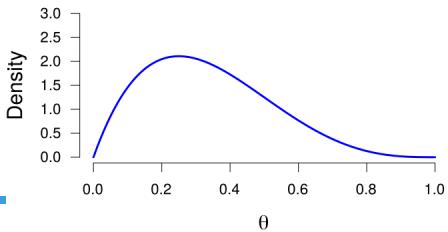


Prior and Posterior Distribution of θ



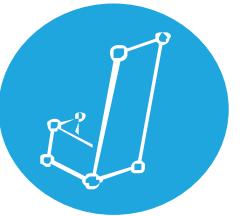
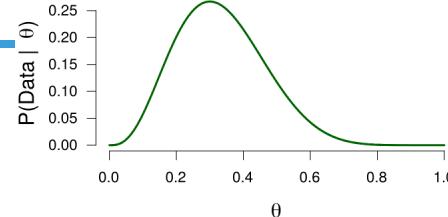
Same lines as on slide 34/35! It shows which values of theta received a decrease/increase in plausibility, because of the data

Beta Distribution ($a = 2$, $b = 4$)



We start with our prior beliefs

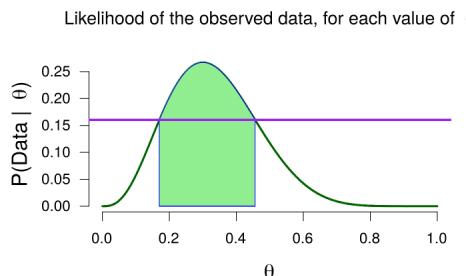
Likelihood of the observed data, for each value of θ



We update those with data

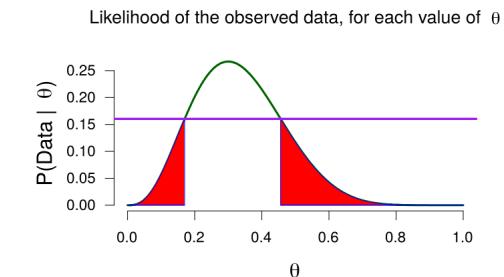
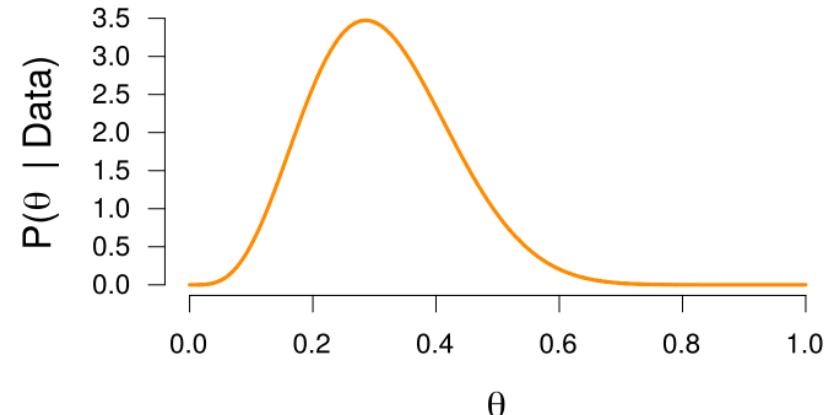
Values of θ that predicted the data better than average receive a boost in plausibility (i.e., their updating ratio > 1)

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

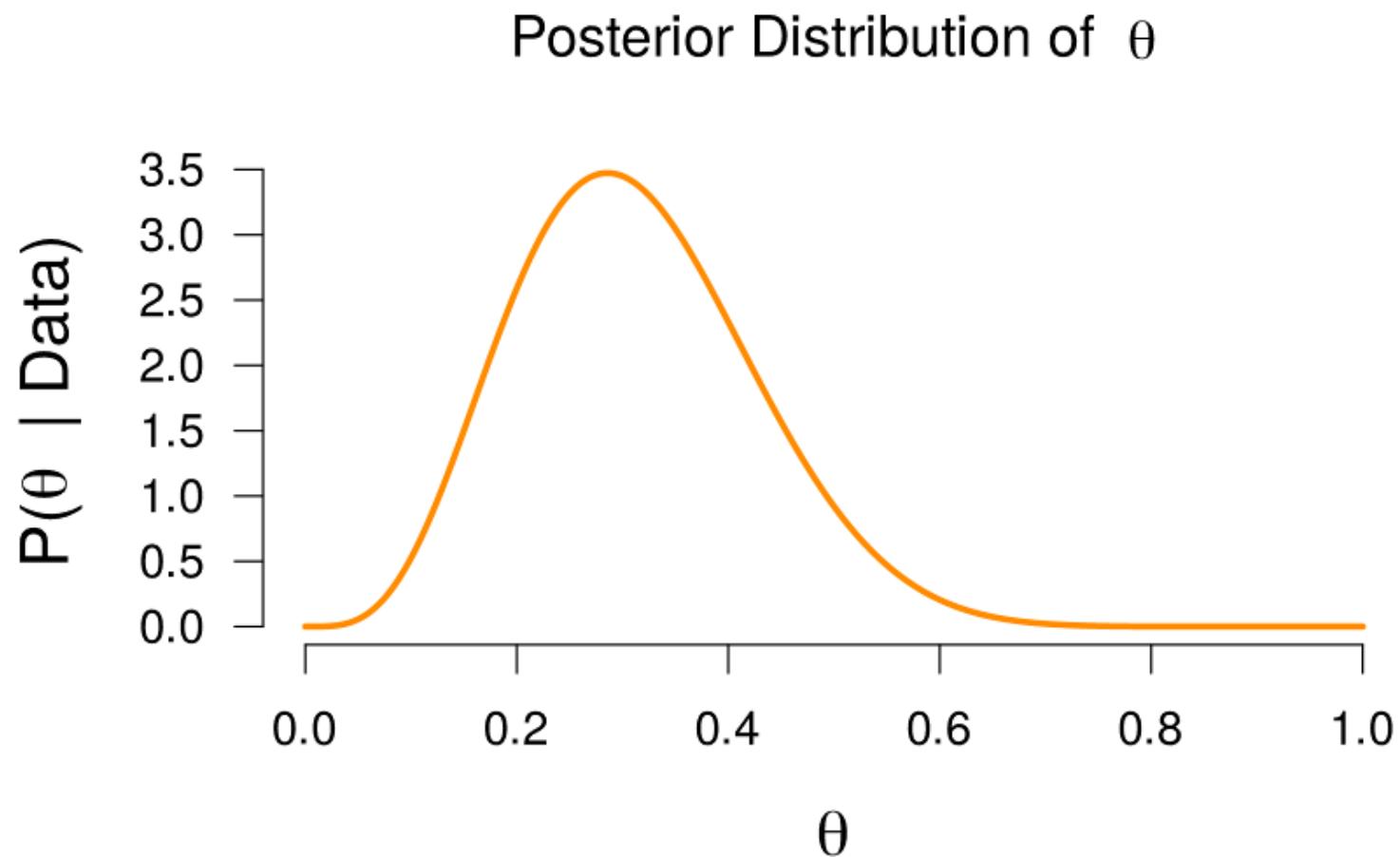
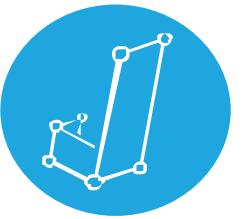


We end with our posterior beliefs

Posterior Distribution of θ

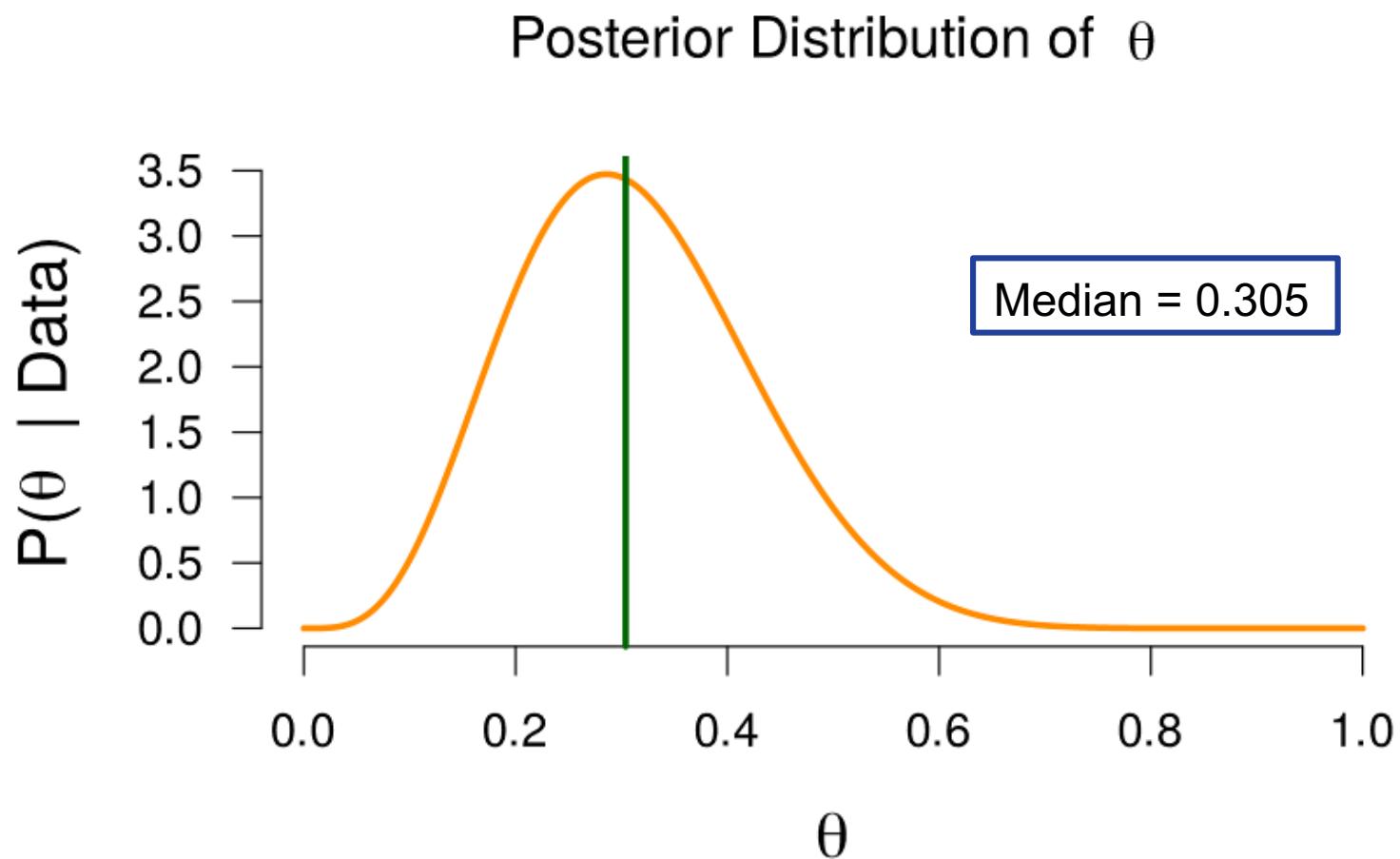
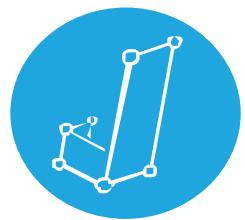


Estimating a Proportion: Posterior Distribution



We can look at the posterior distribution to do parameter **estimation!** This would be the best guess of this model

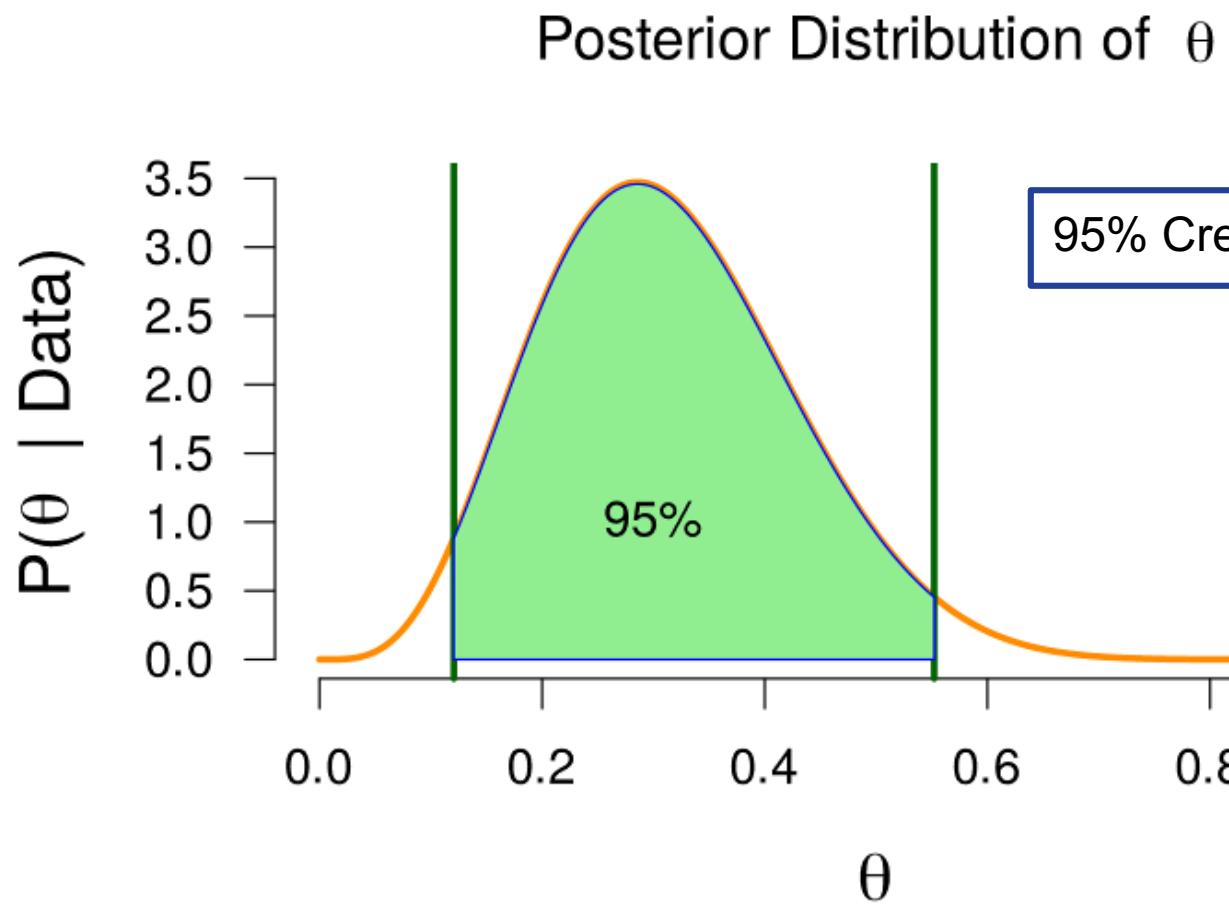
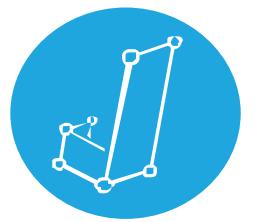
Estimating a Proportion: Posterior Distribution



We can look at the posterior distribution to do parameter **estimation!** This would be the best guess of this model

We can make a point estimate, and take the **posterior median or mean**

Estimating a Proportion: Posterior Distribution



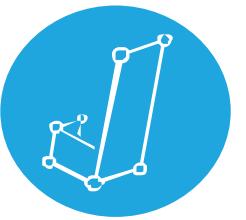
We can make an interval estimate, and take the **central Credible Interval**

95% Credible interval = (0.12, 0.55)

To obtain the $x\%$ central credible interval, we take $x\%$ of the most central posterior mass, and see which 2 points are the thresholds

How to interpret a $x\%$ central credible interval?

Under this model, there is a $x\%$ probability that the true proportion is between 0.12 and 0.55



Live Demonstration

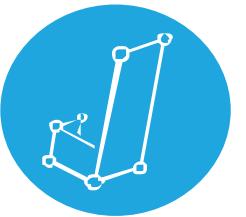


VS

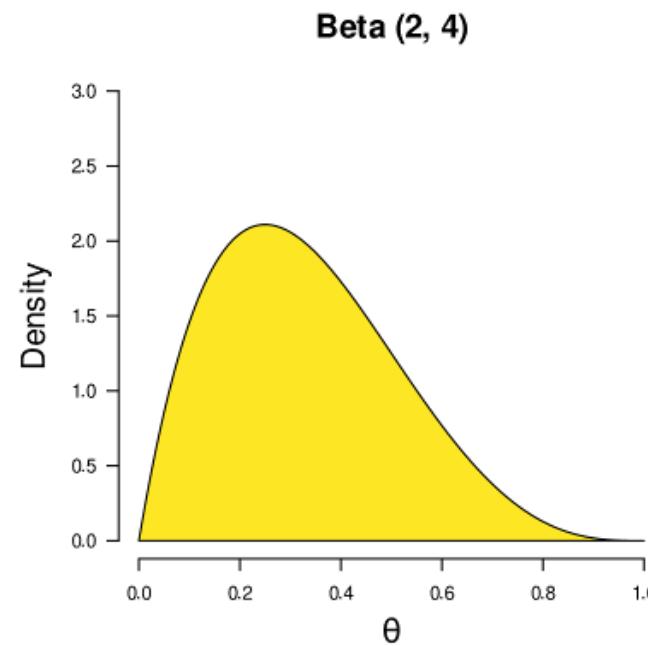
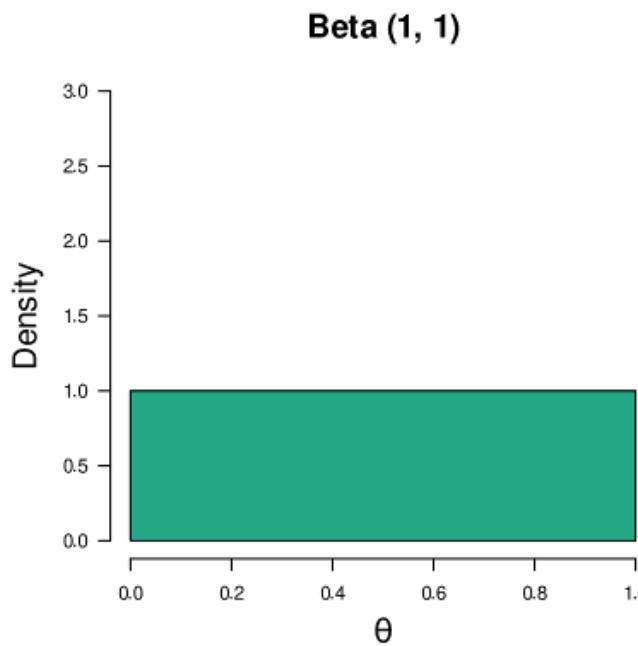


- <https://maglit.me/bayesian-intro-shiny>
- JASP → Learn Bayes →
Binomial Estimation

