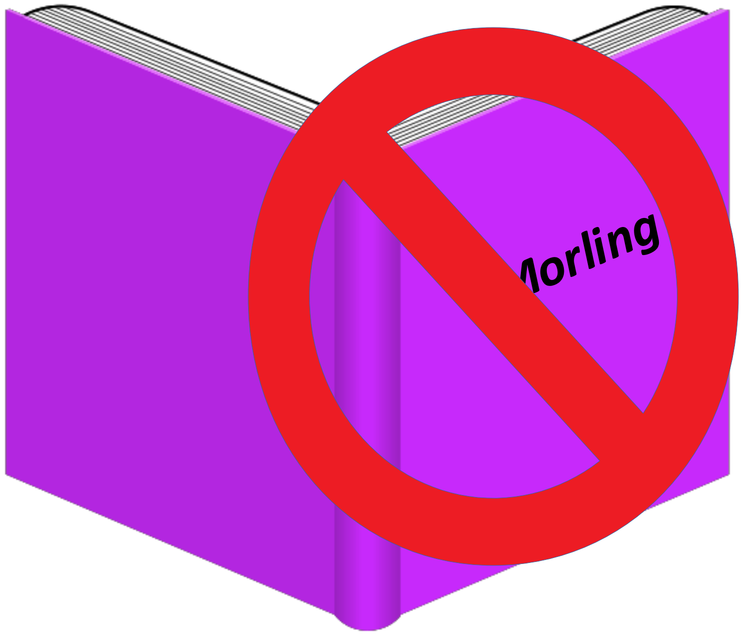


Research Methods and Statistics

Lecture 20: Introduction to Bayesian Statistics

Johnny van Doorn

Pictures source: pixabay

What Are We Doing?

We set an alpha (i.e., desired type 1 risk) and determine the hypotheses

We observe data, compute test statistic and corresponding p-value

$p > \alpha$

Fail to reject H_0

= absence of evidence

$p < \alpha$

Reject H_0

= evidence against H_0

- Black/white reasoning is dangerous and arbitrary (see bonus slide of lecture 16)
- We cannot gain evidence **for** H_0 (i.e., evidence of absence)

"… is associated with a myriad of negative consequences including reduced optimism and task performance (Porath & Erez, 2007), reduced working memory and attention (Erez et al., 2015), and higher levels of stress (Adams & Webster, 2013)."

Anna Kaminska & Devin G. Ray (2023) Interpersonal memory failure in the workplace: The effect of memory and hierarchy on employee's affective commitment, The Journal of Social Psychology

The Next 3 Lectures

- Today (Chapters 1-4)
  - Introduction to Bayesian Estimation of a Proportion
- Thursday Dec 4 (Chapters 1-4)
  - Bayesian Hypothesis Testing of a Proportion
- Thursday Dec 11(Chapter 5)
  - Bayesian Inference for Correlation and T-Test
  - Statistics in the Wild

van Doorn, J. (2024). A Brief Introduction to Bayesian Inference: From Tea to Beer.

Two important notes

- Tuesday Dec 9: No lecture due to the demonstration
  - Lecture recording on Canvas, adjusted exam material
- Thursday Dec 11: Different lecture location
  - Pathe Tuschinschki

Today

- Bayesian statistics
  - What are models?
  - How do models learn from data?
  - Live demonstration
- Recap
  - Practical stuff & next week
  - Example exam question

Stanford | News Service

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

Fair Coins Tend To Land On The Same Side They Started: Evidence From 350,757 Flips

What is Bayesian Inference?

What are the tools we need?

What is Bayesian Inference?

An alternative framework to "frequentist statistics"

Not based on studying what happens on repeated sampling

Different way of parameter estimation and hypothesis testing

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

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

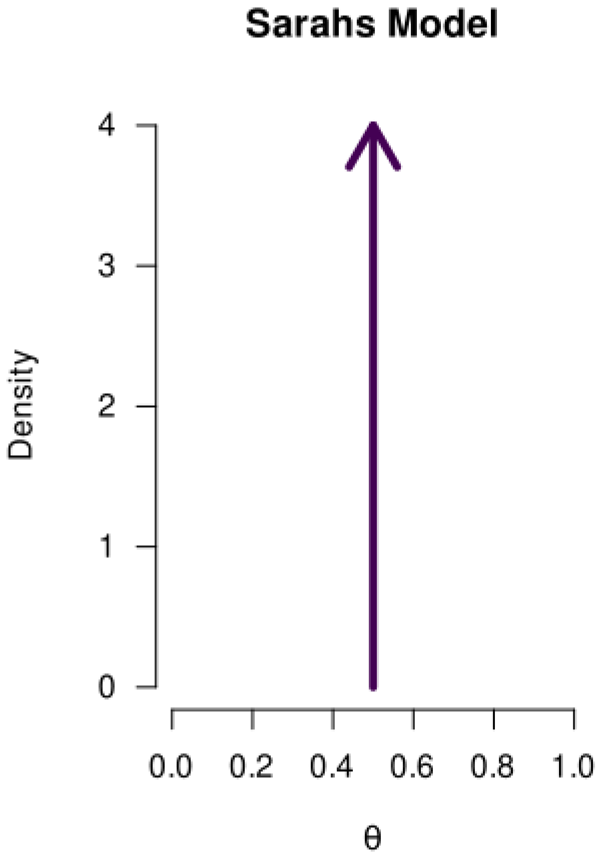
For instance, for flipping a coin:
A binomial model with θ = 0.5

What is the theoretical implication of this model?

What are Statistical Models?

A binomial model with θ = 0.5

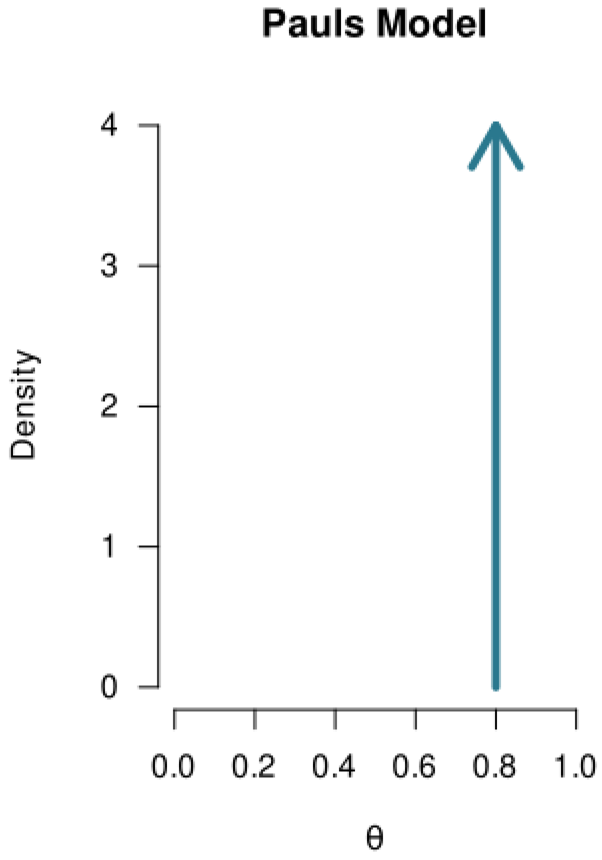
We can reflect a model's statement by means of a probability distribution



What are Statistical Models?

A binomial model with θ = 0.8

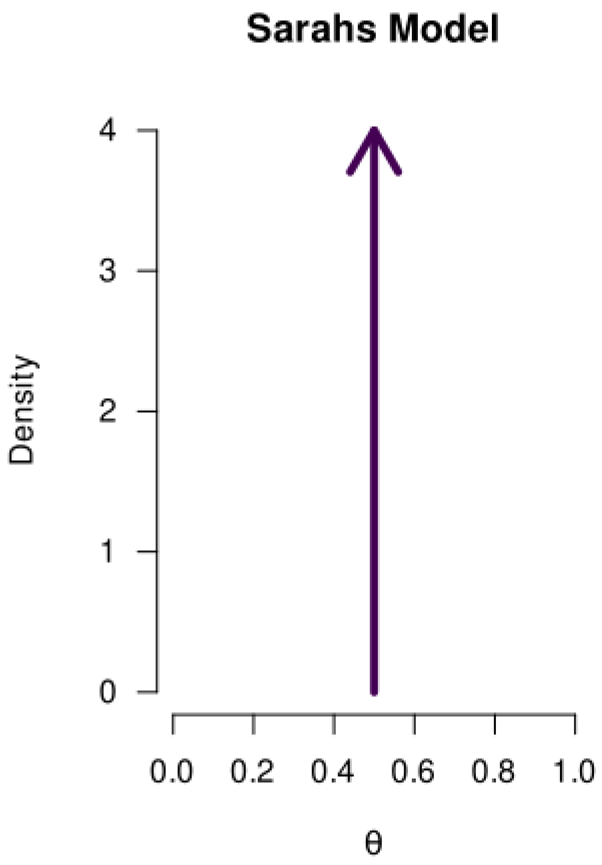
We can reflect a model's statement by means of a probability distribution

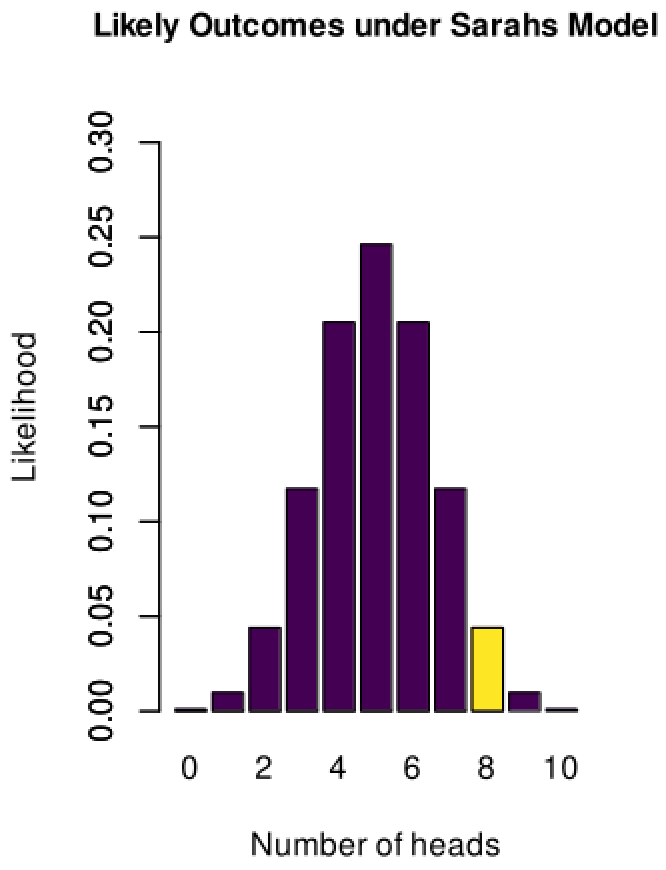


Statistical Models Make Predictions

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

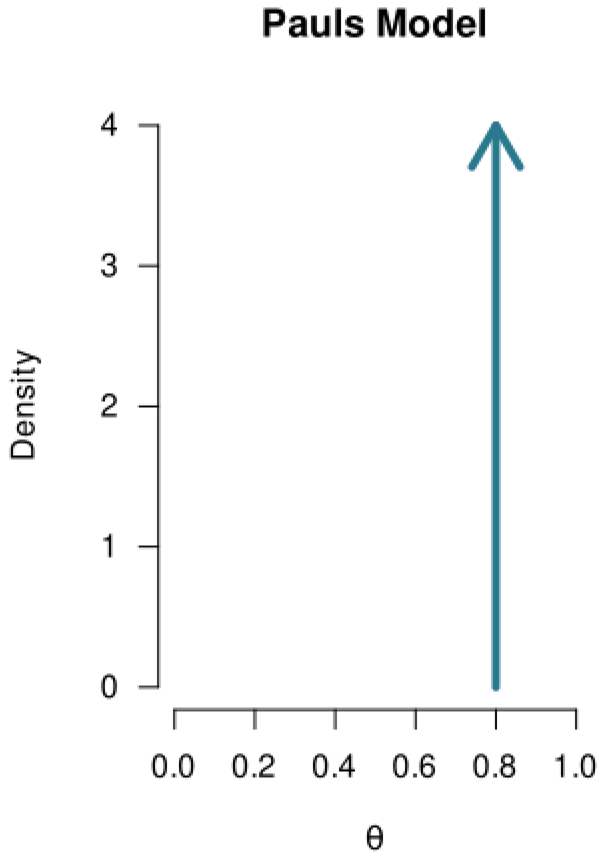
Based on what these models claim about θ, certain outcomes are more/less likely

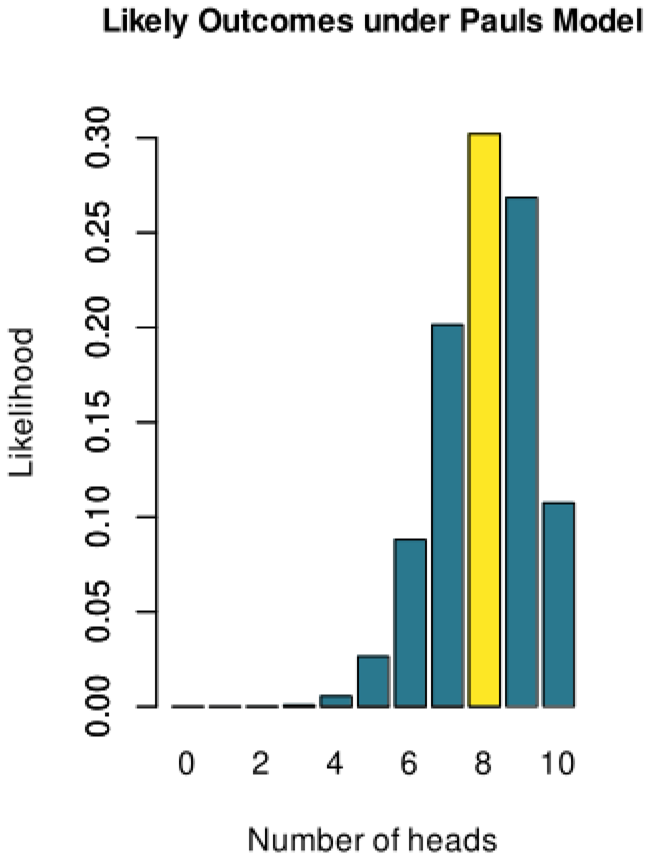
The yellow bar indicates how likely an outcome of 8/10 heads is, under Paul's model

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

To create this figure, we take the binomial formula, and fill in θ = 0.8

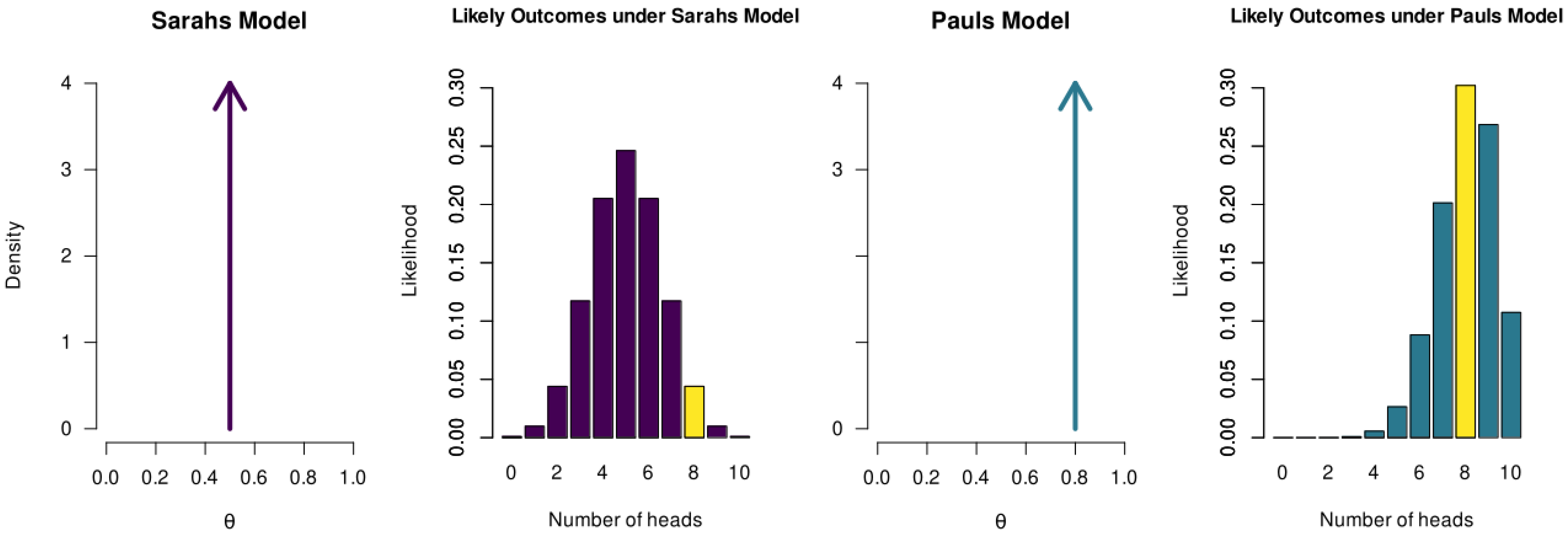
For instance, for an outcome of k=8 heads, the formula gives 0.302



