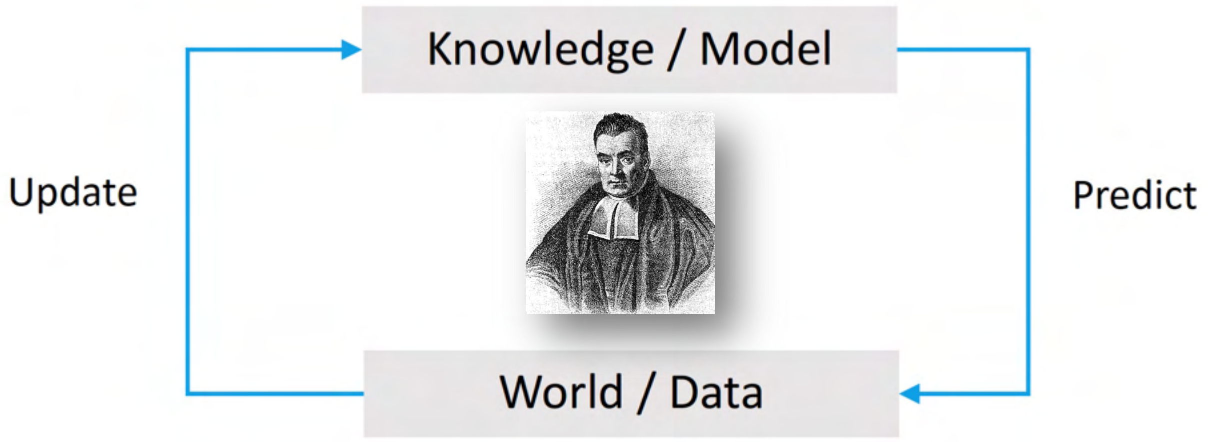


Intro to Bayesian Statistics in R

Angelika Stefan

a.m.stefan@uva.nl

R-Ladies Amsterdam, September 28, 2021

About me

University of Amsterdam

PhD candidate Psychological Methods

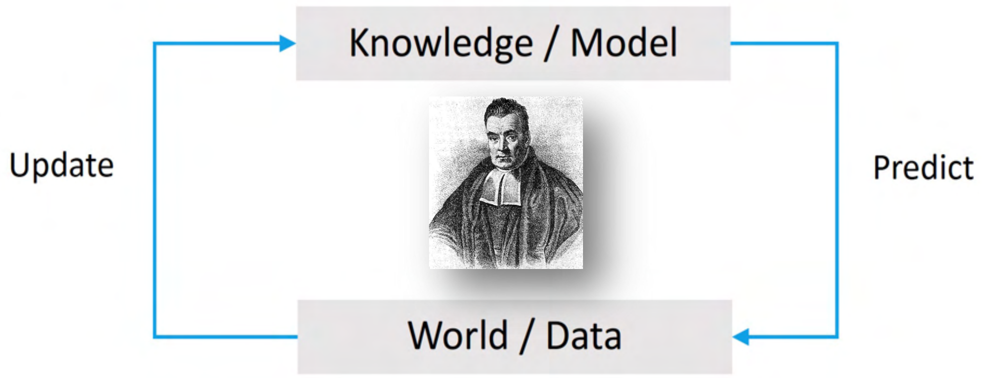
LMU Ludwig-Maximilians-Universität München

Economic, Organizational, & Social Psychology M.Sc.

Psychology B. Sc.

Cultural Anthropology

About the workshop



Understanding the basics of Bayesian statistics

Let’s get started!

Bayesian statistics is getting more and more popular

REVIEW ARTICLE

Front. Psychol., 08 August 2014 | https://doi.org/10.3389/fpsyg.2014.00765

The Bayesian boom: good thing or bad?

Ulrike Hahn*

Department of Psychological Sciences, Centre for Cognition, Computation, and Modelling, Birkbeck, University of London, London, UK

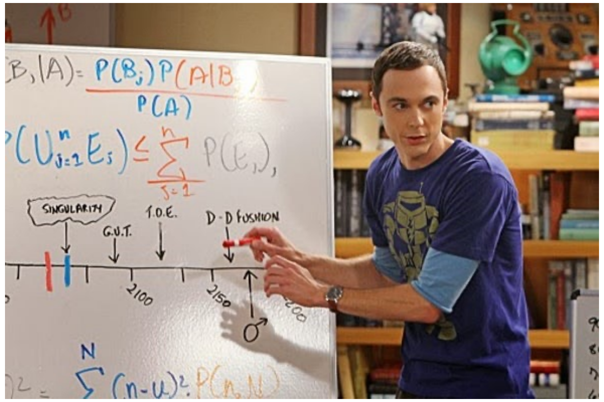
Published: 04 April 2017

Introduction to Bayesian Inference for Psychology

Alexander Etz & Joachim Vandekerckhove

Psychonomic Bulletin & Review **25**, 5–34(2018) | Cite this article

14k Accesses | **24** Citations | **129** Altmetric | Metrics



Psychonomic Bulletin & Review

February 2018, Volume 25, Issue 1, pp 35–57 | Cite as

Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications

Authors | Authors and affiliations

Eric-Jan Wagenmakers, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Jonathon Love, Ravi Selker, Quentin F. Gronau, Martin Šmíra, Sacha Epskamp, Dora Matzke, Jeffrey N. Rouder, Richard D. Morey

Bayesian Versus Orthodox Statistics: Which Side Are You On?

Zoltan Dienes

First Published May 18, 2011 | Review Article | Find in PubMed | Check for updates

https://doi.org/10.1177/1745691611406920

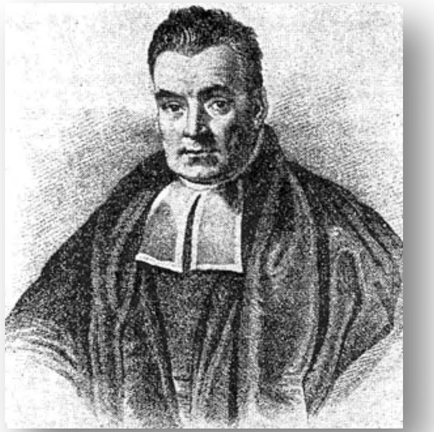
Article information ⌄

Altmetric 31

Abstract

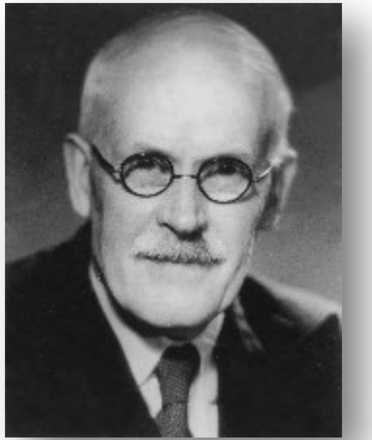
Researchers are often confused about what can be inferred from significance tests. One problem occu

Bayesian statistics is nothing new



Reverend Thomas Bayes (1701 – 1761)

- Presbyterian minister
- Studied logic and theology in Edinburgh
- Bayes theorem



Sir Harold Jeffreys (1891 – 1989)

- Geophysicist, mathematician, astronomer
- Professor for astronomy at Cambridge
- Bayes factor

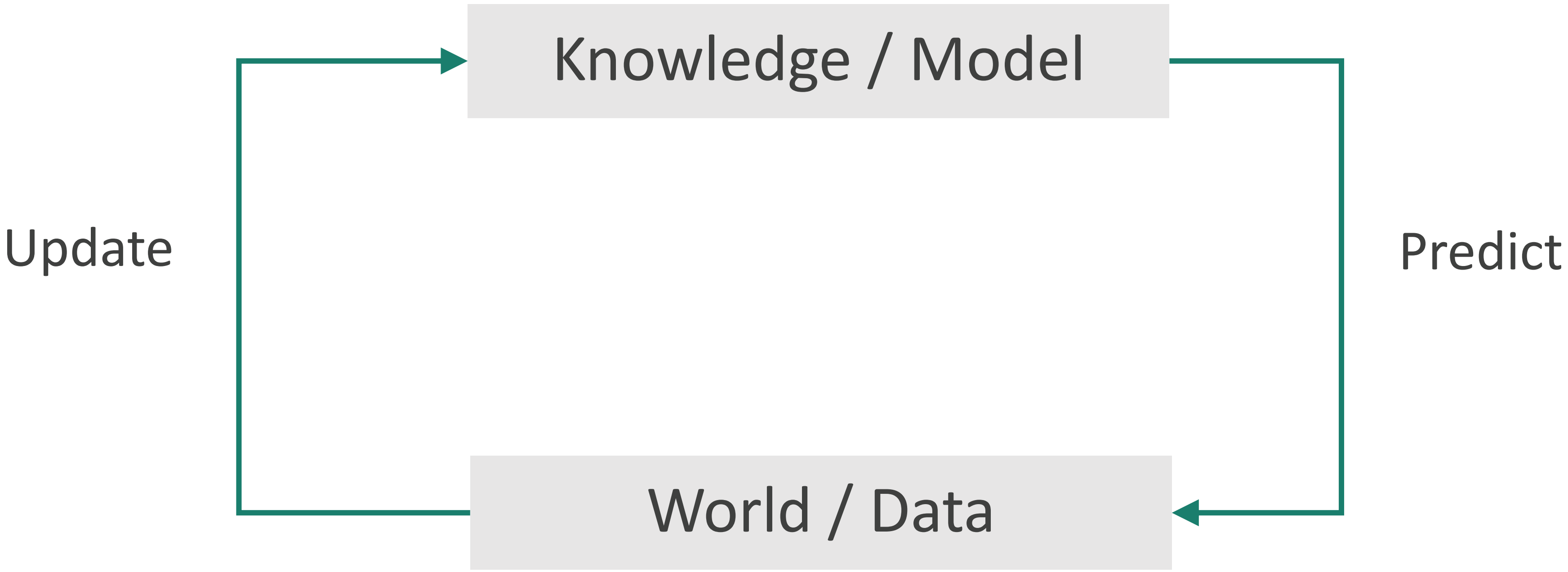
Bayesian statistics is about quantifying uncertainty

"Chance then exists not in nature, and cannot coexist with knowledge; it is merely an expression for our ignorance of the causes, and our consequent inability to predict the result."

William S. Jevons, 1873

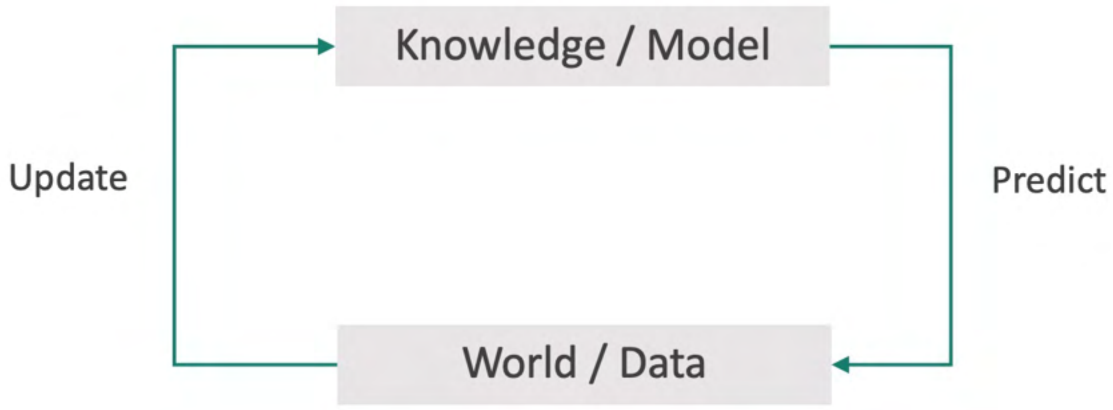
"It's probably B…"

Bayesian statistics is about updating knowledge



Bayesian statistics is about updating knowledge

Parameter Estimation



Hypothesis Testing /
Model Comparison

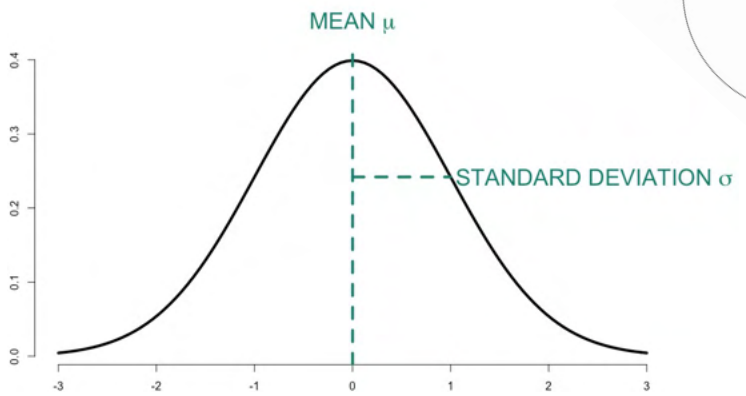
Knowledge updating in parameter estimation

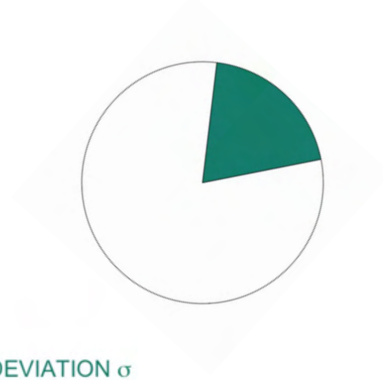
What is a parameter?

