What is Bayesian Statistics?

Alireza Akhondi-Asl

MSICU Center For Outcomes

Department of Anesthesiology, Critical Care and Pain Medicine

Learning Objectives

- What is Bayes' Rule?
- Mechanism of belief update in Bayesian statistics?
- What are the Differences with the frequentist statistics?

Probability Interpretations

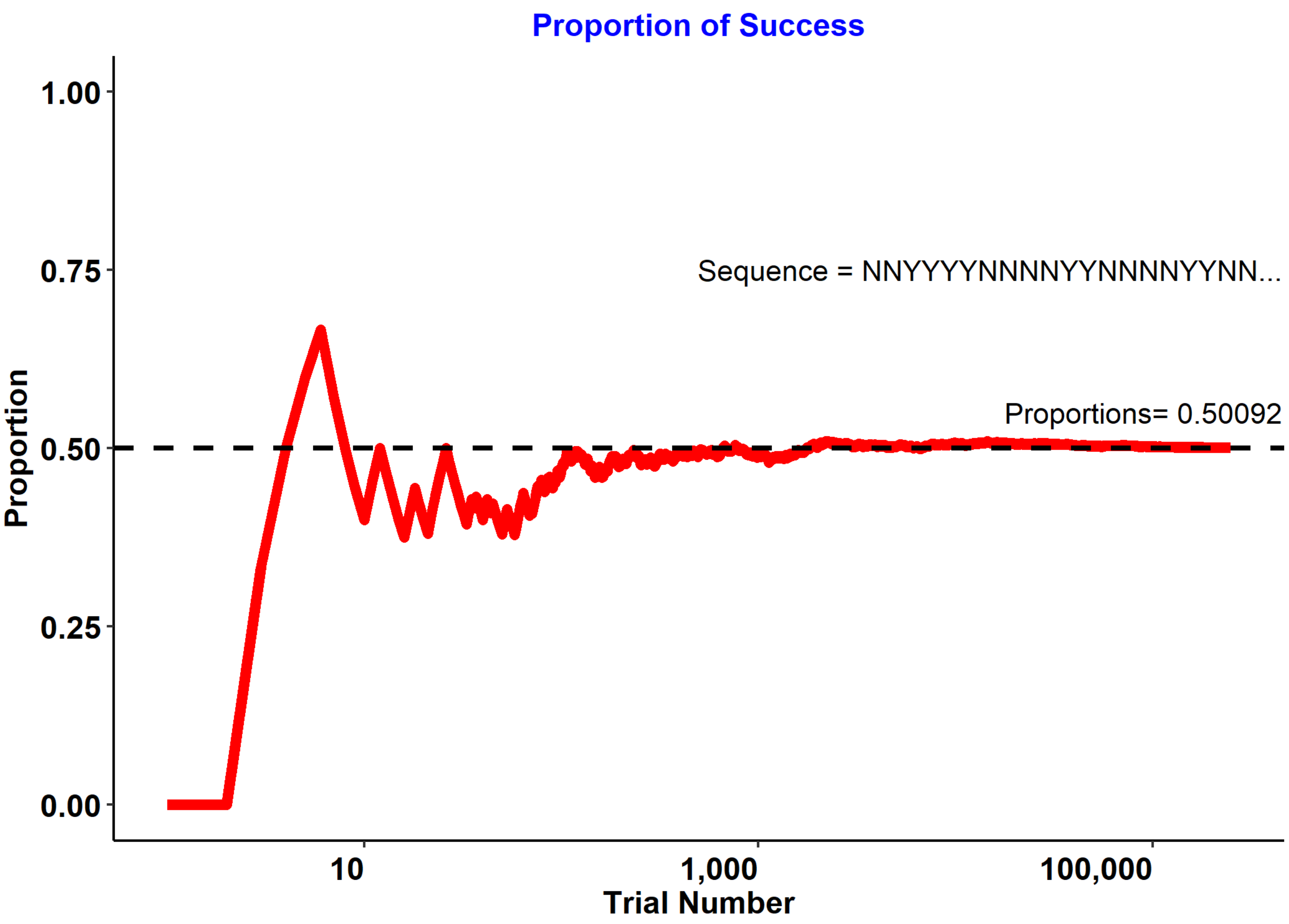
Frequentist

Subjective

Frequentist Probability

- Relative Frequency of an Event in long run:

$$P(e) = \lim_{n \to \infty} \frac{\#times\ e\ happend}{n}$$



Subjective Probability

It is inside the head probability

- How strongly do you believe that a patient is going to survive?
- The probability of the Democrats winning the 2024 US presidential election.

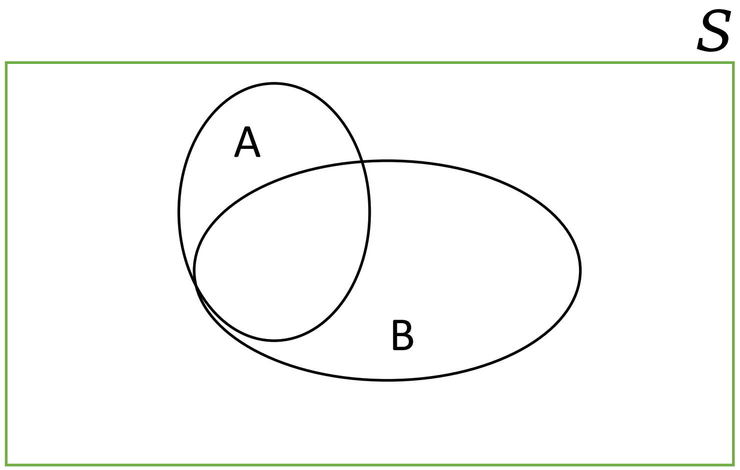
Sometimes it is hard to quantify our belief.

- Thinking about a fair bet.
- Comparing with other events with clear probabilities

We should be coherent.

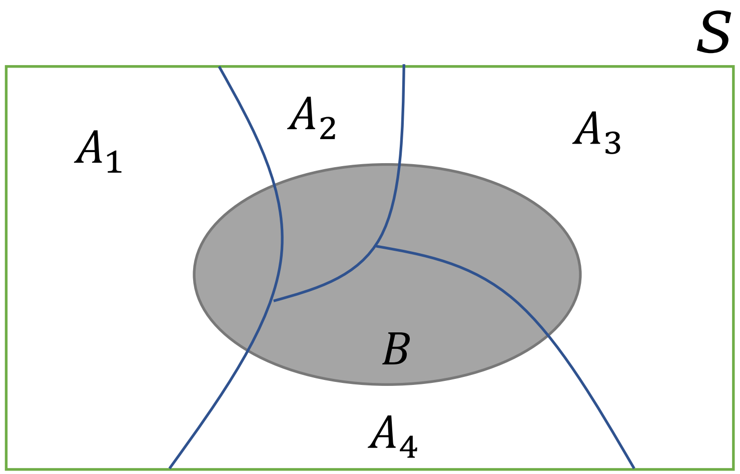
Conditional probability

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad \text{, with } P(A) > 0$$



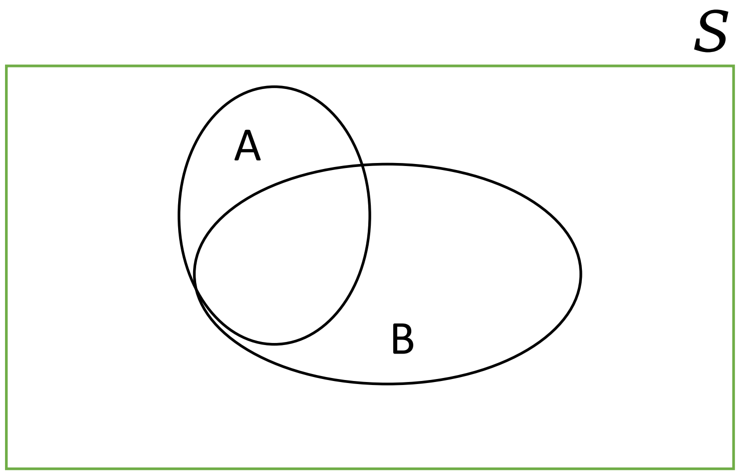
Law of total probability

If $A_1, A_2, \ldots, A_4$ is a partition of the sample space, then for any event B we have:



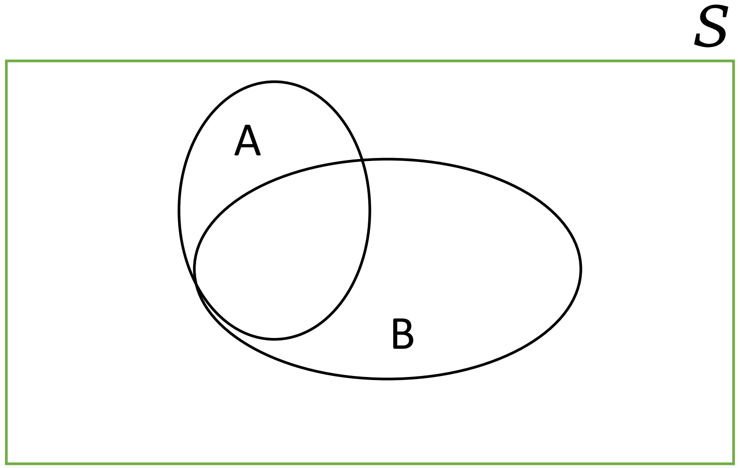
$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + P(B|A_4)P(A_4) \\ &= \sum_i P(B|A_i)P(A_i) \end{aligned}$$

Bayes' Rule



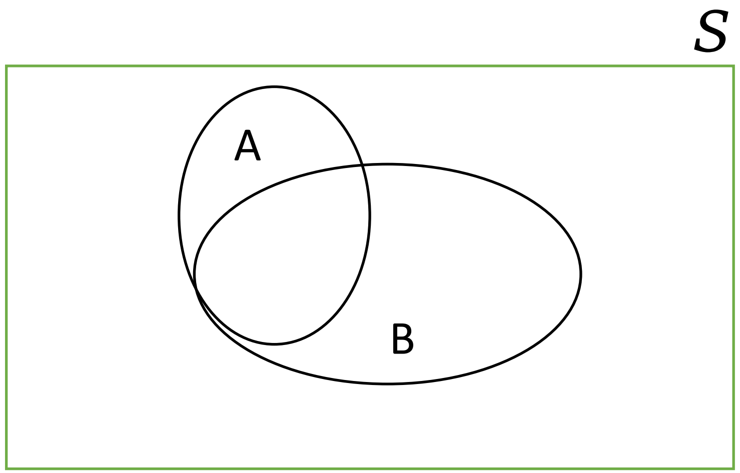
$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Bayes' Rule



$$P(A \cap B) = \boxed{P(A|B)P(B) = P(B|A)P(A)}$$

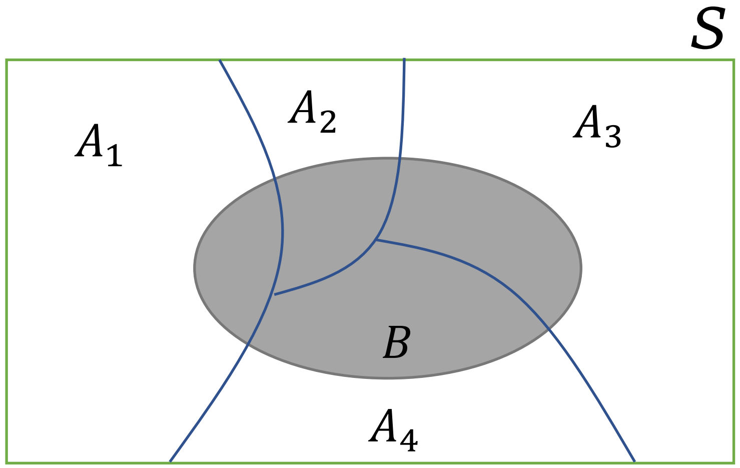
Bayes' Rule



$$P(A \cap B) = \boxed{P(A|B)P(B) = P(B|A)P(A)}$$

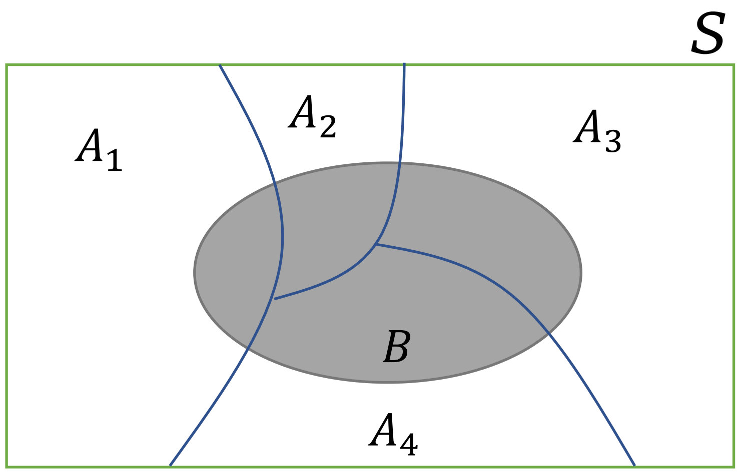
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' Rule



$$P(A_1),\ P(A_2),\ P(A_3),\ P(A_4)$$

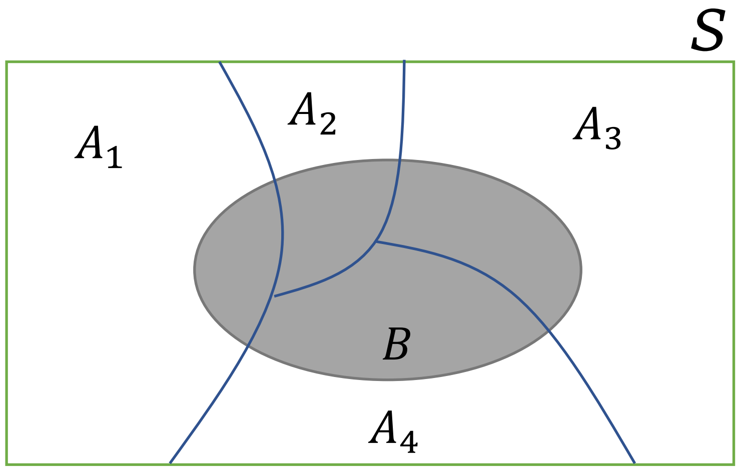
Bayes' Rule



$$P(A_1),\ P(A_2),\ P(A_3),\ P(A_4)$$

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

Bayes' Rule



$$P(A_1),\ P(A_2),\ P(A_3),\ P(A_4)$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

Medical Test

- A certain disease affects about 1 out of 1000 people in a population.
- P(☹)=0.001
- P(☺)=0.999
- There is a test to check whether the person has the disease. The test has very high sensitivity and specificity. In particular, we know that:
- P(T+|☹)=0.98
- P(T+|☺)=0.01

Medical Test

If you test positive for this disease, what are the chances that you have the disease?

A) 98 Percent

B) Less than 10 percent

Medical Test

$$P(☹|T+)=\frac{P(T+|☹)P(☹)}{P(T+)}$$

Medical Test

$$P(☹|T+)=\frac{P(T+|☹)P(☹)}{P(T+|☹)P(☹)+P(T+|☺)P(☺)}$$

Medical Test

$$P(☹|T+)=\frac{P(T+|☹)P(☹)}{P(T+|☹)P(☹)+P(T+|☺)P(☺)}=\frac{0.98\times0.001}{0.98\times0.001+0.01\times0.999}=0.089$$

Medical Test

$$P(☹|T+) = \frac{P(T+|☹)P(☹)}{P(T+|☹)P(☹)+P(T+|☺)P(☺)} = \frac{0.98\times0.001}{0.98\times0.001+0.01\times0.999} = 0.089$$

The test **updates your chances** of having the diseases from 0.001 to 0.089.

Medical Test

If you test positive for this disease, what are the chances that you have the disease?