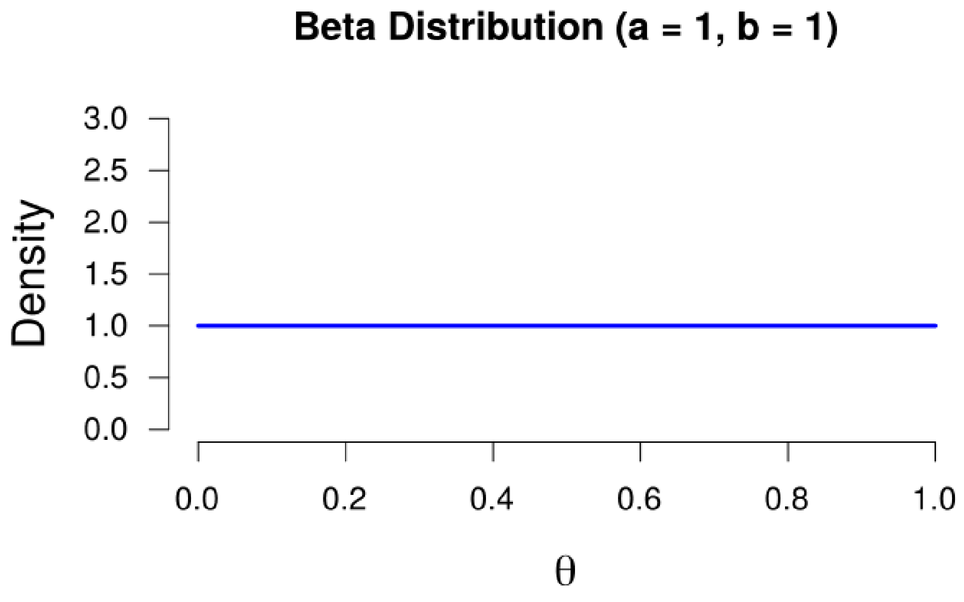


Statistical Models Make Predictions

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



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

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

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

Models Can Also State a Range of Values

Introducing the beta distribution:

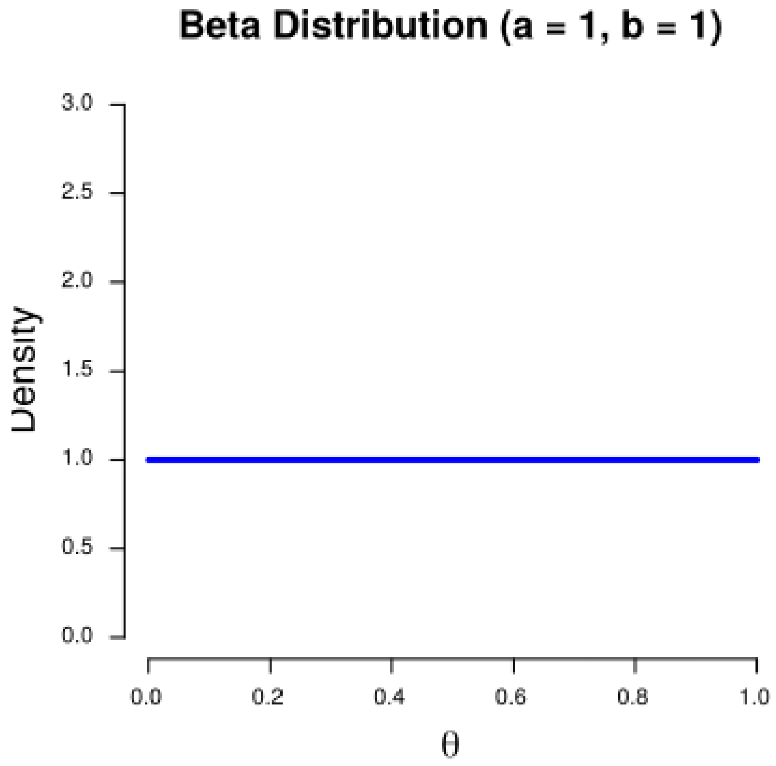
- It ranges from 0 to 1
- Its shape is determined by two values: a and b
- If a and b equal 1, it is uniform

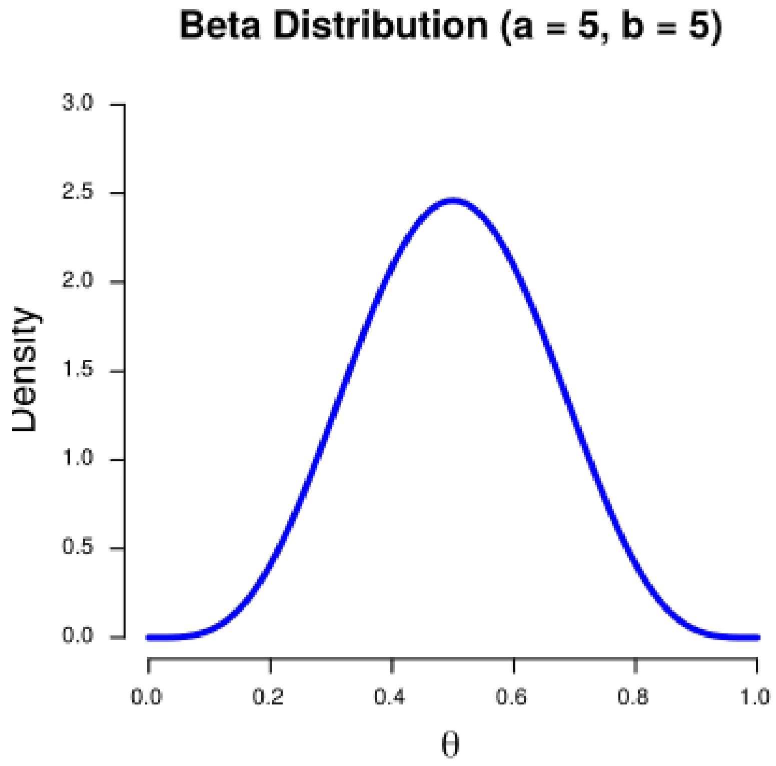
Some apps that let you shape a beta distribution

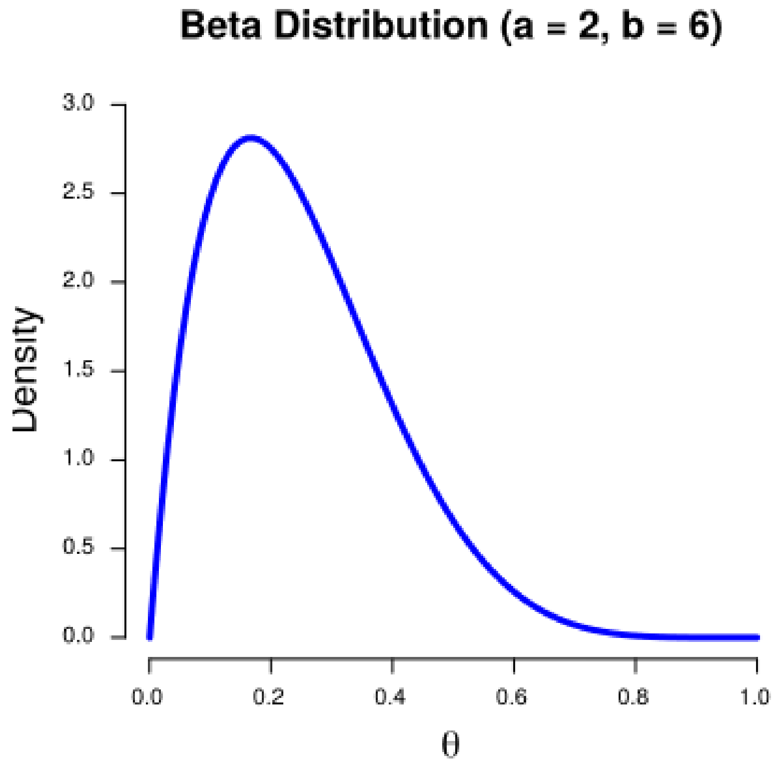
https://researchmethodsuva.shinyapps.io/test/

https://maglit.me/antivichmfy

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

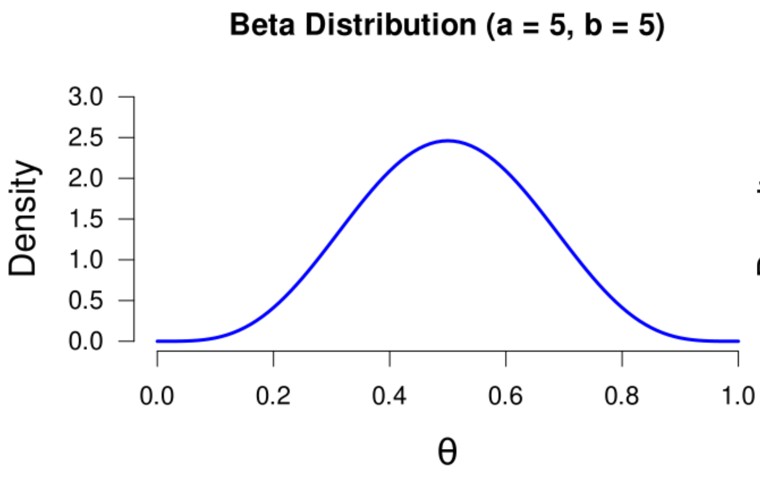
A prior distribution that reflects the idea that values below 0.5 are more plausible (i.e., the coin is biased towards tails)

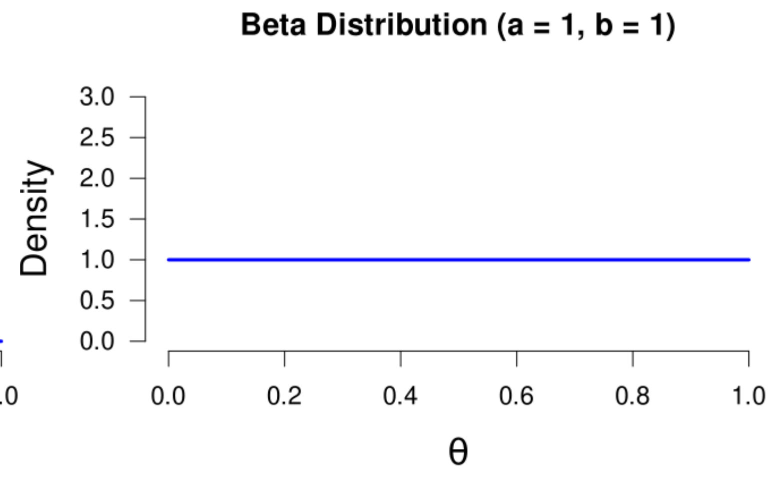
Beta Distribution Interpretation

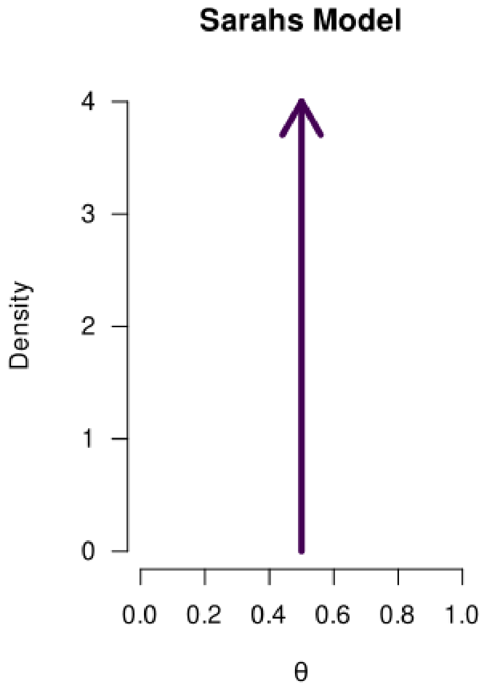
In the context of a prior distribution for a proportion, the a and b can be interpreted as previously observed heads and tails.

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!





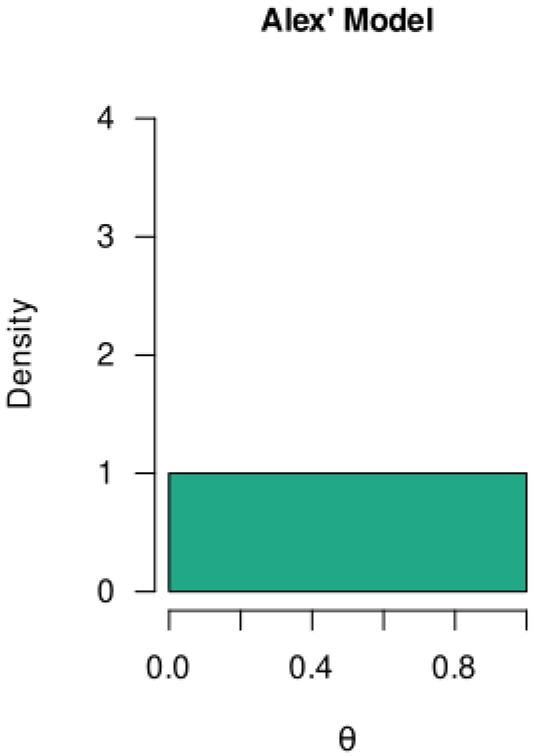


Models Can Also State a Range of Values

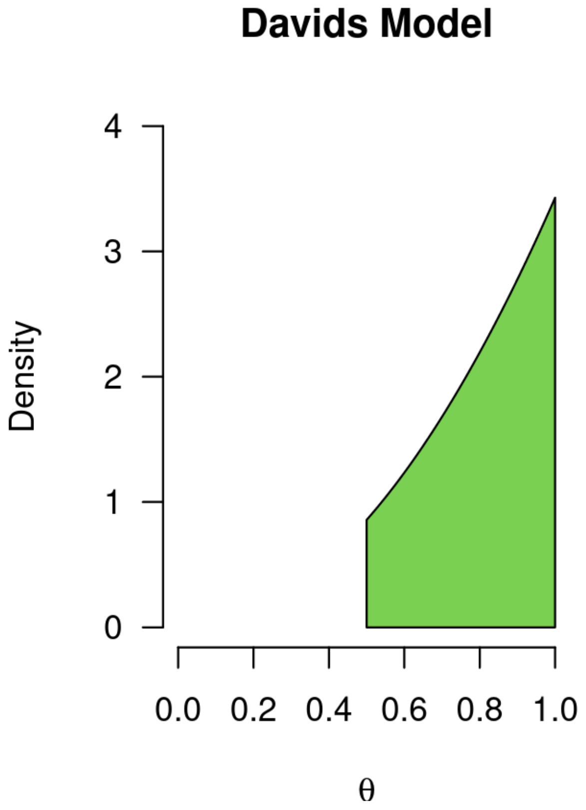
We can reflect a model's statement by means of a probability distribution

A binomial model with θ∽ Beta(1, 1)

A binomial model with θ ∽ Beta(3, 1)

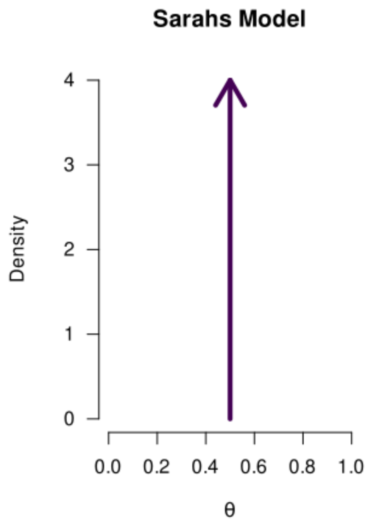


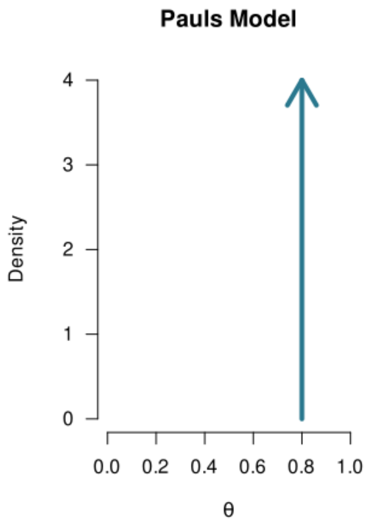
What are the theoretical implications of these models?