Any measured quantity of a statistical population that summarizes or describes an aspect of the population.

Examples:

- Mean
- Standard deviation
- Variance
- Proportion





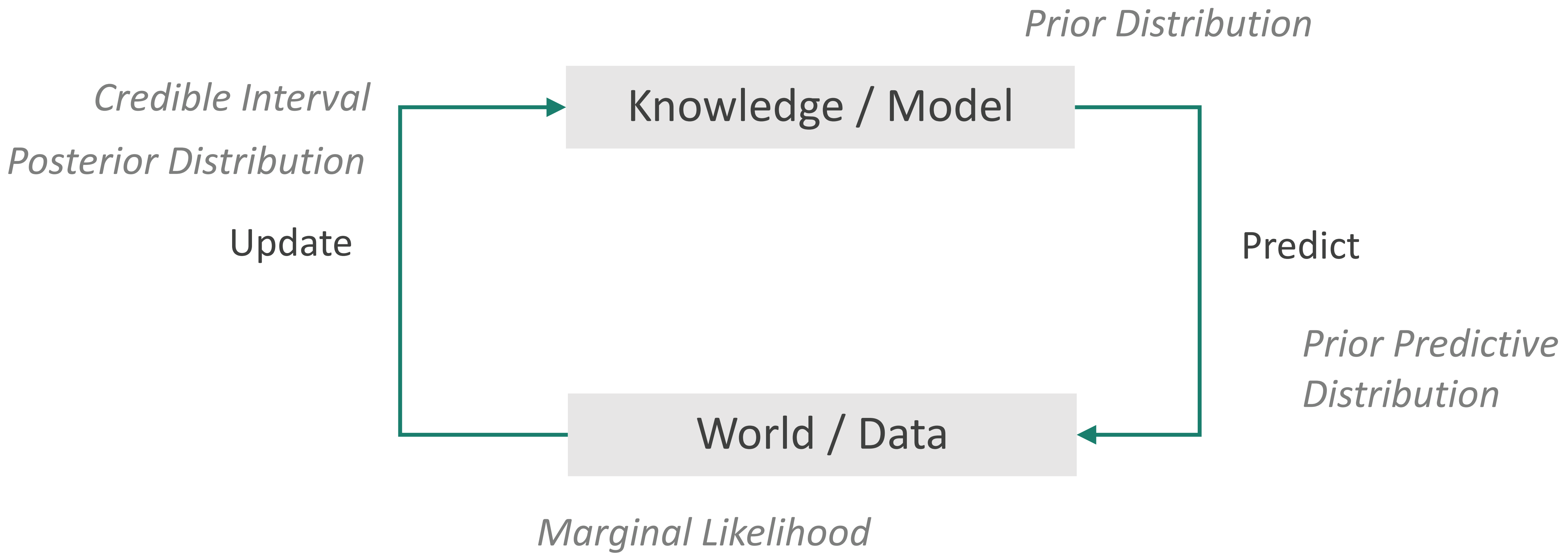
Knowledge updating in parameter estimation

What is a parameter?

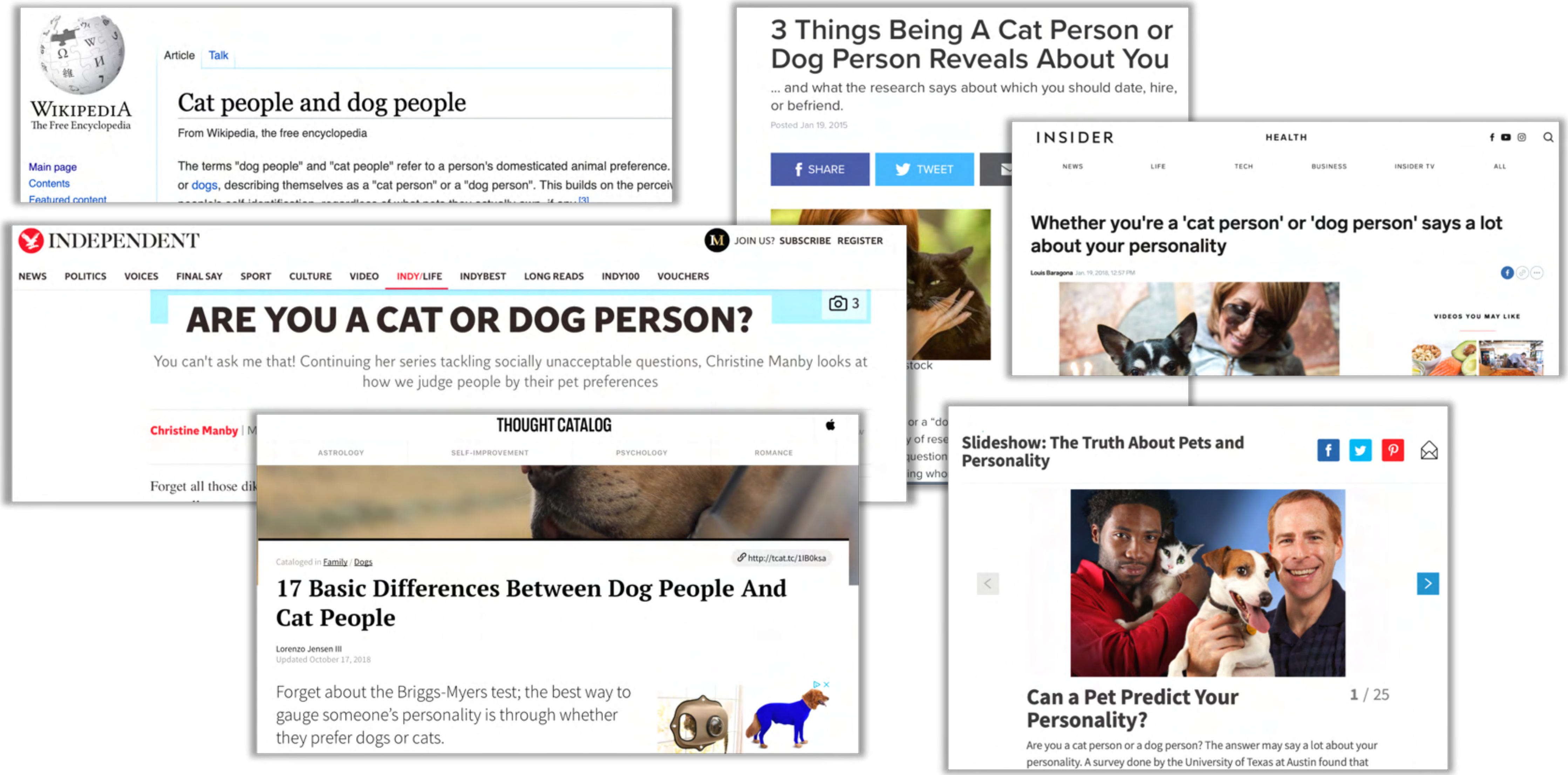
Any measured quantity of a statistical population that summarizes or describes an aspect of the population.

We do not know the population parameter

Bayesian statistics is about updating knowledge



Running example



Parameter of interest

People can be classified as either a cat person or a dog person.

We are interested in the **proportion of dog people** in the Netherlands.