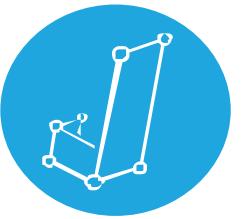
Testing a Proportion: Model Comparison



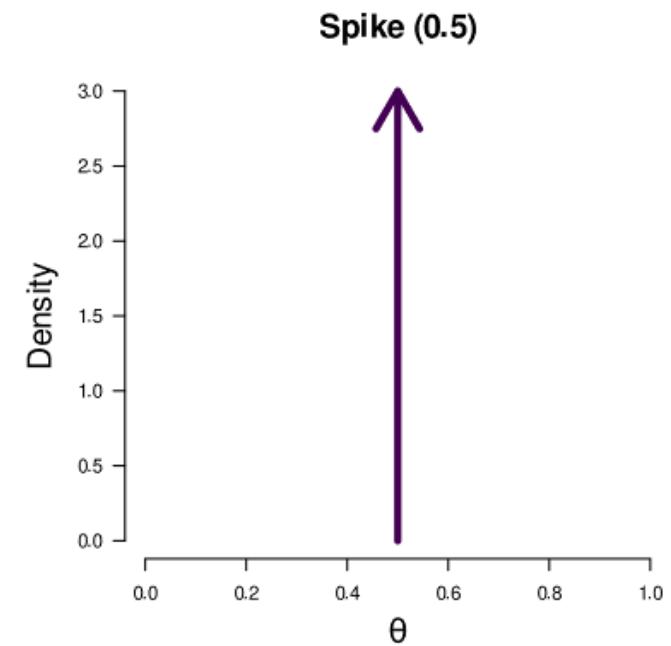
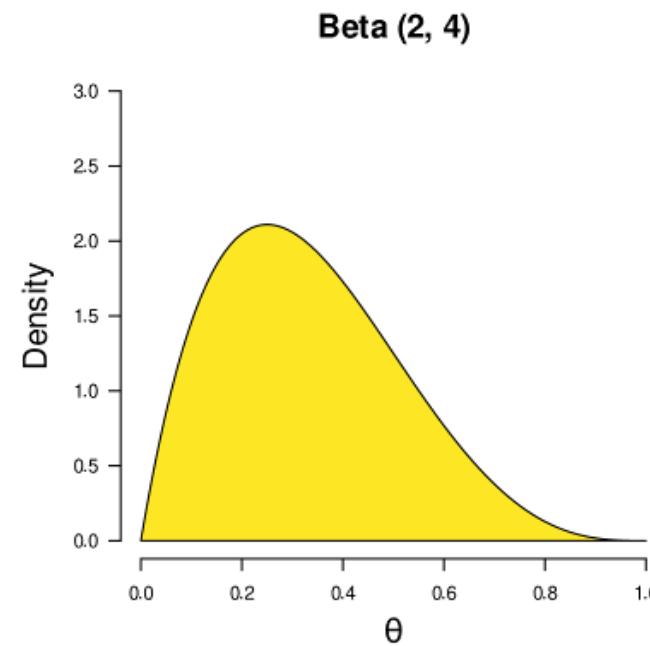
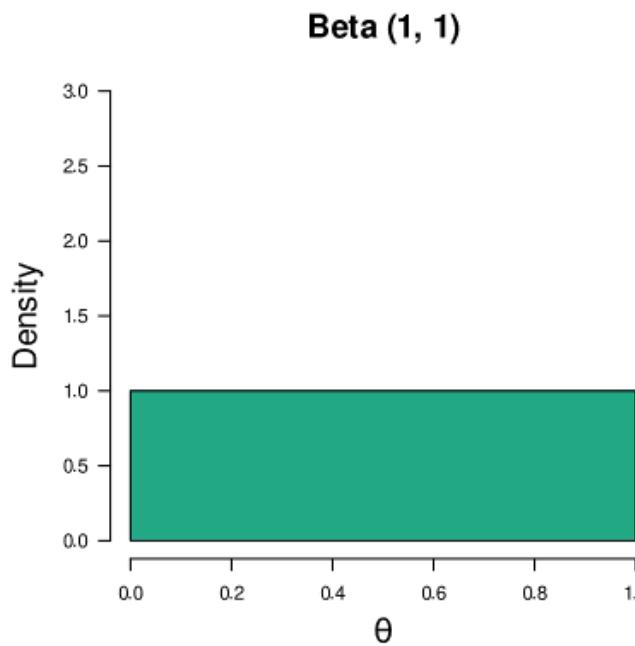
- We have seen we can specify all sorts of models that reflect all sorts of statements about the population proportion



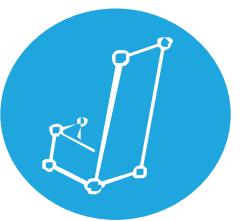
Testing a Proportion: Model Comparison



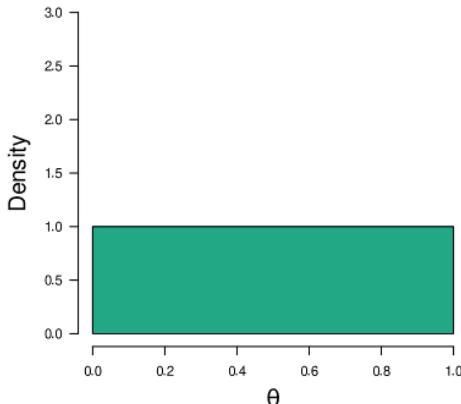
- We have seen we can specify all sorts of models that reflect all sorts of statements about the population proportion



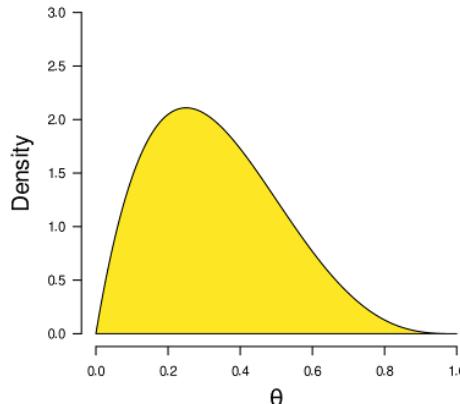
What Do the Models Predict?



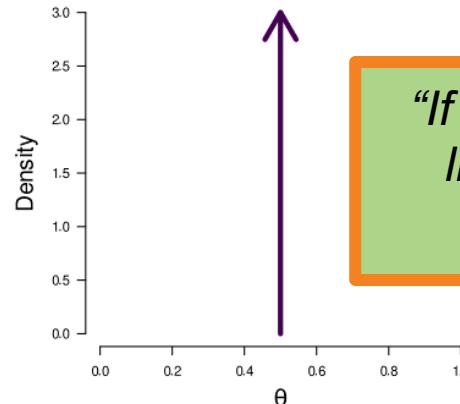
Beta (1, 1)



Beta (2, 4)

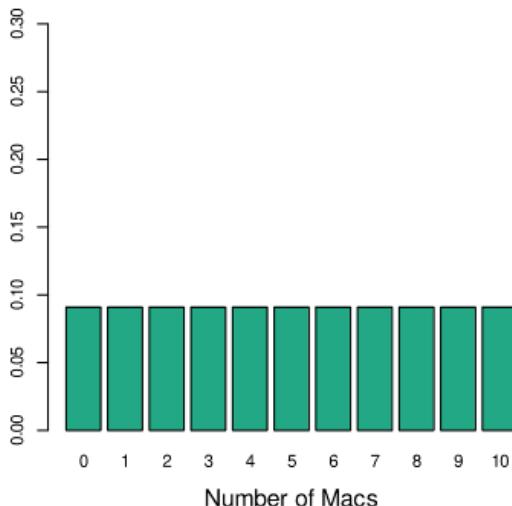


Spike (0.5)

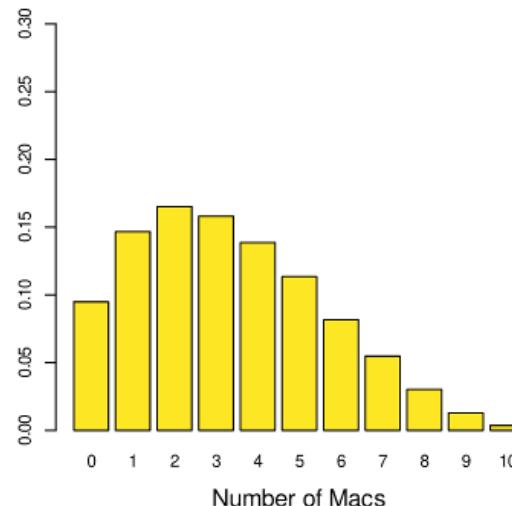


"If you had 100€ to bet on likely values of θ , how would you divide it?"

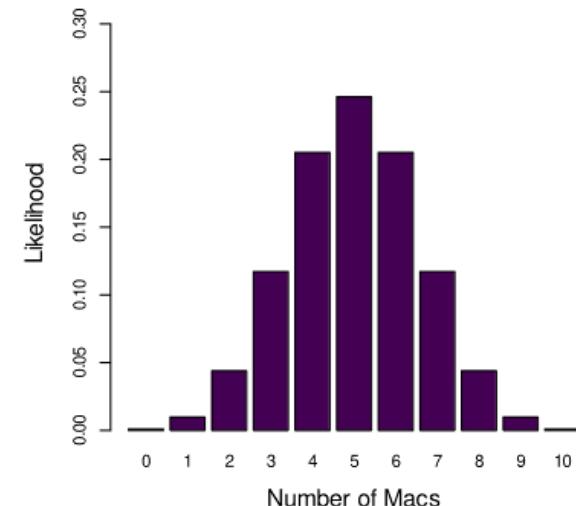
Likely Outcomes



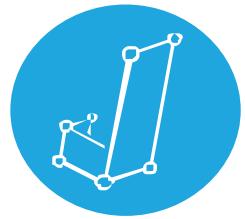
Likely Outcomes



Likely Outcomes

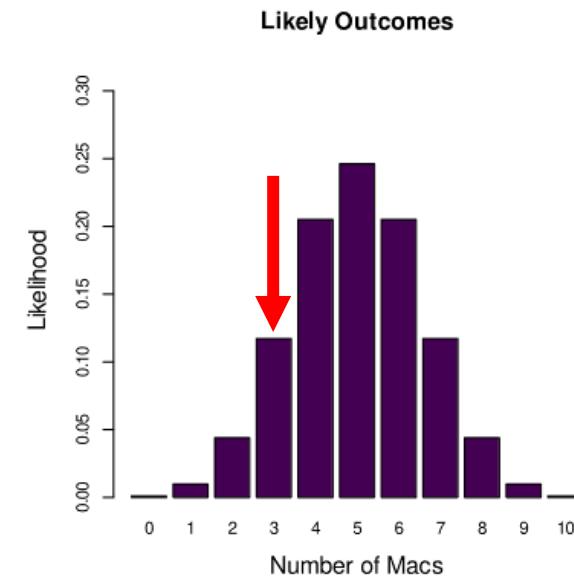
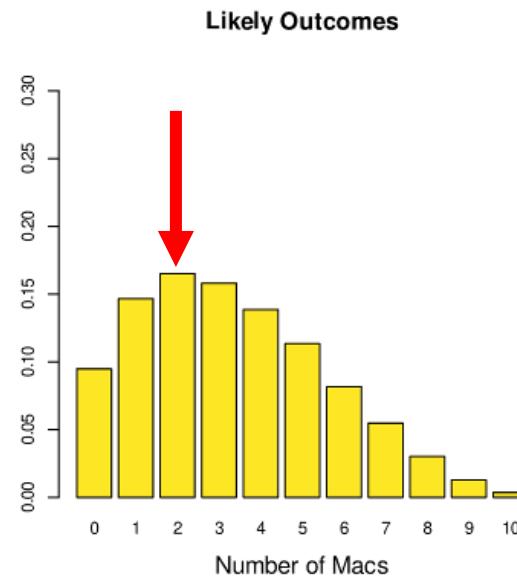
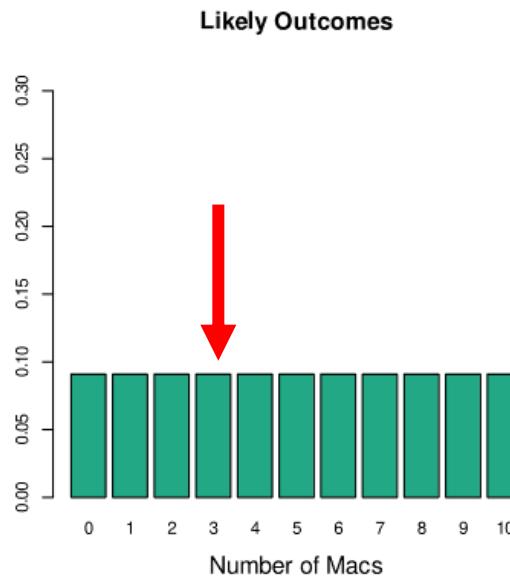


How Well Did The Models Predict The Data?

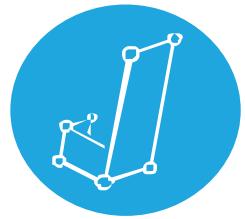


- The models placed their bets, we observed 3 Macs
- How much the model bet on the observed outcome equals the quality of its prediction
- The ratio of two such bars is known as the **Bayes factor**

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

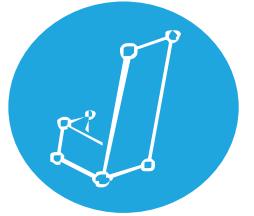


Bayes Factor



- The workhorse of Bayesian hypothesis testing
- It quantifies which of two models predicted the data better, and how much better
- Usually, we pit two models against each other:
 - The null model (H_0), with a spike at the point of testing
 - (e.g., 0.5 for proportion, 0 for correlation or difference in means)
 - The alternative model (H_1), which predicts multiple values (dictated by its prior distribution)
 - The alternative model can represent a one-sided or two-sided hypothesis

Bayes Factor Interpretation



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

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

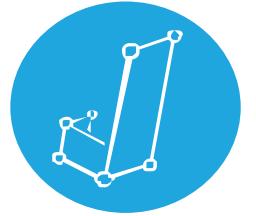
- **Example 2: $BF_{10} = 1/5 \rightarrow BF_{01} = 5$**

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

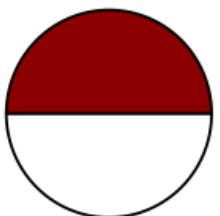
- **Example 3: $BF_{10} = 1$**

The data are equally likely under H_1 and under H_0

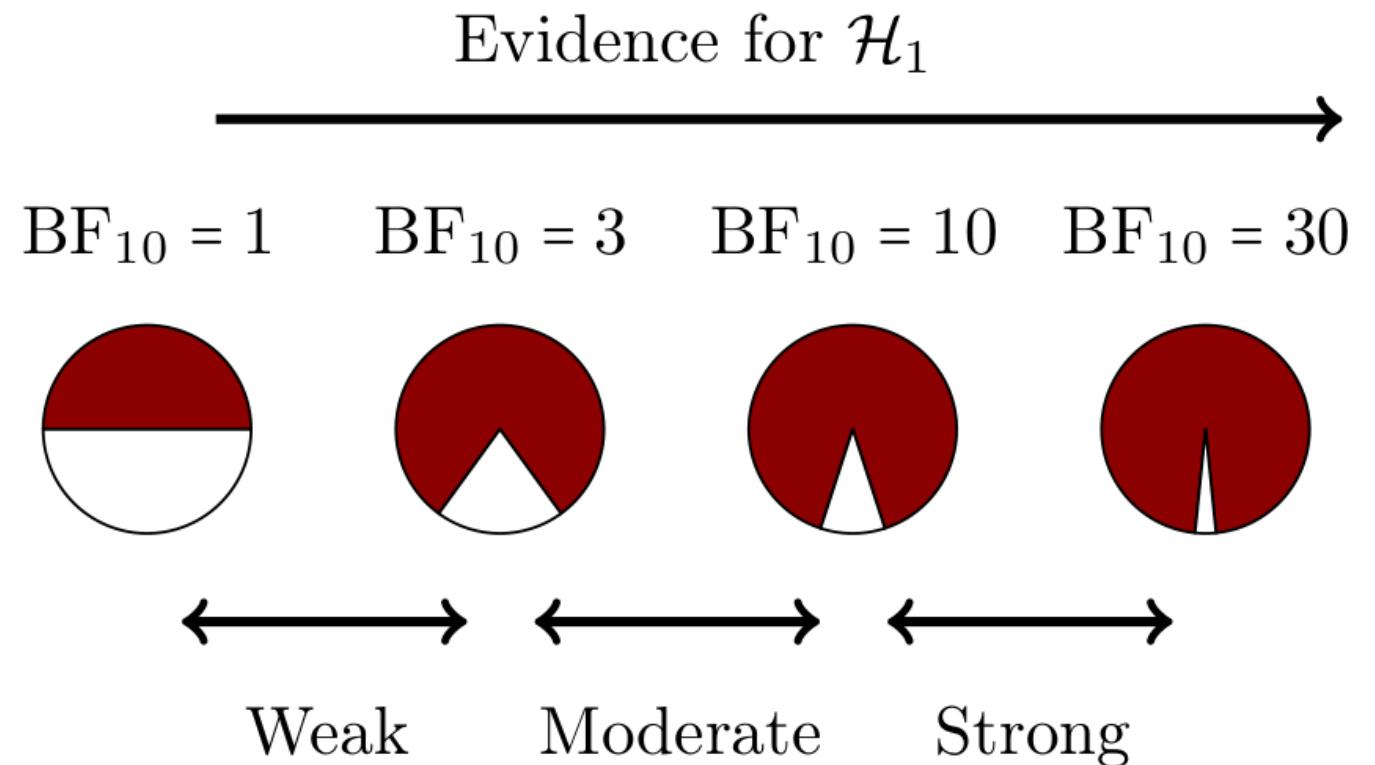
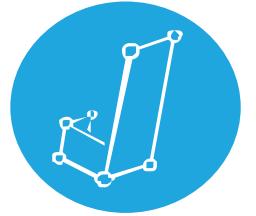
Bayes Factor Interpretation



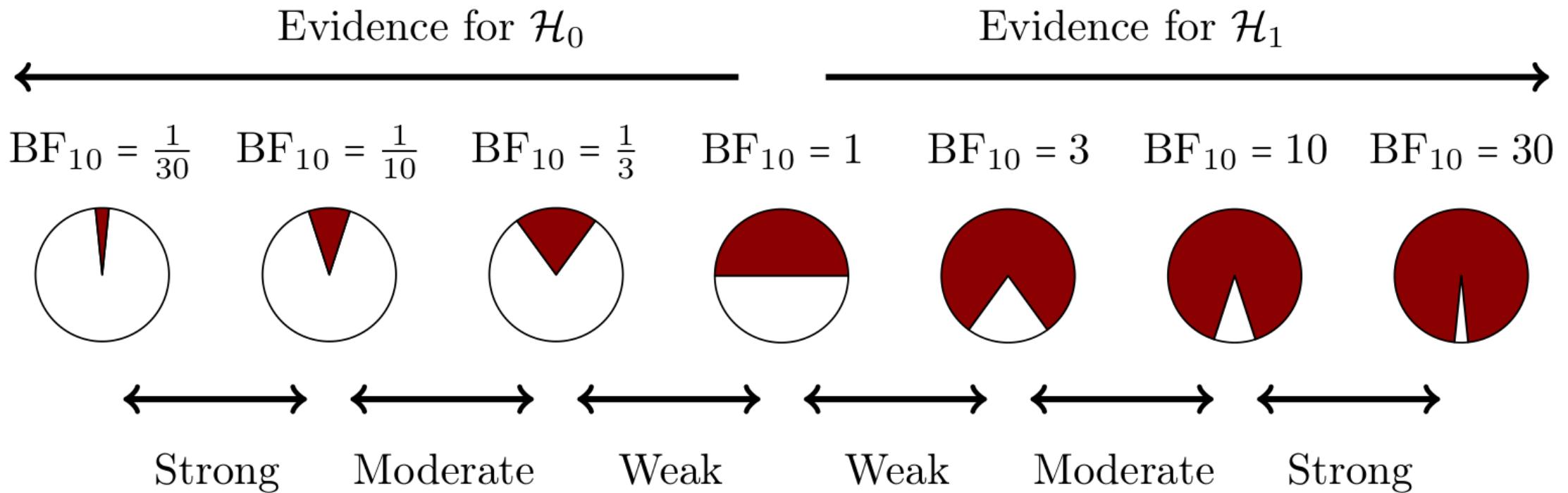
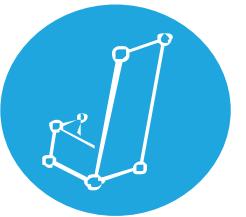
$$BF_{10} = 1$$