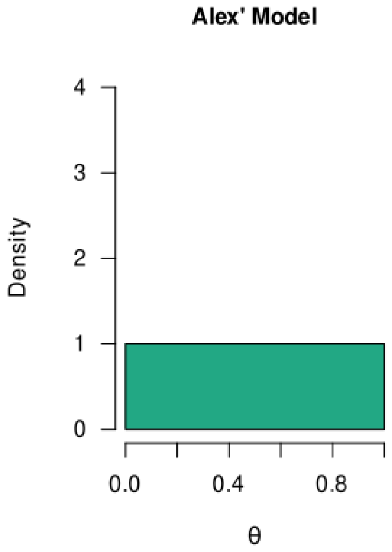


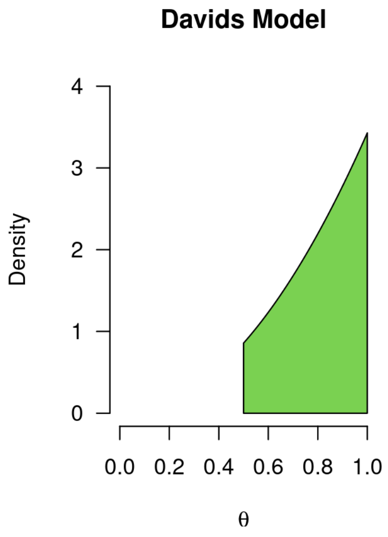
It is truncated below 0.5, so Davids model only postulates values > 0.5

So, We Have All Sorts of Models...





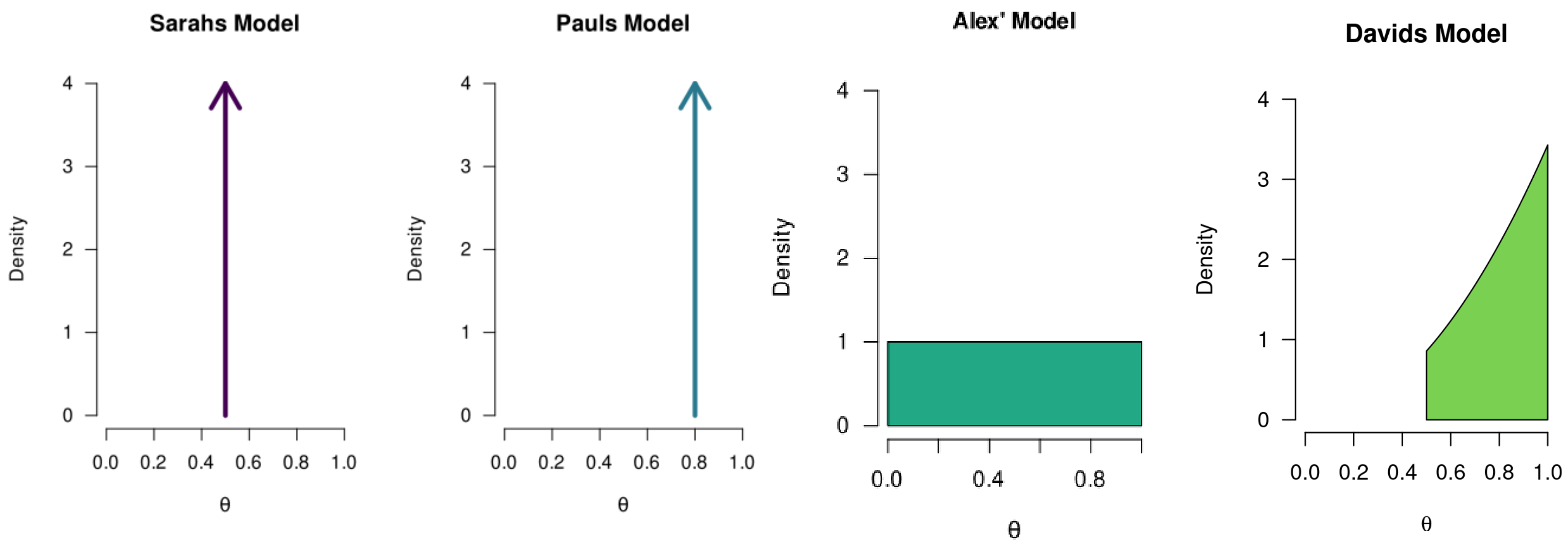




On Thursday we discuss comparing these models to each other

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

Today

- Bayesian statistics
  - What are models?
  - **How do models learn from data?**
  - Live demonstration
- Recap
  - Practical stuff & next week
  - Example exam question

Bayes' Theorem

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

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:

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

Applying Bayes' Theorem to Statistical Inference

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

Our knowledge about *ϑ*, **before** seeing the data
Also known as the "prior beliefs"

Applying Bayes' Theorem to Statistical Inference

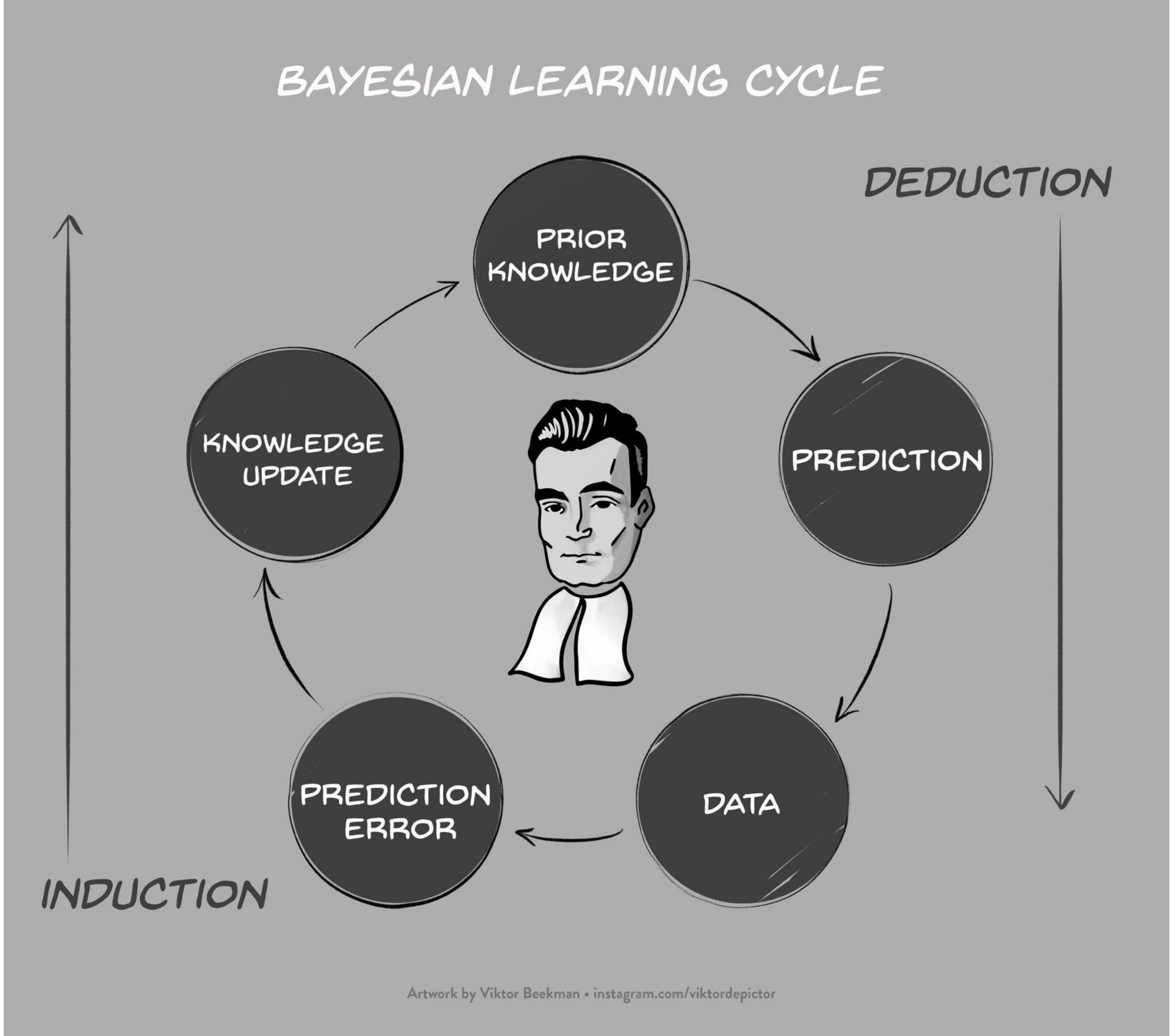
$$\underbrace{P(\theta \mid \text{data})}_{} = P(\theta)\ \frac{P(\text{data} \mid \theta)}{P(\text{data})}$$

Our knowledge about ϑ, **after** seeing the data
Also known as the "posterior beliefs"

Applying Bayes' Theorem to Statistical Inference

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

How well did each value of θ predict the data, compared to all other values of theta?
Also known as the "predictive updating factor"



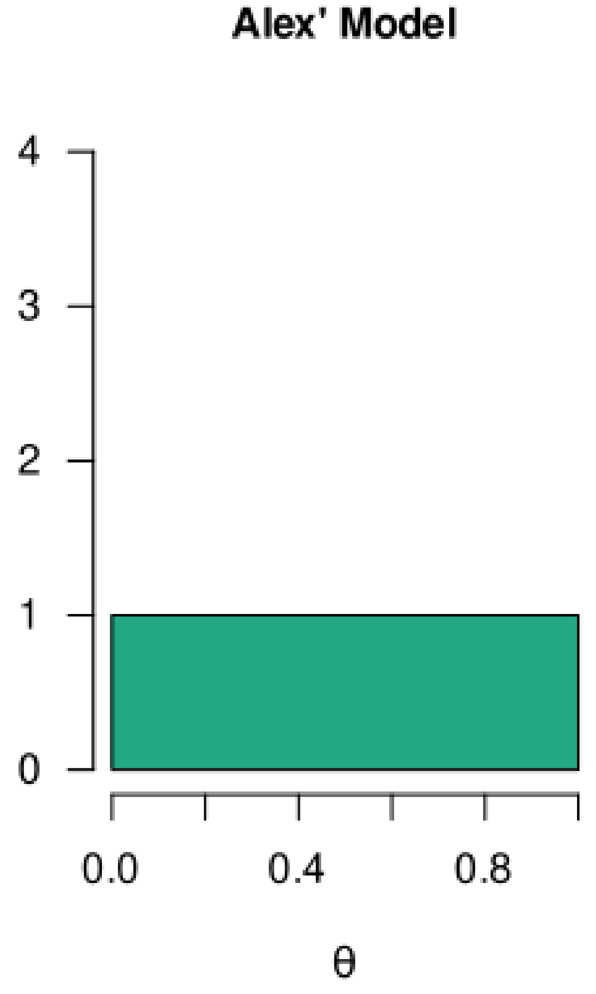
Source:
https://www.bayesianspectacles.org

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!

- For now, we focus on Alex' model



Estimating a Proportion: Data

We observe the following data (n = 10):

2 tails

8 heads

Our *statistic* is the observed proportion: 8/10 = 0.8

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

For instance, how likely is our observed data, if θ equals 0.7?

This is again determined by the general binomial model

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

Estimating a Proportion: Predictive Updating Factor

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

For instance, how likely is our observed data, if θ equals 0.7?

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

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

We take the binomial formula, and fill in θ = 0.7

$$P(x = 8 \mid \theta = 0.7) =$$

Estimating a Proportion: Predictive Updating Factor

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

For instance, how likely is our observed data, if θ equals 0.7?

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

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

$$P(x = 8 \mid \theta = 0.7) = \frac{10!}{8!(10-8)!}$$

Estimating a Proportion: Predictive Updating Factor

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

For instance, how likely is our observed data, if θ equals 0.7?

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

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

$$P(x = 8 \mid \theta = 0.7) = \frac{10!}{8!(10-8)!}0.7^8(1-0.7)^{10-8}$$

Estimating a Proportion: Predictive Updating Factor

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

For instance, how likely is our observed data, if θ equals 0.7?

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

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

$$P(x = 8 \mid \theta = 0.7) = \frac{10!}{8!(10-8)!}0.7^8(1-0.7)^{10-8} = 0.2335$$

Estimating a Proportion: Predictive Updating Factor

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

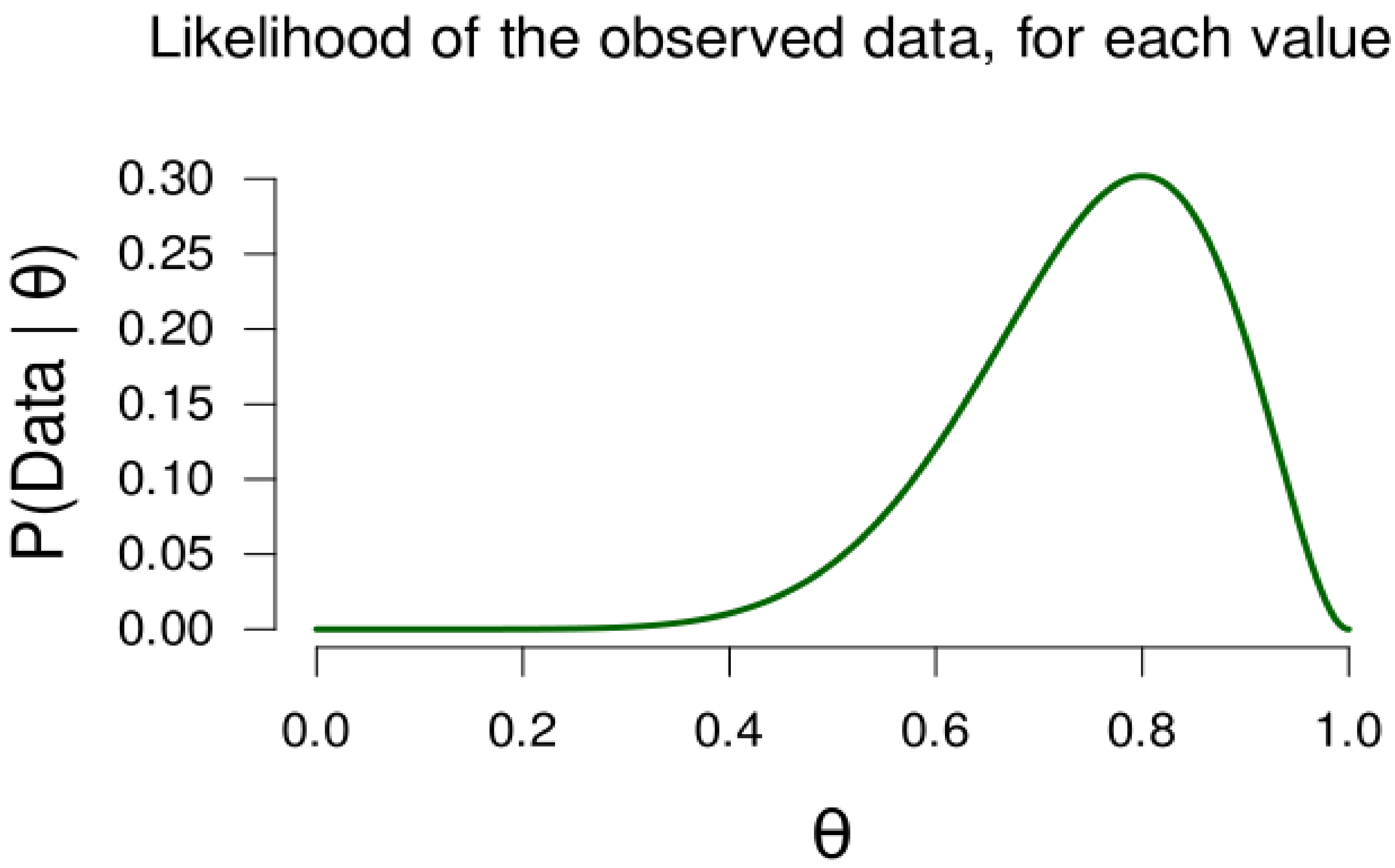
For instance, how likely is our observed data, if θ equals 0.7?