$\boldsymbol{\theta} = 0$

0% dog people

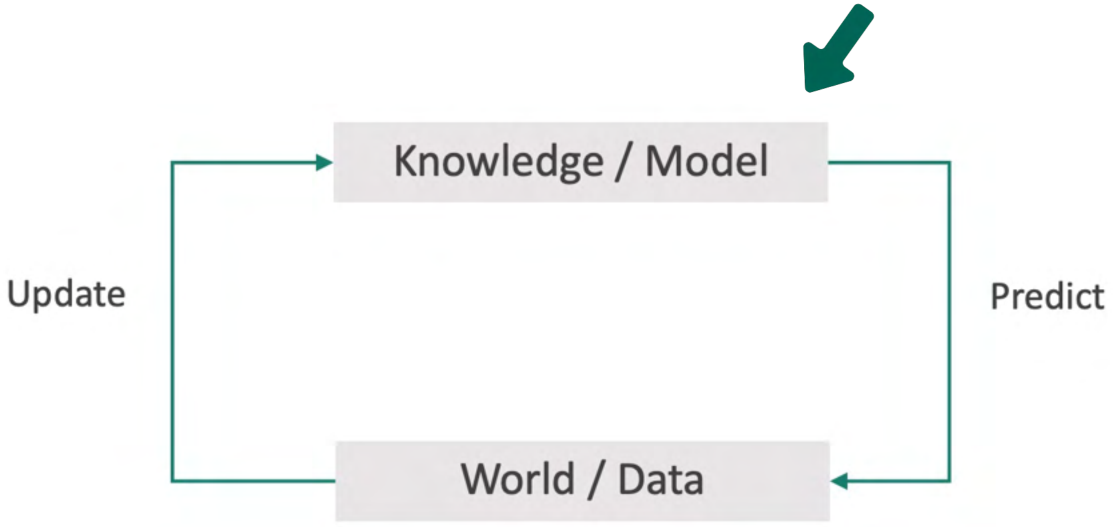
$\boldsymbol{\theta} = 1$

100% dog people

Prior Distribution

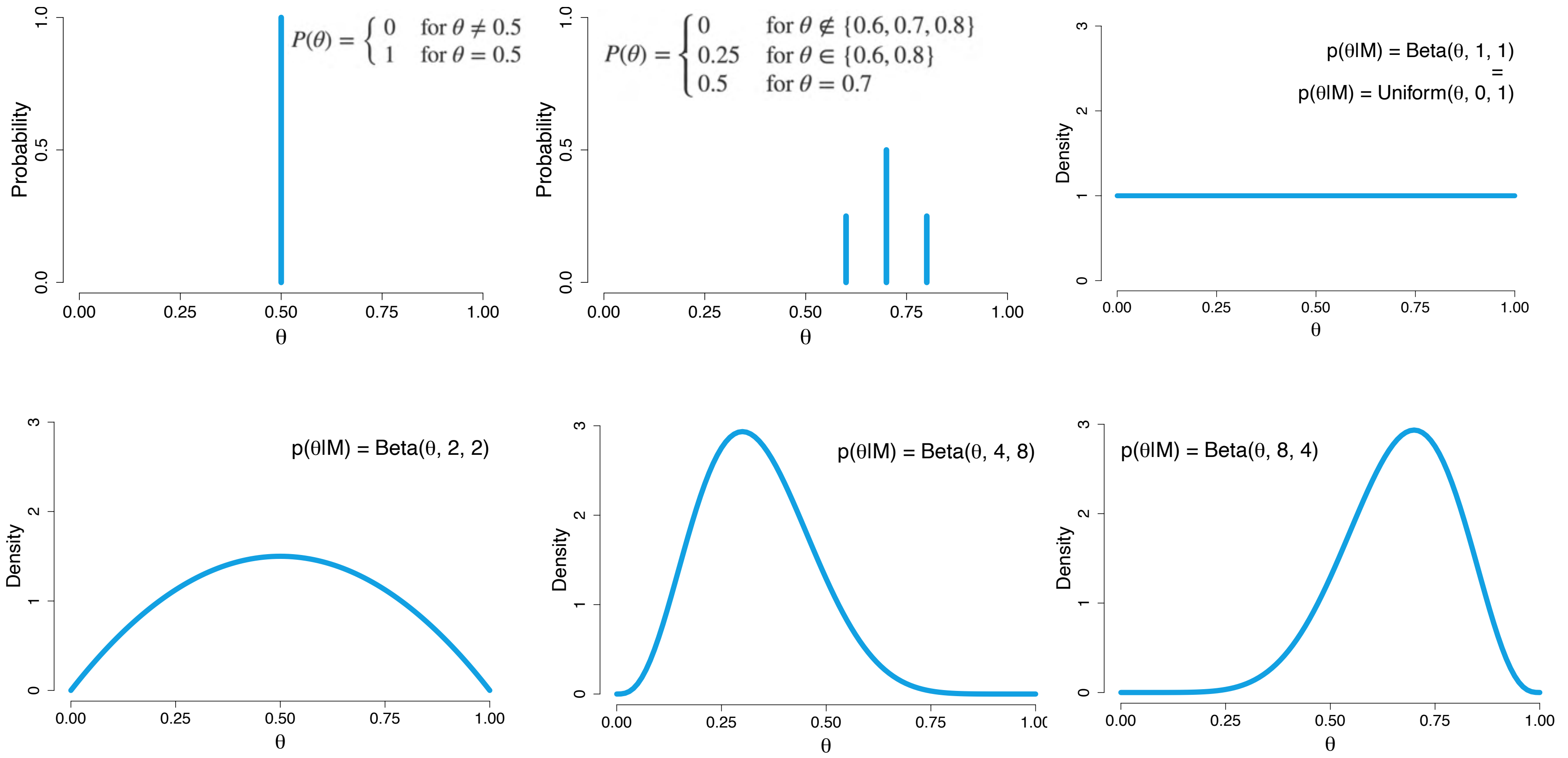
$$p(\theta)$$

- Probability (density) function
- Quantifies uncertainty about a parameter *before* data collection
  - Wide prior: high uncertainty
  - Narrow prior: only few values considered likely



Prior Distribution

Different Prior Distributions



Choose your prior!

```r
# Change the shape parameters to find the prior distribution that represents
# your belief best.
# Remember: theta = 0 = 0% dog people; theta = 1 = 100% dog people

shape1 <- 1
shape2 <- 1

curve(dbeta(x, shape1, shape2),
      bty="l", xlab=bquote(theta), ylab="Density", cex.lab=1.5)
```

The Likelihood

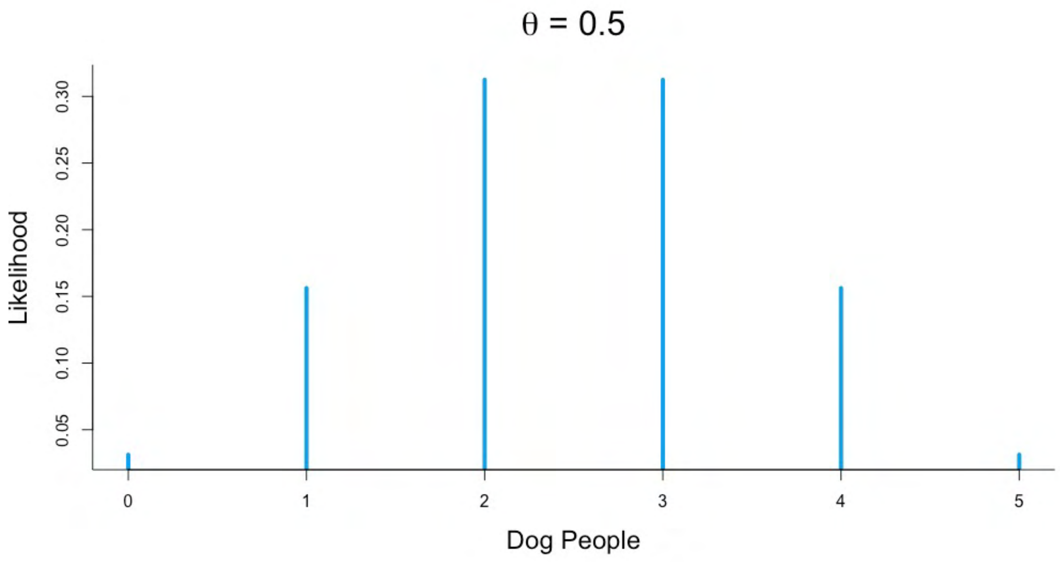
$p(\boldsymbol{X}|\theta)$

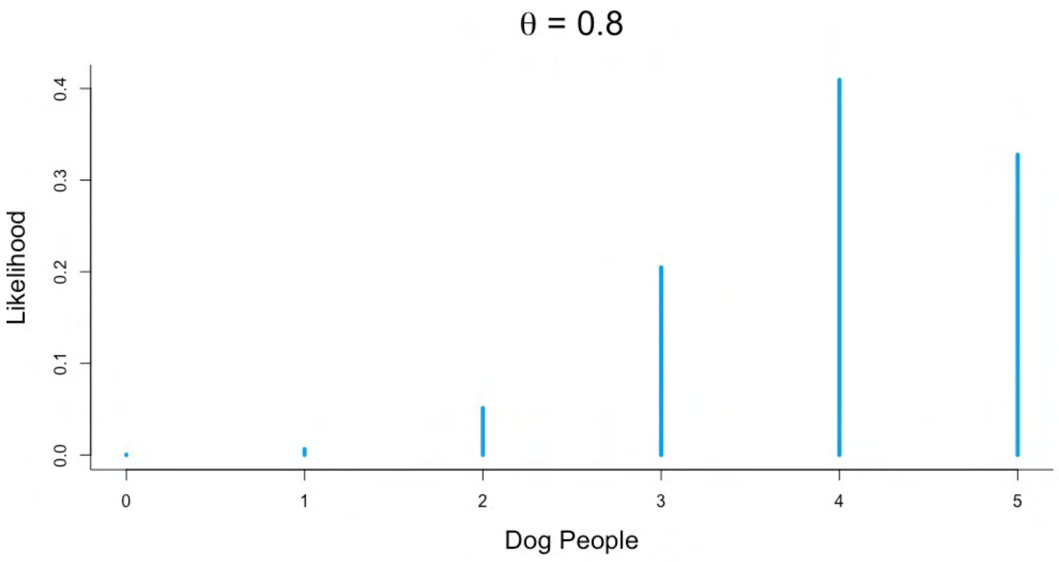
Shows how plausible data are, given a fixed parameter value

The number of successes in a sequence of *n* independent trials, each asking a yes-no question, follows a Binomial distribution with success parameter $\boldsymbol{\theta}$.

The Likelihood

$p(\boldsymbol{X}|\theta)$



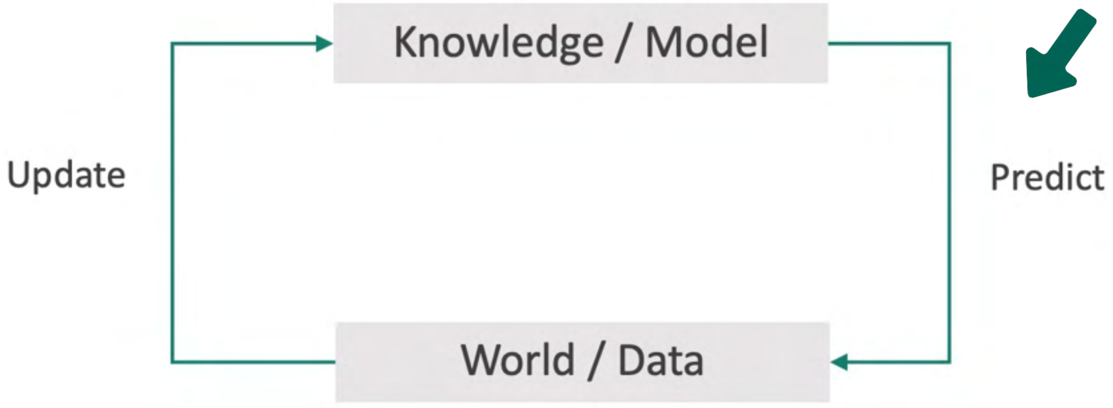



```
dbinom(0:n_observations, size = n_observations, prob = 0.5)
```

```
dbinom(0:n_observations, size = n_observations, prob = 0.8)
```

Prior Predictive Distribution $p(X)$

- Makes a prediction about the plausibility of data
- Assuming that a model (prior + likelihood) is correct, what predictions does it make about possible data?



Prior Predictive Distribution

$p(\boldsymbol{X})$

Computation of the prior predictive distribution

$$p(X) = \sum_{\theta} p(\theta) \cdot p(X \mid \theta)$$

Weigh the likelihood of possible observations with the prior across all possible parameter values

Prior Distribution: Shows how plausible parameter values are before seeing the data

Likelihood: Shows how plausible data are, given a fixed parameter value

Prior Predictive Distribution

$p(\boldsymbol{X})$

Computation of the prior predictive distribution

$$p(X) = \int_0^1 p(\theta) \cdot p(X \mid \theta) d\theta$$

Weigh the likelihood of possible observations with the prior across all possible parameter values

Prior Distribution: Shows how plausible parameter values are before seeing the data

Likelihood: Shows how plausible data are, given a fixed parameter value

Prior Predictive Distribution

$p(\boldsymbol{X})$

Computation of the prior predictive distribution

$$p(X) = \int_0^1 p(\theta) \cdot p(X \mid \theta) d\theta$$

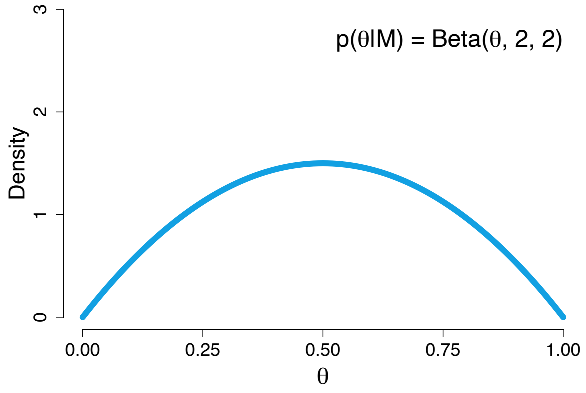
Weigh the likelihood of possible observations with the prior across all possible parameter values

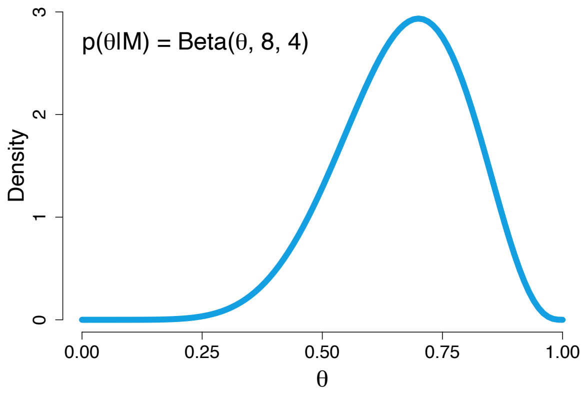
```
integrand <- function(theta){
  dbinom(i,n_observations,theta) * dbeta(theta,shape1,shape2)
}

integrate(integrand, lower = 0, upper = 1)$value
```