A) 98 Percent

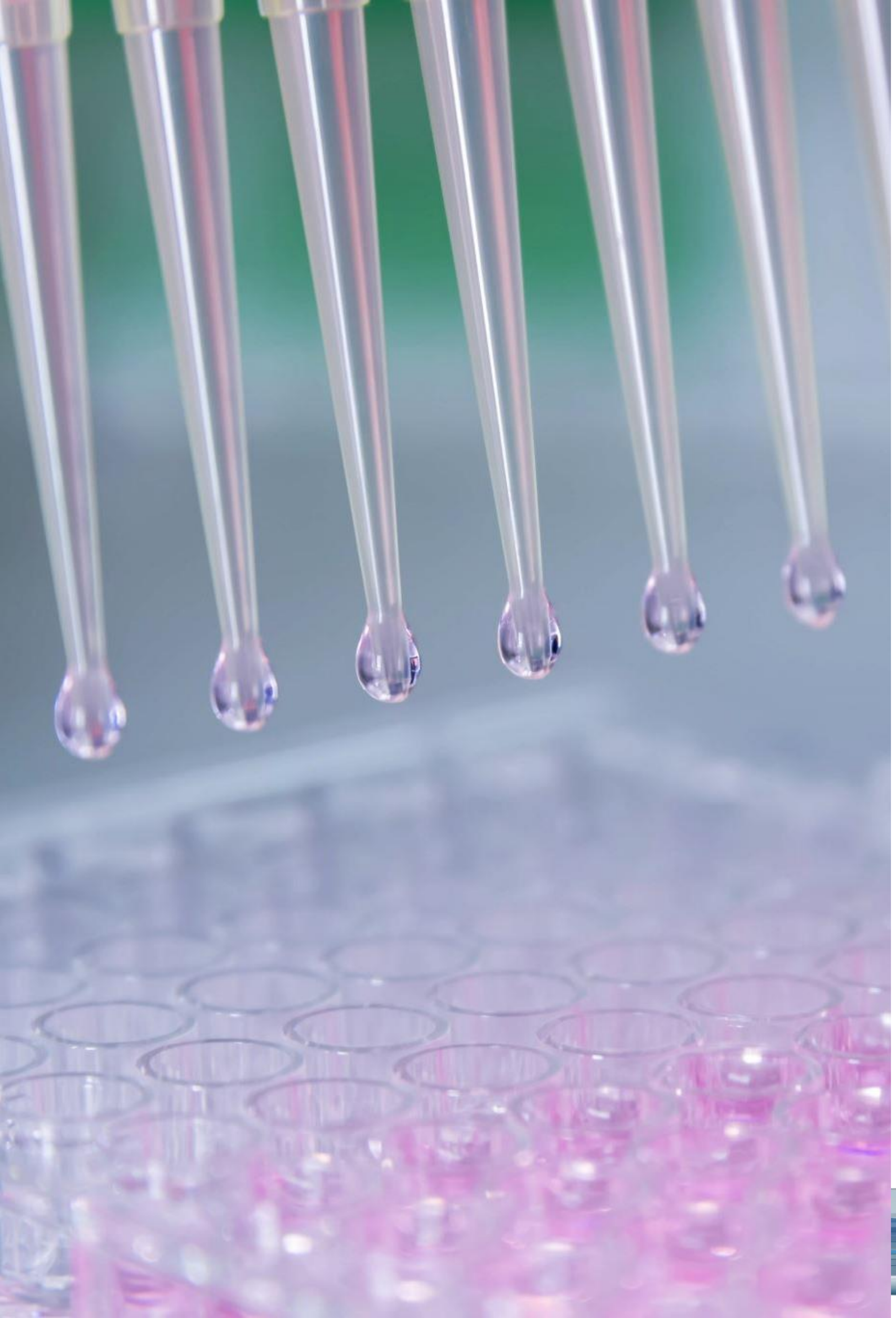
✓ B) Less than 10 percent

Richard Royall's Three Questions

What does present evidence tell?

What should we believe?

What should we do?



Medical Test Paradox

- **A second independent test with the same accuracy is done and it is positive again. What are the chances that you have the disease?**
  - A) More than 90 percent
  - B) Less than 10 percent

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

Medical Test Paradox

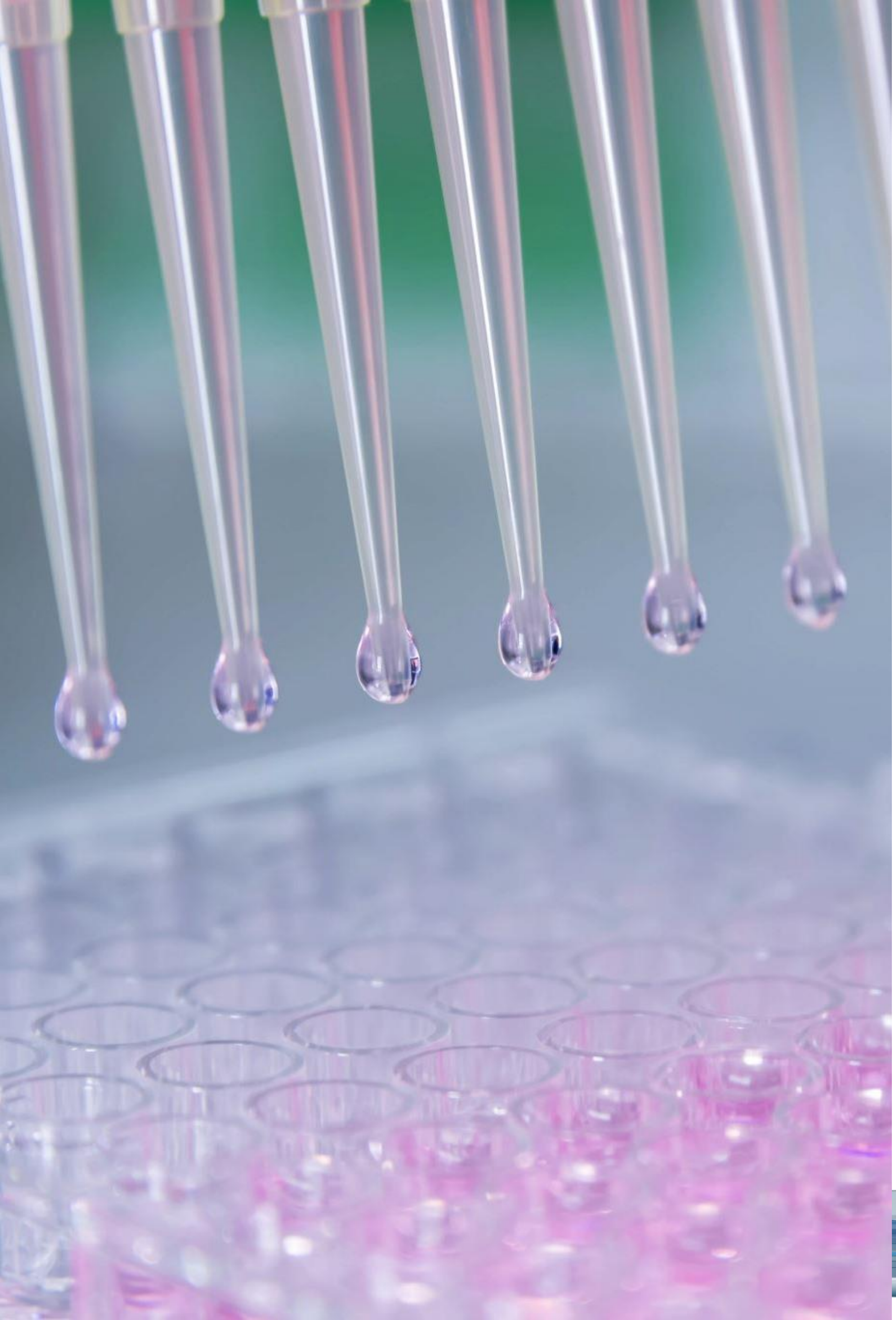
- $P(☹|T+)=\frac{P(T+|☹)P(☹)}{P(T+|☹)P(☹)+P(T+|☺)P(☺)}$
- $P(☹)=$?
- $P(☺)=$?

Medical Test Paradox

- P(☹|T+)=$\frac{\text{P(T+|☹)P(☹)}}{\text{P(T+|☹)P(☹)+P(T+|☺)P(☺)}}$
- P(☹)= 0.089
- P(☺)= 0.911

Medical Test Paradox

- $P(☹|T+)=\frac{P(T+|☹)P(☹)}{P(T+|☹)P(☹)+P(T+|☺)P(☺)}=\frac{0.98\times0.089}{0.98\times0.089+0.01\times0.911}=0.906$
- $P(☹)=0.089$
- $P(☺)=0.911$



Medical Test Paradox

- **A second independent test with the same accuracy is done and it is positive again. What are the chances that you have the disease?**
  - A) More than 90 percent
  - B) Less than 10 percent

Statistical Analysis

Frequentist

Bayesian

Likelihoodist

Frequentist

- The most popular method for statistical inference
- Parameters are fixed but unknown constants
  - We cannot make any probability statement about the parameters
- Probabilities are long-run relative frequencies from the repeated experiments
- Data is assumed to be random
  - Randomness is due to sampling from a fixed population.
  - The uncertainty is due to sampling variation.