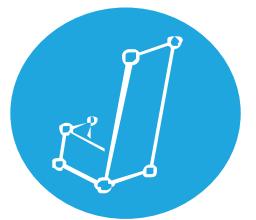
Bayes Factor Interpretation



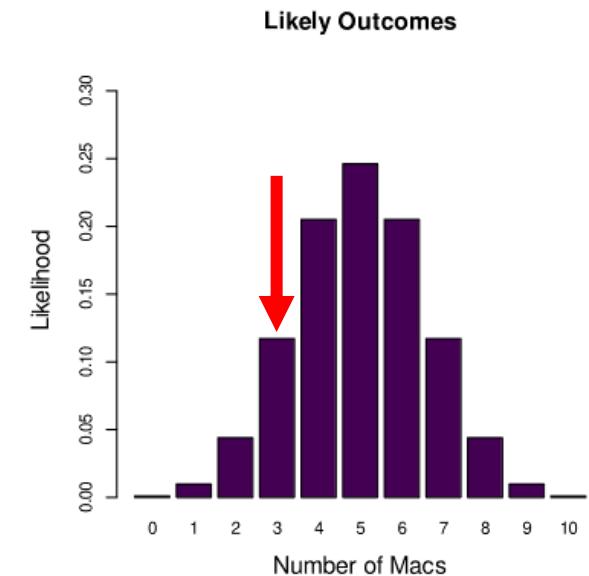
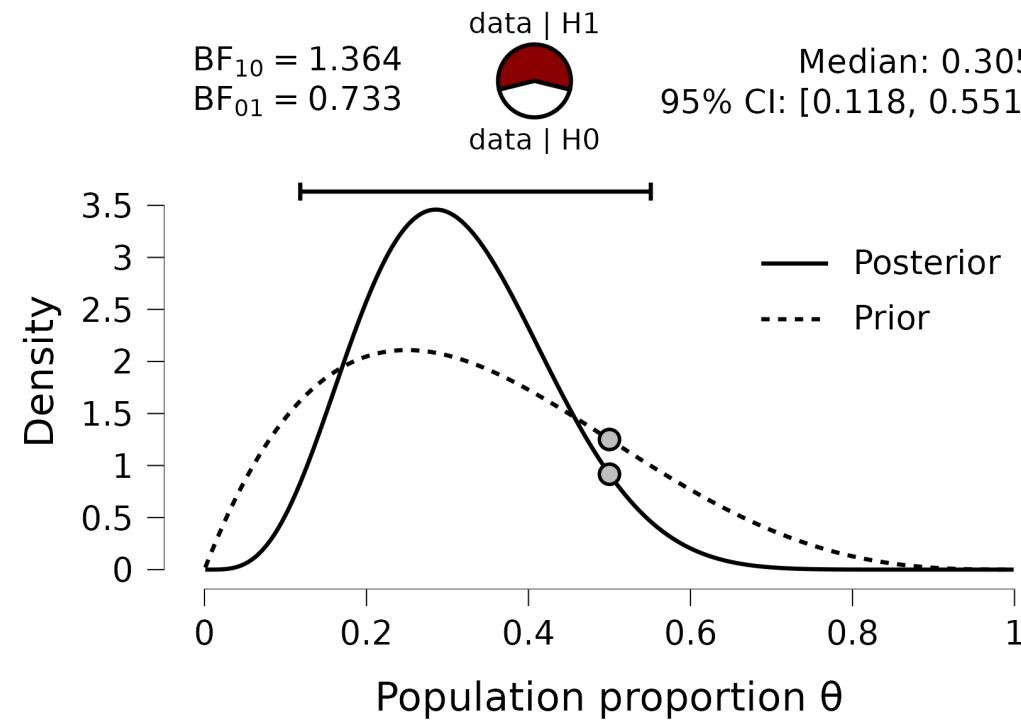
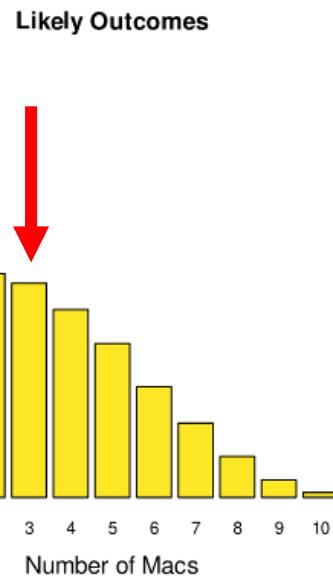
Bayes Factor Interpretation



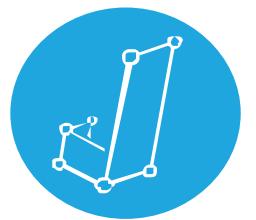
How Well Did The Models Predict The Data?



- For comparing a point model to a range model, we can use the **Savage-Dickey density ratio** as a computational shortcut to obtain the Bayes Factor
 - $\text{BF}_{10} = \text{prior density at point of testing} / \text{posterior density at point of testing}$

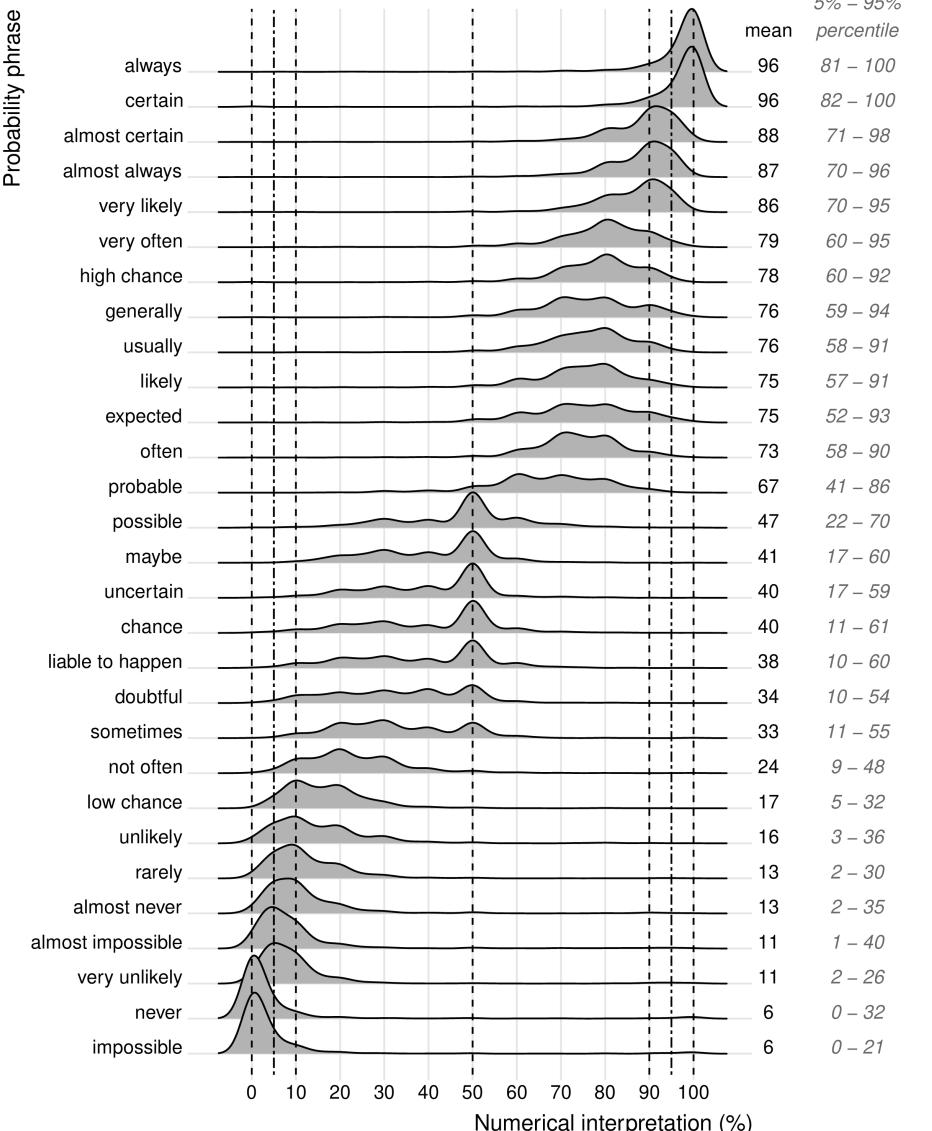


Bayes Factor Interpretation - Caution



But!

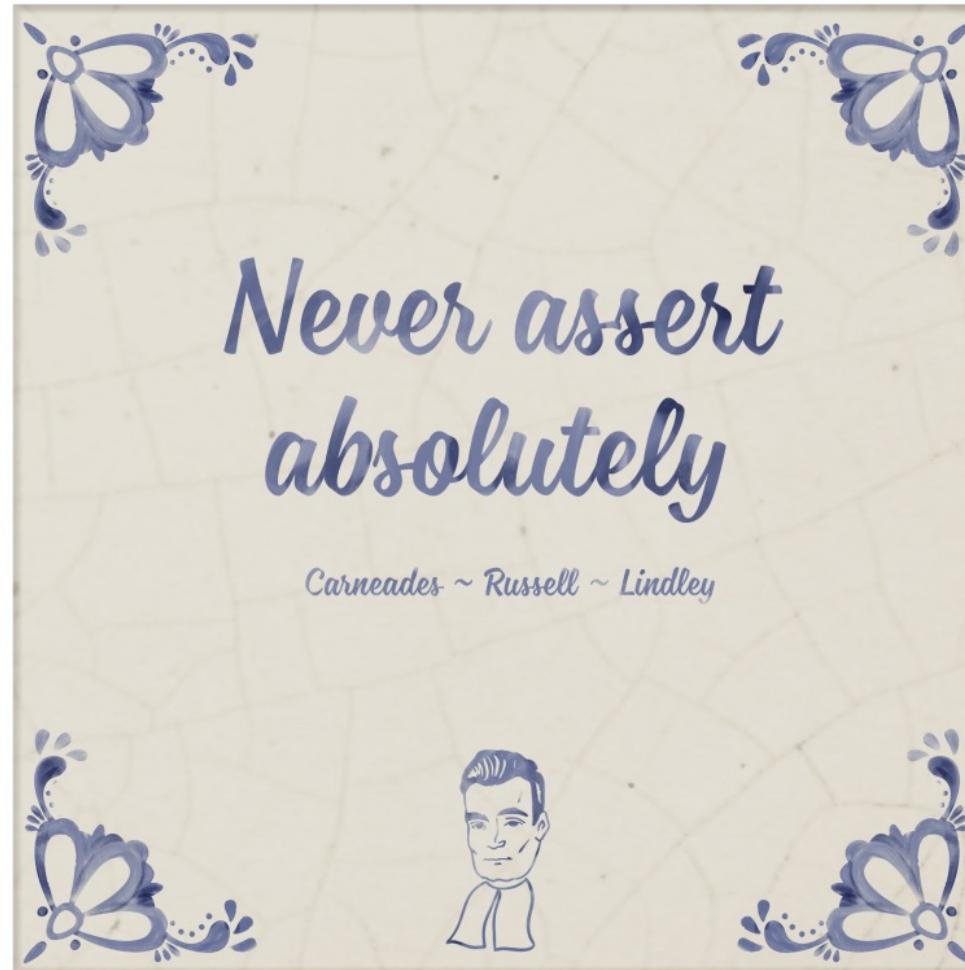
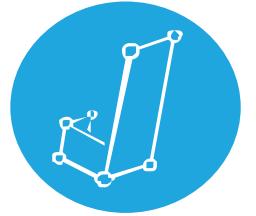
- The general challenge of translating between words and numbers:
- People have different interpretations of probabilistic words/phrases



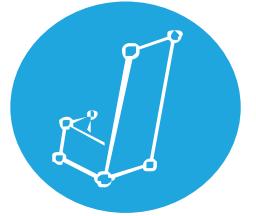
Source:

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.

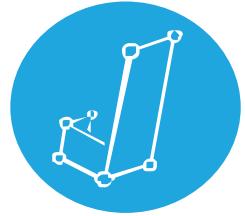
Bayes Factor Interpretation - Caution



Bayes Factor Interpretation - Caution

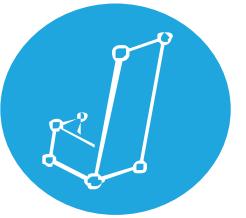


Bayes Factor Features

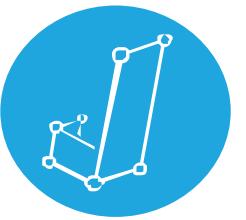


- Parsimony
 - A built in Occam's Razor (e.g., one-sided vs two-sided)
- Sensitivity Analysis
 - Investigate the impact of the prior distribution
- Sequential Analysis
 - Monitor evidence as data accumulate

Bayes & Frequentism



Bayes	Frequentism
Compare the predictive quality of two hypotheses → Invites continuous reasoning	Minimize type 1 error (incorrectly reject H_0) → Invites binary reasoning
Can monitor the results as data accumulate	Can only look at the data once: looking twice increases the risk of type 1 error
Check assumptions	Check assumptions
Still incomplete (but steadily expanding)	Has wide range of types of analyses
Sampling plan: can monitor the results and stop when satisfied	Have to specify sampling plan before, based on hypothesized effect size



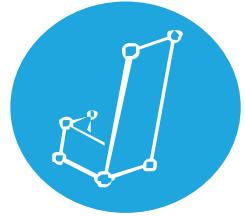
Live Demonstration



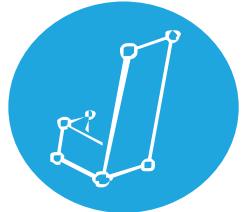
VS



Try It Yourself!

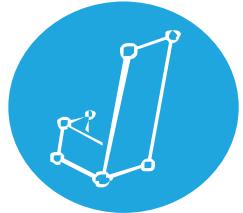


- ▶ Try your new-found knowledge on your own data, or try some of the example exercises on the OSF
 - Or here:
<https://johnnydoorn.github.io/BayesBookdown/exercises.html>
- ▶ Bayesian T-test/ANOVA?



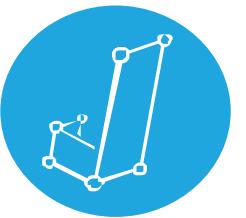
Choosing the Prior Distribution

- Before the experiment / looking at the data (so there are no calculations)
- Informed by previous experiments/knowledge or kept uninformative
(e.g., uniform prior on the population proportion)
- One-sided / two-sided determined by what you want to test
(e.g., do you want to know if there is a difference, or a positive difference?)
- Assess robustness of result to choice of prior distribution



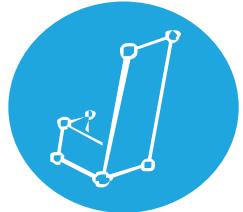
Choosing the Prior Distribution

- The prior distribution is on the same *domain* as the parameter of interest:
 - For a proportion: $[0, 1]$
 - For a correlation: $[-1, 1]$
 - For a difference in means: $[-\infty, \infty]$



Bayesian Correlation

$$P(\rho \mid \text{data}) = P(\rho) \frac{P(\text{data} \mid \rho)}{P(\text{data})}$$



Bayesian Correlation: Hypothesis

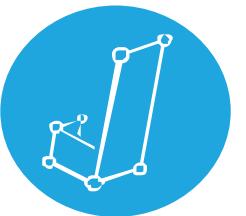
$$\mathcal{H}_0 : \rho = 0$$

$$\mathcal{H}_1 : \rho \neq 0$$

$$\mathcal{H}_+ : \rho > 0$$

$$\mathcal{H}_- : \rho < 0$$

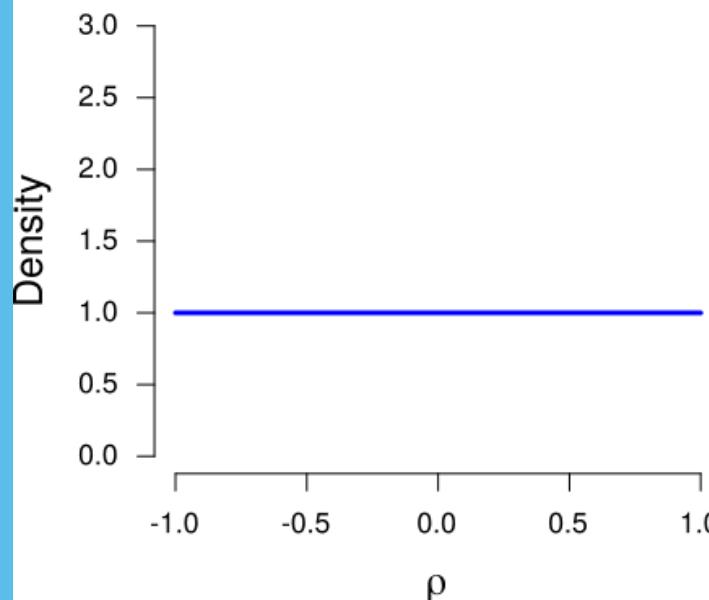
“Alternative Hypothesis”



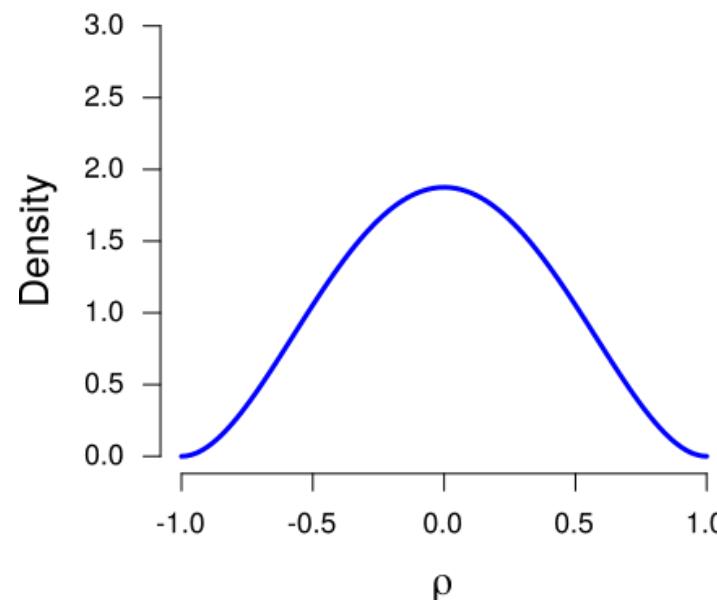
Bayesian Correlation: Prior Distribution

The prior distribution is on the same domain as the parameter of interest: we can take the **stretched Beta distribution!**
It is the same as Beta, but then stretched to the domain $[-1, 1]$