Prior Predictive Distribution

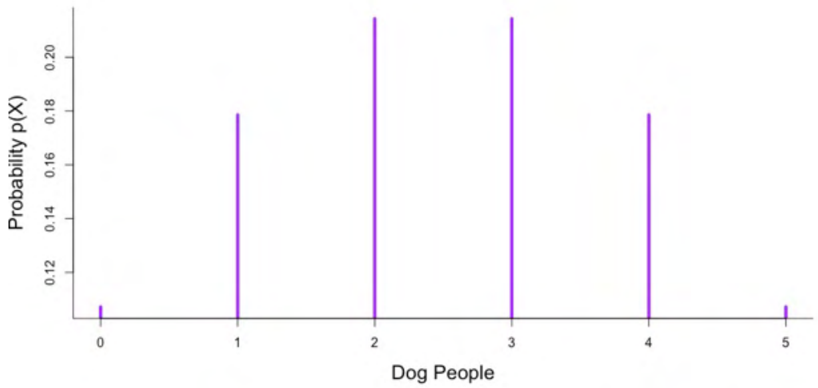
$p(\boldsymbol{X})$

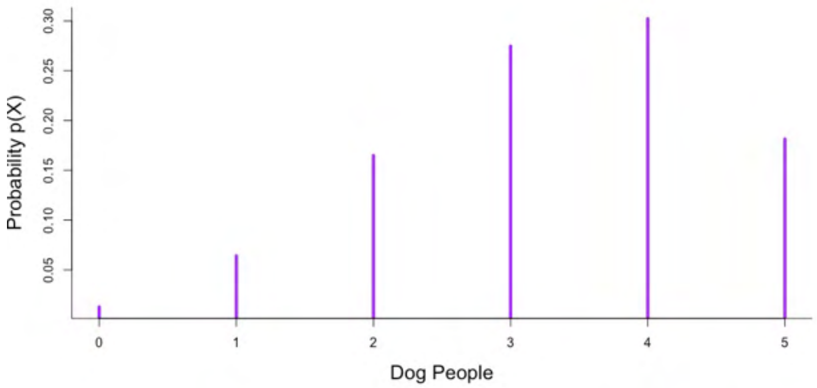
Prior





Prior Predictive





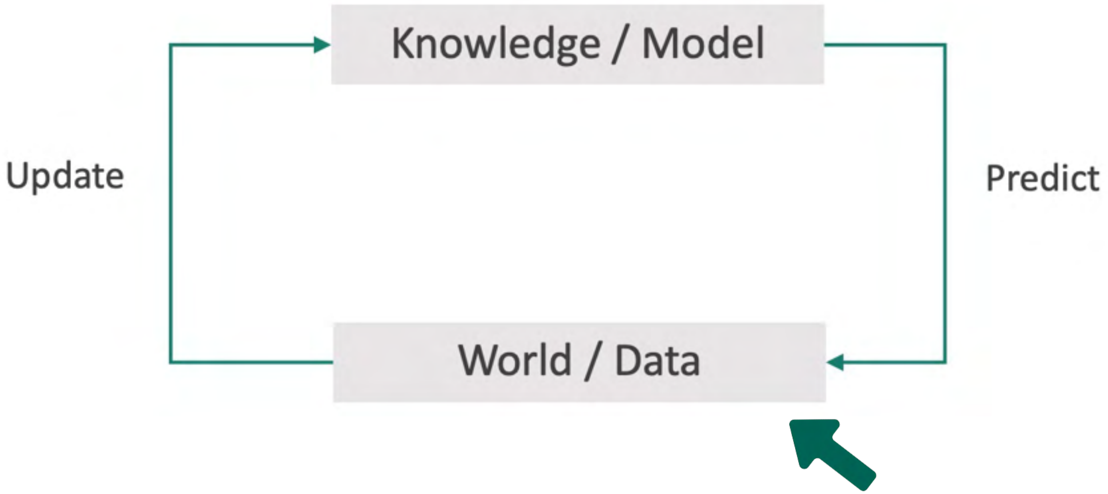
Yay, data!

Our fictional data:

In a random sample of N = 5 we observe:

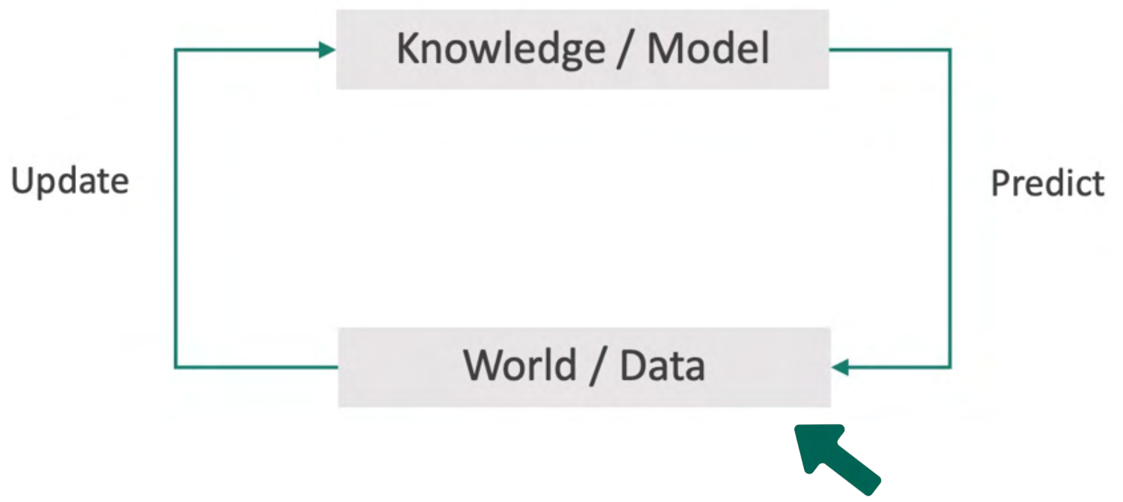
x = 3 Dog people
5-x = 2 Cat people

vs.



Marginal Likelihood $p(x)$

- How plausible are the *observed* data under the model
- Evaluation of the prior predictive distribution at the observed data

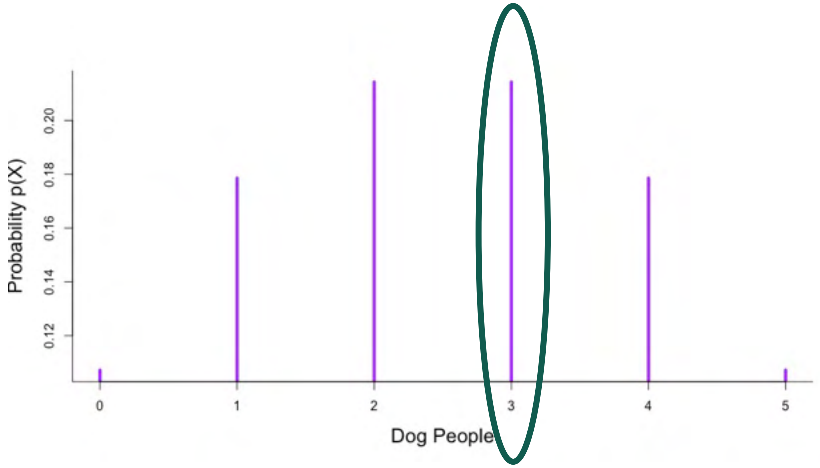


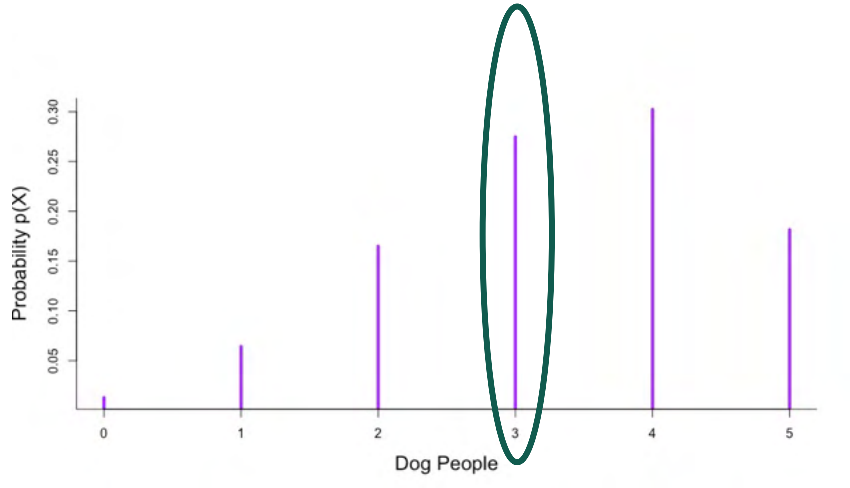
Marginal Likelihood

$p(x)$

Prior

Predictive

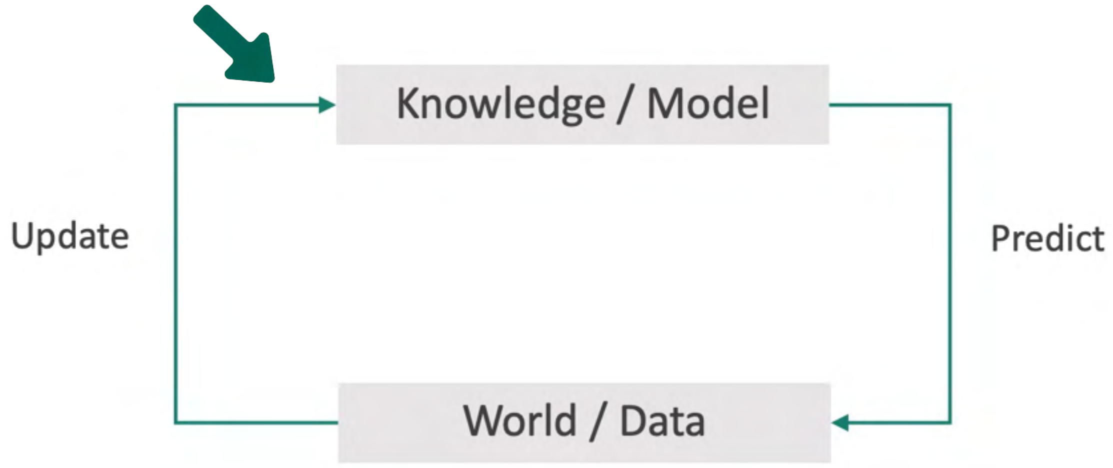




Posterior Distribution

$p(\theta \mid x)$

- Probability (density) function
- Quantifies uncertainty about a parameter within a specific model *after* the data collection



Posterior Distribution

$$p(\theta \mid x)$$

How to get from prior to posterior distribution

Bayes theorem: $$\underbrace{p(\theta \mid x)}_{\text{Knowledge } \textit{after} \text{ data collection}} = \underbrace{p(\theta)}_{\text{Knowledge before data collection}} \cdot \underbrace{\frac{p(x \mid \theta)}{p(x)}}_{\text{Updating factor}}$$

Posterior Distribution

$p(\theta \mid x)$

How to get from prior to posterior distribution

Likelihood

Bayes theorem: $p(\theta \mid x) = p(\theta) \cdot \dfrac{p(x \mid \theta)}{p(x)}$