Frequentist

- $P(Data|\theta)$
- Maximum Likelihood Estimation
- P-values : $P(Data|\theta = \theta_0)$
- Confidence Intervals, Effect Size
- No probability statement about θ

Bayesian

Probability is interpreted as "degree of subjective belief".	• The events do not need to be repeatable.
We don't know the value of parameters and therefore, we consider them to be random variables.	• Epistemic uncertainty. • Parameters are probabilistic in nature.
Since we have observed data, it is fixed.	
We Update our prior belief based on observed data. The updated belief is called posterior belief	• We use Bayes' rule to calculate posterior.

Bayesian

- Update our belief in a parameter using new evidence or data.
  - Based on Bayes' rule

$$P(\theta|Data) = \frac{P(Data|\theta)P(\theta)}{P(Data)}$$

Bayesian

- Update our belief in a parameter using new evidence or data.
  - Based on Bayes' rule

$$P(\theta|Data) = \frac{P(Data|\theta)\overset{\textbf{Prior}}{\boxed{P(\theta)}}}{P(Data)}$$

Prior

Bayesian

- Update our belief in a parameter using new evidence or data.
  - Based on Bayes’ rule

$$P(\theta|Data) = \frac{\overbrace{\boxed{P(Data|\theta)}}^{\text{Likelihood}}\overbrace{\boxed{P(\theta)}}^{\text{Prior}}}{P(Data)}$$

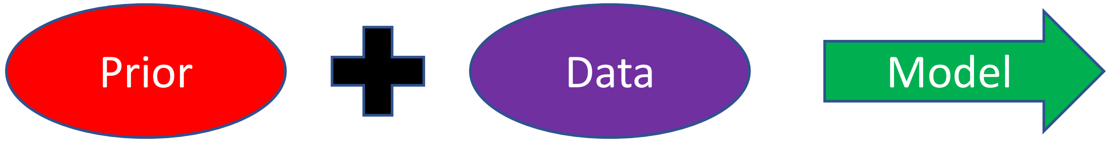
Prior

Model

Bayesian

- Update our belief in a parameter using new evidence or data.
    - Based on Bayes' rule

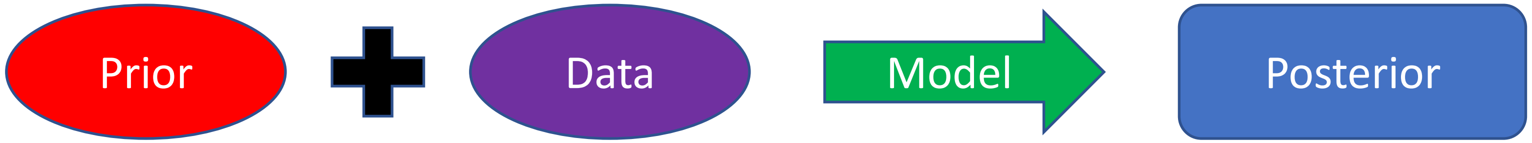
$$P(\theta|Data) = \frac{\overbrace{\boxed{P(Data|\theta)}}^{\text{Likelihood}}\overbrace{\boxed{P(\theta)}}^{\text{Prior}}}{\underbrace{\boxed{P(Data)}}_{\text{Evidence}}}$$



Bayesian

- Update our belief in a parameter using new evidence or data.
  - Based on Bayes' rule

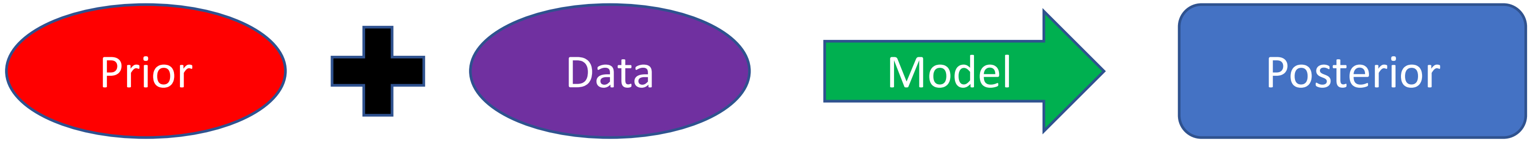
$$\overset{\text{Posterior}}{\boxed{P(\theta|Data)}} = \frac{\overset{\text{Likelihood}}{\boxed{P(Data|\theta)}}\overset{\text{Prior}}{\boxed{P(\theta)}}}{\underset{\text{Evidence}}{\boxed{P(Data)}}}$$



Bayesian

- Update our belief in a parameter using new evidence or data.
  - Based on Bayes' rule

$$\overset{\text{Posterior}}{\boxed{P(\theta|Data)}} \propto \overset{\text{Likelihood}}{\boxed{P(Data|\theta)}}\overset{\text{Prior}}{\boxed{P(\theta)}}$$



Posterior Distribution

Summarizes everything we know

Model Comparison

Prediction

- Posterior predictive distribution

Hypothesis Testing

- Region of practical equivalence

Example

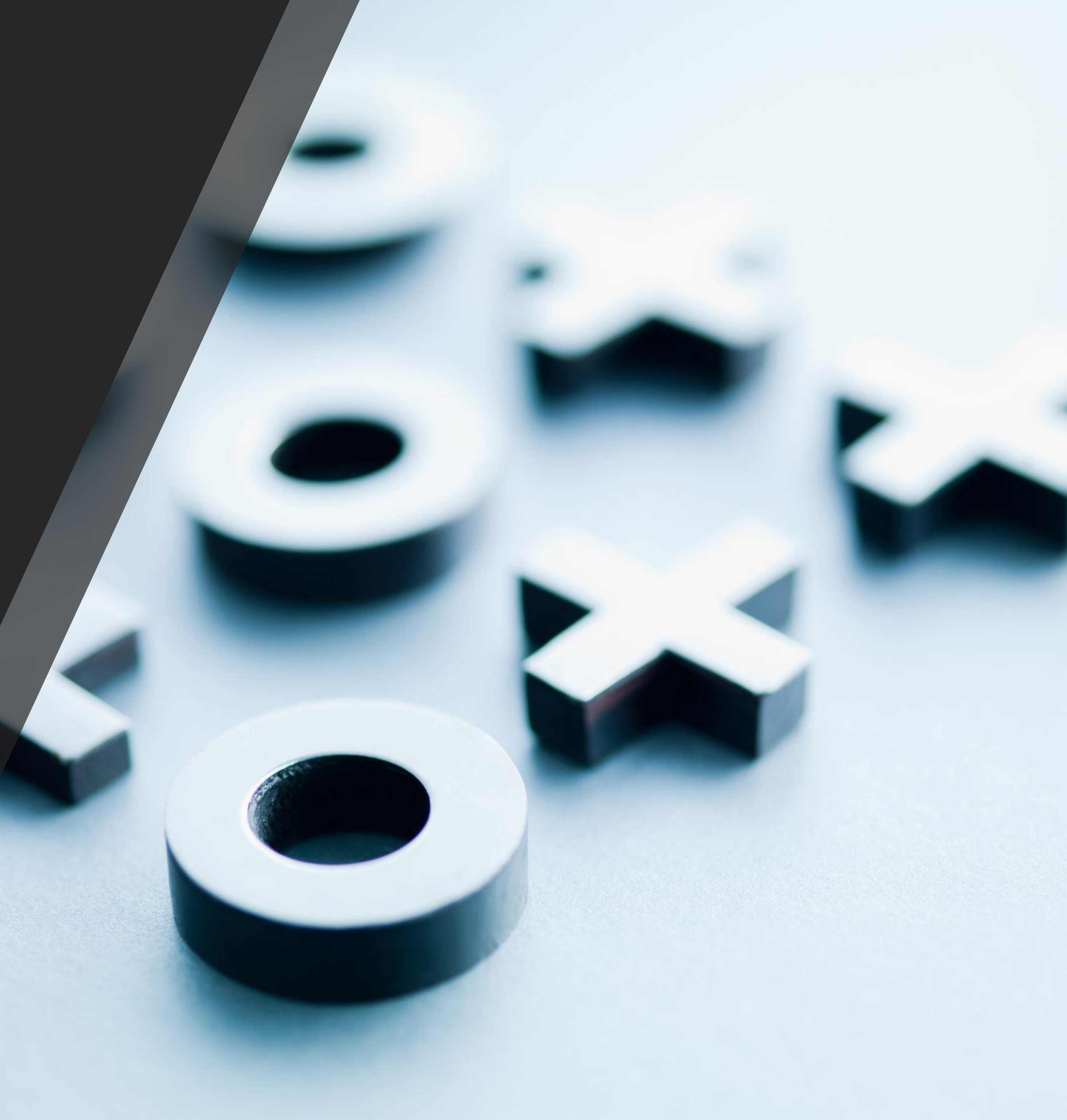
A new treatment approach is proposed. We would like to infer about the success rate of this treatment.