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

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

We do this for each possible value of θ (i.e., between 0 and 1)

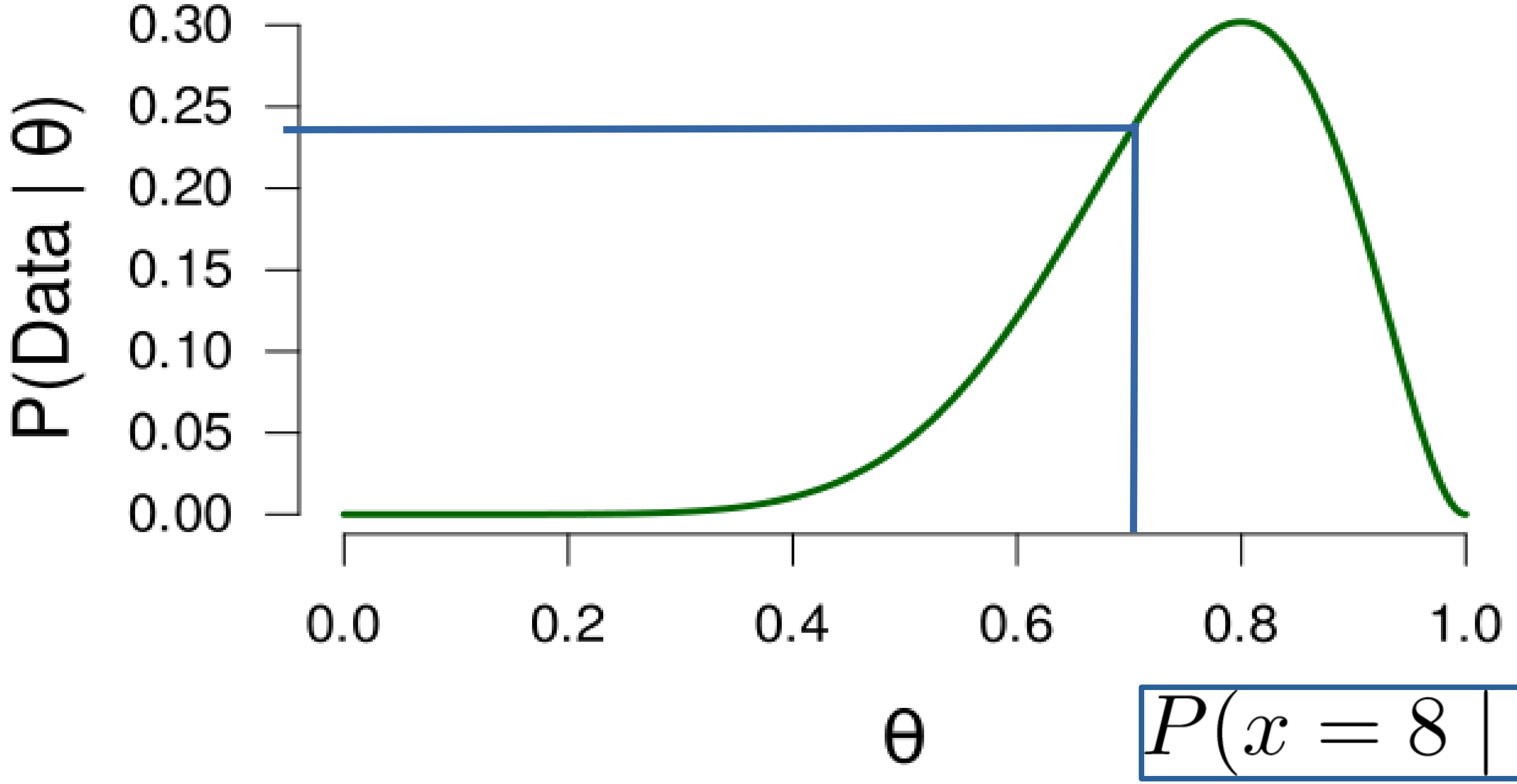
Estimating a Proportion: Predictive Updating Factor



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

Estimating a Proportion: Predictive Updating Factor

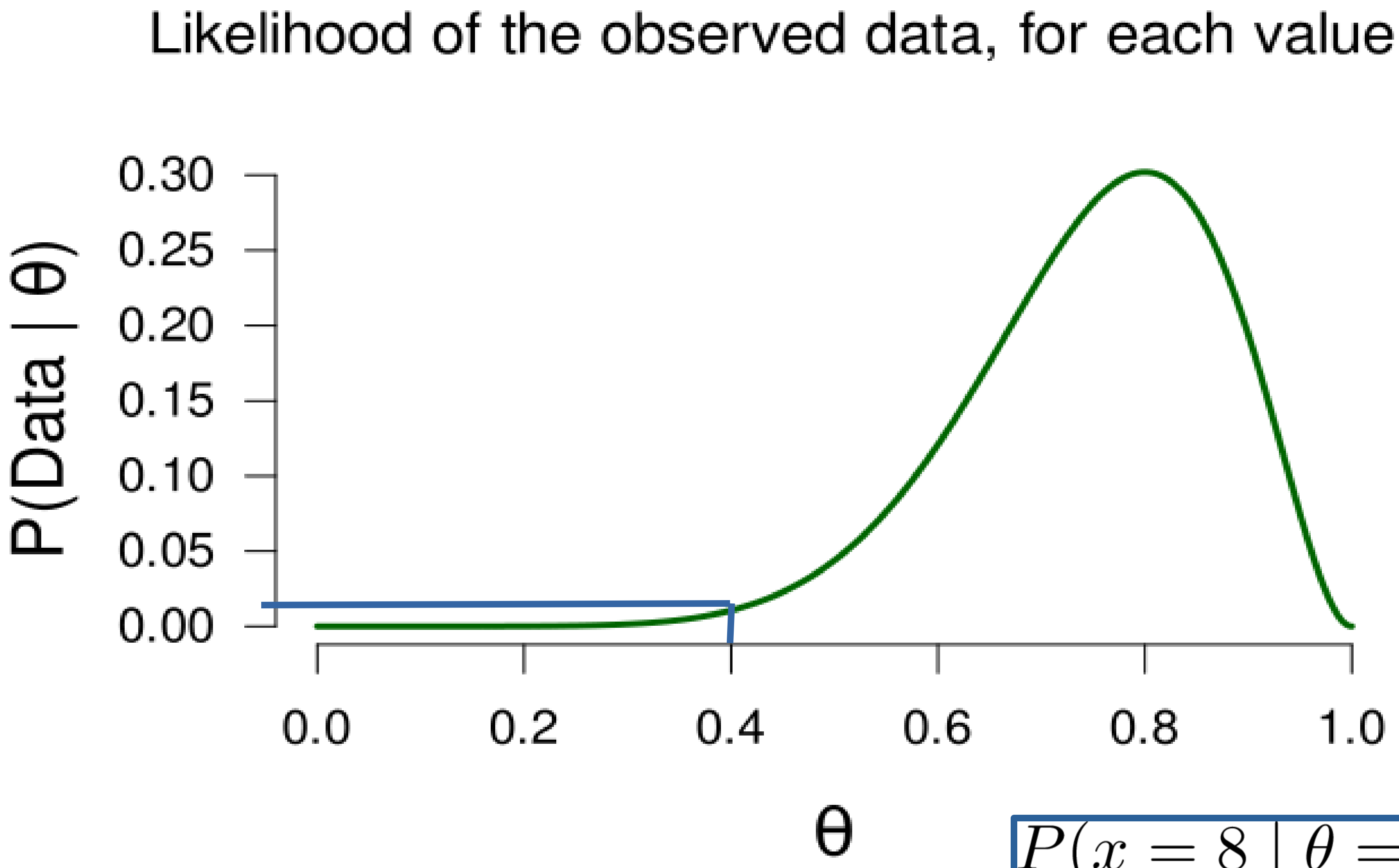
Likelihood of the observed data, for each value of θ



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

$$\boxed{P(x=8 \mid \theta=0.7) = 0.2335}$$

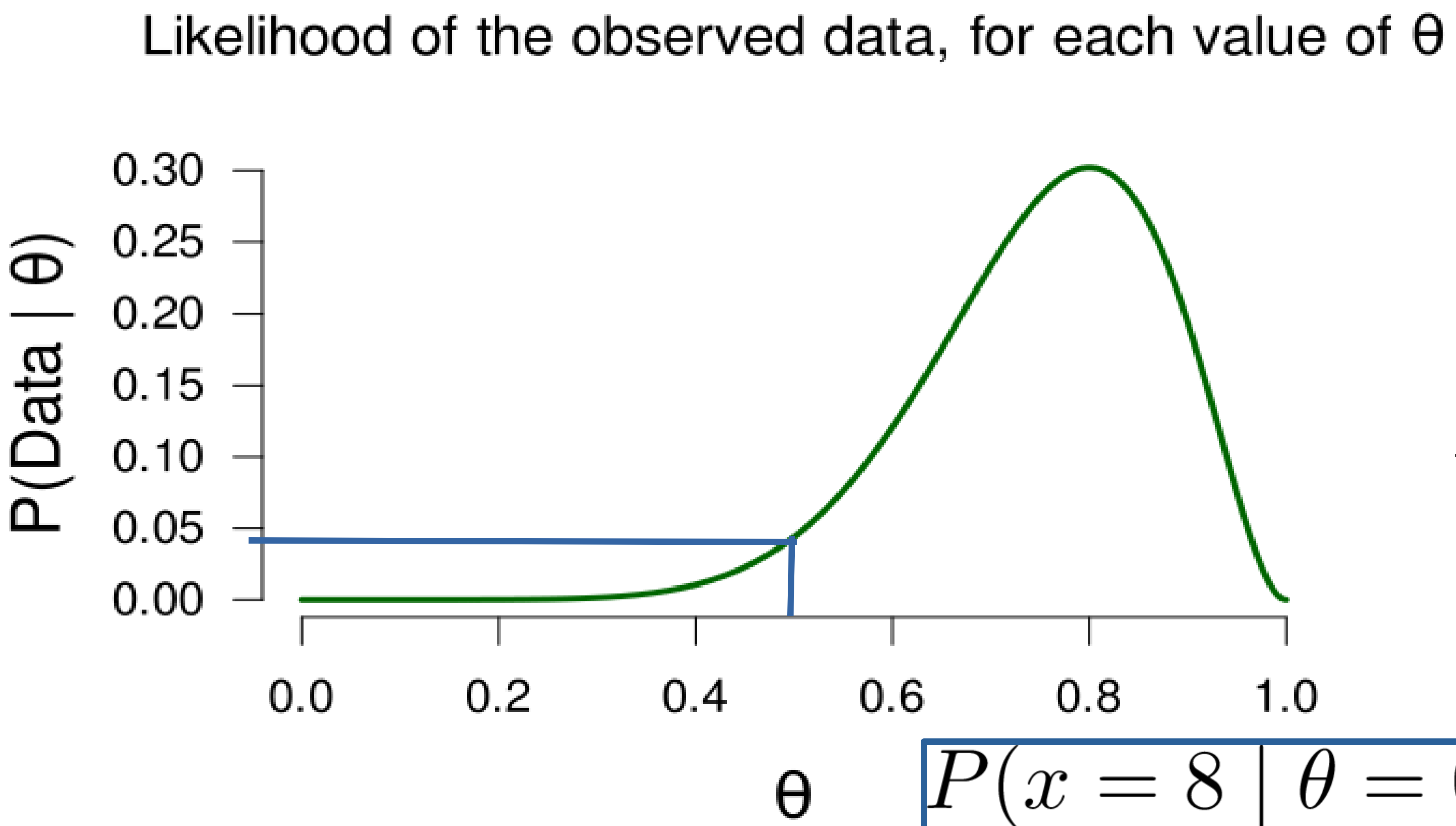
Estimating a Proportion: Predictive Updating Factor



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

$$\boxed{P(x = 8 \mid \theta = 0.4) = 0.0106}$$

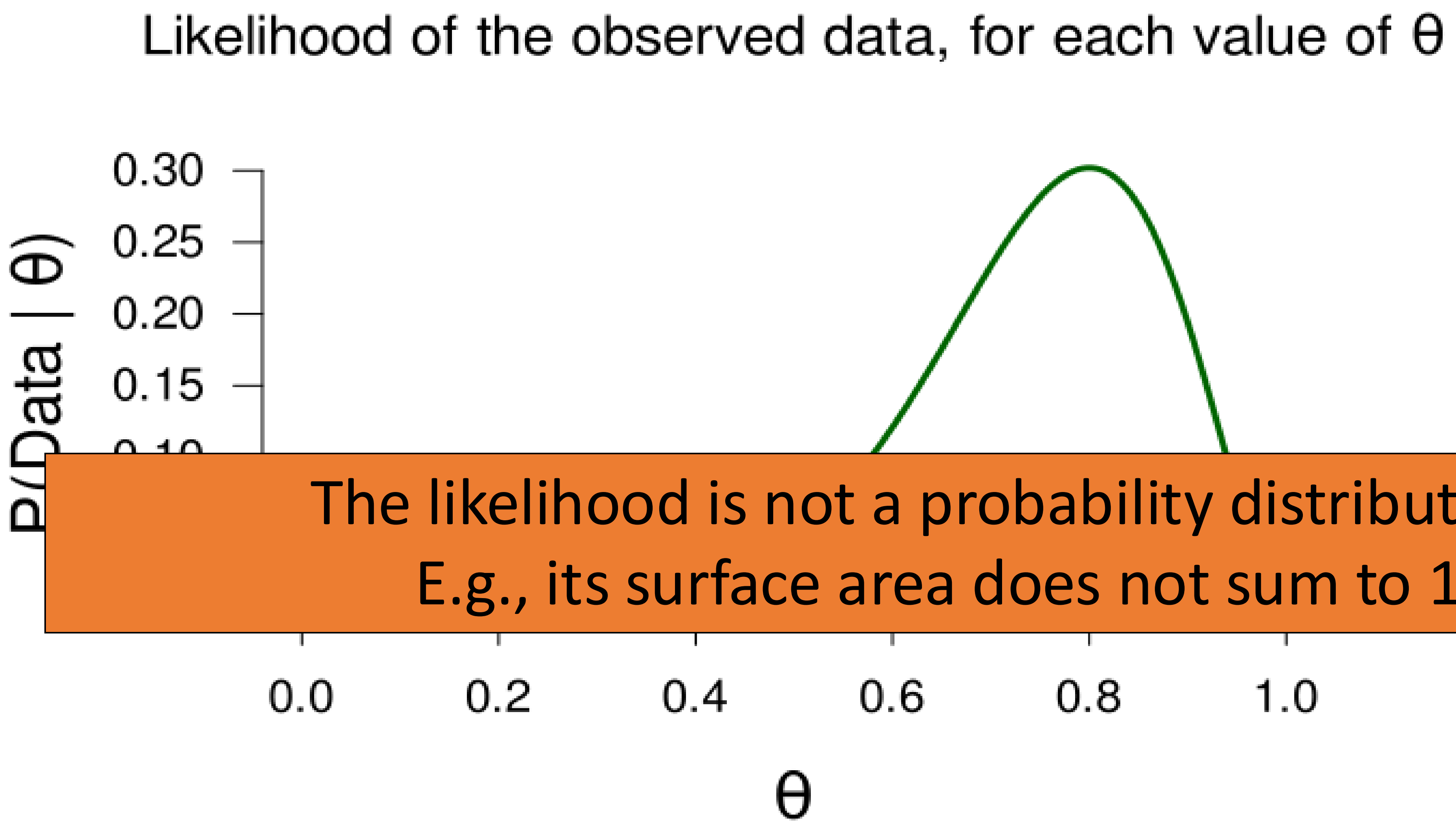
Estimating a Proportion: Predictive Updating Factor



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

$$\boxed{P(x = 8 \mid \theta = 0.5) = 0.0439}$$

Estimating a Proportion: Predictive Updating Factor



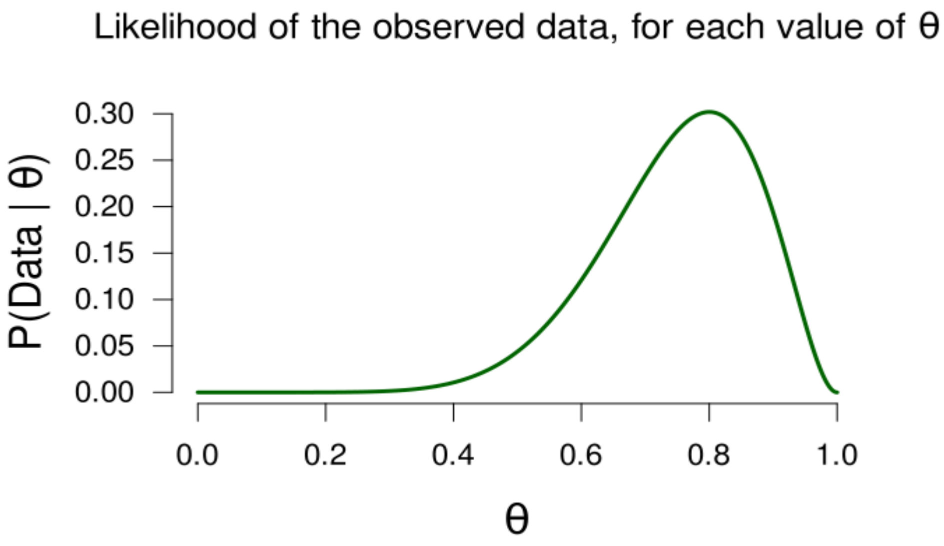
The likelihood is not a probability distribution!
E.g., its surface area does not sum to 1

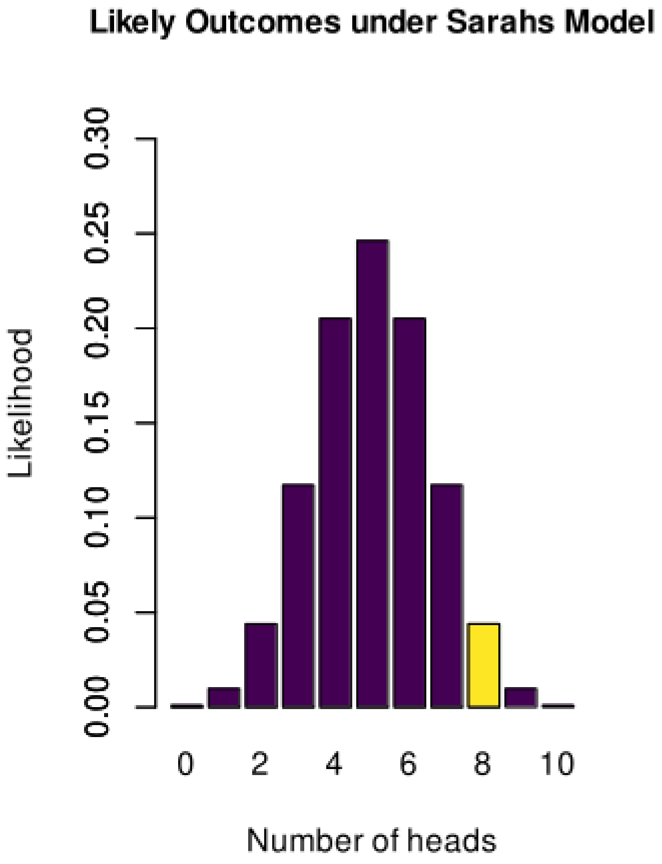
Estimating a Proportion: Predictive Updating Factor