Posterior Distribution

Prior Distribution

Marginal Likelihood

Posterior Distribution

$p(\theta \mid x)$

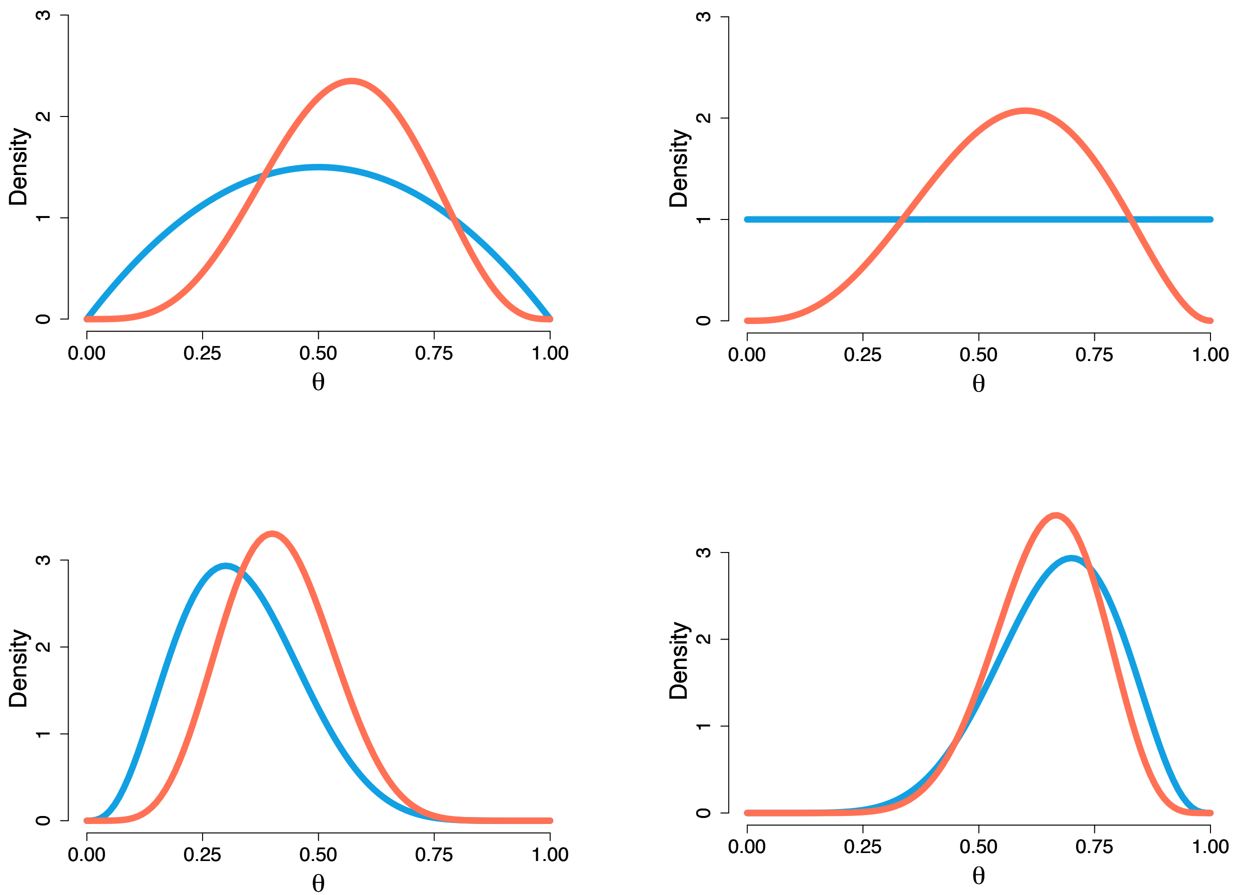
Prior and posterior distribution

```
posterior <- function(theta){
  dbinom(sum(dogpeople),length(dogpeople),theta) * dbeta(theta,shape1,shape2) / ML
}

curve(posterior(x),
      bty="l", xlab = bquote(theta), ylab = "Density", cex.lab = 1.5)
```

Posterior Distribution

$p(\theta \mid x)$

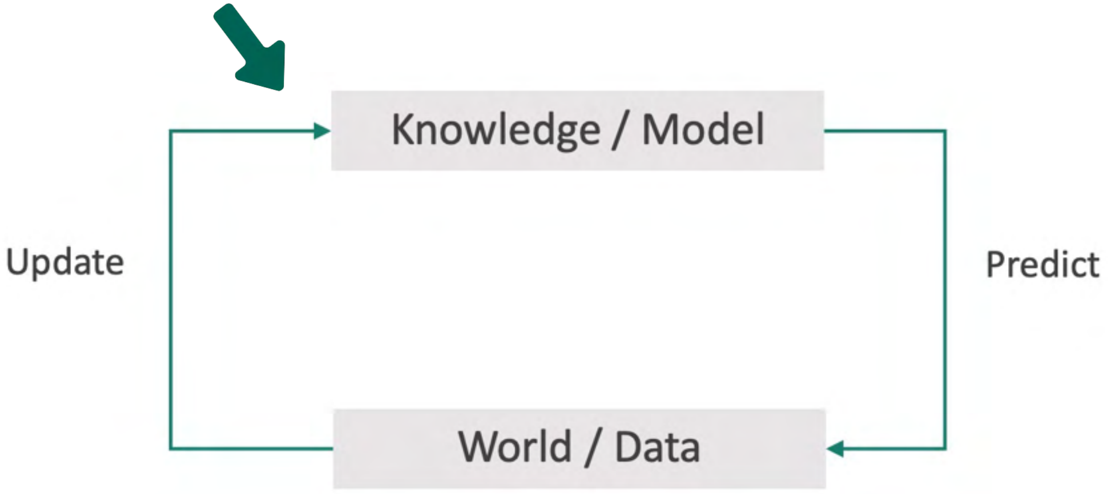
Prior and posterior distribution



(Data: 3 dog people, 2 cat people, blue: prior, red: posterior)

Credible Interval

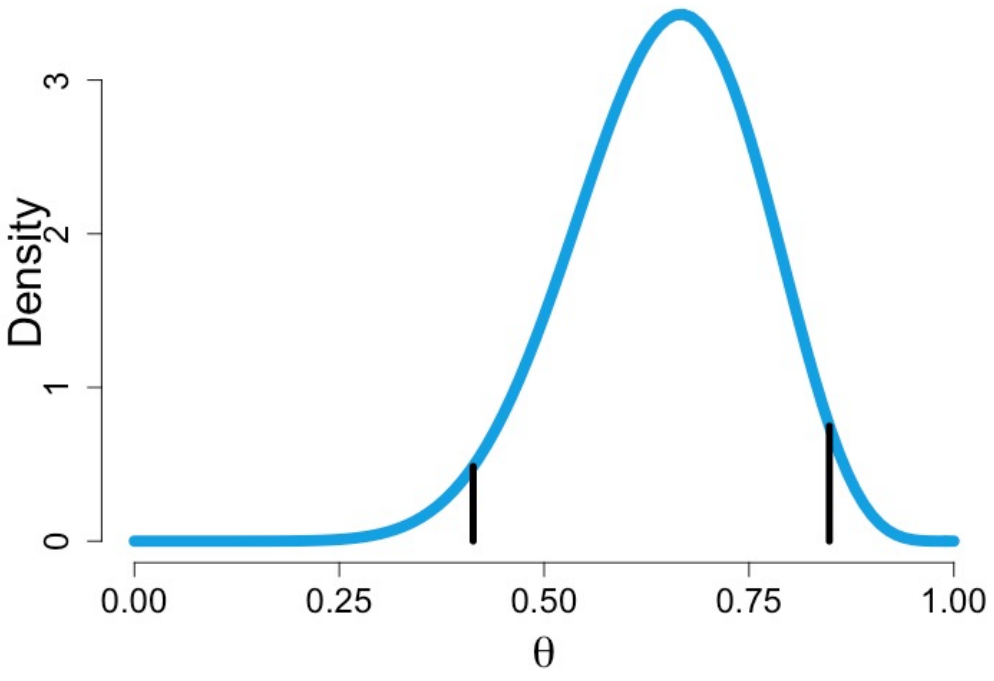
- With a probability of x%, the parameter lies within this interval
- Defined based on the posterior distribution



Credible Interval

Computation of the credible interval

Central Credible Interval

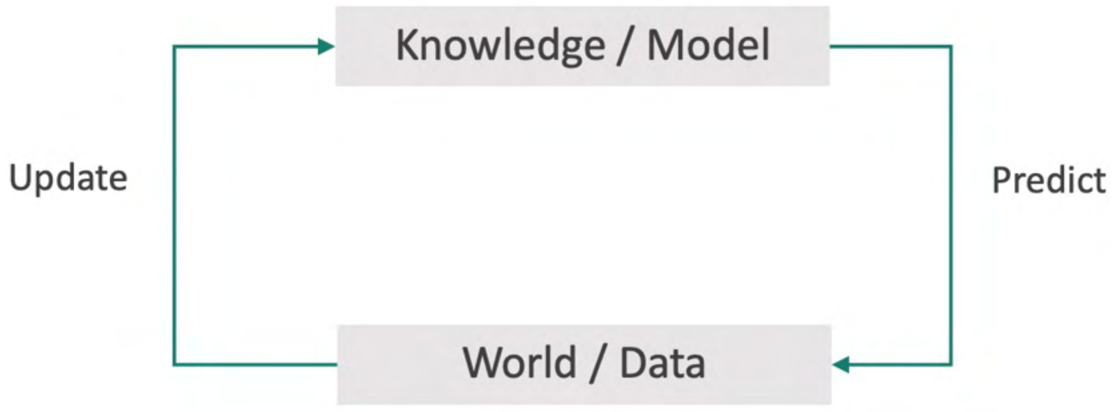


The 95% CI contains the central 95% of the posterior distribution

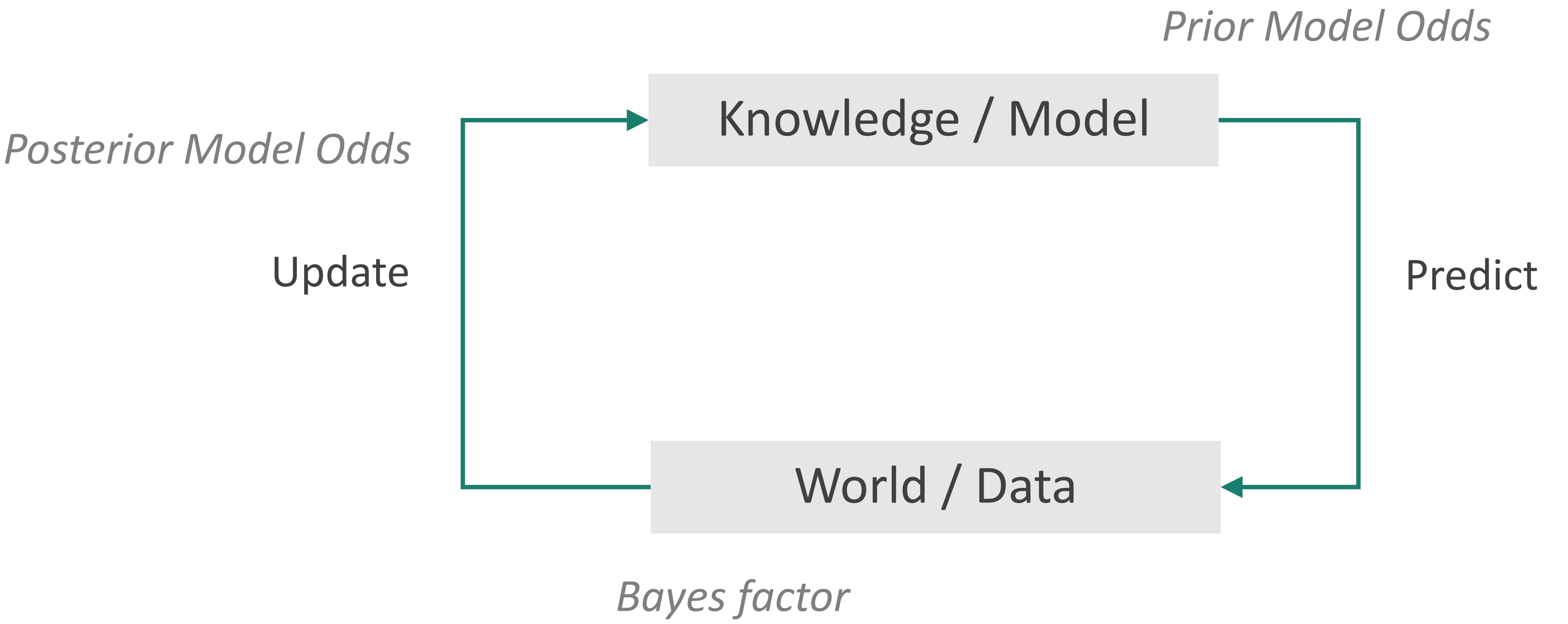
Bayesian statistics is about updating knowledge