We observe results of treatment of N patients.

Likelihood

- Since the outcome is binary and samples are independent, for a fixed number of trials, N, we can use binomial distribution to describe our data generation model:

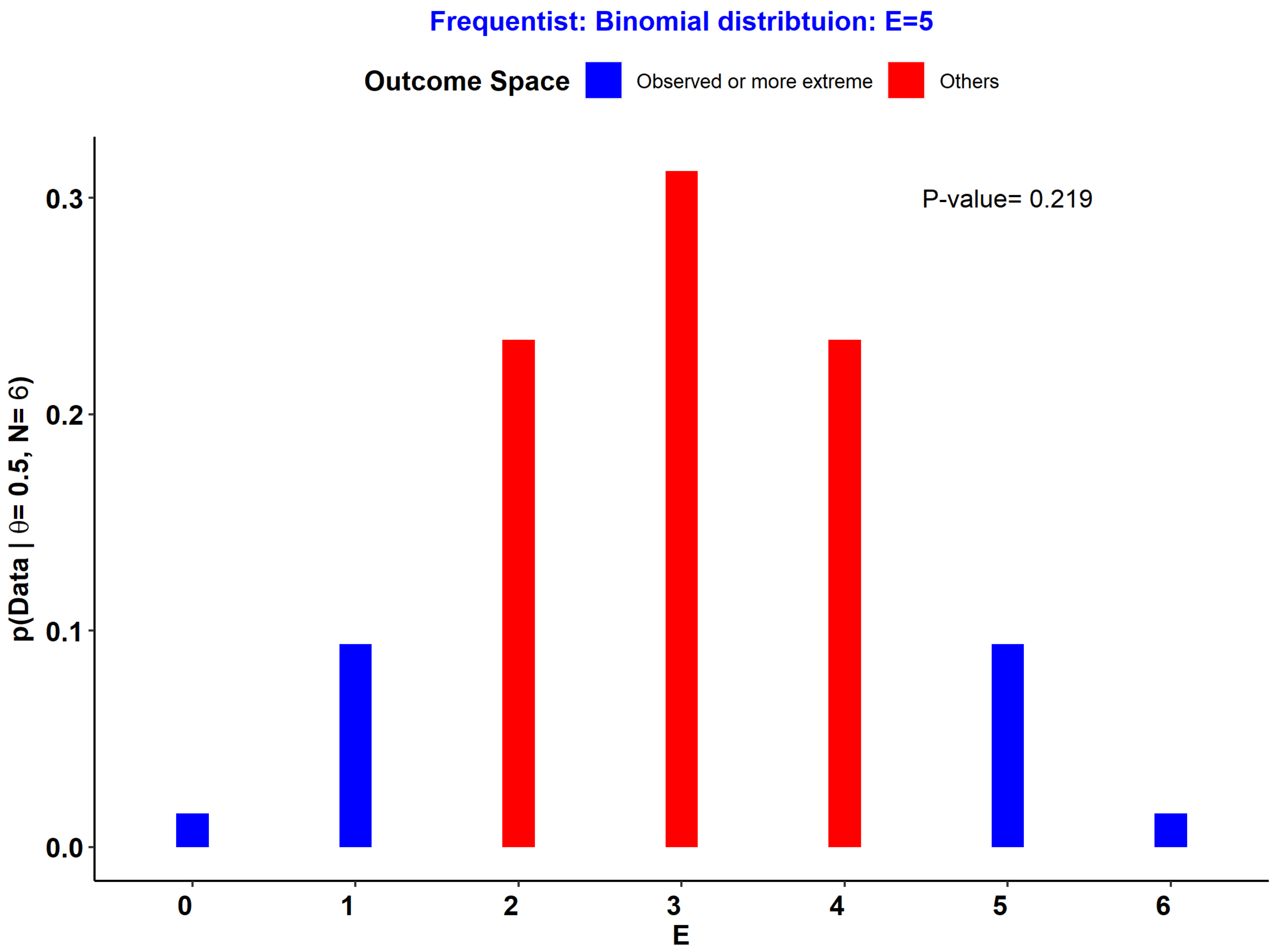
$$p(Data|\theta) = p(E|N,\theta) = \binom{N}{E}\theta^E(1-\theta)^{N-E}$$