Both figures give the likelihood, but differ in their x-axis

This figure fixes θ (θ =0.5) and varies k

This figure fixes k (heads = 8) and varies θ





Estimating a Proportion: Predictive Updating Factor

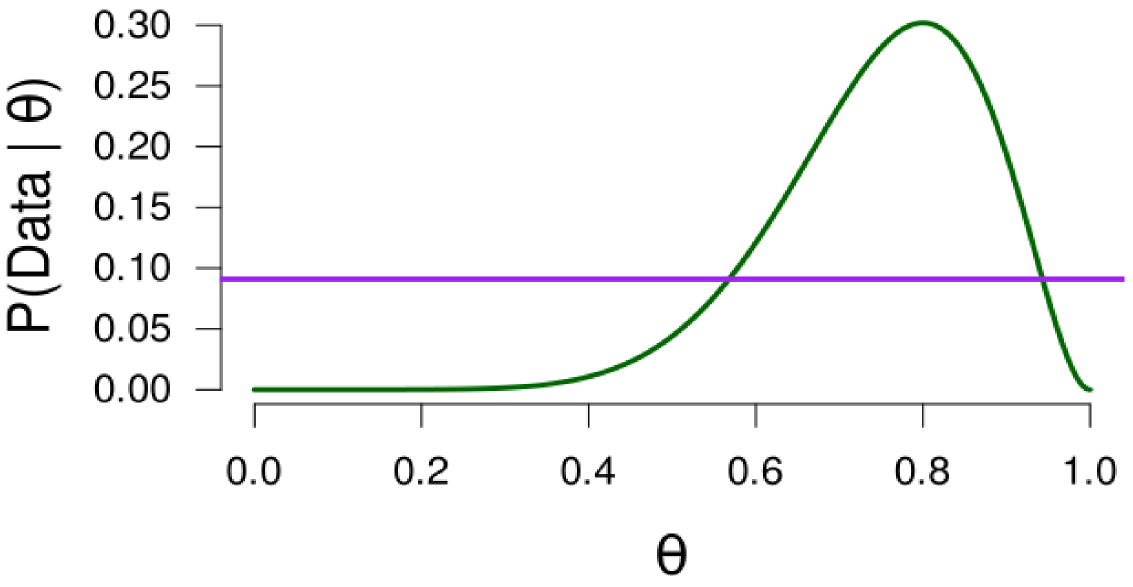
$$P(\text{data})$$

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)

Usually this is very difficult and we need computers to compute this. In our case, the marginal likelihood equals 0.091

Likelihood of the observed data, for each value of θ



Estimating a Proportion: Predictive Updating Factor

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

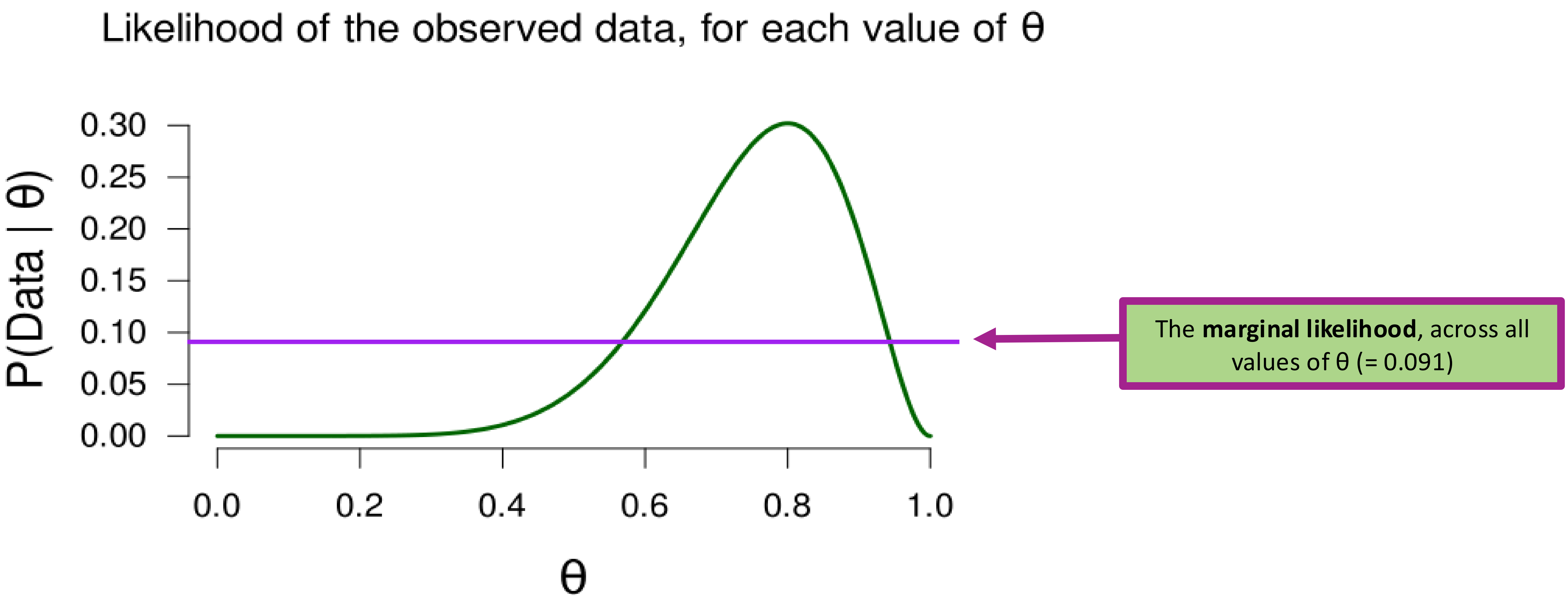
$$\frac{P(\text{data} \mid \theta)}{P(\text{data})}$$

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

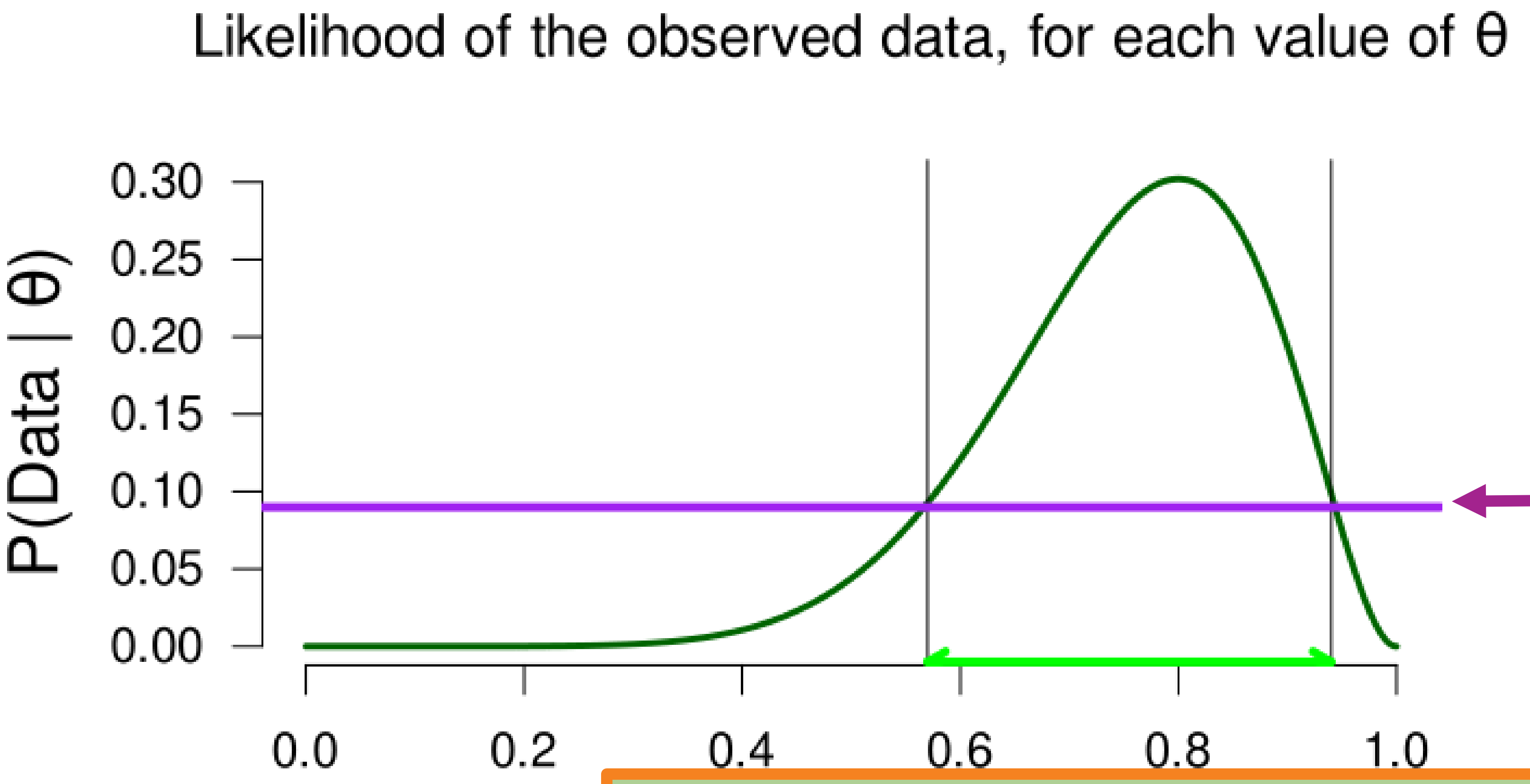
Taken together, this ratio tells us how well each value of θ predicted the data, **relative** to all other values!

Now we know that the marginal (i.e., average) likelihood is 0.091. This means that when the likelihood of the data for a specific value of θ is greater than 0.091, 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



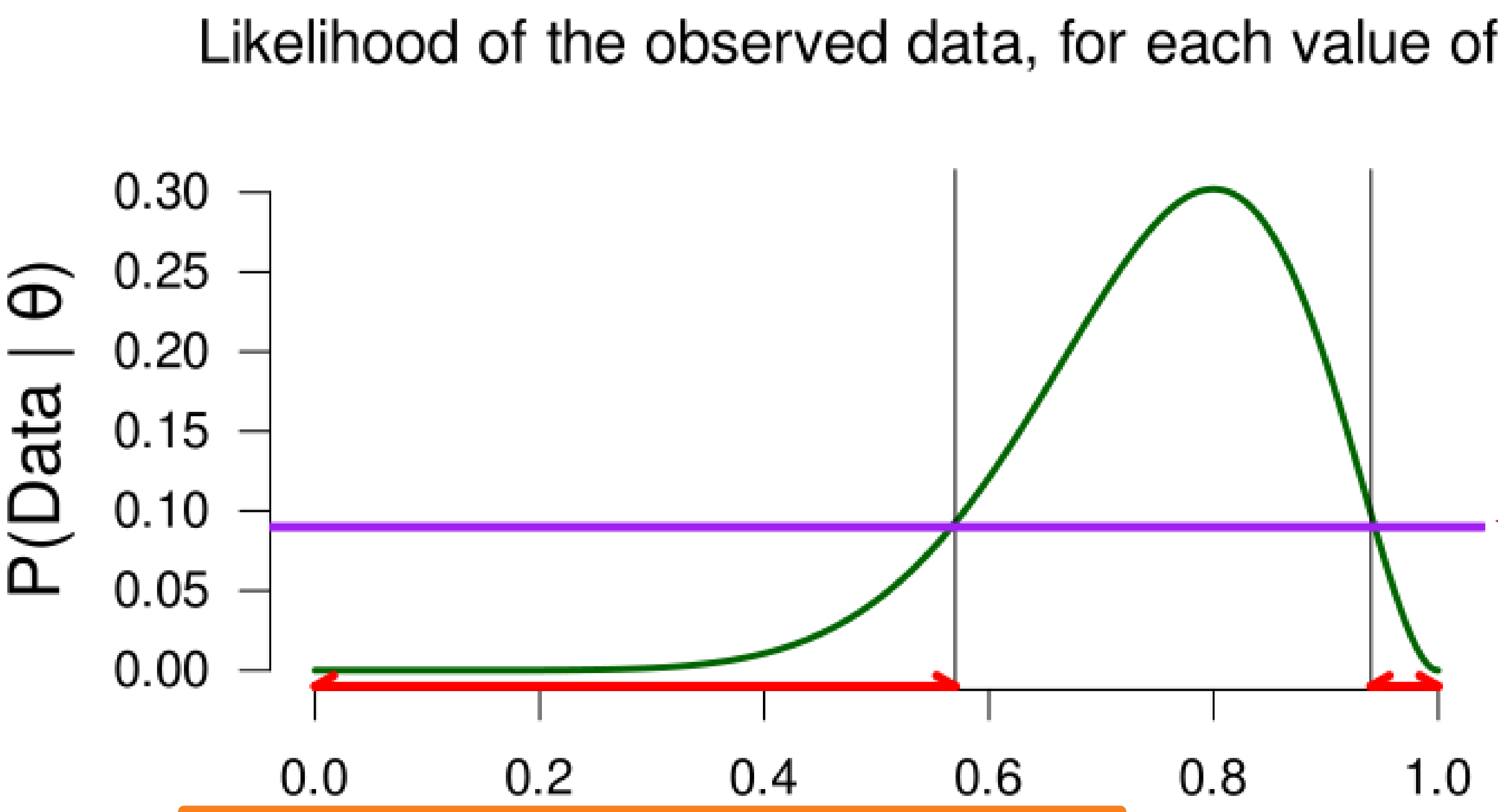
Estimating a Proportion: Predictive Updating Factor



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

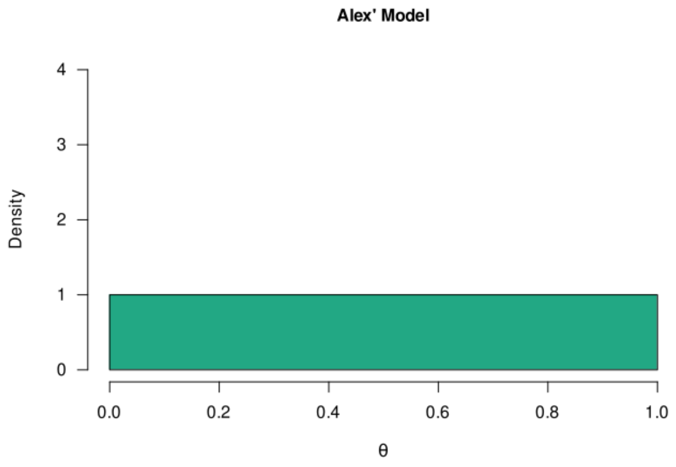


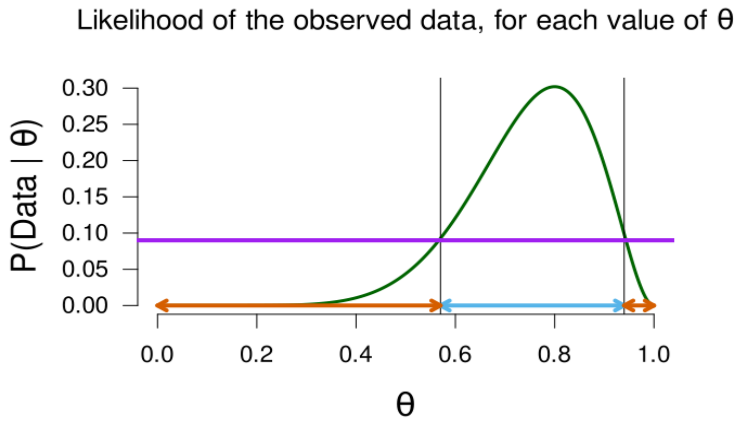
The **marginal likelihood**, across all values of θ (= 0.091)

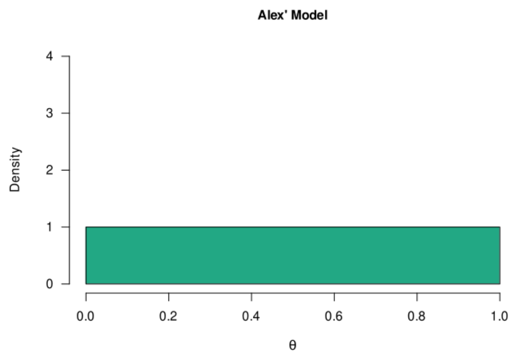
Values of θ that predicted the data worse than average (i.e., their likelihoods are **less** than the marginal likelihood)