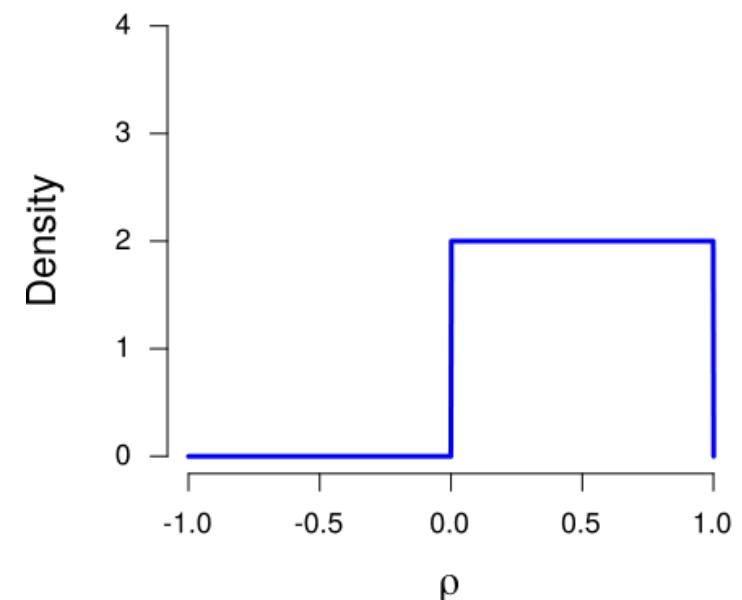
Stretched Beta Distribution ($a = 1, b = 1$)



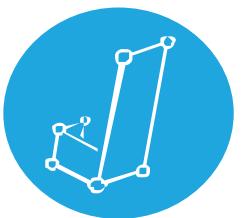
Stretched Beta Distribution ($a = 3, b = 3$)



Truncated Stretched Beta Distribution
($a = 1, b = 1$)



Bayesian Correlation: Prior Distribution

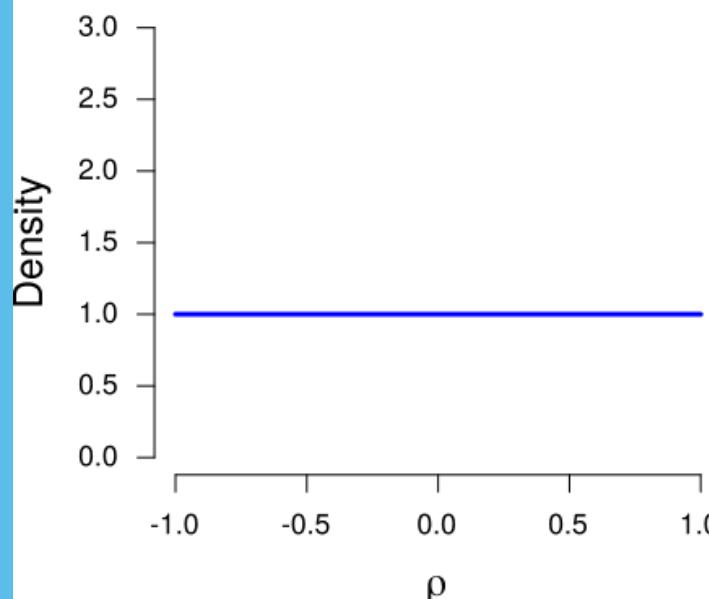


A prior distribution that reflects the belief that all values of ρ are equally plausible, **a priori**, we call this an *uninformative prior*

A prior distribution that reflects the belief that values of ρ close to 0 are more plausible (i.e., the correlation will be around 0 or not be very strong), **a priori**

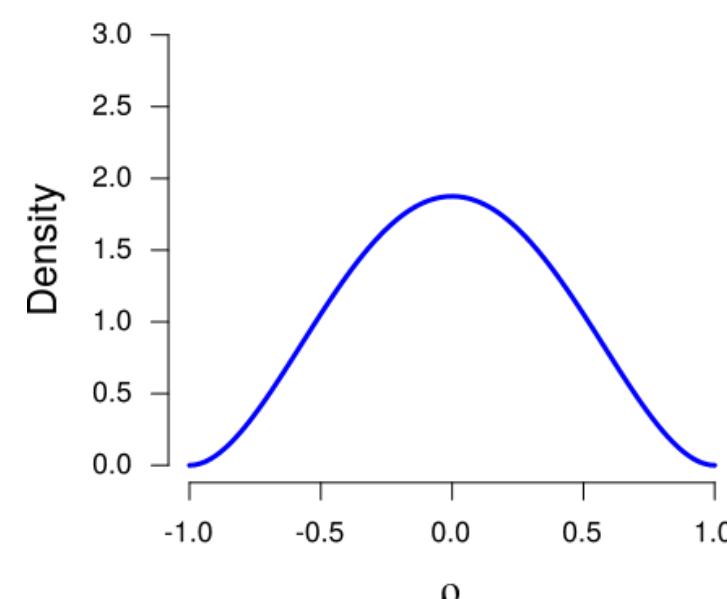
A prior distribution that reflects the belief that only positive values of ρ are possible and that those values of ρ are equally plausible (i.e., there will be a positive correlation), **a priori**

Stretched Beta Distribution ($a = 1, b = 1$)



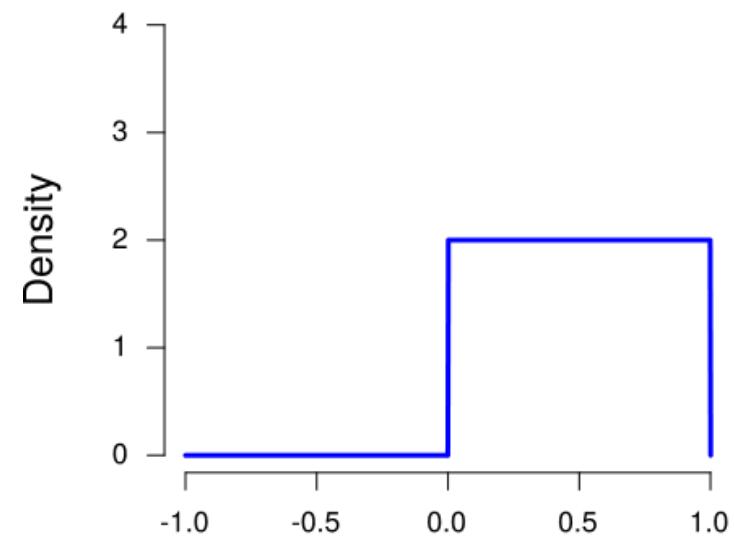
$$\mathcal{H}_1 : \rho \neq 0$$

Stretched Beta Distribution ($a = 3, b = 3$)

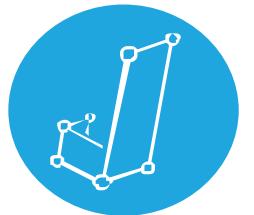


$$\mathcal{H}_1 : \rho \neq 0$$

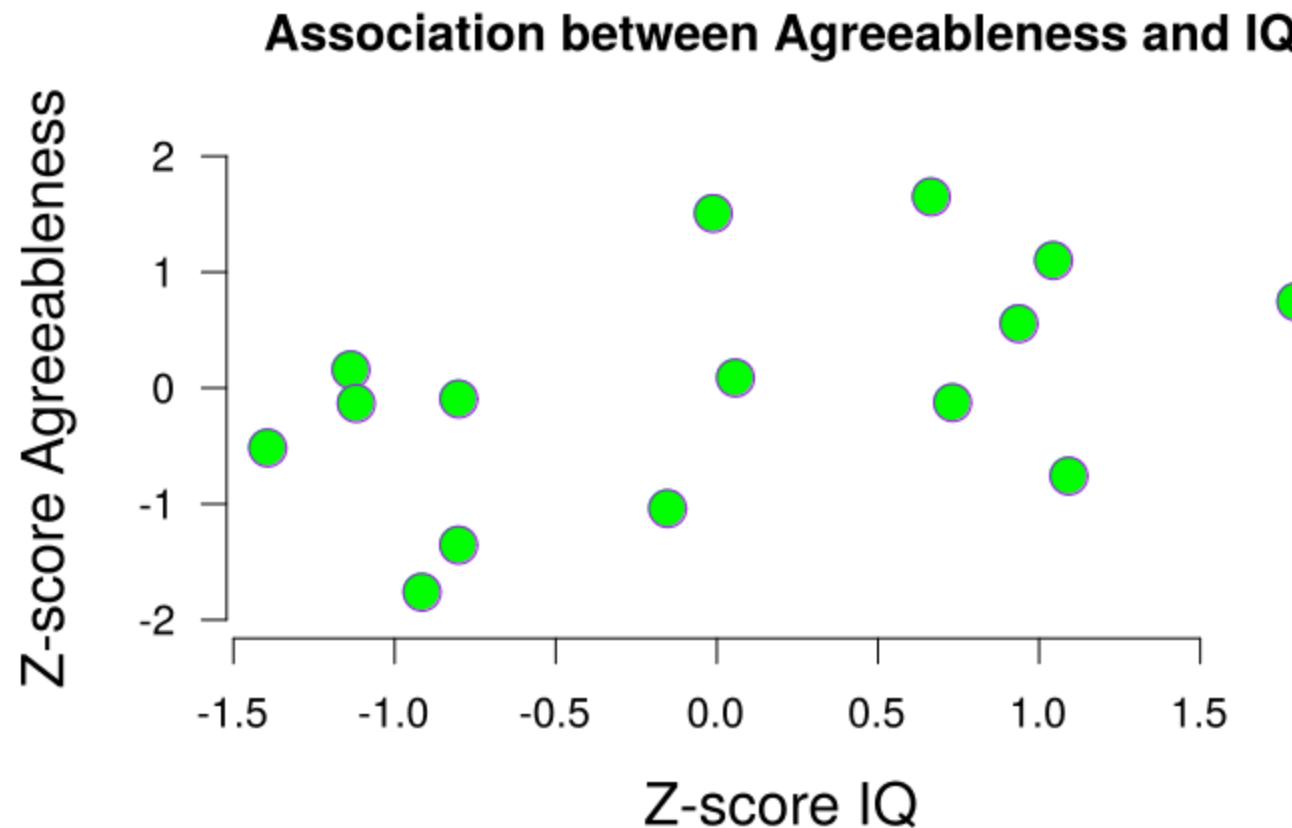
Truncated Stretched Beta Distribution ($a = 1, b = 1$)

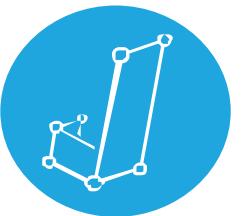


$$\mathcal{H}_+ : \rho > 0$$

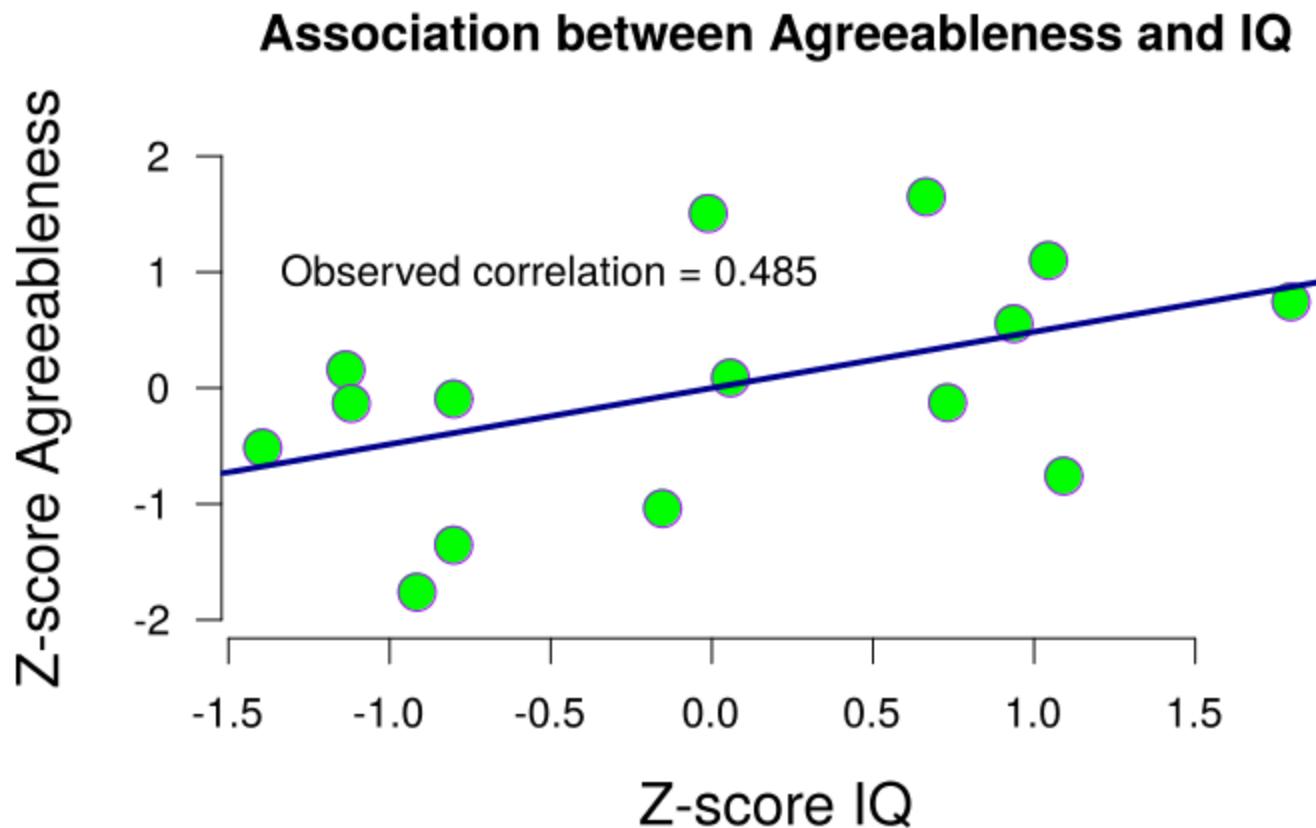


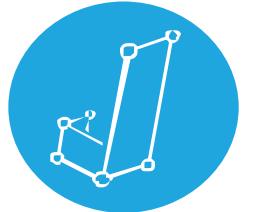
Bayesian Correlation: Data





Bayesian Correlation: Data

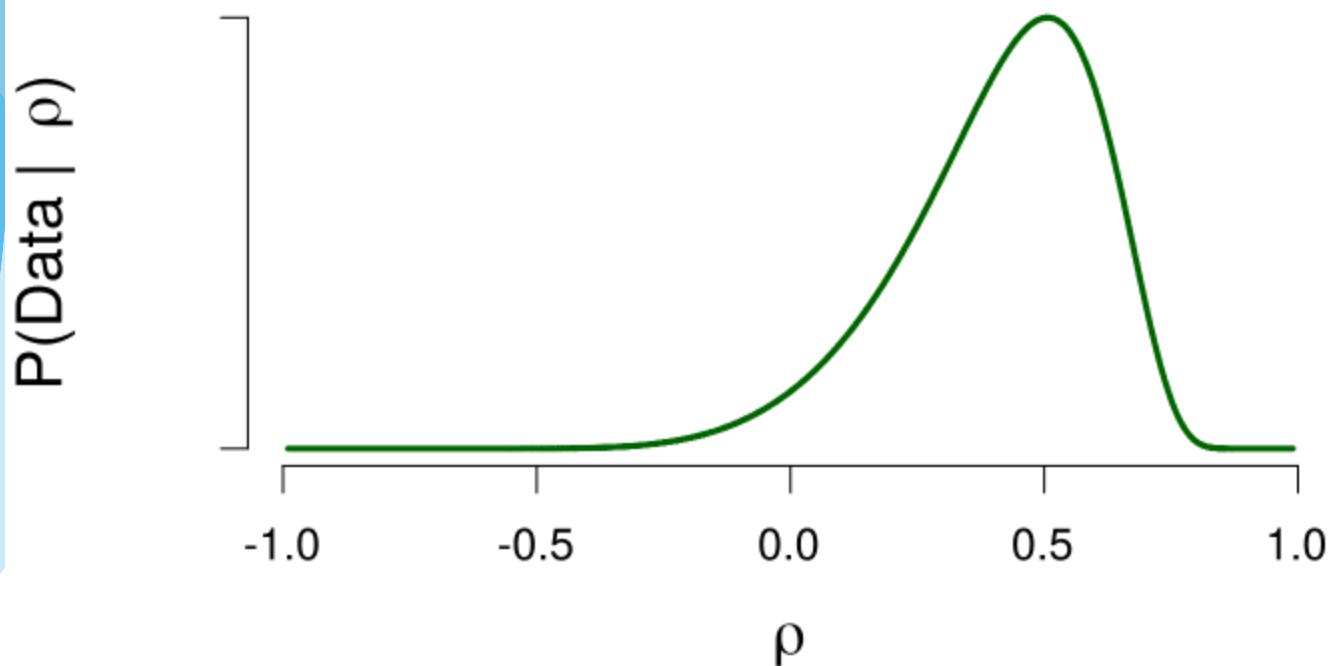




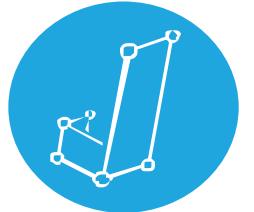
Bayesian Correlation: Likelihood

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

Likelihood of the observed data, for each value of ρ



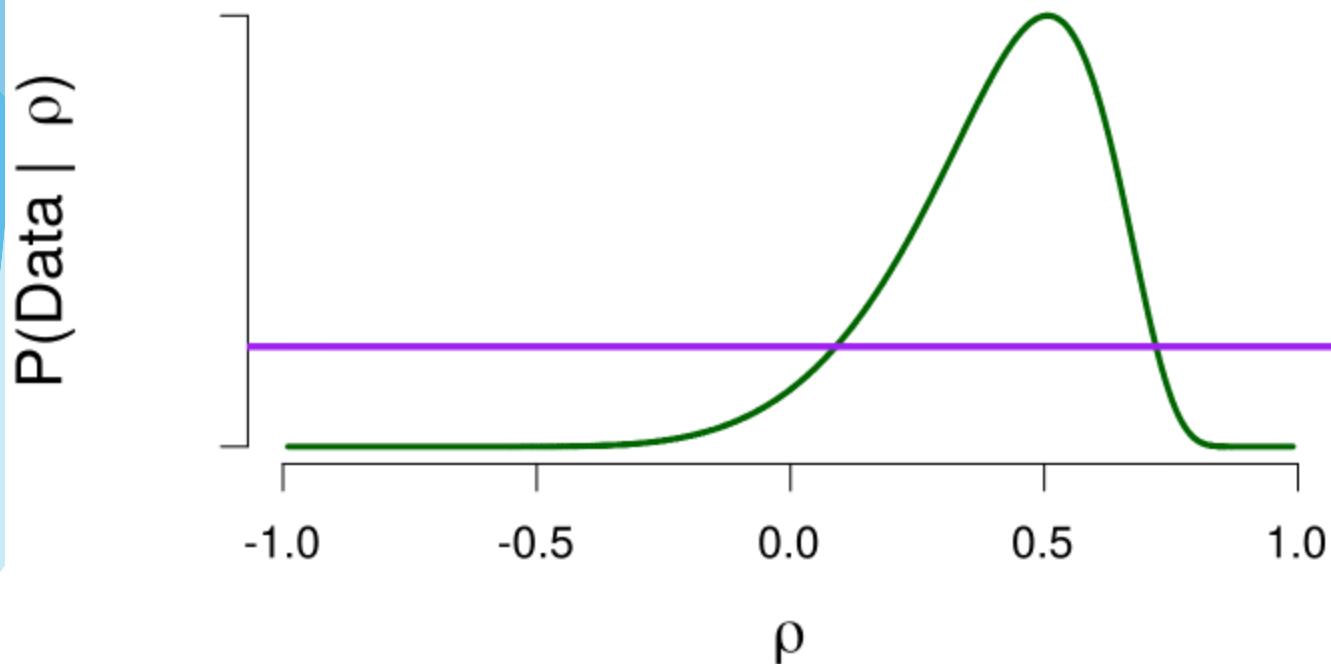
We see that the data are likely for values of ρ close to 0.5.
This makes sense, because the observed correlation (i.e., the data) is equal to 0.485!