Parameter Estimation

Hypothesis Testing / Model Comparison



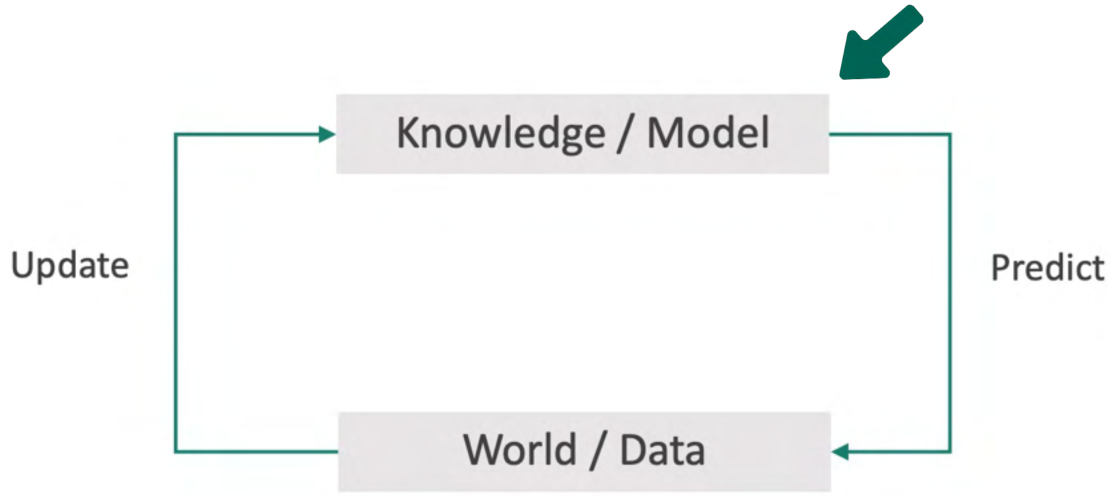
Bayesian statistics is about updating knowledge



Prior Model Odds

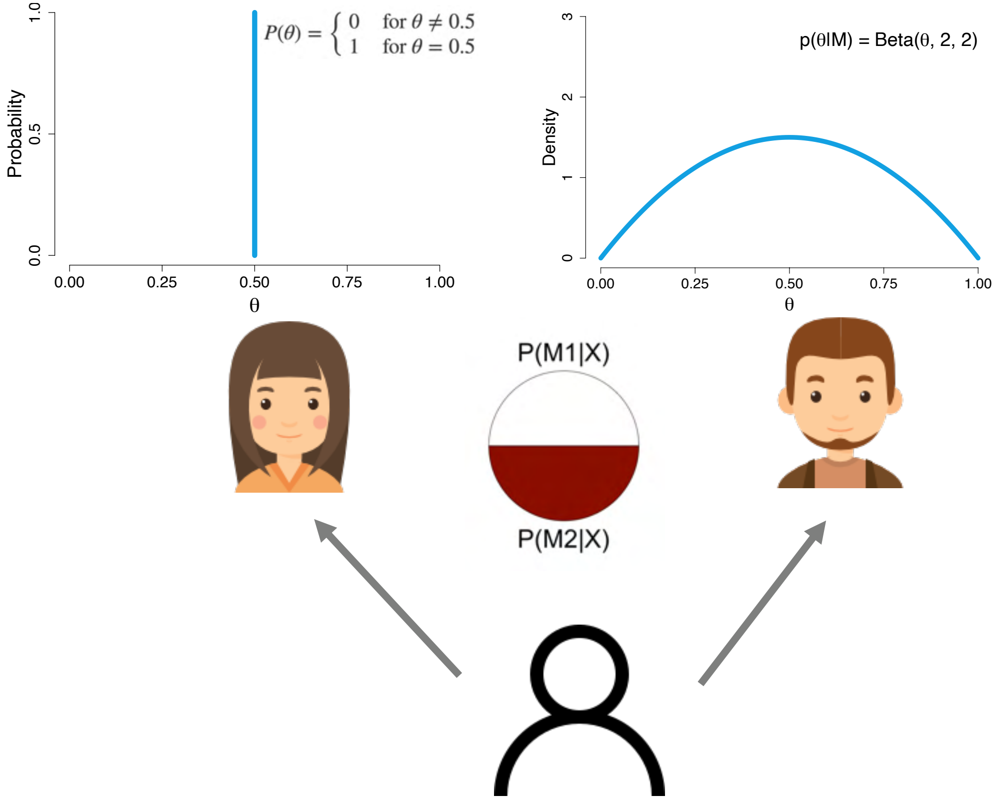
$$p(M_1)/\,p(M_2)$$

- Compares the relative plausibility of two models *before* data collection

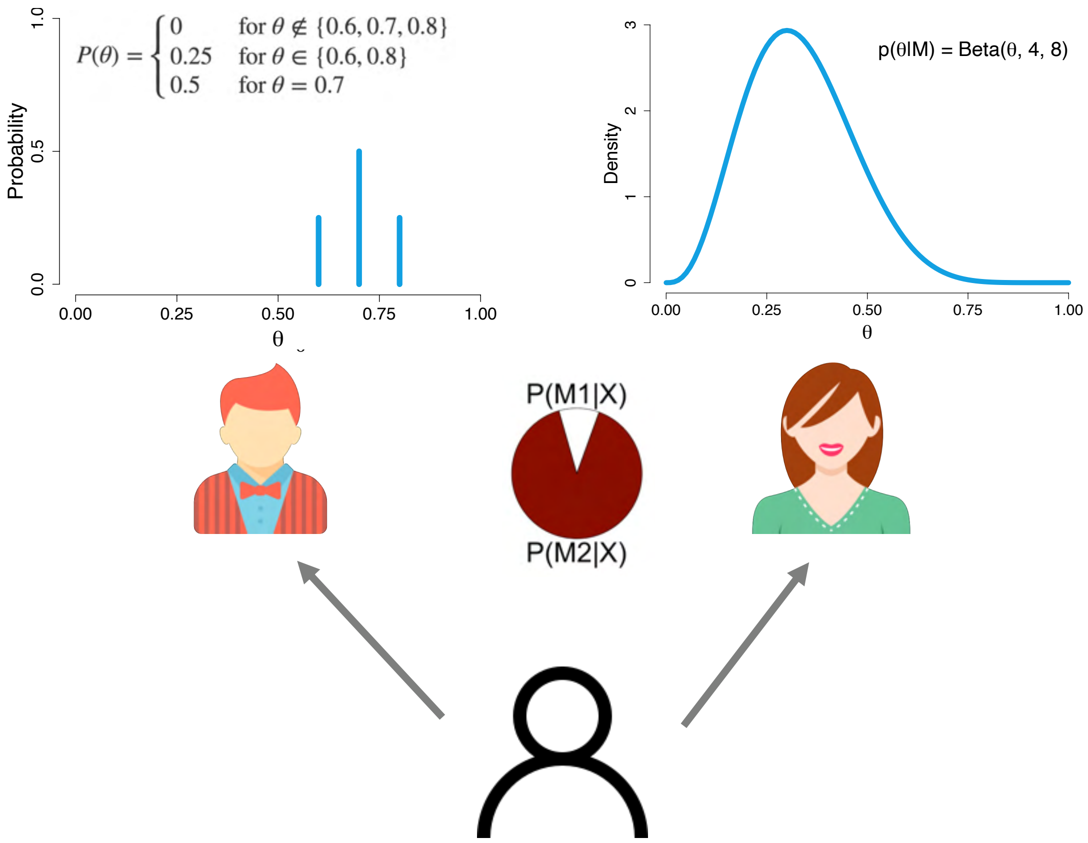


Prior Model Odds

$$p(M_1)/\,p(M_2)$$



"p(M_1):p(M_2) = 1"



"p(M_1):p(M_2) = 1:9"

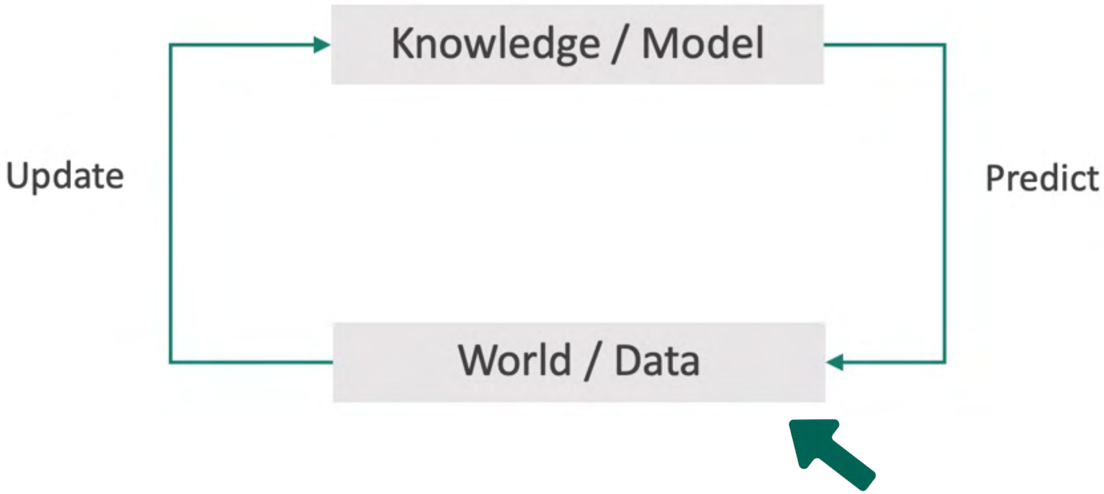
Yay, data!

Our fictional data:

In a random sample of N = 5
we observe:

x = 3 Dog people
5-x = 2 Cat people

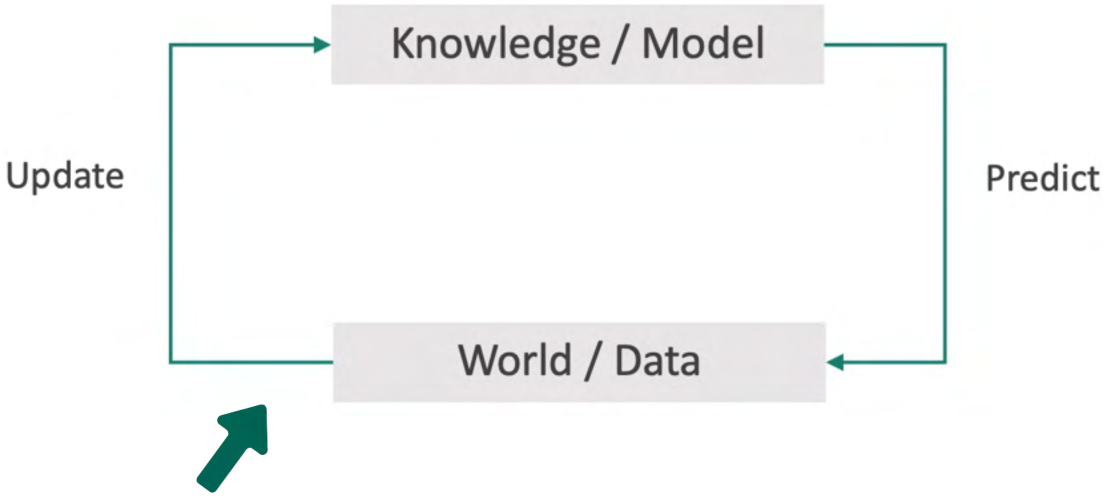
vs.