Estimating a Proportion: Posterior Distribution

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

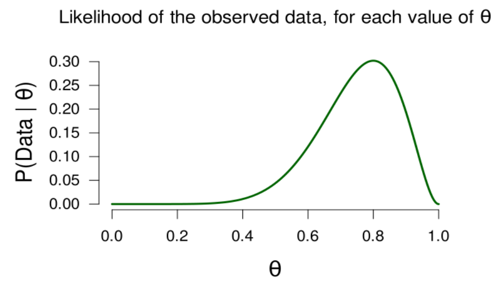






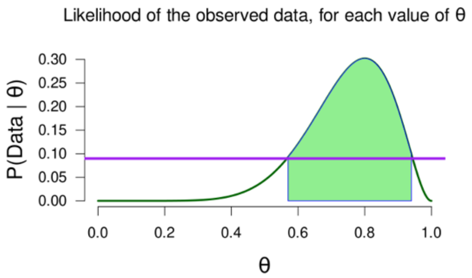
We start with our prior beliefs

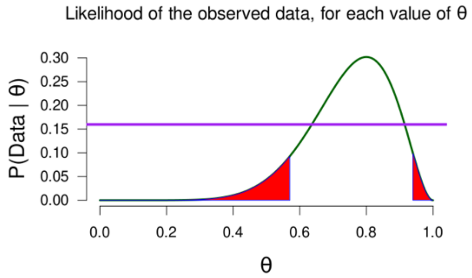
We update those with data



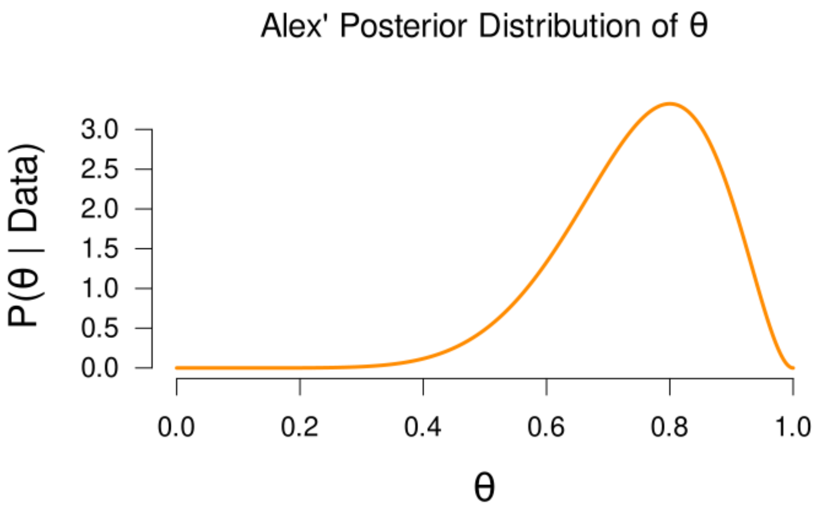
Values of θ that predicted the data better than average receive a boost in plausibility (i.e., their updating ratio >1

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

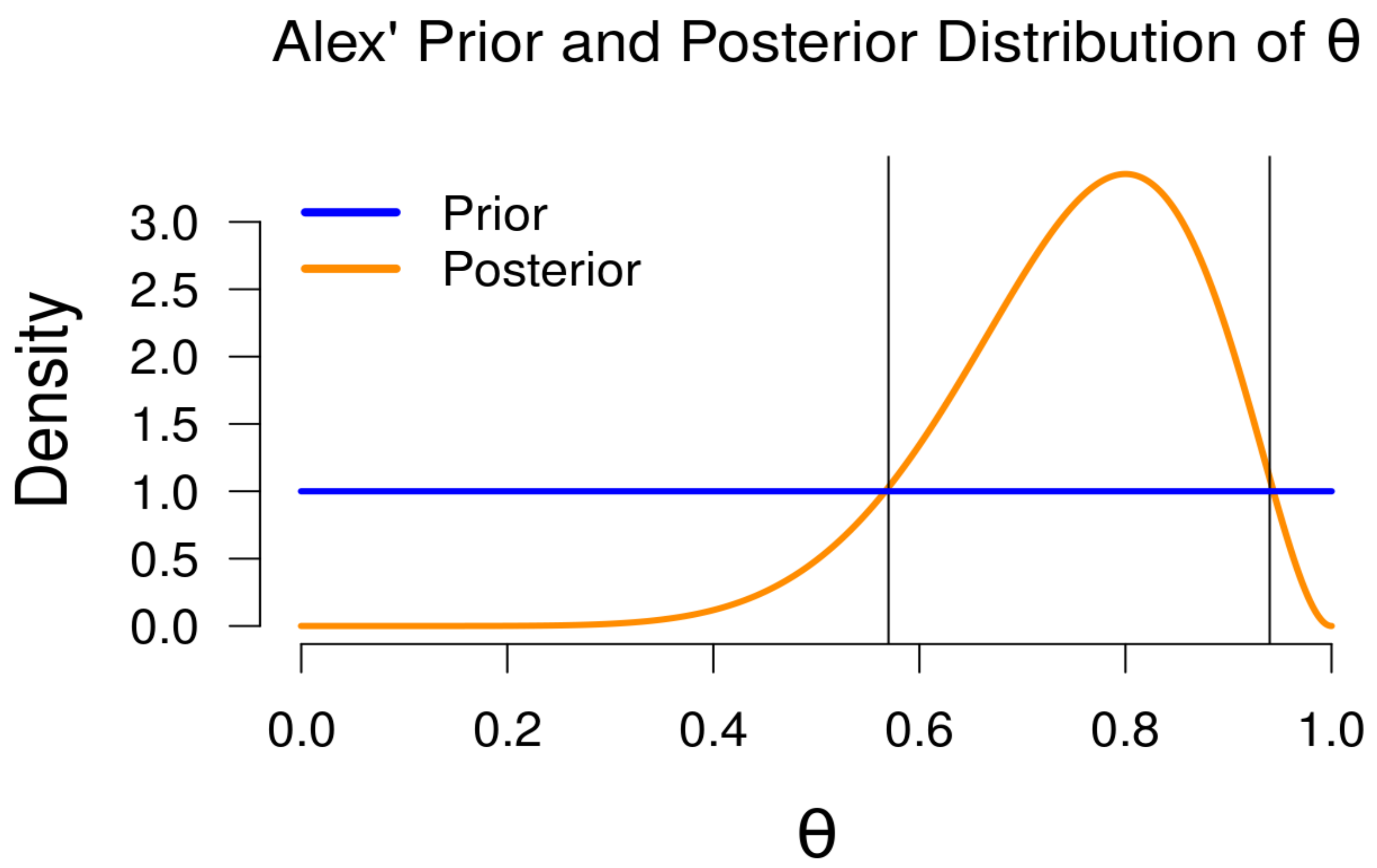




We end with our posterior beliefs

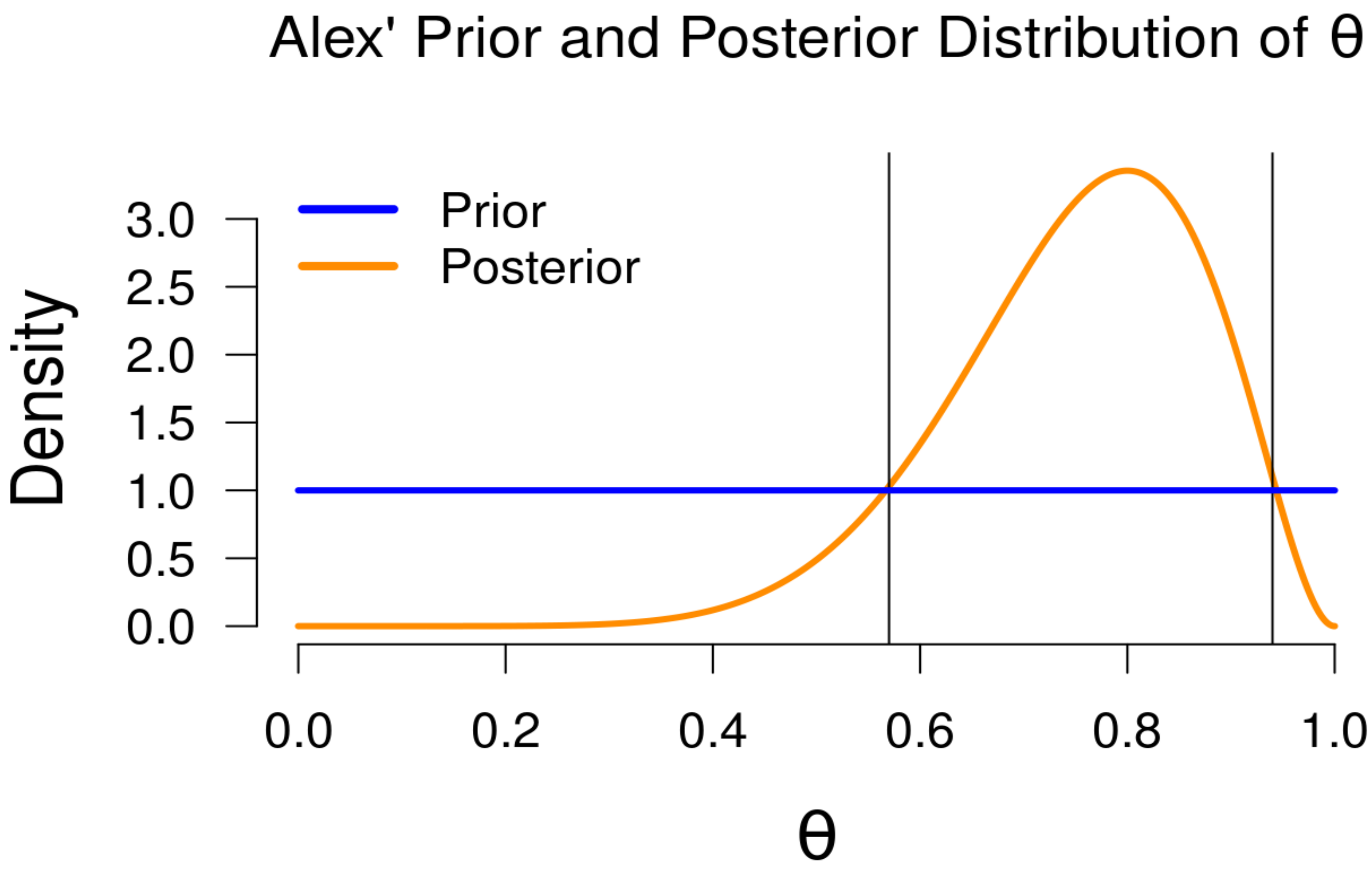


Estimating a Proportion: Posterior Distribution



Same lines as on slide 48/49! It shows which values of theta received a decrease/increase in plausibility, because of the data

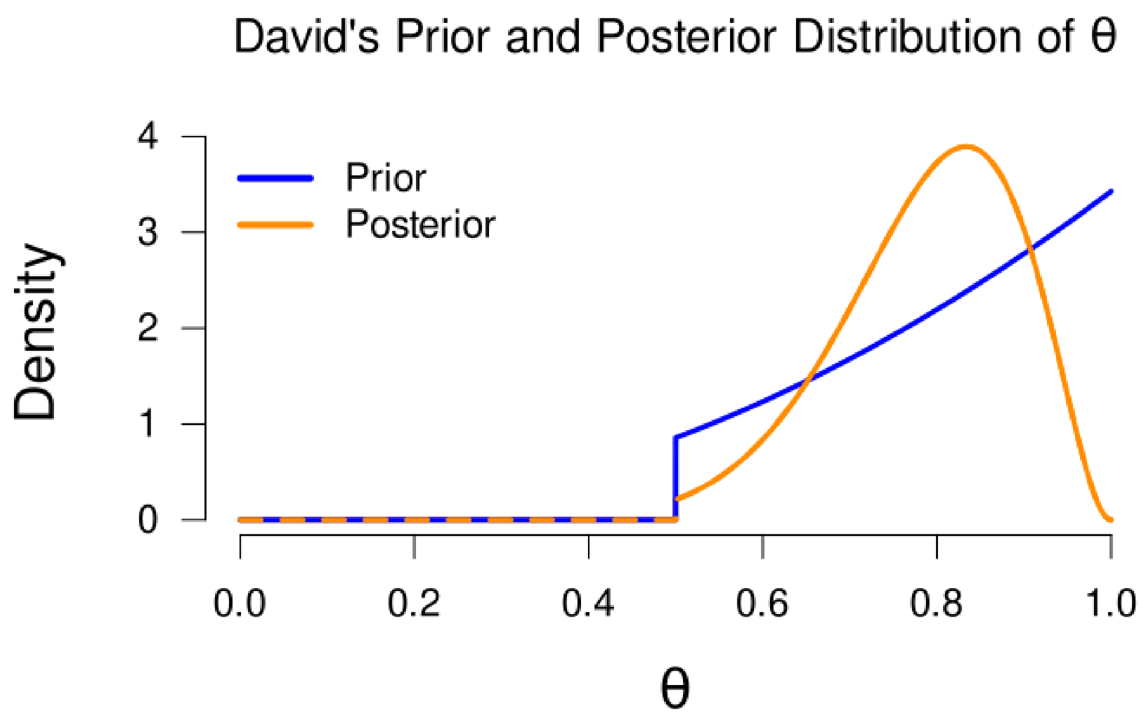
Estimating a Proportion: Posterior Distribution



The posterior distribution is a probability distribution!

Estimating a Proportion: Posterior Distribution

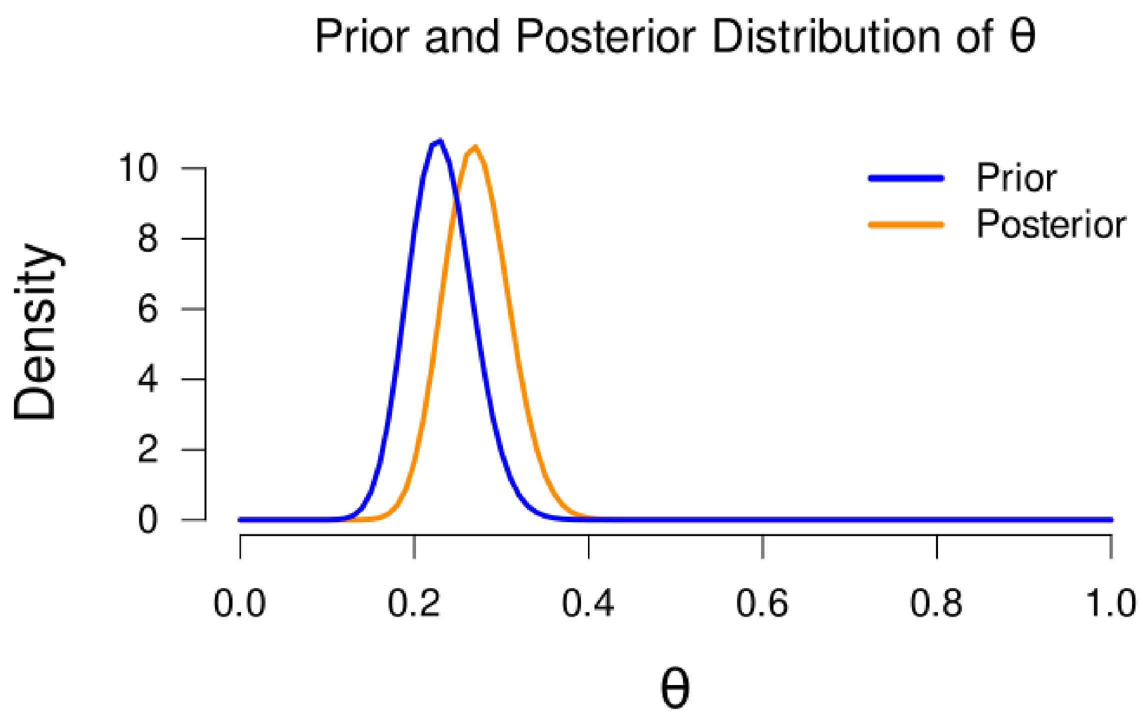
What about other models/prior distributions?



David's Model ($a = 3$, $b = 1$, truncated)
Because all prior mass below 0.5 is set to 0, also all posterior mass below 0.5 will be equal to 0

Estimating a Proportion: Posterior Distribution

What about other models/prior distributions?