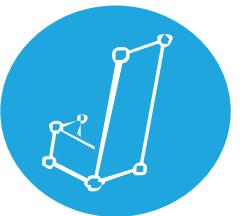
Bayesian Correlation: Likelihood

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

Likelihood of the observed data, for each value of ρ

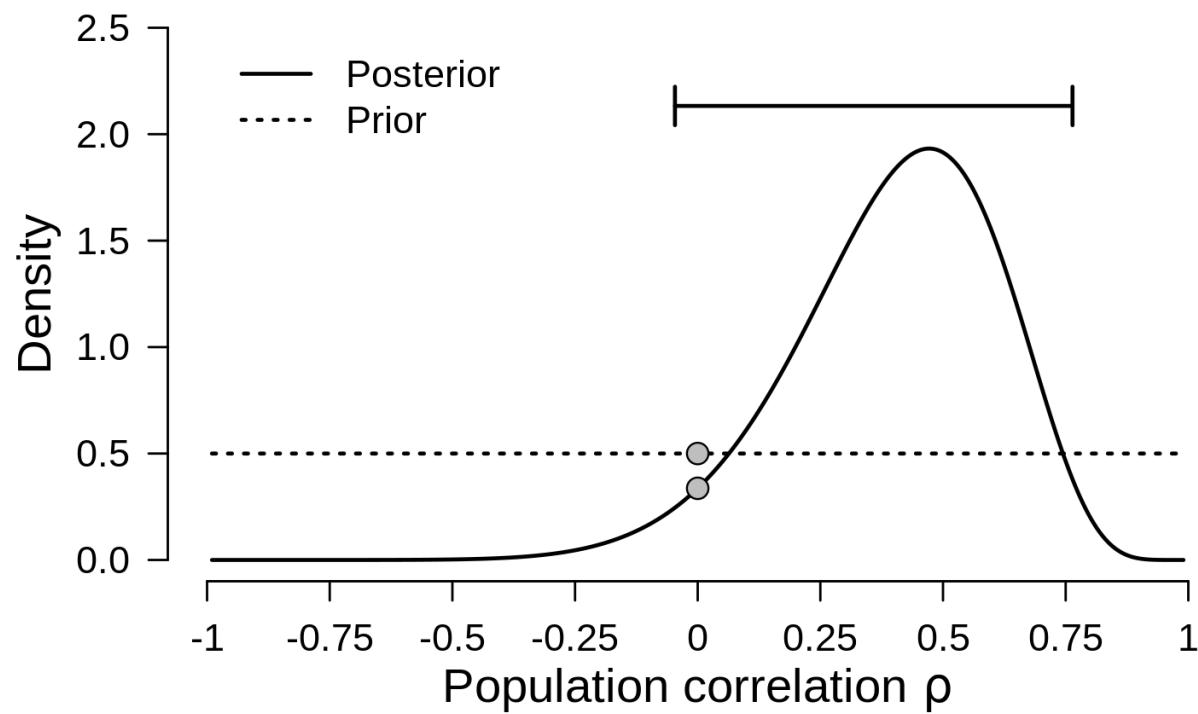


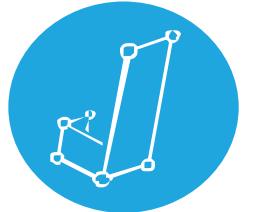
The **marginal likelihood**,
across all values of ρ



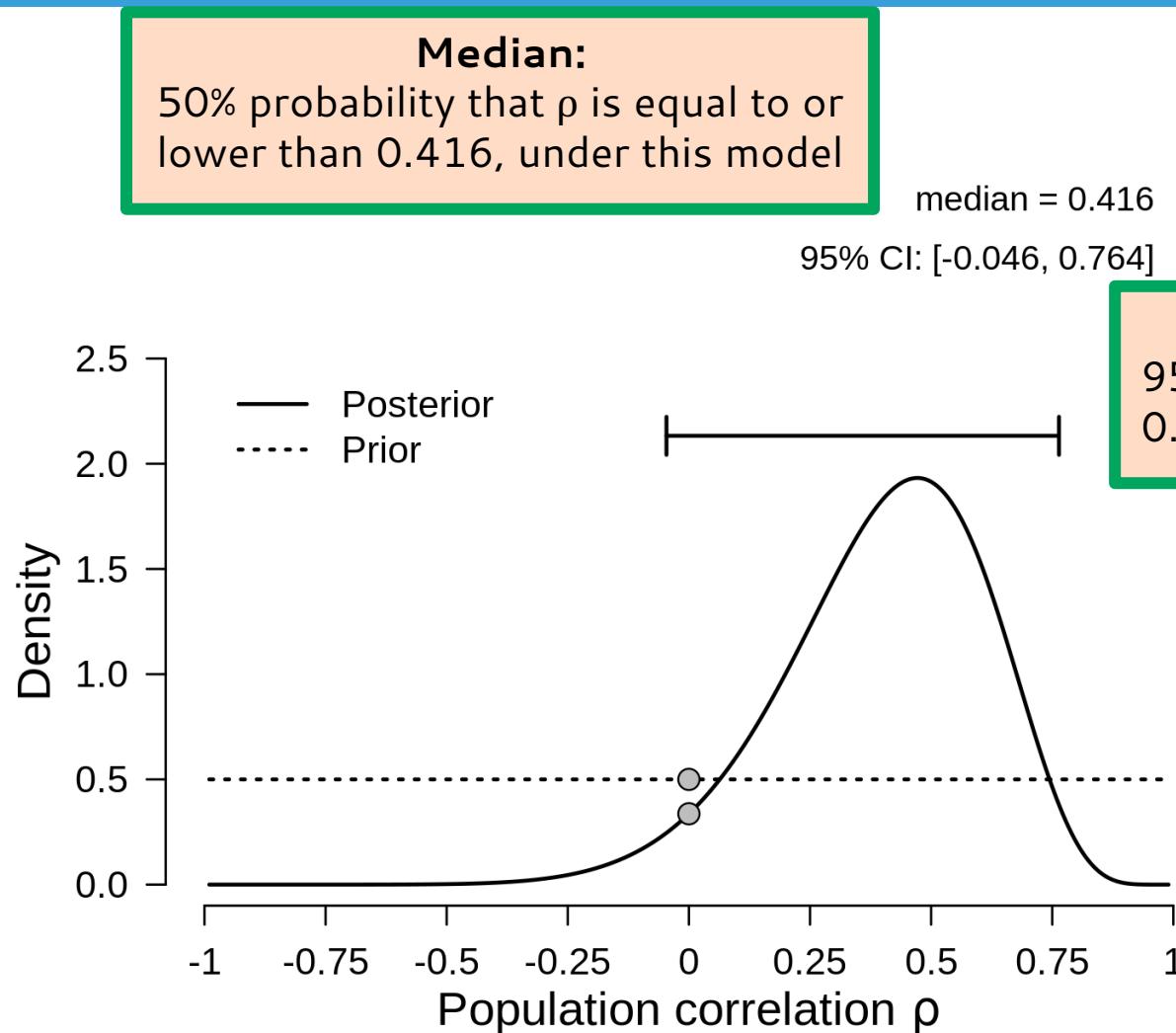
Bayesian Correlation: Posterior Distribution

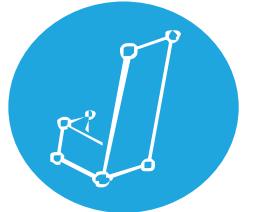
$$P(\rho \mid \text{data})$$



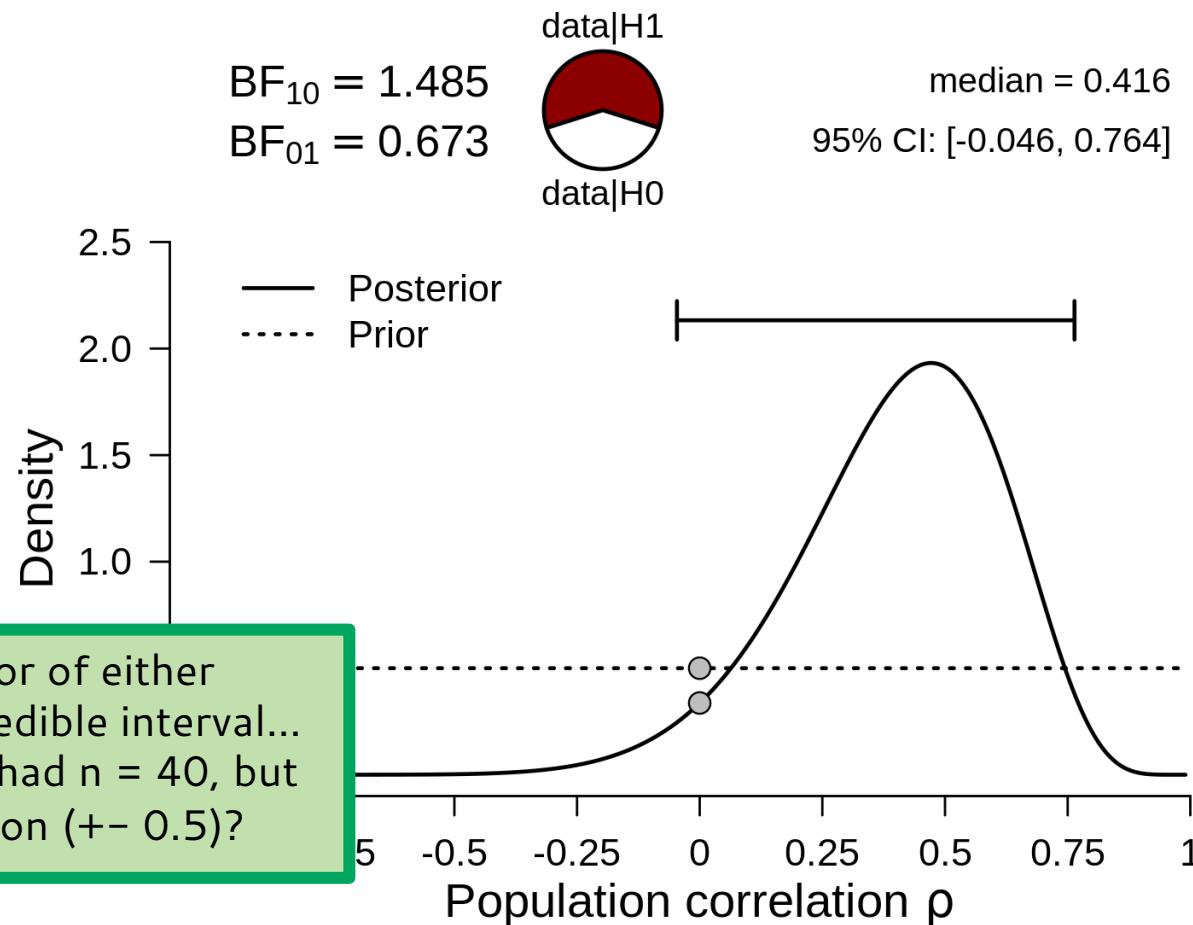


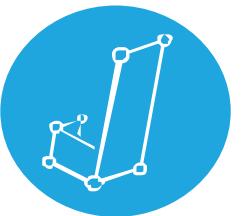
Bayesian Correlation: Posterior Distribution



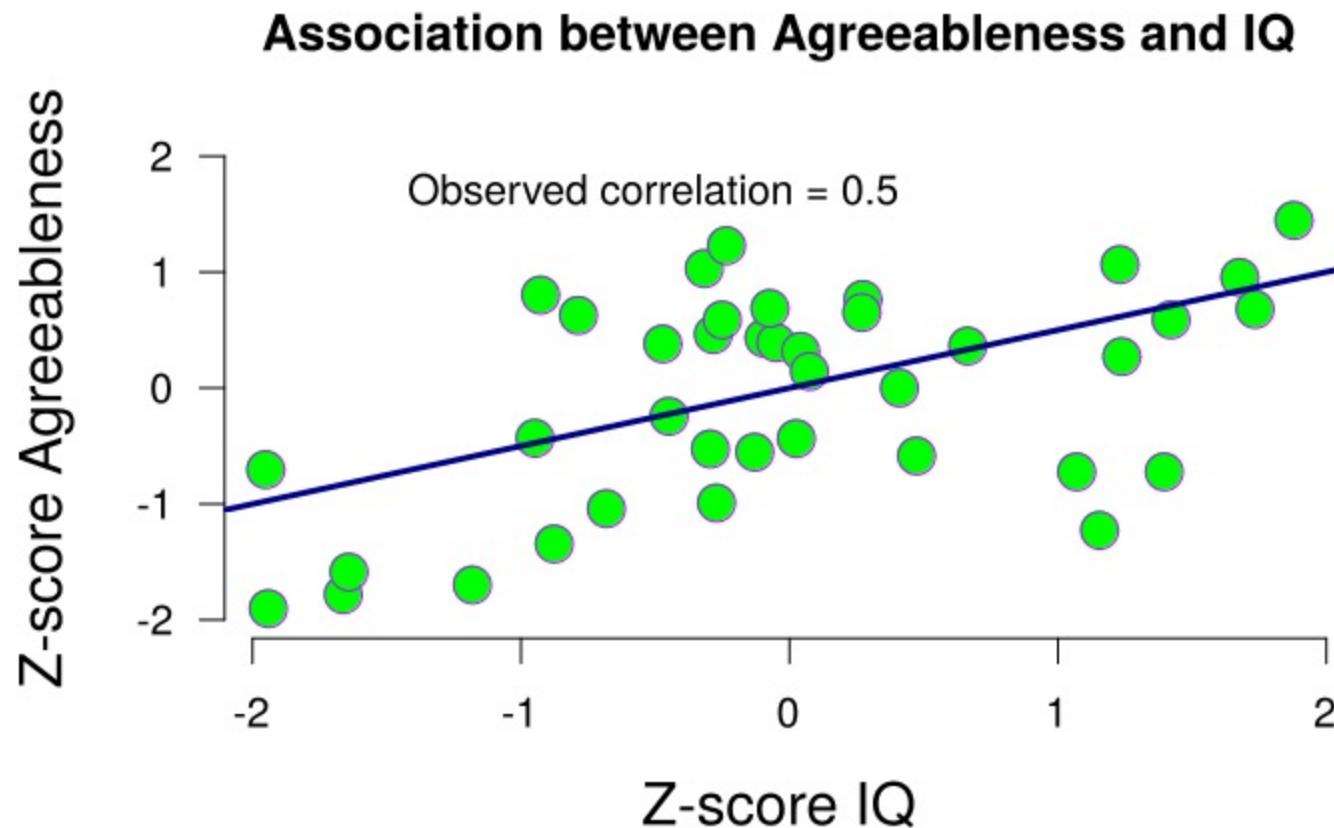


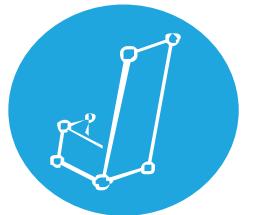
Bayesian Correlation: Bayes Factor





Bayesian Correlation: Data

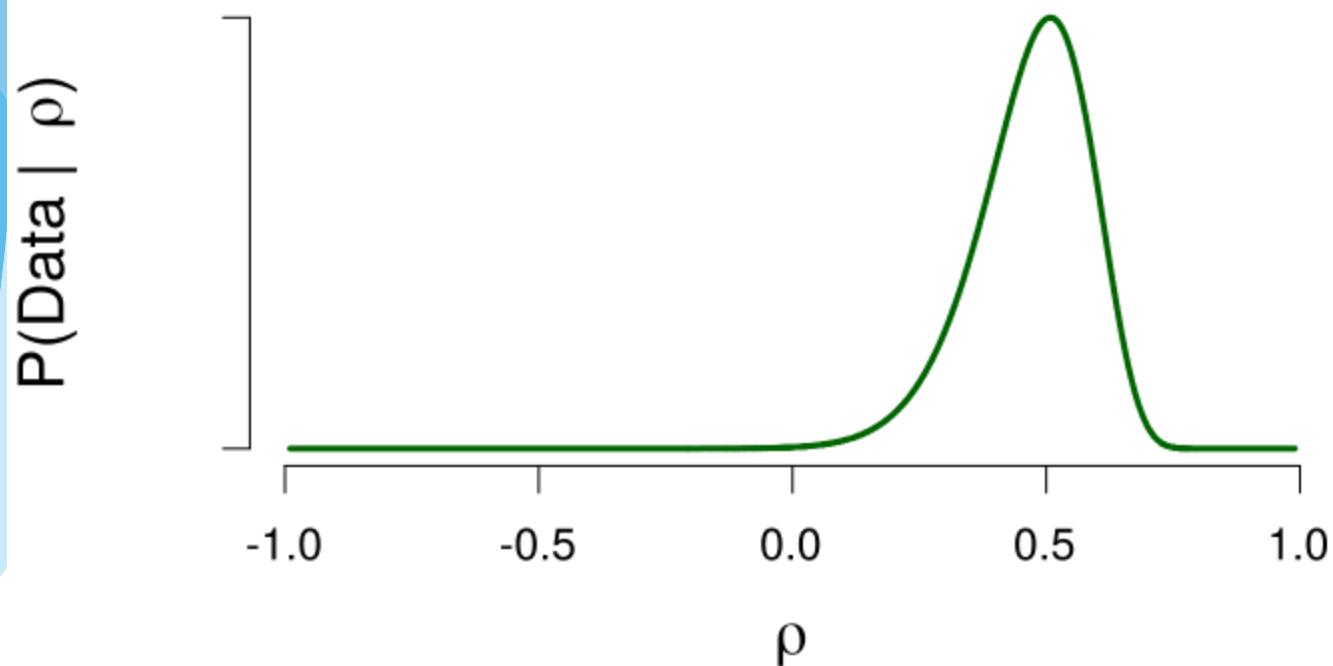




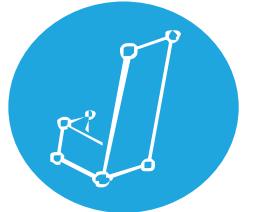
Bayesian Correlation: Likelihood

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

Likelihood of the observed data, for each value of ρ



The likelihood grew more narrow, compared to when we had a lower sample size!