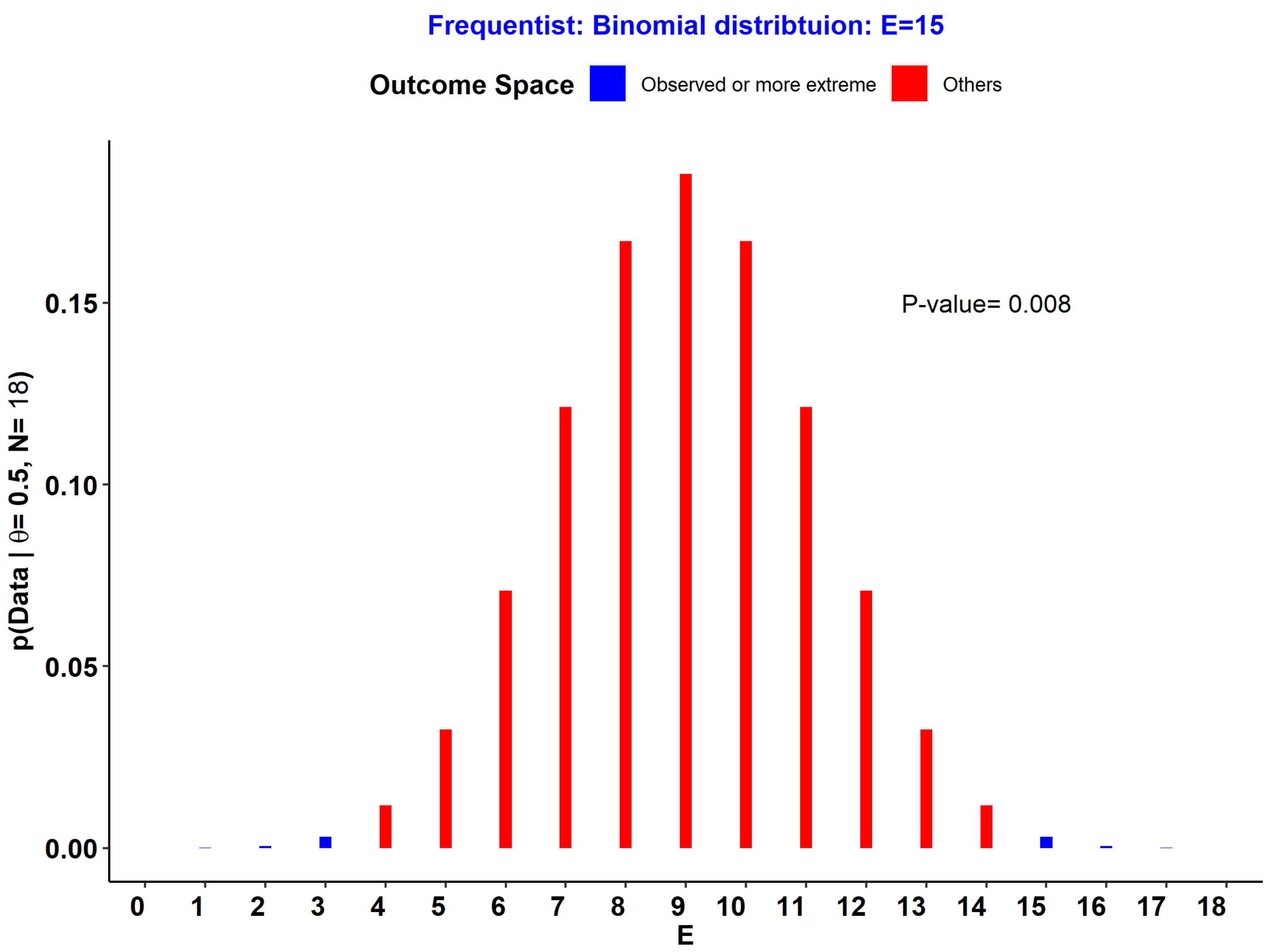
Example: Frequentist

Our Null Hypothesis is that $\theta_0 = 0.5$

Frequentist
N=6
E=5
CI: 0.36-1.0
$\hat{\theta}$ =0.833



Frequentist
N=18
E=15
CI: 0.59-0.96
$\hat{\theta}$ =0.833



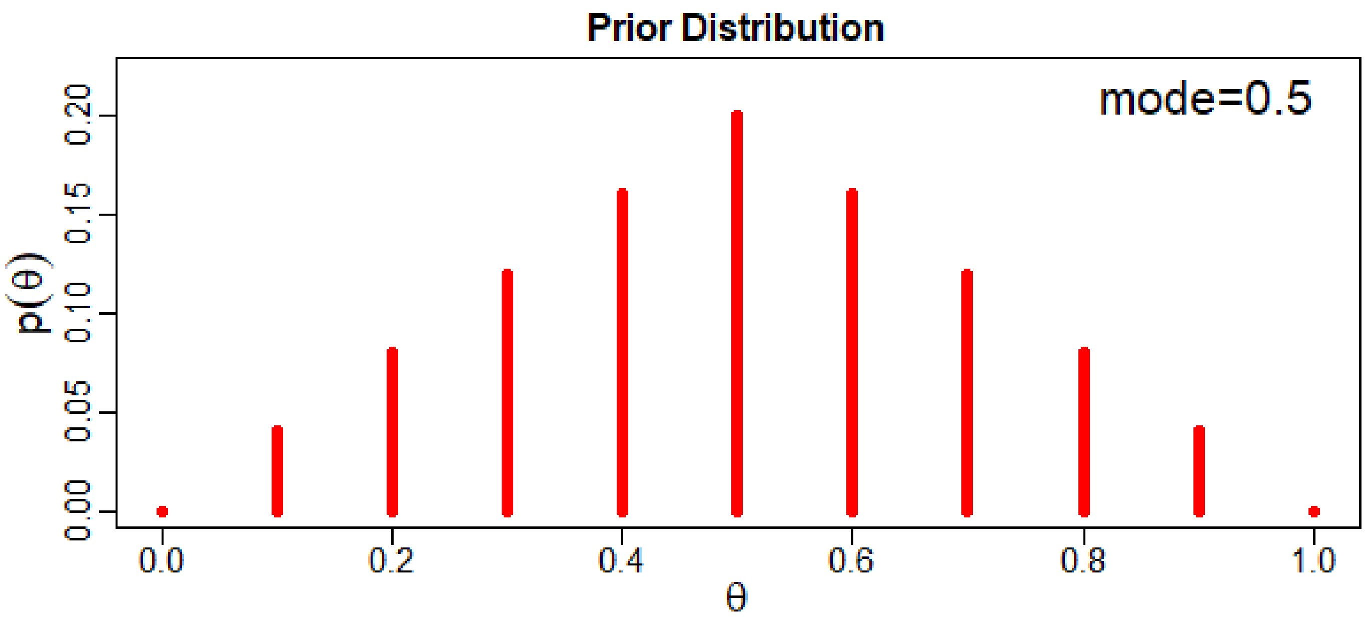
Example: Bayesian

Let's assume that we believe the success rate is around 50%.

This is our prior belief before observing any data.

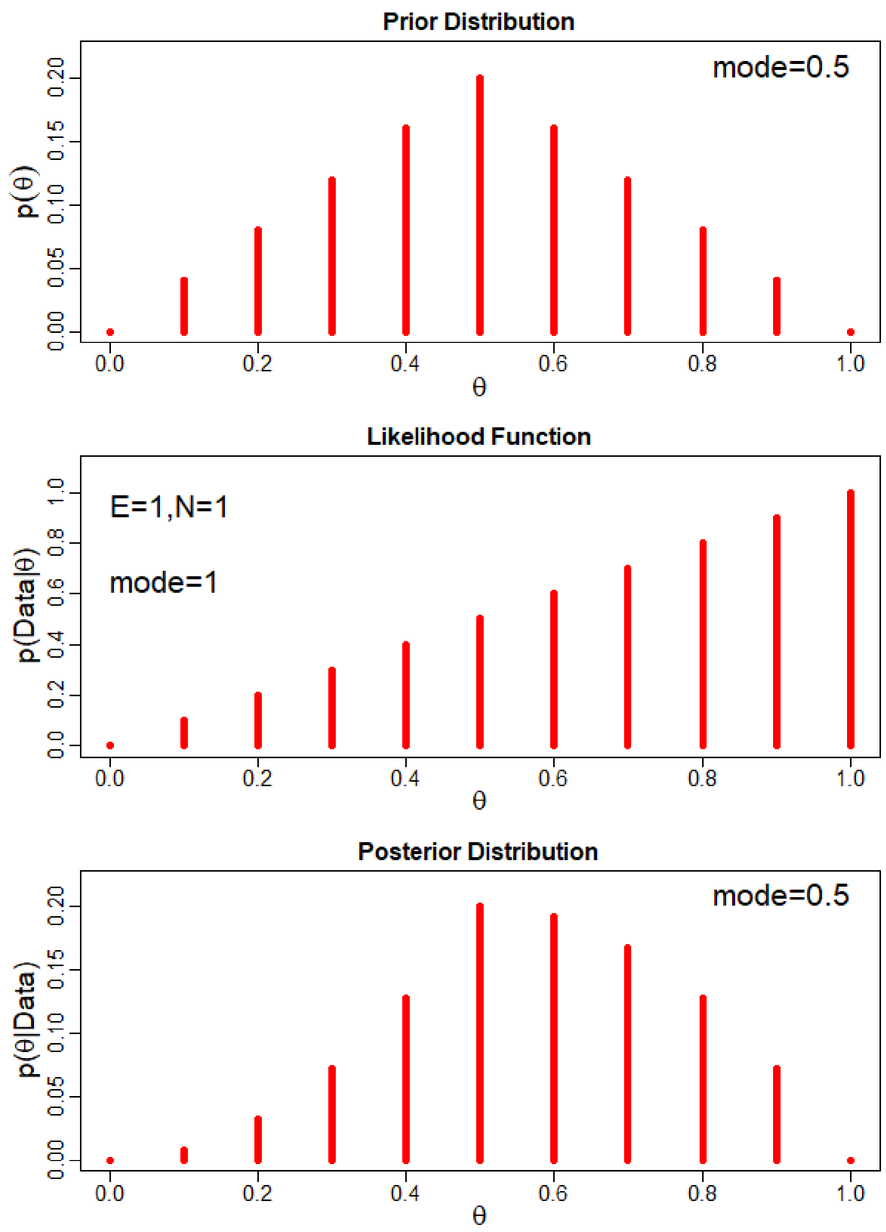
We update our belief after observing each outcome.

Prior



Prior Distribution
mode=0.5
p(θ)
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
θ