Bayes factor

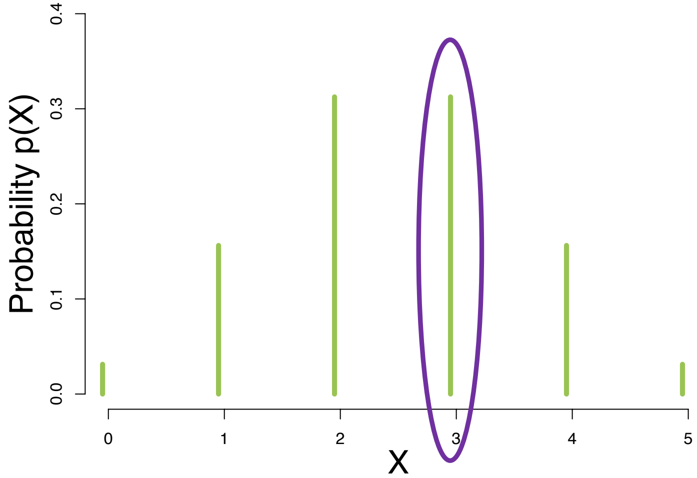
$$p(x|M_1)/\ p(x|M_2)$$

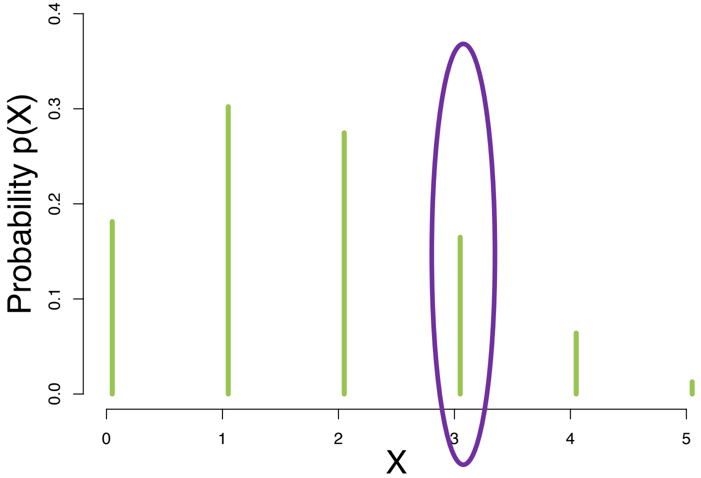
- Tells you how much more likely the observed data are under one model than under another model
- Can be interpreted as degree of relative evidence for a model

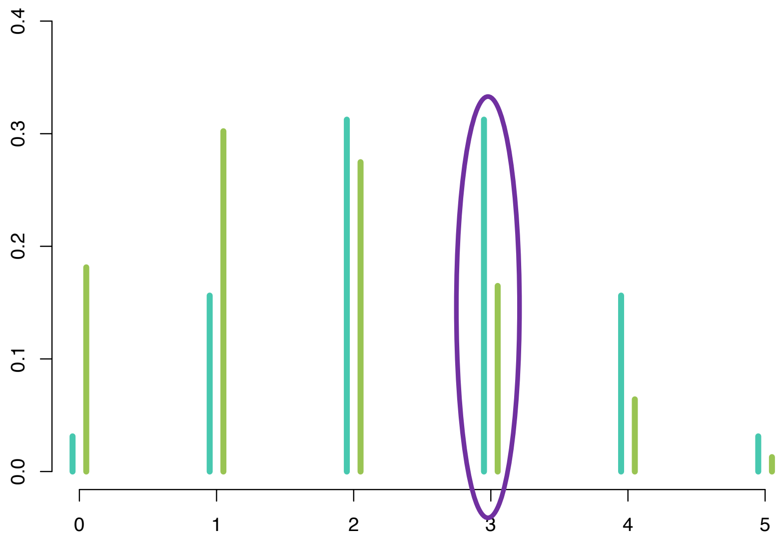


Bayes factor

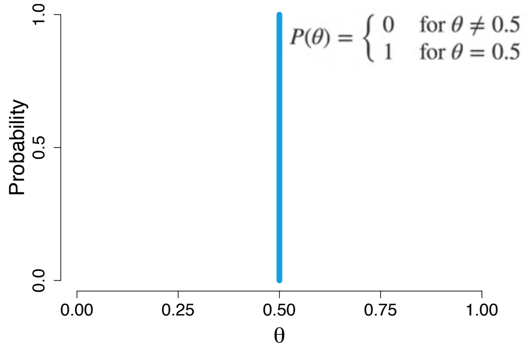
$$p(x|M_1)/\ p(x|M_2)$$

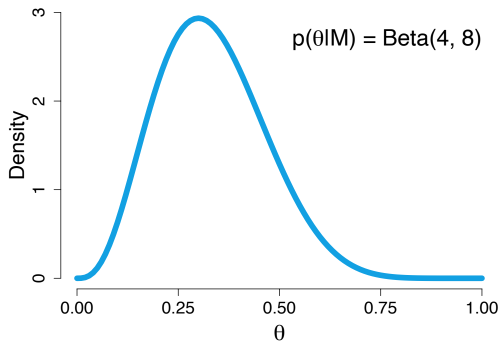






Prior predictive distributions for the binomial models with the following prior distributions on θ





$p(x=3|M_1) = 0.3125$

$p(x=3|M_2) = 0.1648$

$BF_{12} = 1.896$

Bayes factor

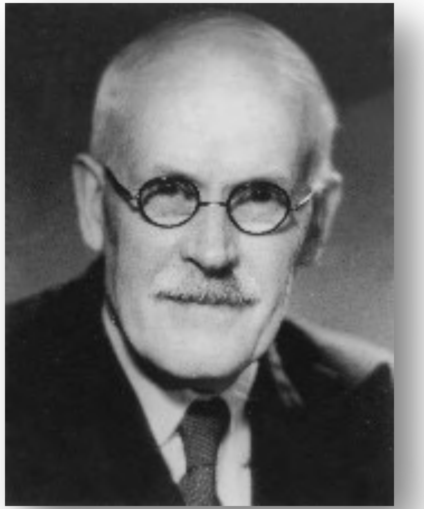
$$p(x|M_1) / \, p(x|M_2)$$

```
# Compute the marginal likelihoods (ML_1 was computed earlier)
ML_1 <- ML
ML_2 <- dbinom(sum(dogpeople), length(dogpeople), prob = 0.5)

# Compute the Bayes factor
BF12 <- ML_1 / ML_2
```

Bayes factor

$$p(x|M_1)/\ p(x|M_2)$$

Bayes factor interpretation



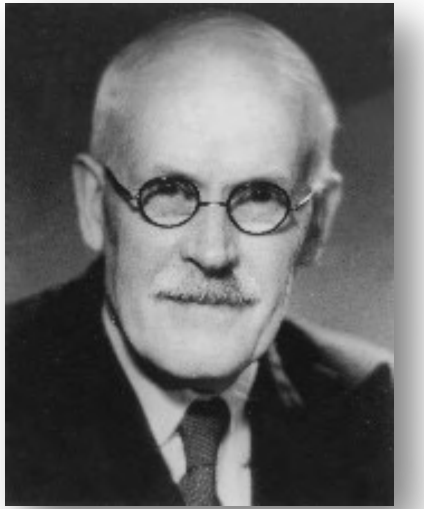
$BF_{12} > 1$: Evidence in favor of model 1
$BF_{12} = 1$: No evidence
$BF_{12} < 1$: Evidence in favor of model 2

Bayes factor

$$p(x|M_1)/\ p(x|M_2)$$

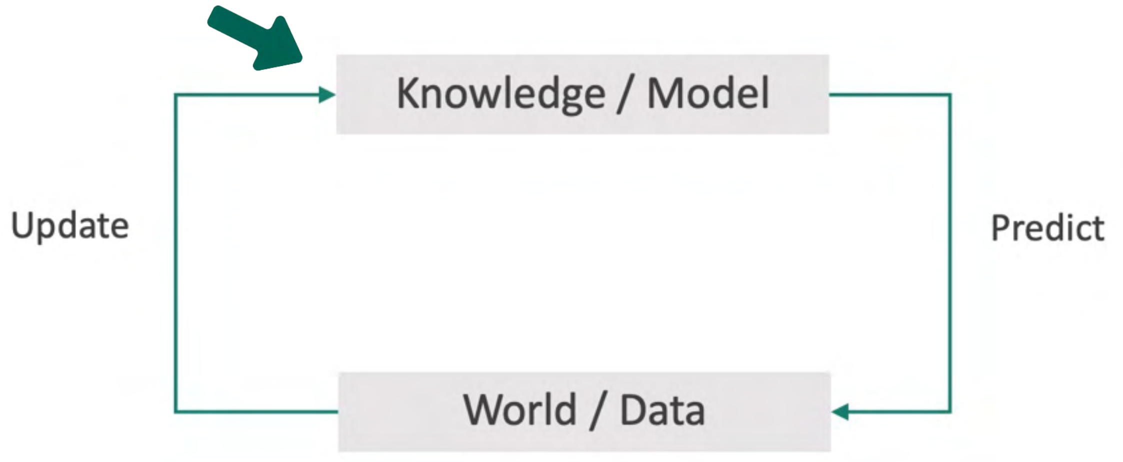
What is a convincing Bayes factor?

Bayes factor	Evidence category
> 100	Extreme evidence for $\mathcal{H}_1$
30 - 100	Very strong evidence for $\mathcal{H}_1$
10 - 30	Strong evidence for $\mathcal{H}_1$
3 - 10	Moderate evidence for $\mathcal{H}_1$
1 - 3	Anecdotal evidence for $\mathcal{H}_1$
1	No evidence
1/3 - 1	Anecdotal evidence for $\mathcal{H}_0$
1/10 - 1/3	Moderate evidence for $\mathcal{H}_0$
1/30 - 1/10	Strong evidence for $\mathcal{H}_0$
1/100 - 1/30	Very strong evidence for $\mathcal{H}_0$
< 1/100	Extreme evidence for $\mathcal{H}_0$



Posterior Model Odds $p(M_1|x)/\ p(M_2|x)$

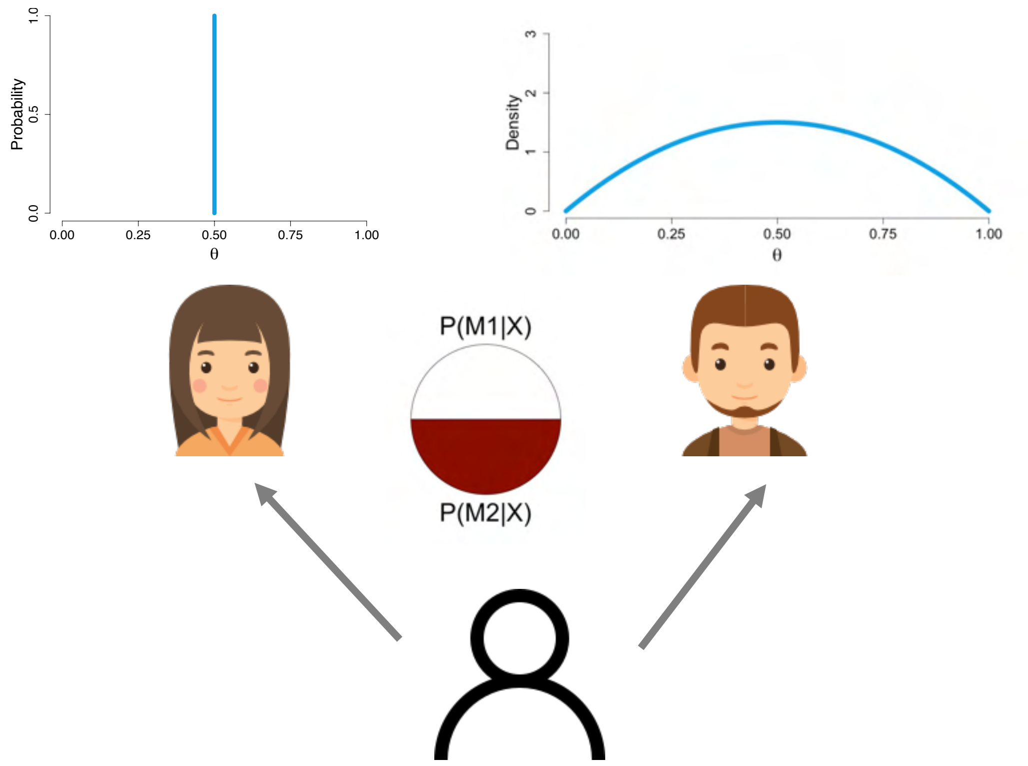
- Compares the relative plausibility of two models *after* data collection