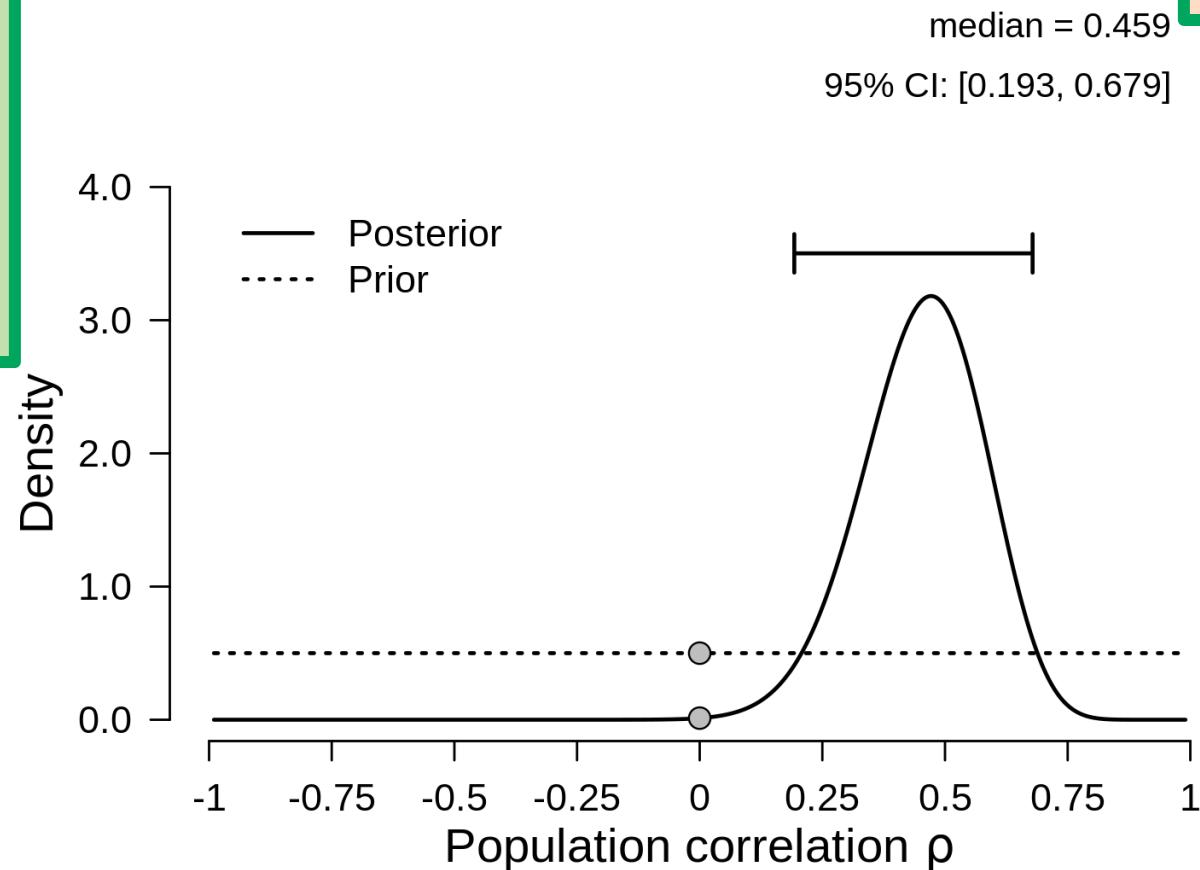
Bayesian Correlation: Posterior Distribution

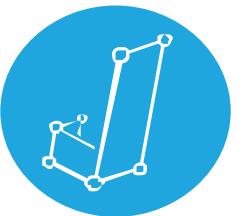
The posterior is also more narrow!

This also leads to a more narrow credible interval (still 95%, but the two numbers are closer together (e.g., 0 is no longer in the interval))

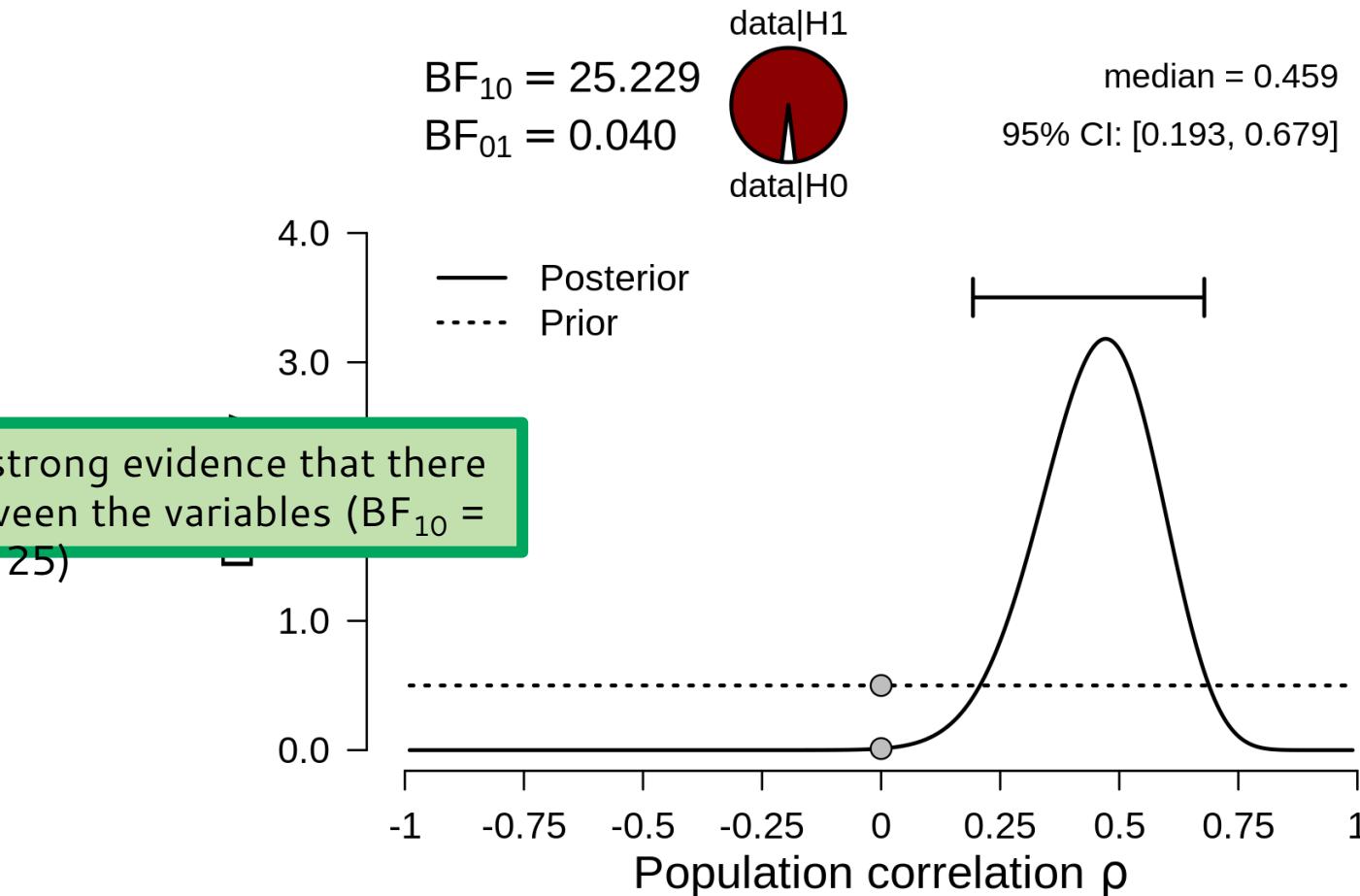
We can make a more specific prediction, with the same certainty (95%)

95% Credible interval:
“95% probability that ρ is between 0.193 and 0.679”

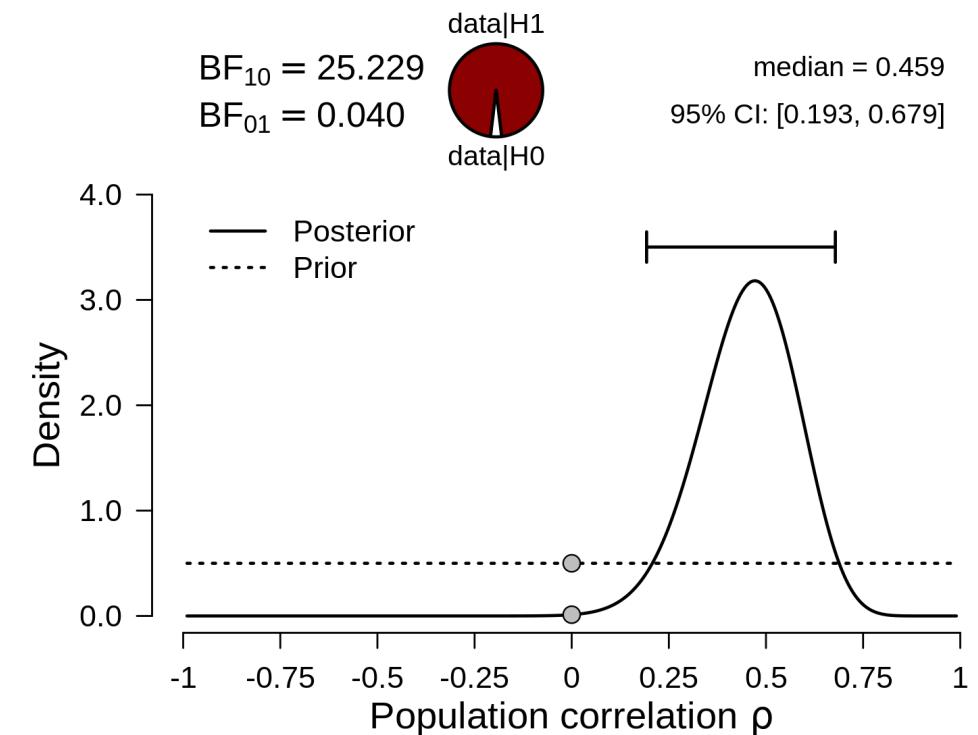
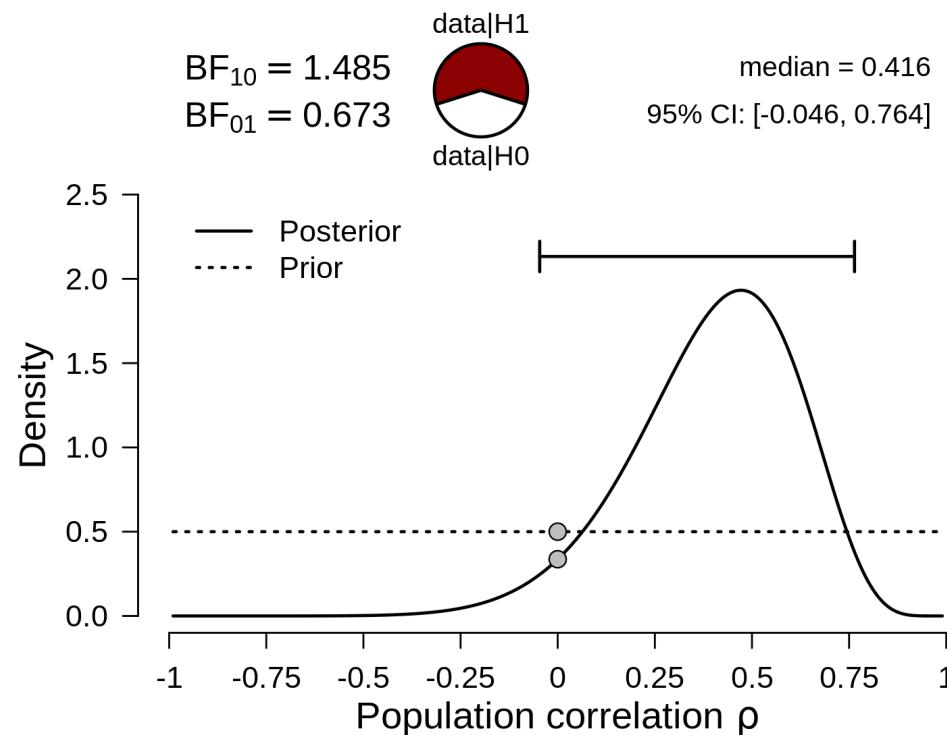
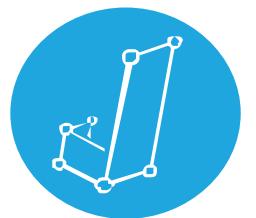




Bayesian Correlation: Bayes Factor



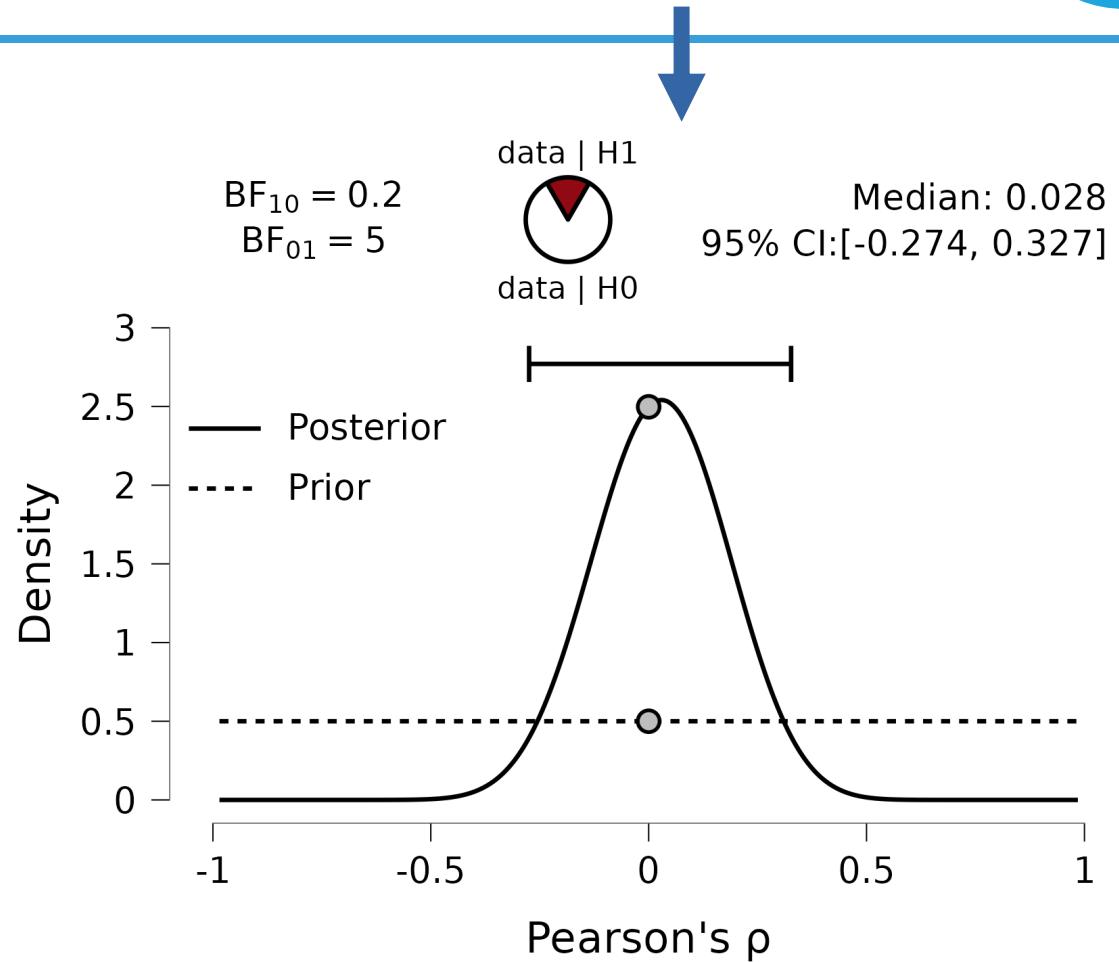
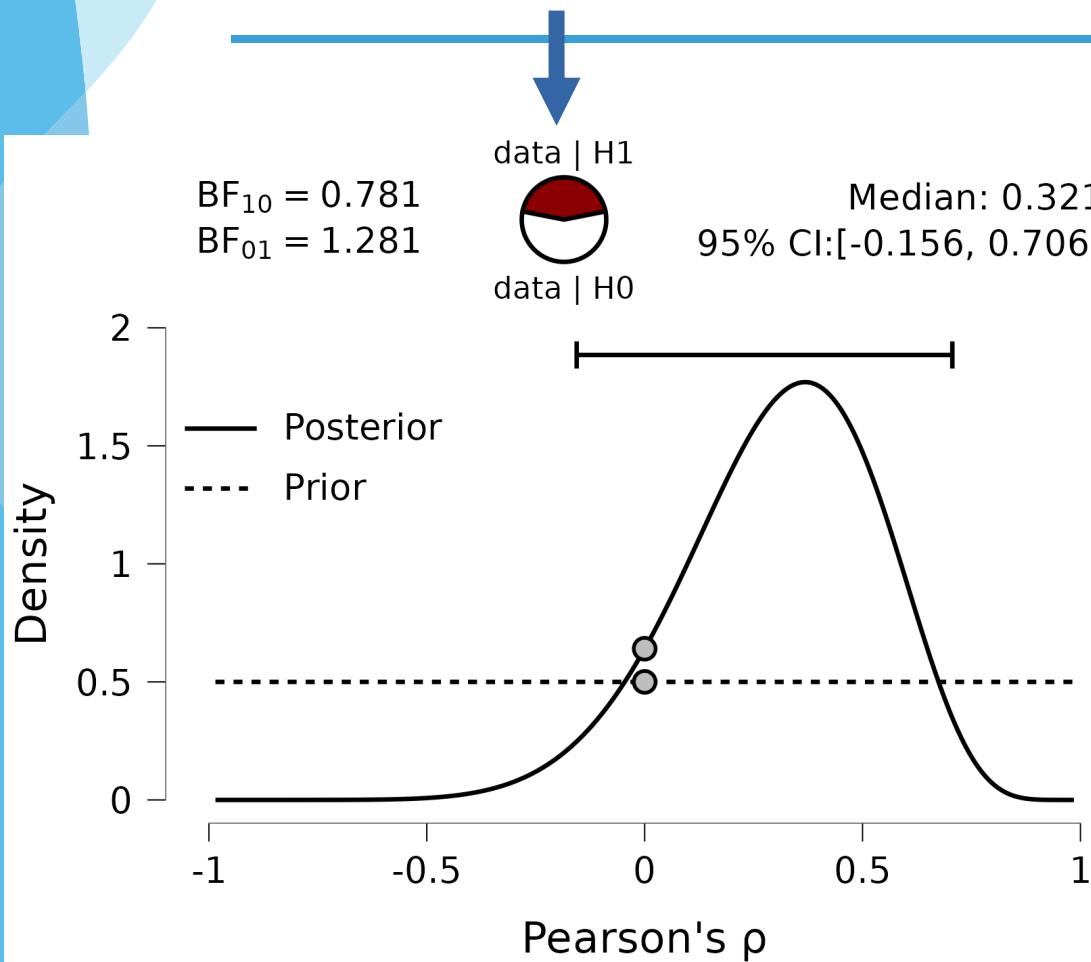
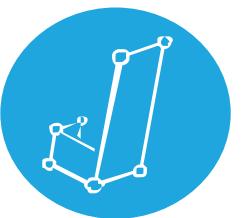
Bayesian Correlation: Bayes Factor



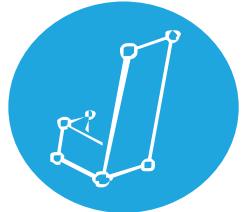
N = 15, observed correlation = 0.485

N = 40, observed correlation = 0.5

“Absence of evidence” vs “evidence of absence”



Both analyses would yield a non-significant p-value – but what does that actually mean?