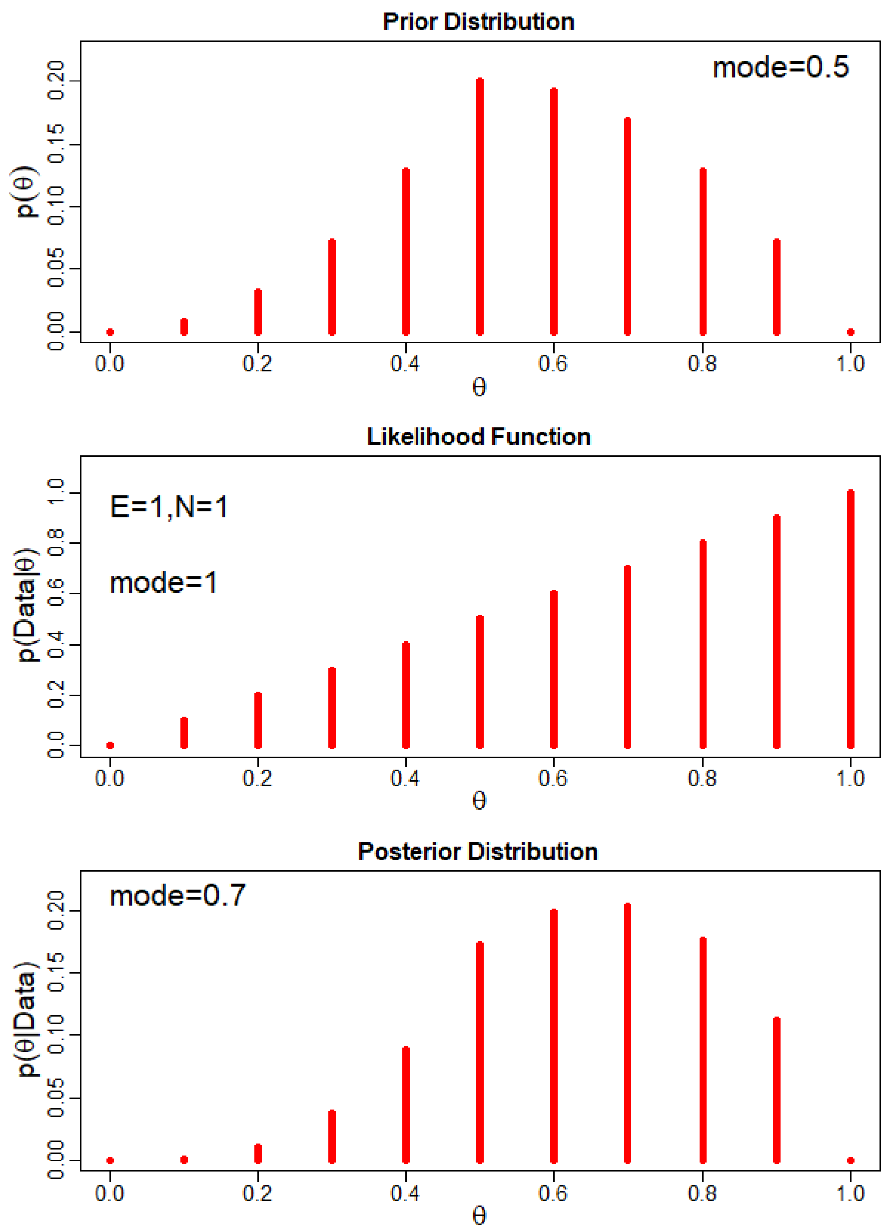
O_1=Y



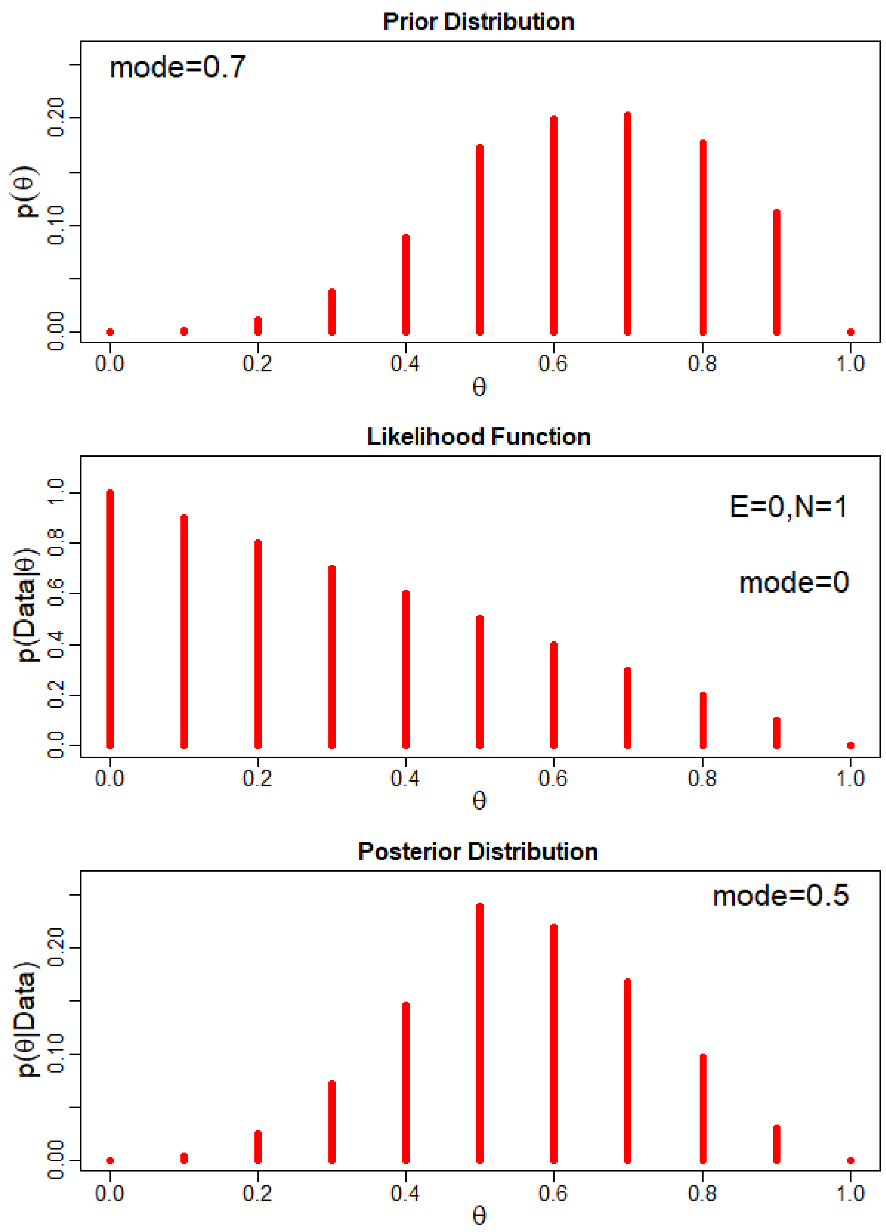
Prior Distribution
mode=0.5
p(θ)
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood Function
E=1,N=1
mode=1
p(Data|θ)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
θ
Posterior Distribution
mode=0.5
p(θ|Data)
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
θ