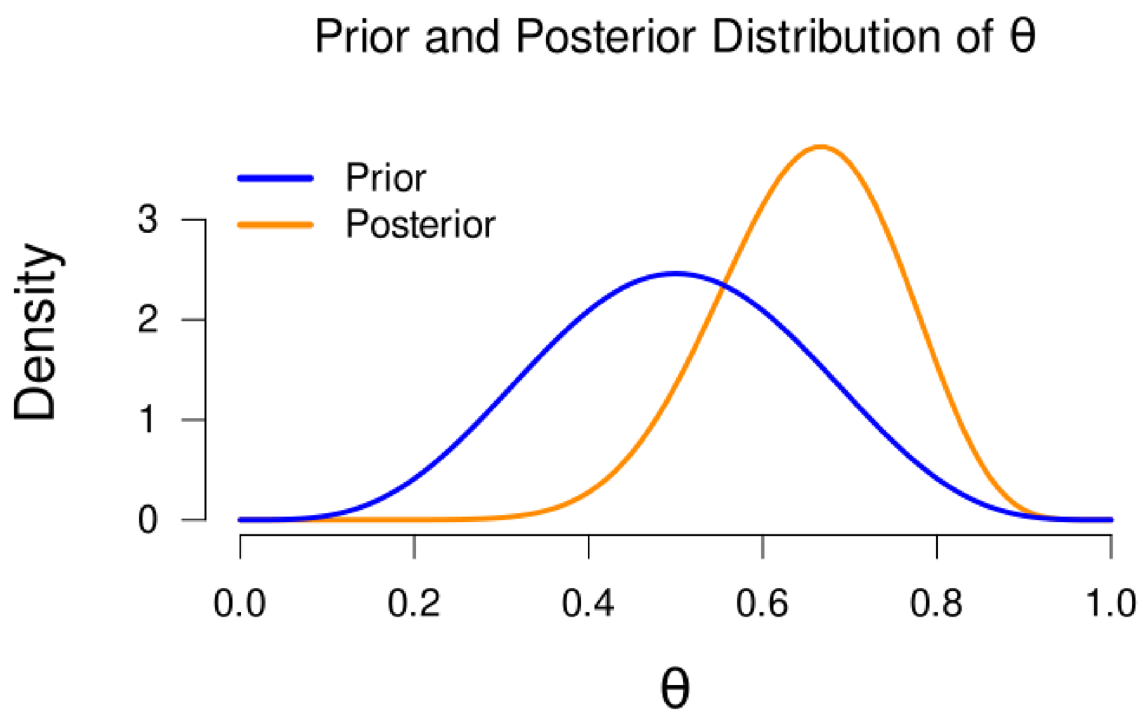
A = 30, b = 100
A very strong prior that assigned more mass to values < 0.5 (coin biased to tails). The prior conviction is so strong that the data cannot overthrow this conviction: the posterior is still situated at values > 0.5

This highlights an important feature of learning: strong beliefs need a lot of data to be convinced otherwise!

This also makes Sarah and Paul very poor learners: they are so convinced of their value that there is no updating

Estimating a Proportion: Posterior Distribution

What about the other models/prior distributions?



a = b = 5
A model that was more certain that the coin is fair

Estimating a Proportion: Posterior Distribution

Easy way of computing the posterior, if we are working with proportions and the beta distribution

If we start with a prior Beta distribution, then the posterior distribution will also be a Beta distribution, with:

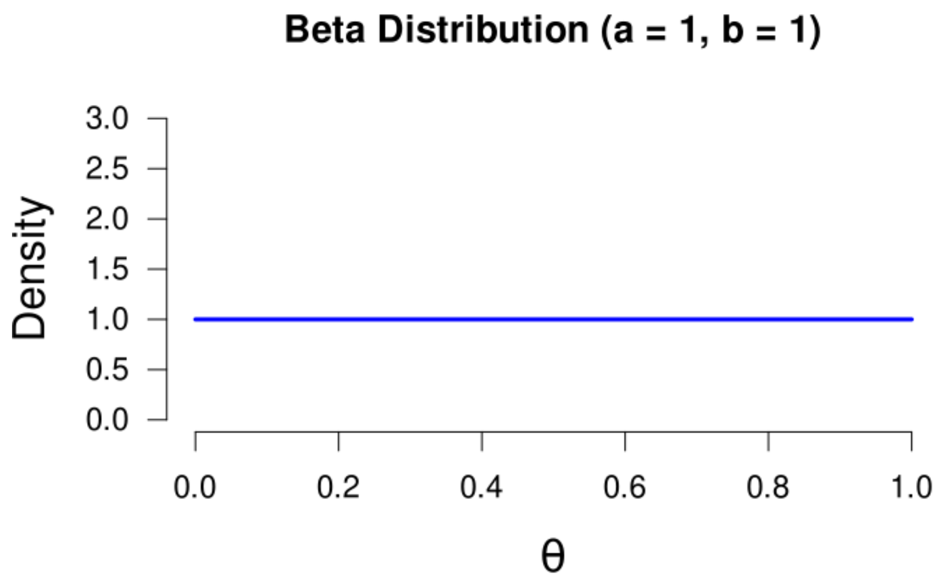
a = a + #successes (in our case, #heads)

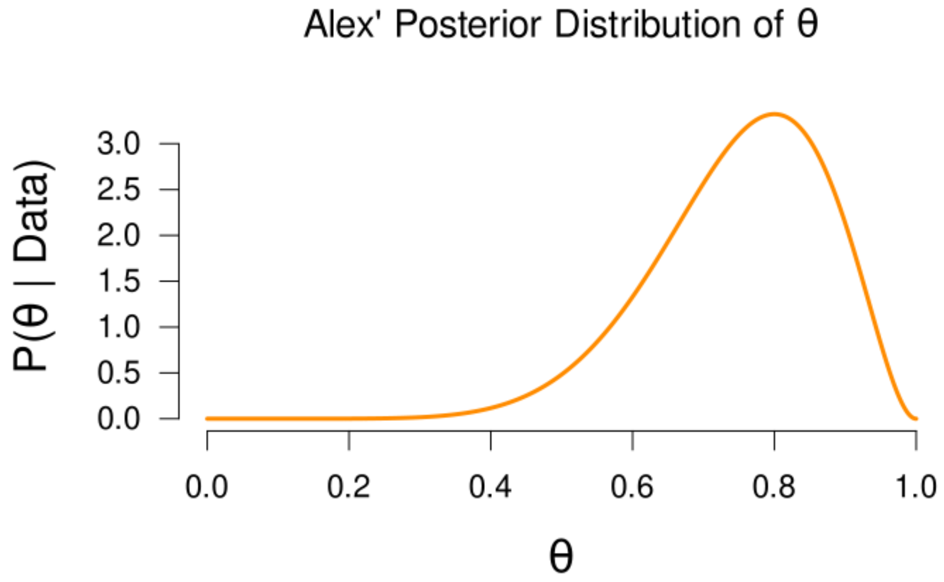
b = b + #failures (in our case, #tails)

Estimating a Proportion: Posterior Distribution

Easy way of computing the posterior, if we are working with proportions and the beta distribution

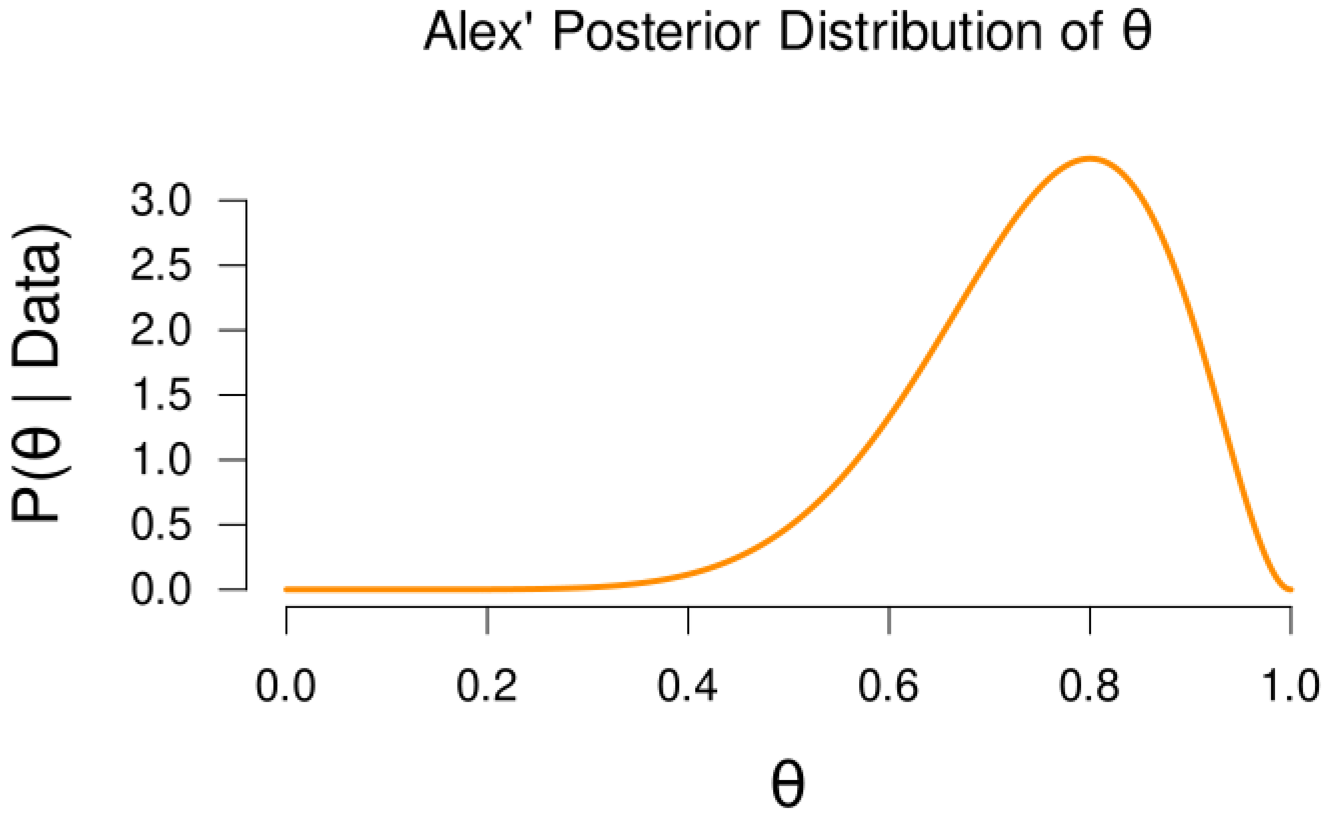
This is a beta distribution with a = 1 + 8
b = 1 + 2