Bayesian T-Test

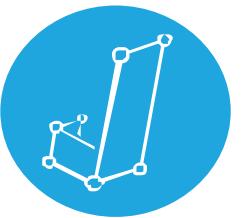
- What about differences between/within groups?
- In psychology, differences between groups are characterized by δ ("delta", also known as Cohen's d): a *standardized* difference between groups

Effect size	d	Reference
Very small	0.01	Sawilowsky, 2009
Small	0.20	Cohen, 1988
Medium	0.50	Cohen, 1988
Large	0.80	Cohen, 1988
Very large	1.20	Sawilowsky, 2009
Huge	2.0	Sawilowsky, 2009

As with the Bayes factor, this is a table to provide some intuition for the magnitude of an effect – there are no hard cut-off values here!

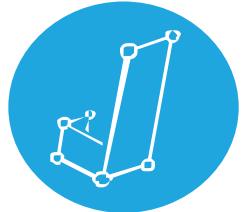
For further reading and formulas

https://en.wikipedia.org/wiki/Effect_size#Difference_family:_Effect_sizes_based_on_differences_between_means



Bayesian T-Test

$$P(\delta \mid \text{data}) = P(\delta) \frac{P(\text{data} \mid \delta)}{P(\text{data})}$$



Bayesian T-Test: Hypothesis

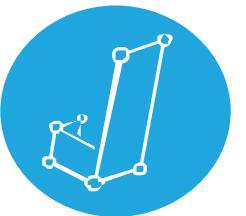
$$\mathcal{H}_0 : \delta = 0$$

$$\mathcal{H}_1 : \delta \neq 0$$

$$\mathcal{H}_+ : \delta > 0$$

$$\mathcal{H}_- : \delta < 0$$

“Alternative Hypothesis”

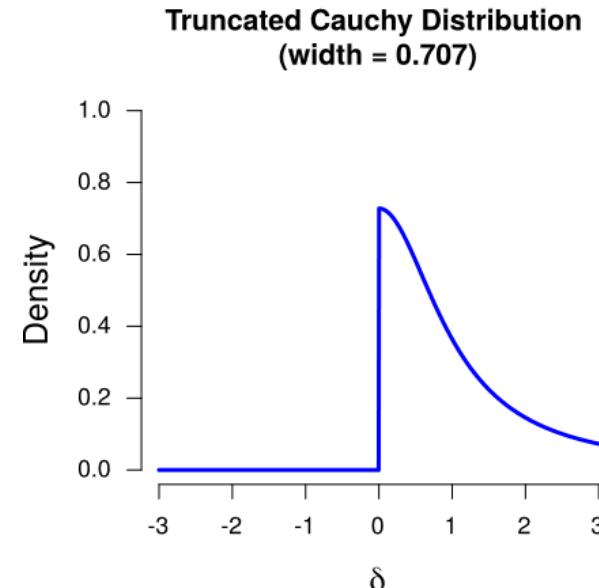
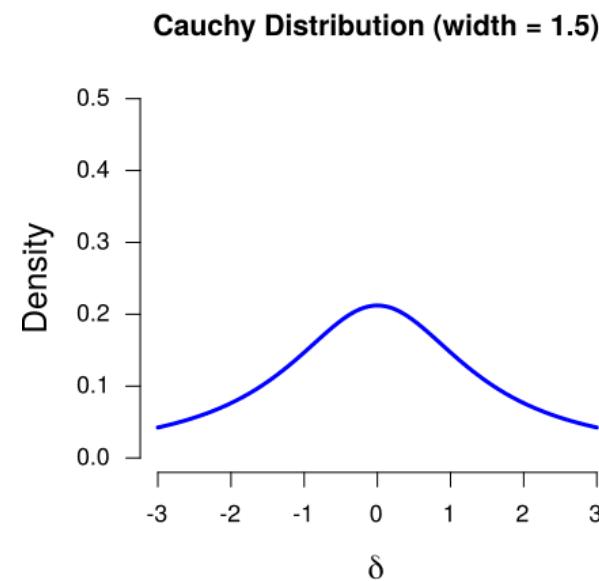
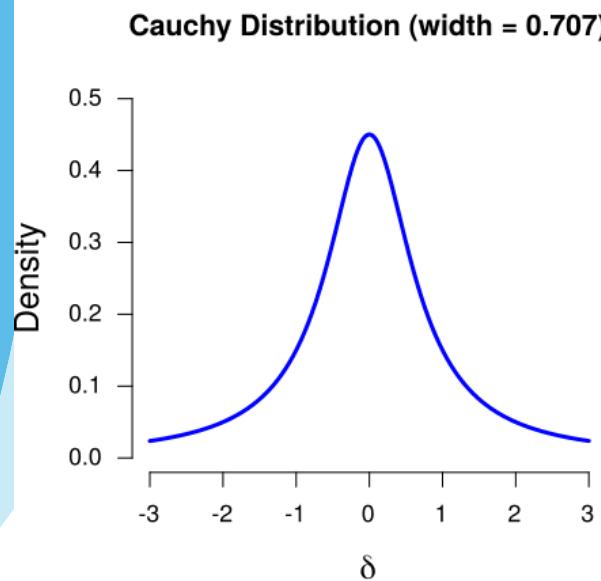


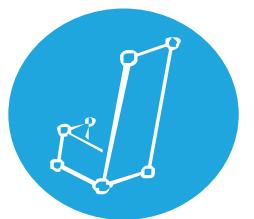
Bayesian T-Test: Prior Distribution

The prior distribution is on the same domain as the parameter of interest: so we need a distribution that is between $[-\infty, \infty]$. A normal distribution could work, but the convention is to use a t-distribution with $df = 1$. This is known as the **Cauchy distribution**

A uniform distribution would not work here, because delta has an infinite domain

The Cauchy distribution is governed by a single shape parameter that determines how wide it is

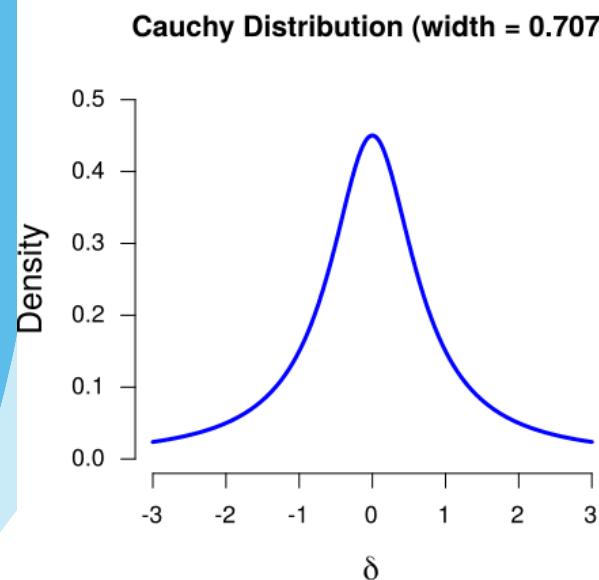




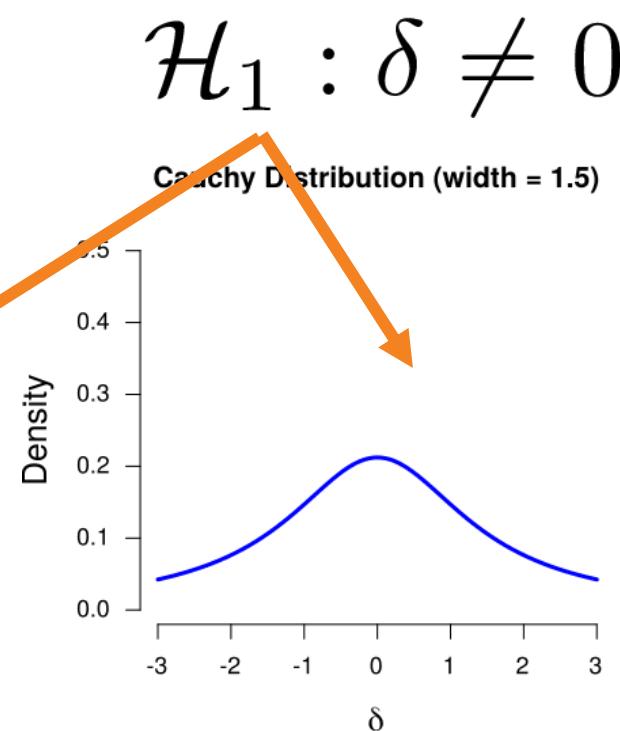
Bayesian T-Test: Prior Distribution

A prior distribution that reflects the belief that 50% of the values of δ are located between -0.707 and 0.707 , **a priori**

This is often selected as the uninformative option. Why?
Unfortunately that is a long and technical story, not suited for this course :(

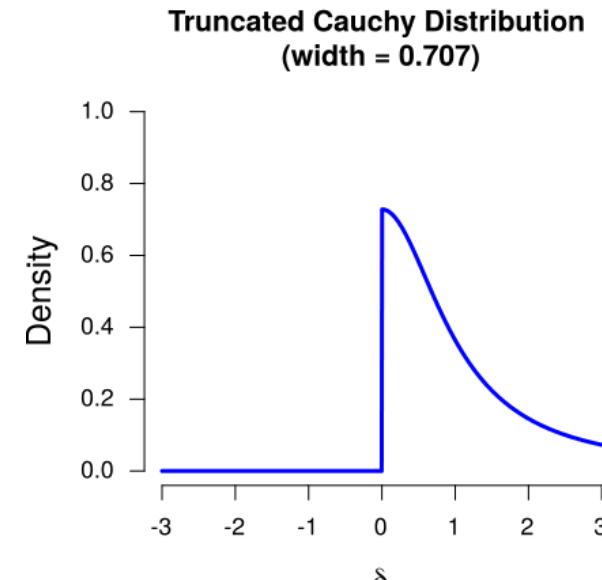


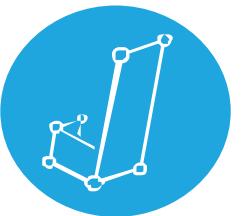
A prior distribution that reflects the belief that 50% of the values of δ are located between -1.5 and 1.5 , **a priori**



A prior distribution that reflects the belief that only positive values of δ are possible and that 50% of the values of δ are between 0 and 0.707 , **a priori**

$\mathcal{H}_+ : \delta > 0$

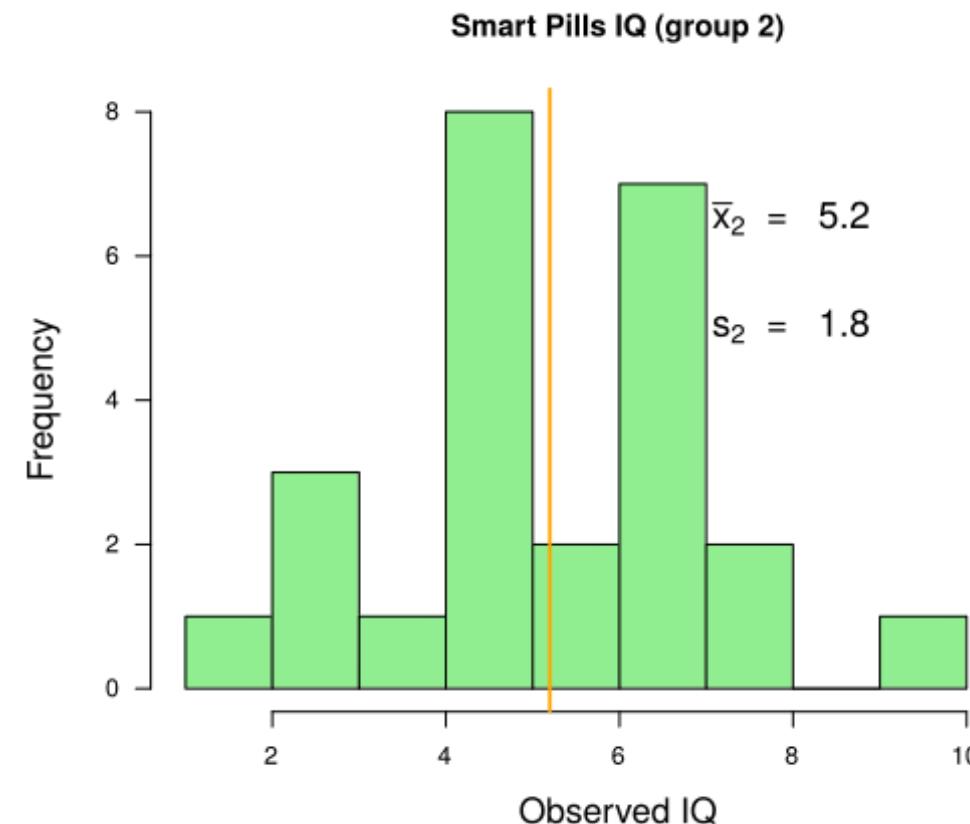
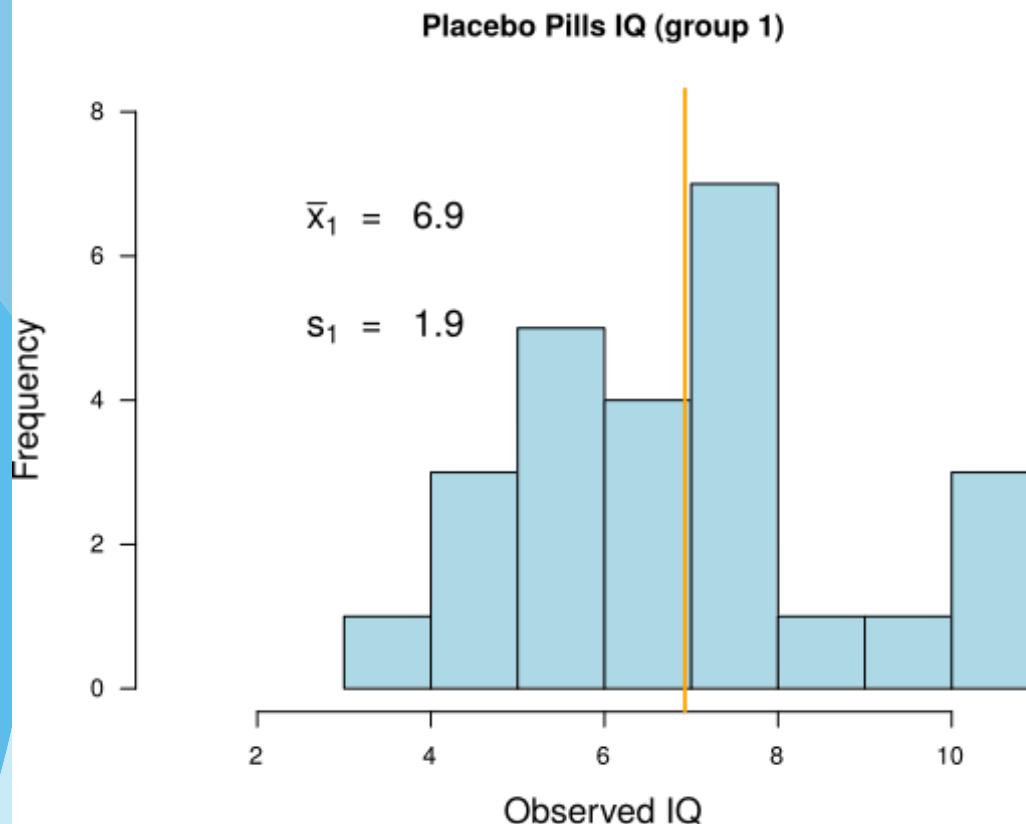


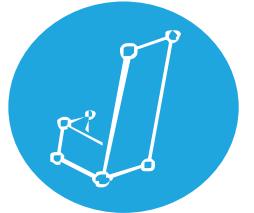


Bayesian T-Test: Data

n = 25

Observed $\delta = 0.95$
($t = 3.27$)

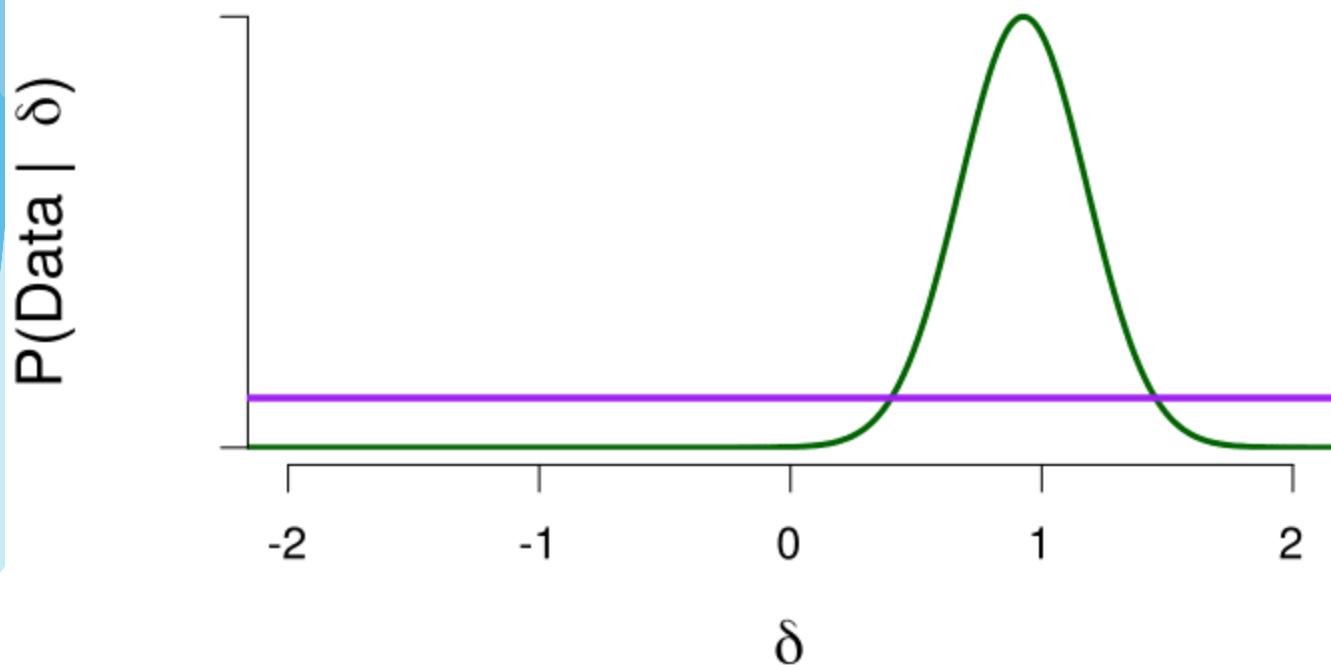




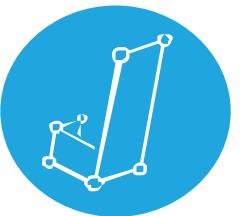
Bayesian T-Test: Likelihood

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

Likelihood of the observed data, for each value of δ



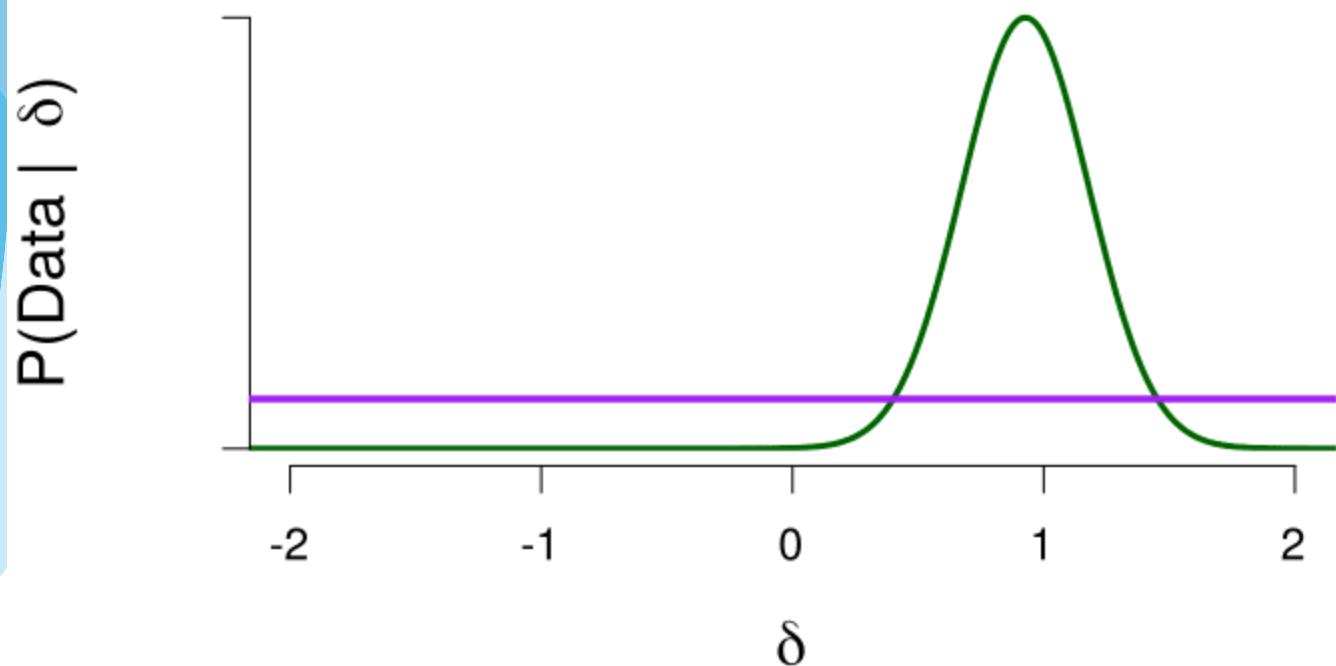
We see that the data are likely for values of δ close to 1.
This makes sense, because the observed delta (i.e., the data) is equal to 0.95!