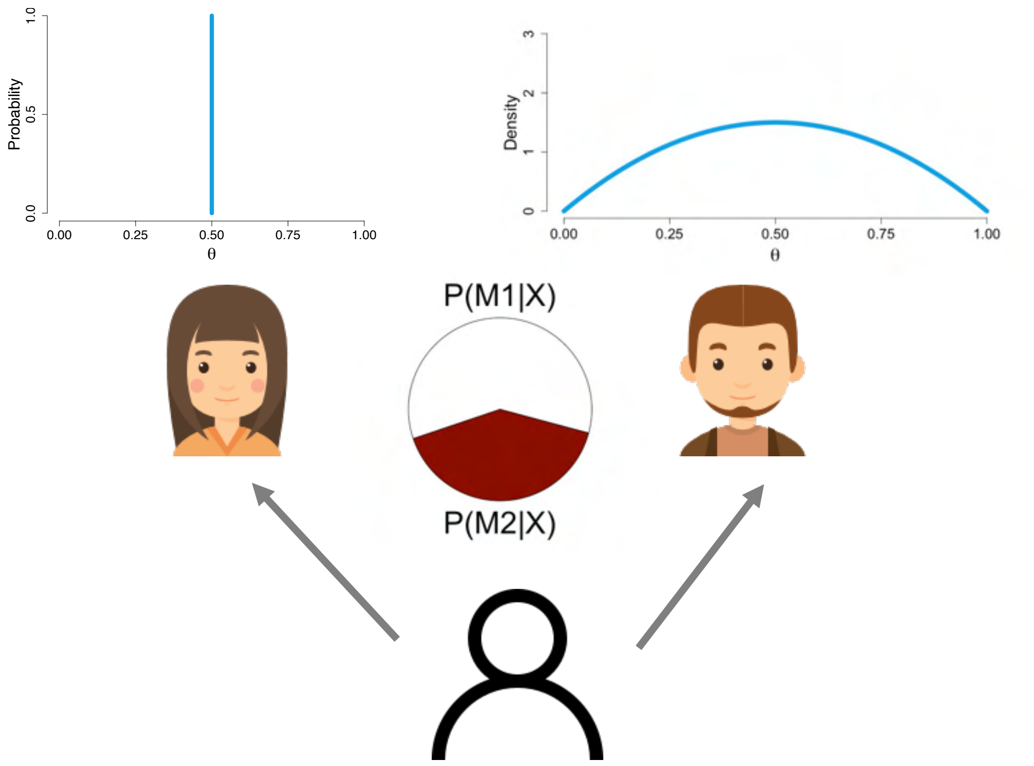
Posterior Model Odds

$$p(M_1|x)/\ p(M_2|x)$$

Posterior Model Odds = Prior odds x BF



$P(M_1):P(M_2) = 1:1$



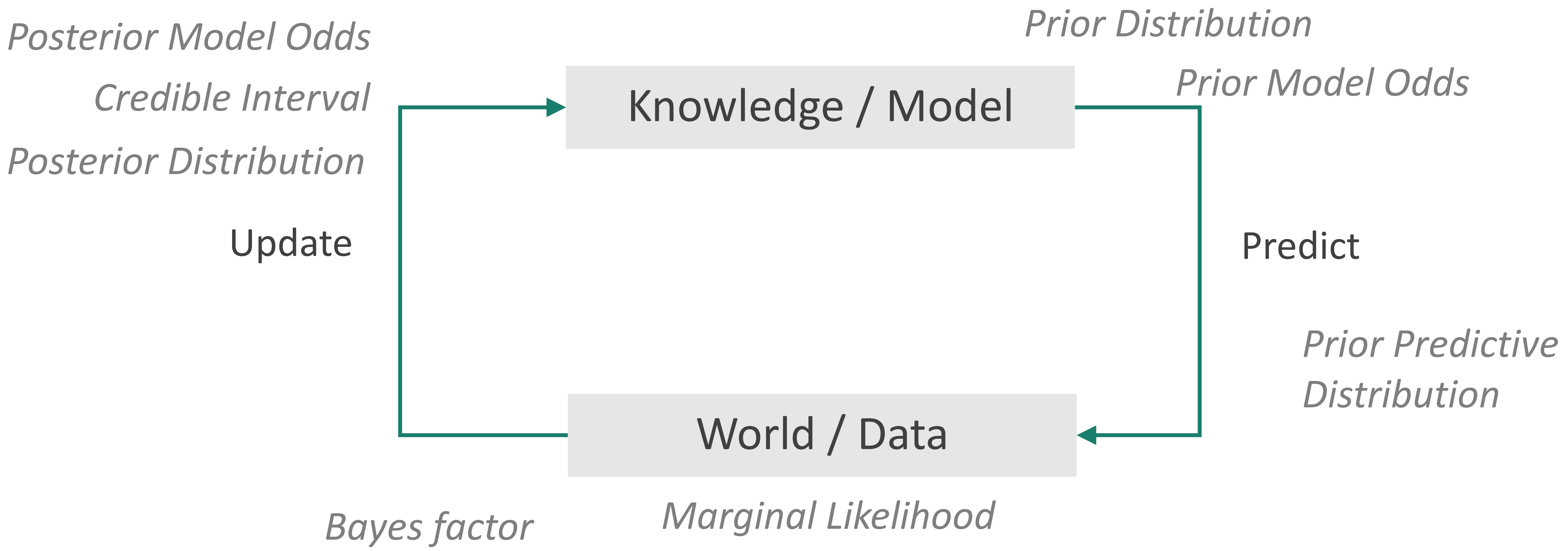
$P(M_1|X):P(M_2|X) = 1:1 * 1.46 = 1.46:1$

Posterior Model Odds

$$p(M_1|x)/\,p(M_2|x)$$

```
# First, define your prior model odds
prior_prob_M1 <- 0.5
prior_prob_M2 <- 0.5
prior_model_odds <- prior_prob_M1 / prior_prob_M2

# Then, update with the Bayes factor
posterior_model_odds <- prior_model_odds * BF12
```

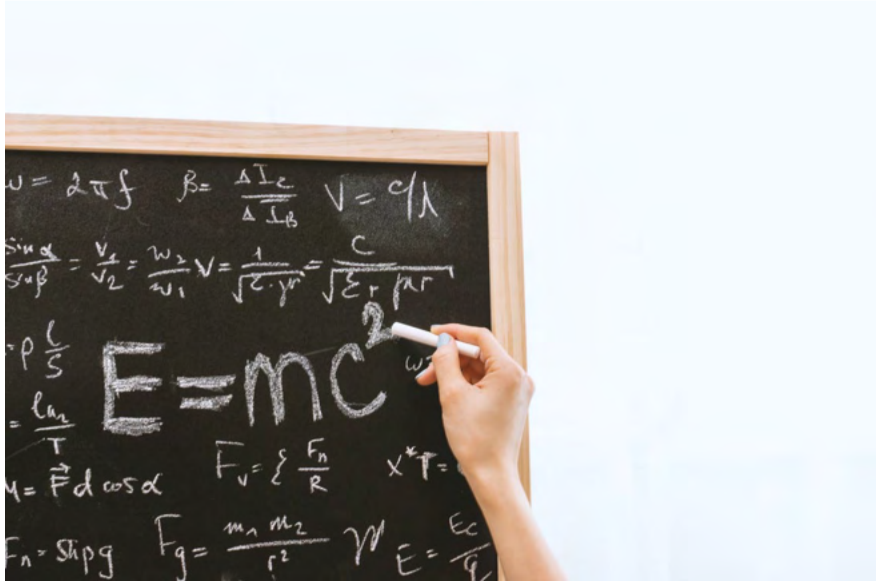
Bayesian statistics is about updating knowledge



But what about more complex cases?

This workshop

My problem



But what about more complex cases?

The good news

- Bayesian statistics can be applied to extremely complex problems (e.g., hierarchical data, many parameters, nonlinear relationships, …)
- Even highly complex modeling uses the same basic concepts (prior distribution, posterior distribution, Bayes factor, credible interval, …)

The bad news

- Complex modeling requires computational solutions to obtain posteriors (think: multi-dimensional integrals)
- These solutions can be computationally intensive

Sounds difficult? Help is near!

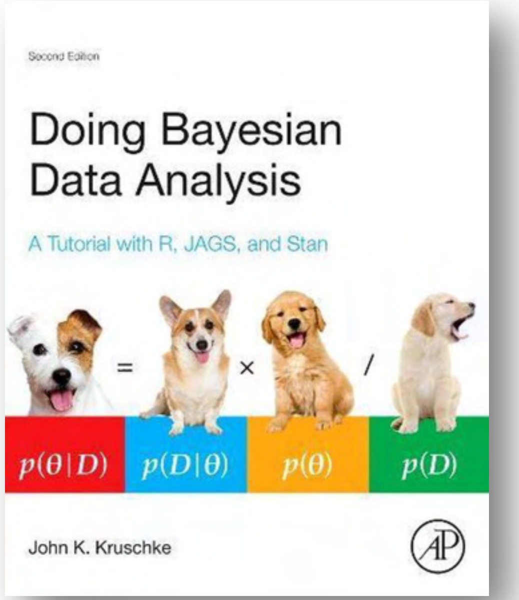
R Packages

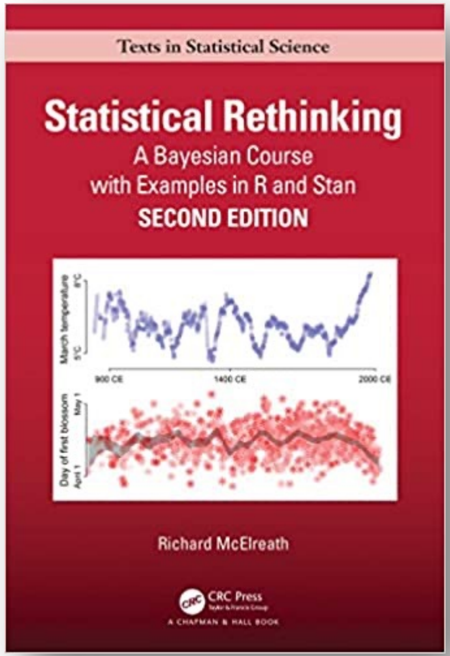
```r
library(BayesFactor)

library(brms)

library(BAS)

library(rstan)
```

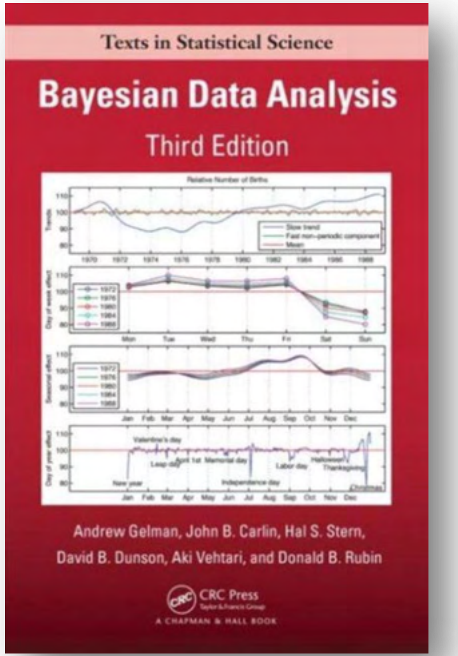
```r
library(bridgesampling)

library(posterior)

library(tidybayes)

library(bayesplot)

library(bayestestR)
```

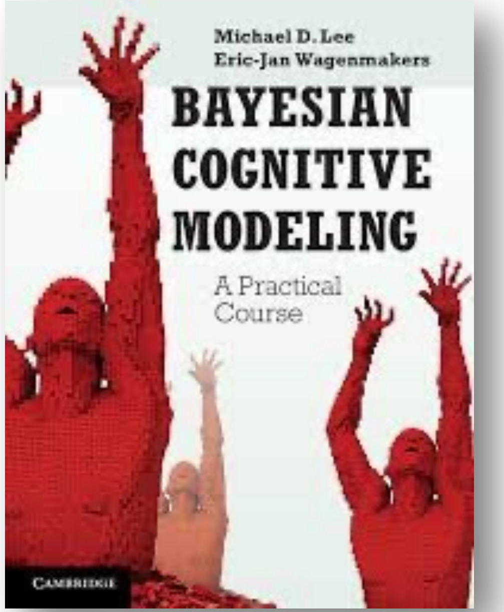
Sounds difficult? Help is near!

Textbooks









The theory of probabilities is basically just common sense reduced to calculus; it makes one appreciate with exactness that which accurate minds feel with a sort of instinct, often without being able to account for it.

Laplace, 1829

Questions, Comments, Issues

E-Mail: a.m.stefan@uva.nl
GitHub: astefan1
Twitter: @ephemeralidea