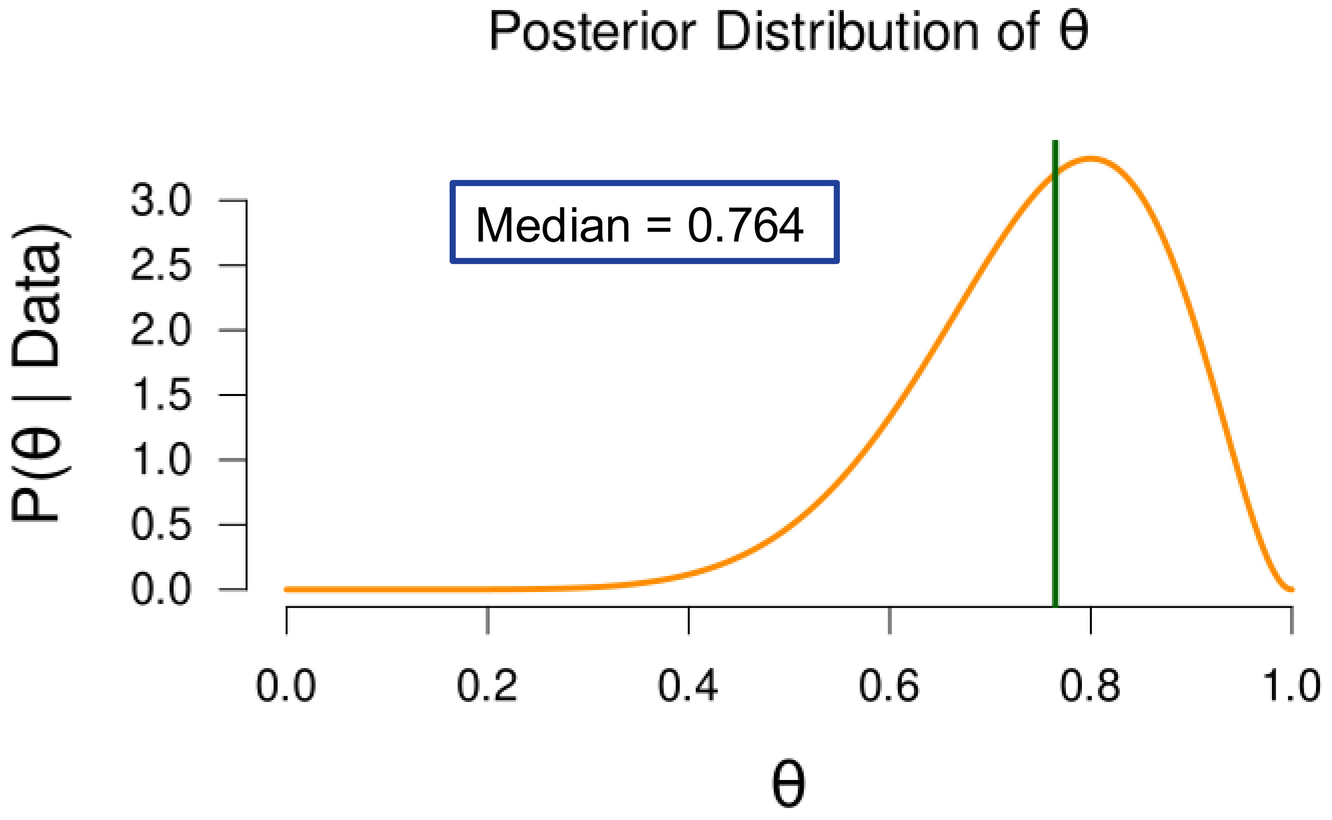
Estimating a Proportion

Now that we have a posterior distribution, we can do **estimation!**



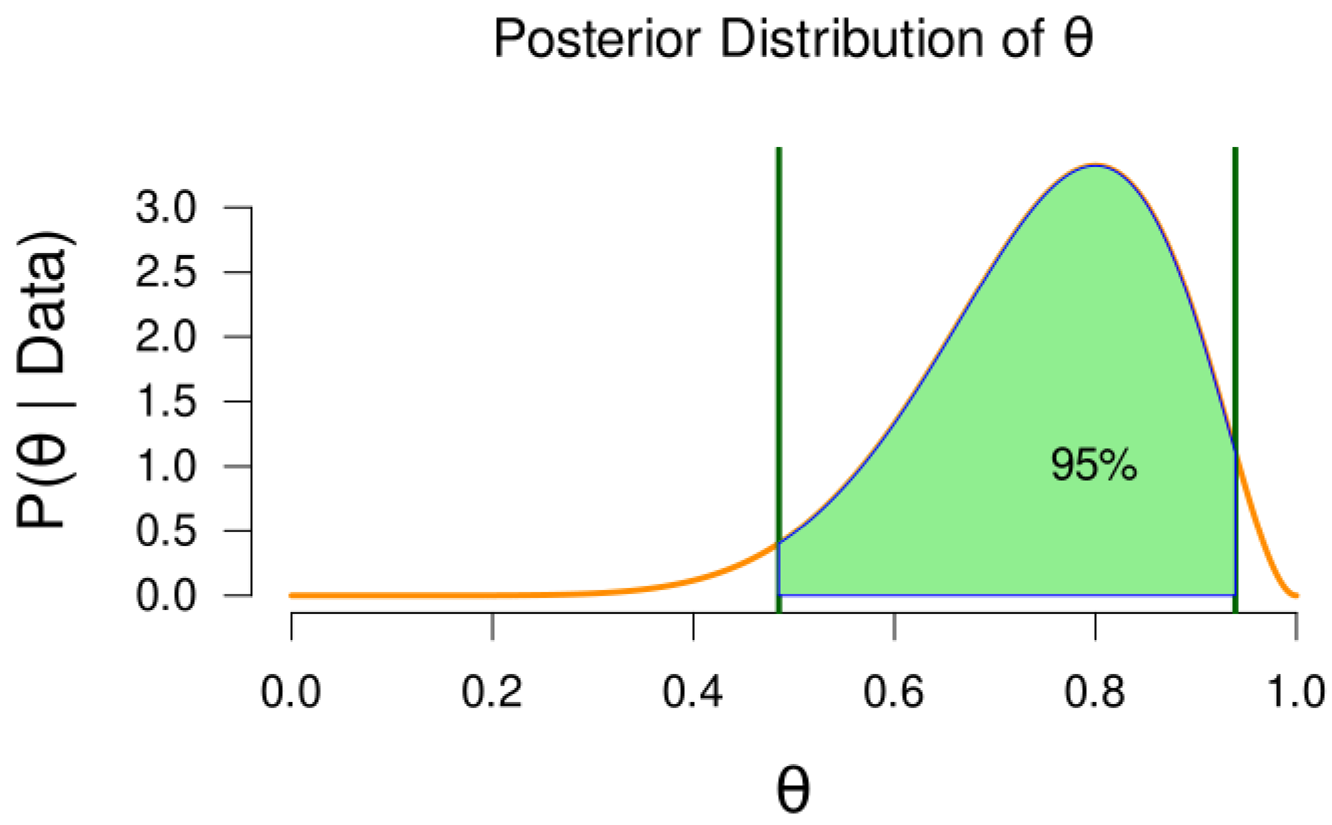
Estimating a Proportion

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



Estimating a Proportion

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

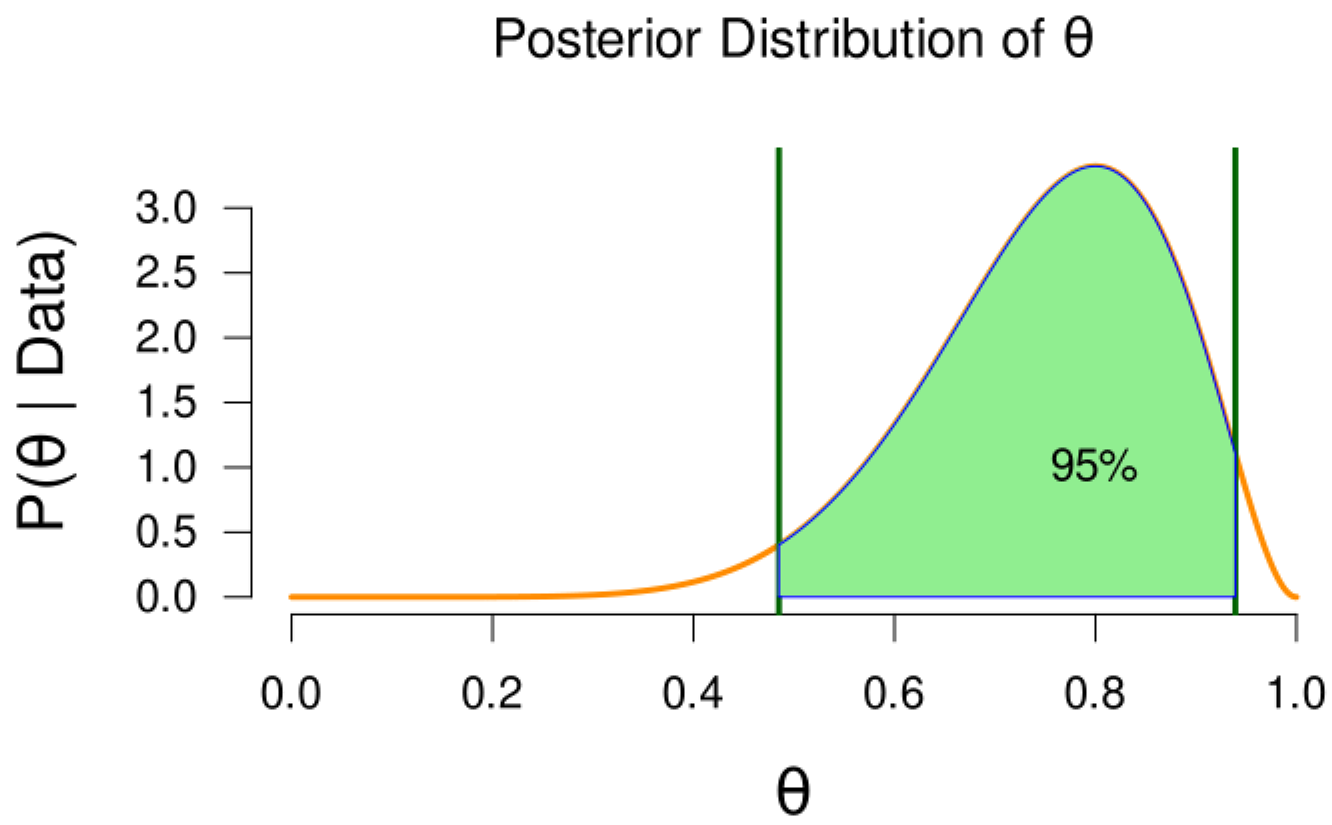


To obtain the x% central credible interval, we take x% of the most central posterior mass, and see which 2 points are the thresholds

95% Credible interval = (0.49, 0.94)

Estimating a Proportion

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



How to interpret a 95% central credible interval?

Under Alex' model, there is a 95% probability that the true proportion is between 0.49 and 0.94

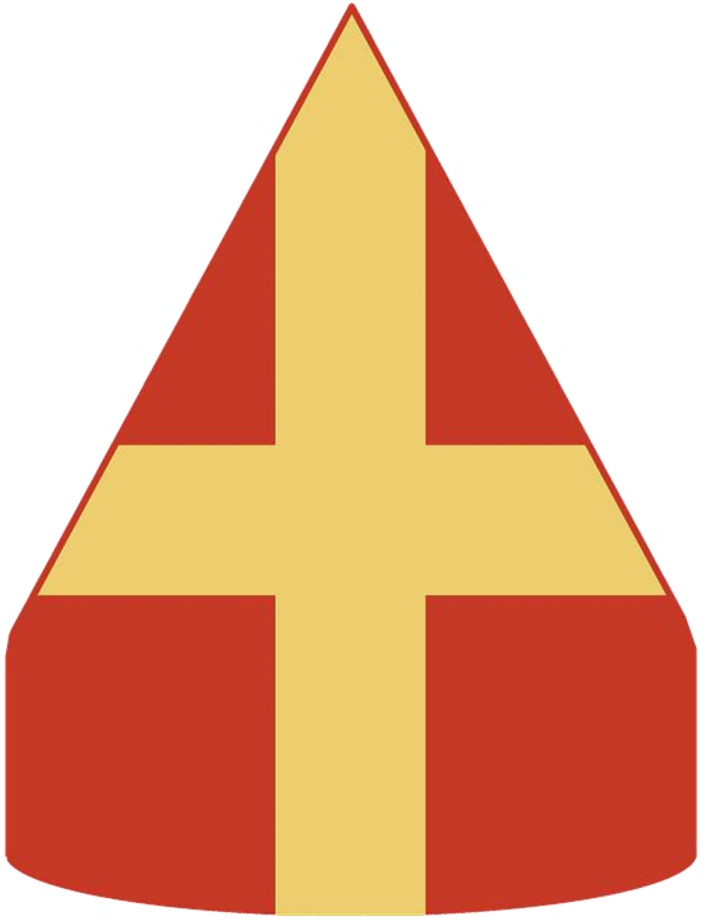
95% Credible interval = (0.49, 0.94)

Today

- Bayesian statistics
  - What are models?
  - How do models learn from data?
  - **Live demonstration**
- Recap
  - Practical stuff
  - Example exam question

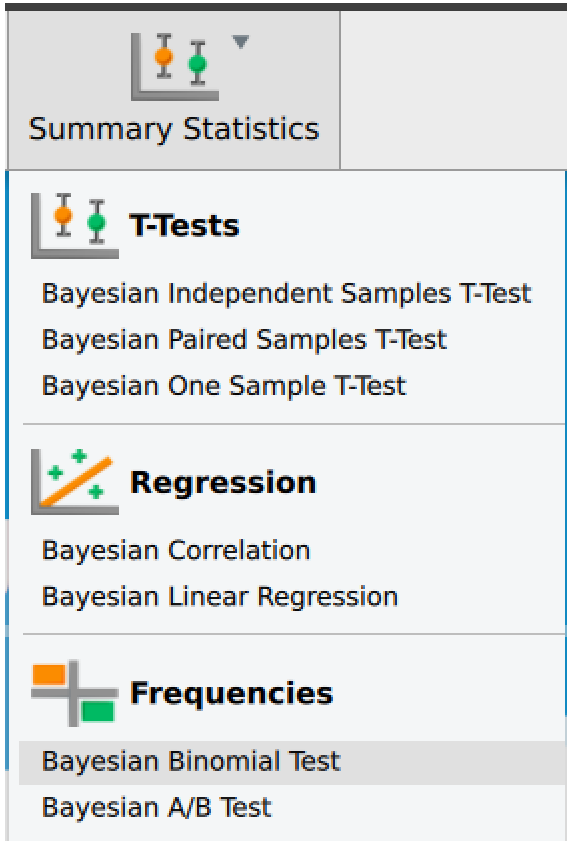
Sinterklaastimation!!

edu.nl/h8mj8



Source:
Pixabay.com

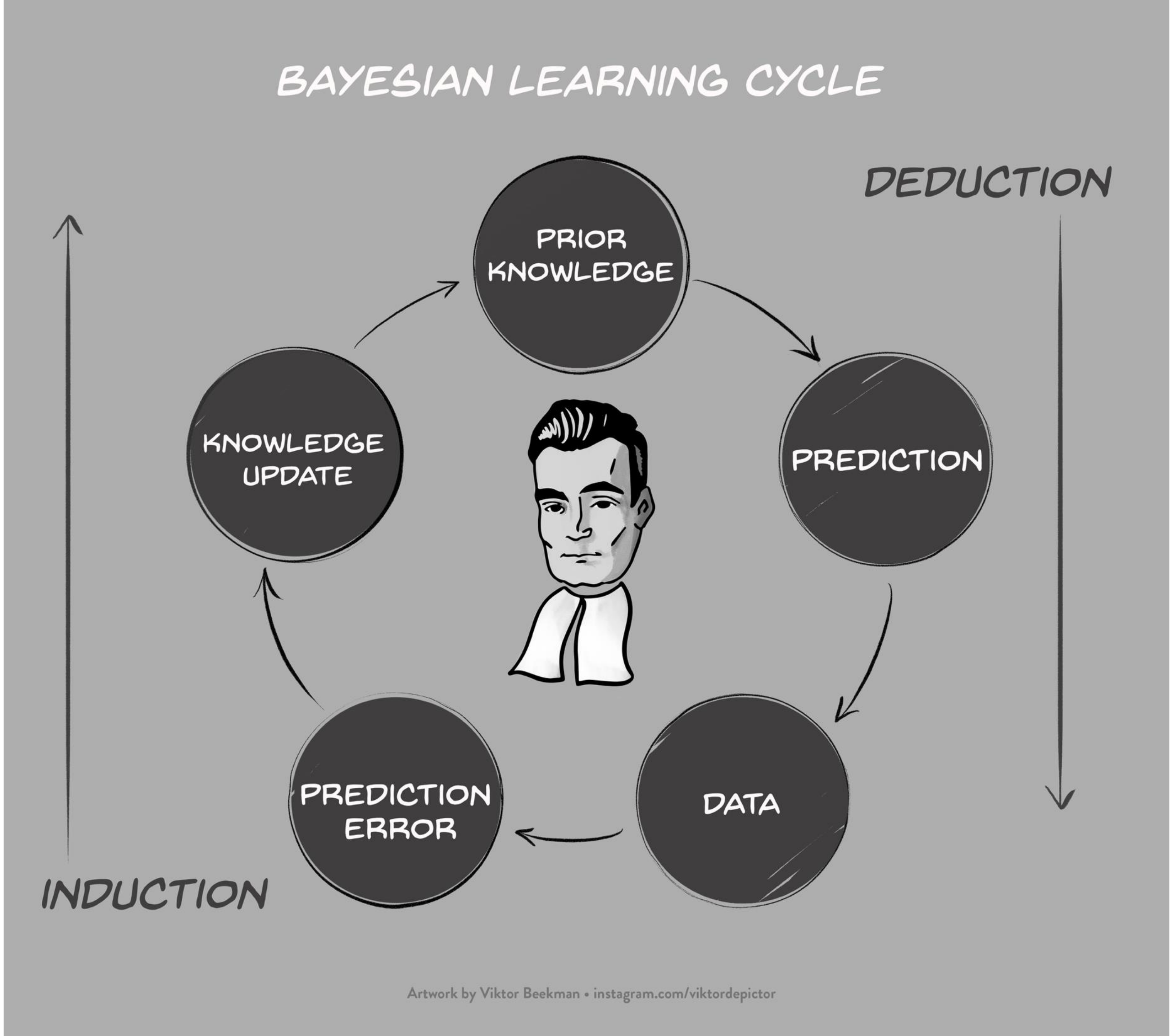
Binomial Analysis in JASP



Enable this extra module first by clicking on the big + the top right corner

Recap

- We have some prior knowledge about a phenomenon
- We formalize this knowledge in the form of a probability distribution
- Then we update this with the information in the data to form posterior knowledge
- A central term here is “predictive quality”: how well does each possible value of the parameter predict the observed data, *compared to the other values*



Source:
https://www.bayesianspectacles.org

Practical Stuff

Course literature:

- van Doorn, J. (2024). A Brief Introduction to Bayesian Inference: From Tea to Beer. Free at https://johnnydoorn.github.io/IntroductionBayesianInference/

Practical Stuff

New software:

- JASP 0.95.4 – Free at https://jasp-stats.org/download/
    - Also available on library computers and https://apps.uva.nl/
    - Instructions included in Exercise chapter and in WA
- Exercises
    - The book and WA include practice in JASP (but also include output, if you just want to practice with interpreting output)
- Exam
    - On the exam you only need to be able to work with the "Summary Statistics" Module (binomial test, t-test, correlation)
    - For example, Question 6 in WA15, Question 6.4 in book
    - Other questions will provide JASP output and ask you to interpret

Practical Stuff

- Exam questions Bayes: more conceptual than calculating
    - Interpret Bayesian results/JASP output
    - Obtain prior/posterior/Bayes factor in JASP's Summary Statistics module
    - Proportional to lectures: about 60% of exam will be Bayes

- Weekly assignment: next week will include Bayes
    - The literature contains exercises in the back
    - Week 16 (trial exam) will contain old Bayesian exam questions

- Exam material: Slides + course literature

Example Exam Question

Which of the following is a probability distribution:

- Posterior distribution
- Likelihood
- Both

Example Exam Question

Which of the following is a probability distribution:

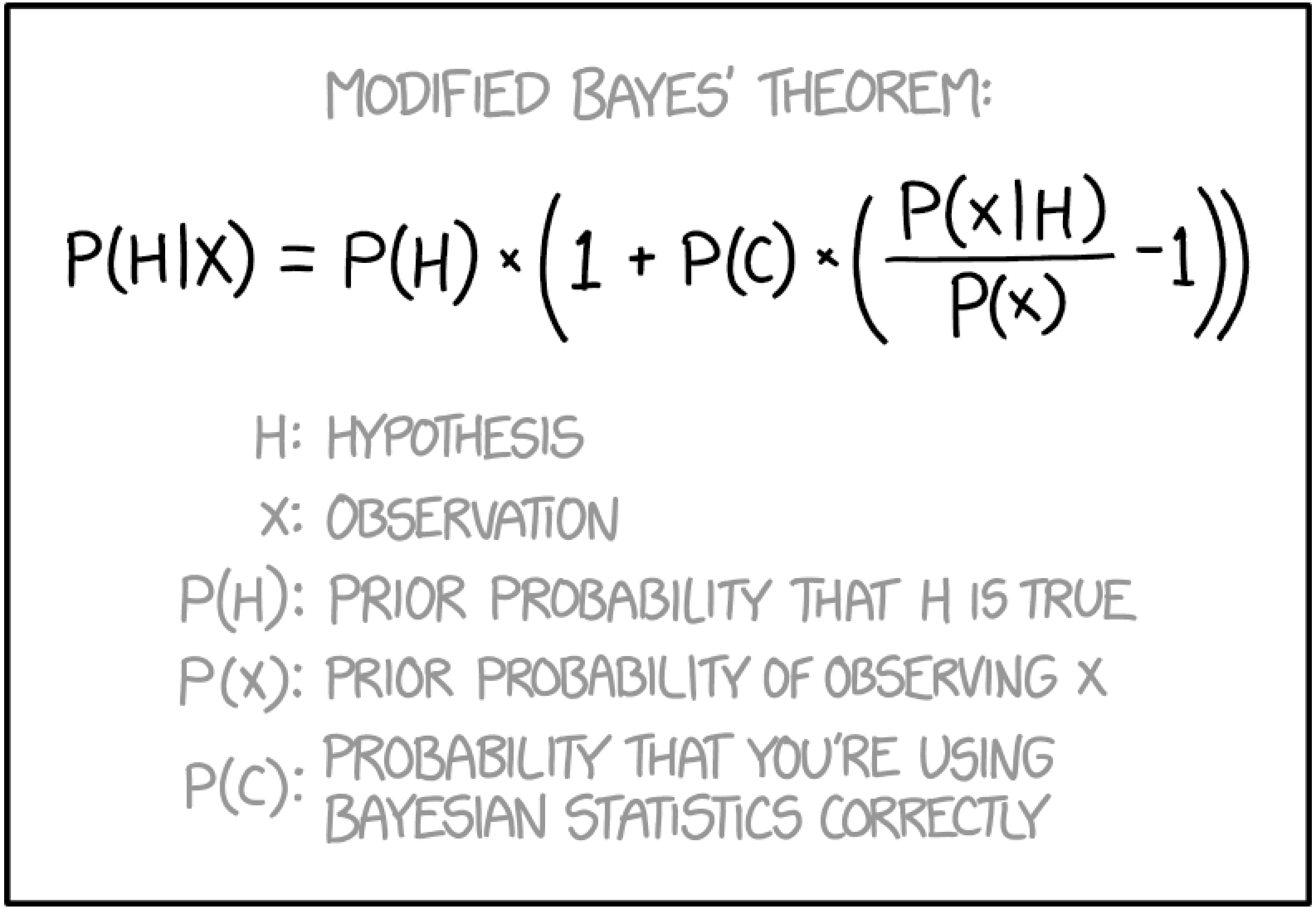
- Posterior distribution
- ~~Likelihood~~
- ~~Both~~

Bayesian Hypothesis Testing

Sneak peek: https://www.youtube.com/watch?v=9TDjifpGj-k

Questions?

Thank you for your attention



https://xkcd.com/2059/

Bonus Book

One of my colleagues has written a **very** introductory text book to introduce some core Bayesian concepts. And it features dinosaurs!