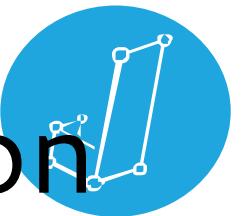
Bayesian T-Test: Likelihood

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

Likelihood of the observed data, for each value of δ

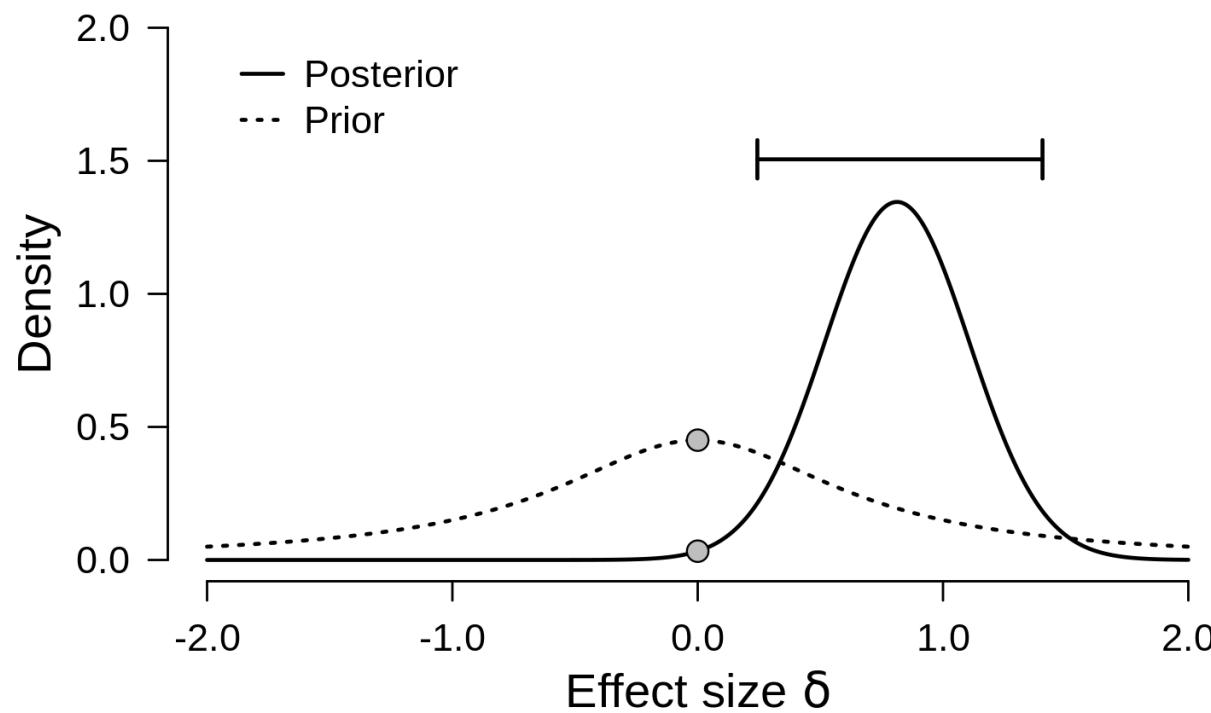


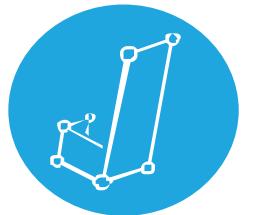
The **marginal likelihood**, across
all values of δ



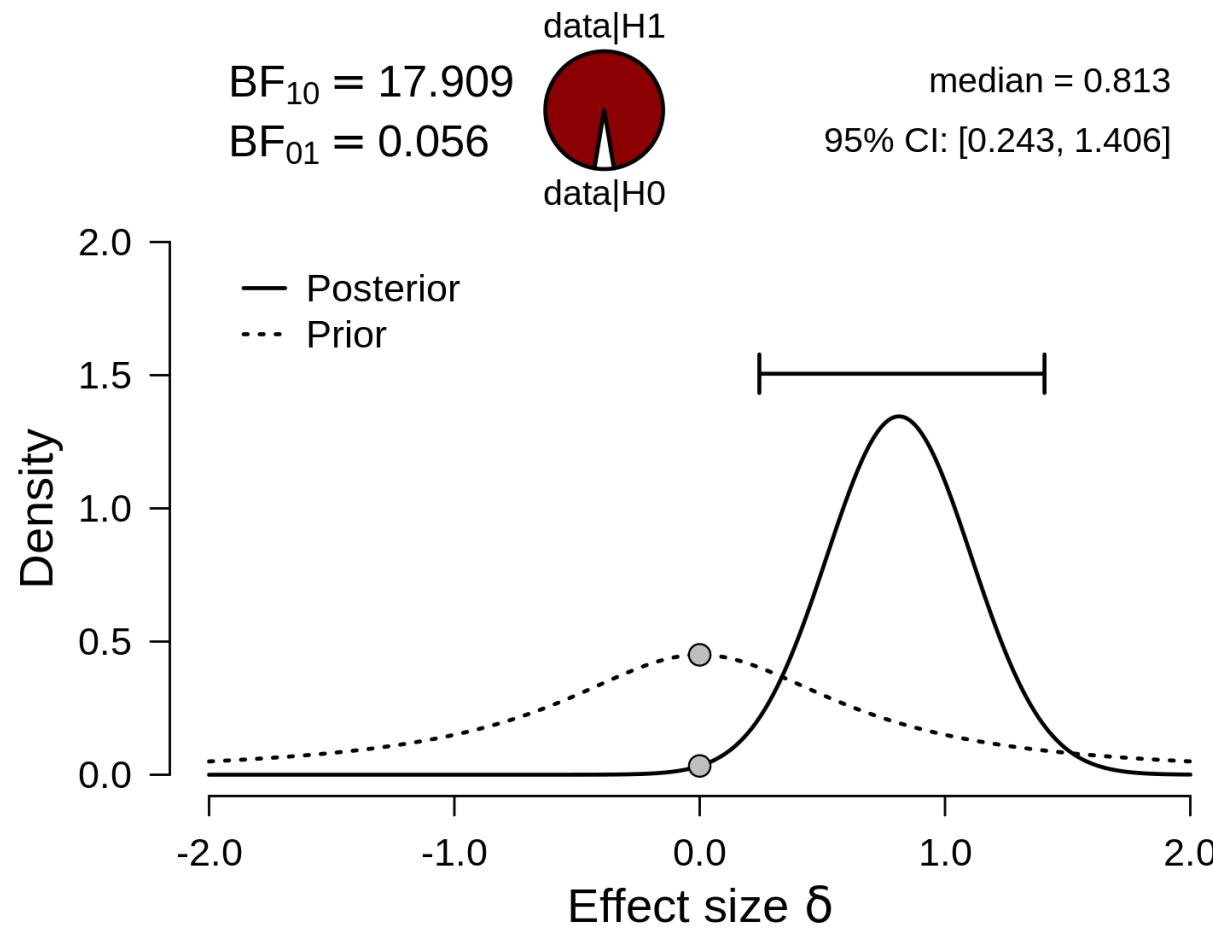
Bayesian T-Test: Posterior Distribution

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

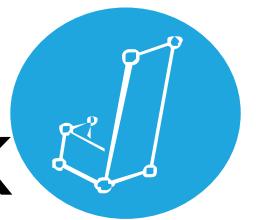




Bayesian T-Test: Bayes Factor



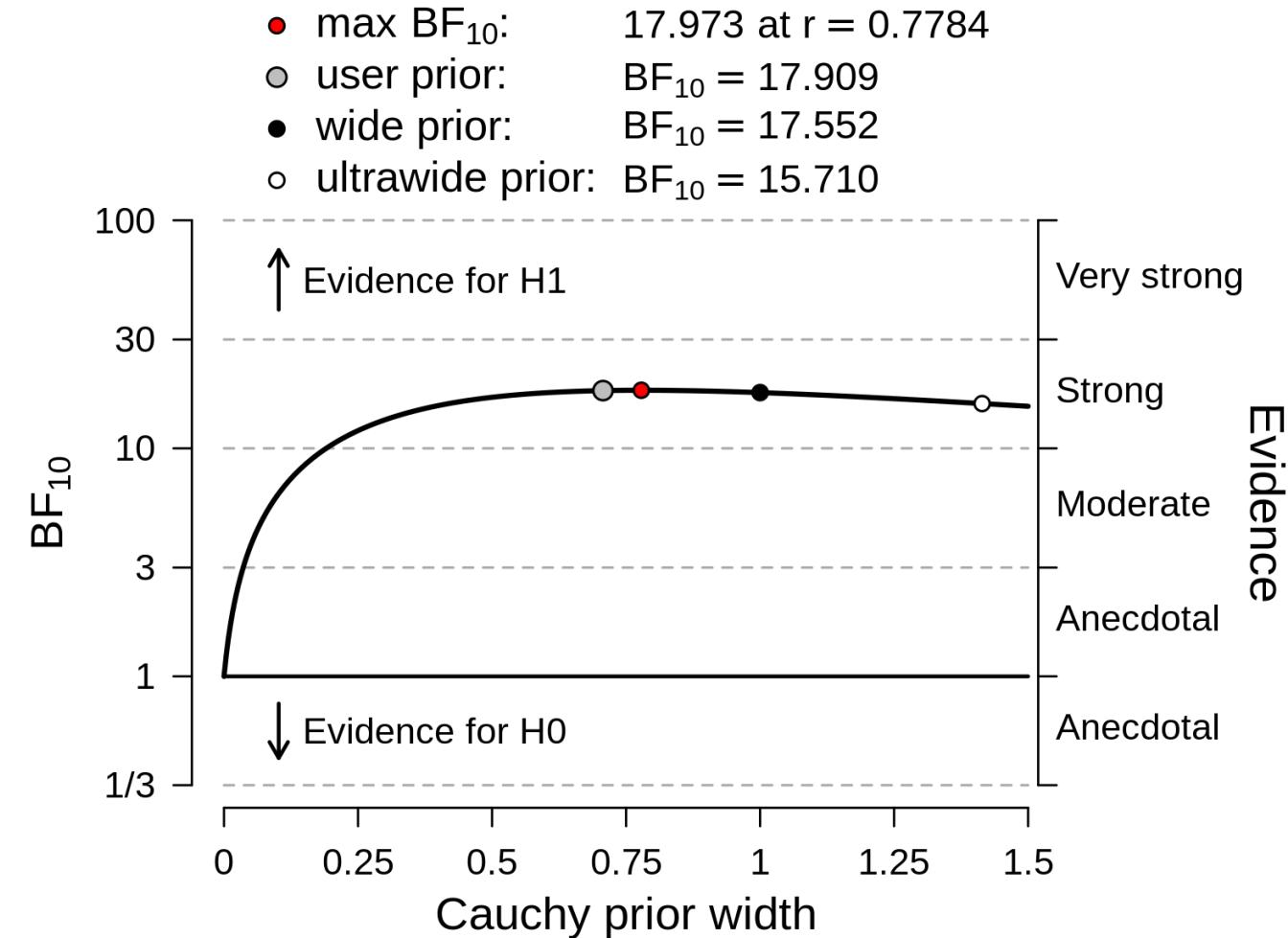
Bayesian T-Test: Robustness Check

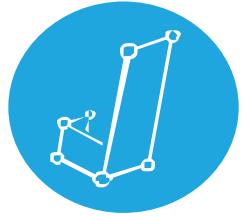


Because choosing the prior setting is fairly subjective, we can explore what would have happened if we had chosen a different value for the prior width. This is known as a **Bayes factor robustness check, or sensitivity analysis**

We see here that the Bayes factor is pretty stable (i.e., flat line) across many values for the prior width. Except for very strong prior settings (0.15 and lower) does the evidence for H_1 decrease

"Here, the Bayes factor does not qualitatively differ for a wide range of prior specifications (0.25 – 1.5)"





Singer Heights

This data set, "Singers", provided "Heights in inches of the singers in the New York Choral Society in 1979. The data are grouped according to voice part.

Variables:

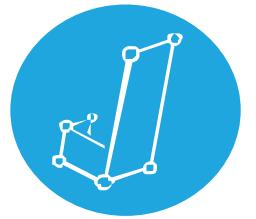
- Height – Height (in inches) of the singers.
- Gender – Gender of the singers (female, male).
- Pitch – Voice pitch of the singers (very low, low, high, very high)

Research question: to what extent does height differ for different vocal pitches and gender?

Source:

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). Graphical Methods for Data Analysis. New York: Chapman & Hall.

Singer Height



MODEL

$$S_i = \alpha + \beta_1 W_i + \beta_2 P_i + \varepsilon_i$$

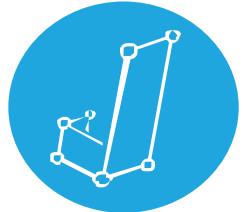
Height = base level + pitch + gender + error

$$S_i = \alpha + \beta_1 W_i + \varepsilon_i$$

Height = base level + pitch + error

$$S_i = \alpha + \beta_2 P_i + \varepsilon_i$$

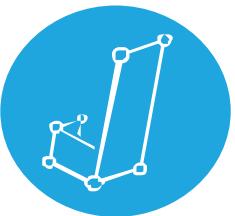
Height = base level + gender + error



Singer Height

Model Comparison ▼

Models	P(M)	P(M data)	BF_M	BF_{10}	error %
Gender + Pitch	0.2000	0.9000	36.0148	1.0000	
Gender + Pitch + Gender * Pitch	0.2000	0.0958	0.4238	0.1064	3.6429
Gender	0.2000	0.0042	0.0167	0.0046	2.0183
Pitch	0.2000	8.8662e-39	3.5465e-38	9.8509e-39	2.0184
Null model	0.2000	1.0914e-40	4.3655e-40	1.2126e-40	2.0183



Singer Height

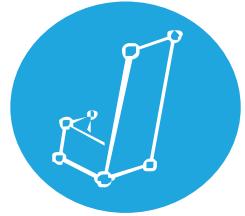
Model Comparison ▼

Models	P(M)	P(M data)	BF _M	BF ₁₀	error %
Gender + Pitch	0.2000	0.9000	36.0148	1.0000	
Gender + Pitch + Gender * Pitch	0.2000	0.0958	0.4238	0.1064	3.6429
Gender	0.2000	0.0042	0.0167	0.0046	2.0183
Pitch	0.2000	8.8662e-39	3.5465e-38	9.8509e-39	2.0184
Null model	0.2000	1.0914e-40	4.3655e-40	1.2126e-40	2.0183

Analysis of Effects - Height

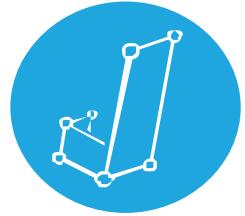
Effects	P(incl)	P(excl)	P(incl data)	P(excl data)	BF _{incl}
Gender	0.6000	0.4000	1.0000	0.0000	∞
Pitch	0.6000	0.4000	0.9958	0.0042	159.7825
Gender * Pitch	0.2000	0.8000	0.0958	0.9042	0.4238

Many Models? Again?



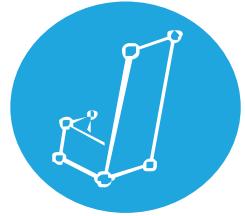
- In ANOVA (and regression) we compare models
- Differs from classical analysis, where we assess effects
- Each model instantiates a theoretically interesting assumption
- All tests of interest are relative
- Remember: "Best" model is:
 - Not necessarily good (check assumptions)
 - Not necessarily much better than second best

Handy Literature



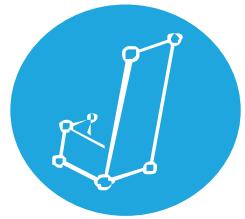
- van Doorn, J., van den Bergh, D., ... & Wagenmakers, E.-J. (2021). [The JASP guidelines for conducting and reporting a Bayesian analysis.](#) Psychonomic Bulletin & Review, 28, 813–826.
- Etz, A., & Vandekerckhove, J. (2018). [Introduction to Bayesian inference for psychology.](#) Psychonomic bulletin & review, 25(1), 5–34.
- Wagenmakers, E.-J.,..., & Morey, R. D. (2018). [Bayesian inference for psychology. Part II: Example applications with JASP.](#) Psychonomic Bulletin & Review, 25, 58–76.
- van den Bergh, D., Van Doorn, J., Marsman, M., Draws, T., Van Kesteren, E. J., Derks, K., ... & Wagenmakers, E. J. (2020). [A tutorial on conducting and interpreting a Bayesian ANOVA in JASP.](#) L'Annee psychologique, 120, 73–96.

Handy Resources



- [How to Use JASP – Inventory of blogs/videos/gifs for frequentist and Bayesian analyses](#)
- [JASP Resources for Teachers: Open-source books](#)
- [Webinar: Theory and Practice of Bayesian Inference Using JASP](#)
- [Step By Step Guide: 1. Bayesian One-Way ANOVA and the full playlist](#)
- JASP on Twitter – <https://twitter.com/JASPStats>
- JASP forum – <https://forum.cogsci.nl/index.php?p=/categories/jasp-bayesfactor>
- Found a bug? Please report on Github: <https://github.com/jasp-stats/jasp-issues/issues>
- More extensive, annual JASP workshop: <https://jasp-stats.org/workshop/>

The Future of JASP



- Feature roadmap
 - ▶ Full syntax mode
 - ▶ Data editing
 - ▶ Selective filtering
 - ▶ Module library
- JASP Foundation