O_2=Y

"Today's posterior is tomorrow's prior"
— Lindley

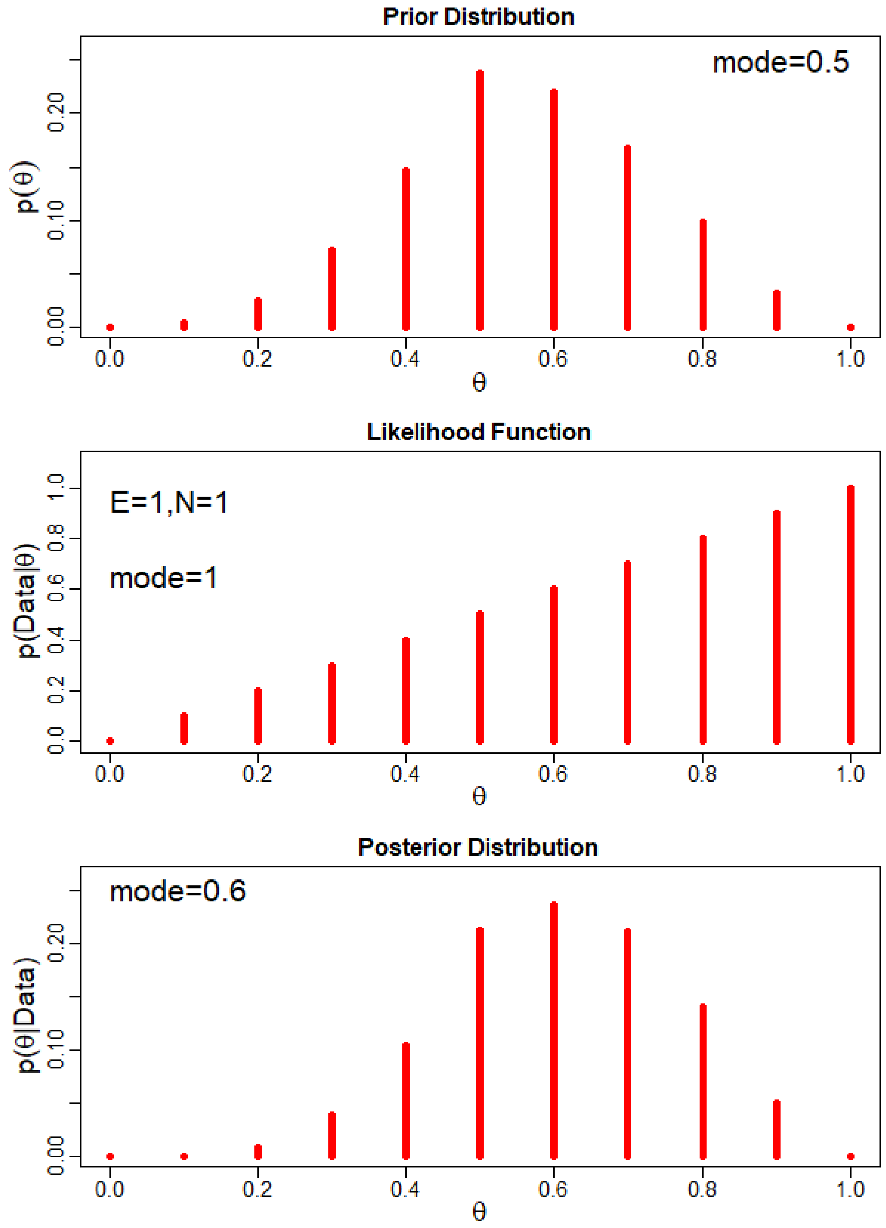


$O_3=N$



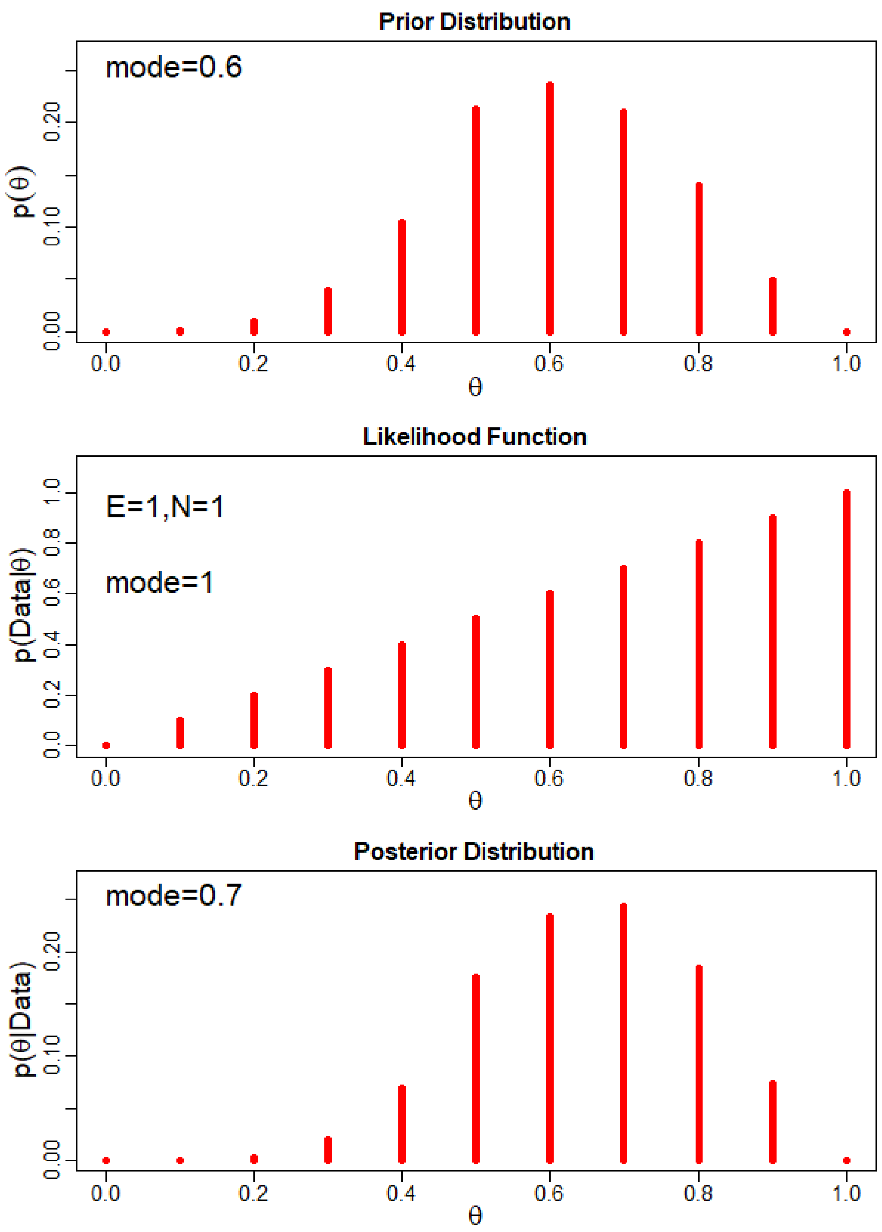
Prior Distribution
mode=0.7
p(θ)
0.00
0.10
0.20
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood Function
E=0,N=1
mode=0
p(Data|θ)
0.0
0.2
0.4
0.6
0.8
1.0
θ
Posterior Distribution
mode=0.5
p(θ|Data)
0.00
0.10
0.20
θ

O_4=Y

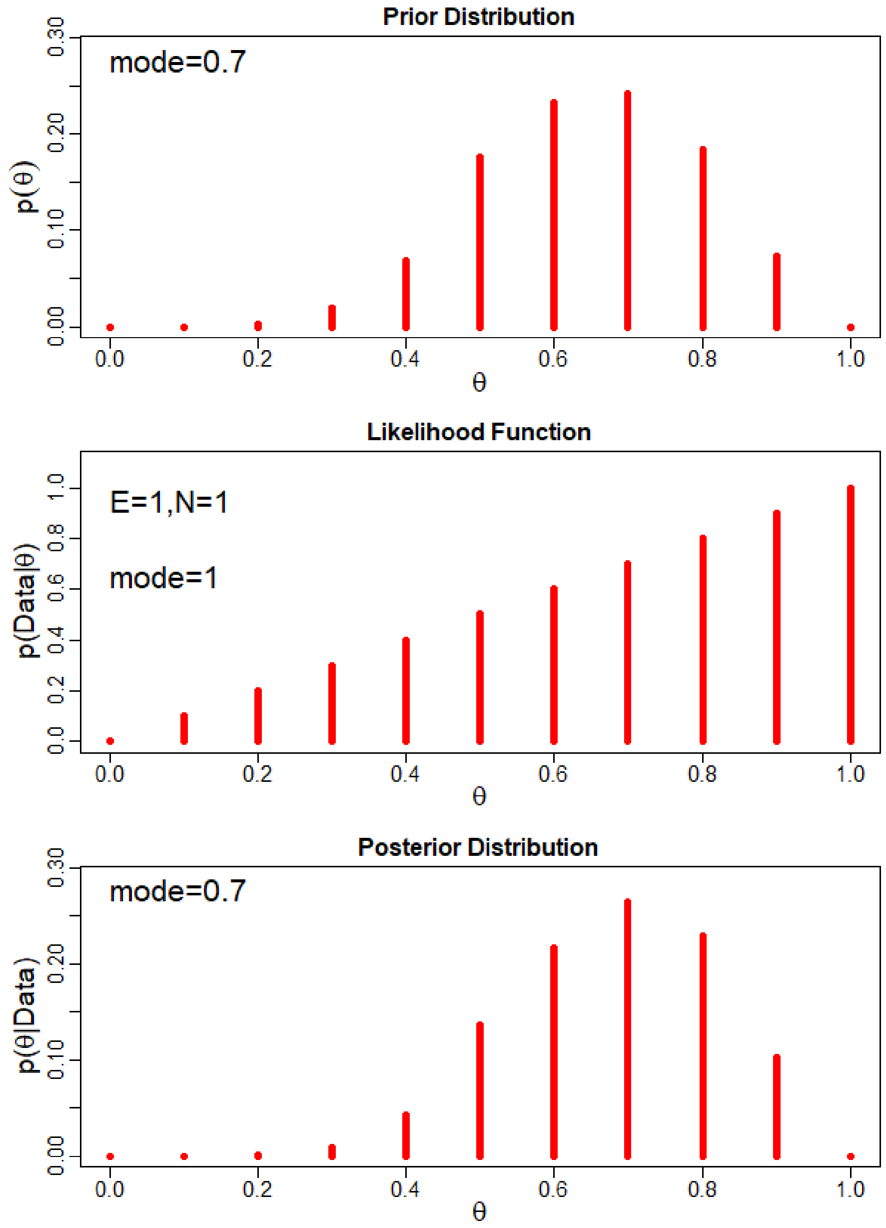


Prior Distribution
mode=0.5
p(θ)
0.00
0.10
0.20
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood Function
E=1,N=1
mode=1
p(Data|θ)
0.0
0.2
0.4
0.6
0.8
1.0
θ
Posterior Distribution
mode=0.6
p(θ|Data)
θ

O_5=Y

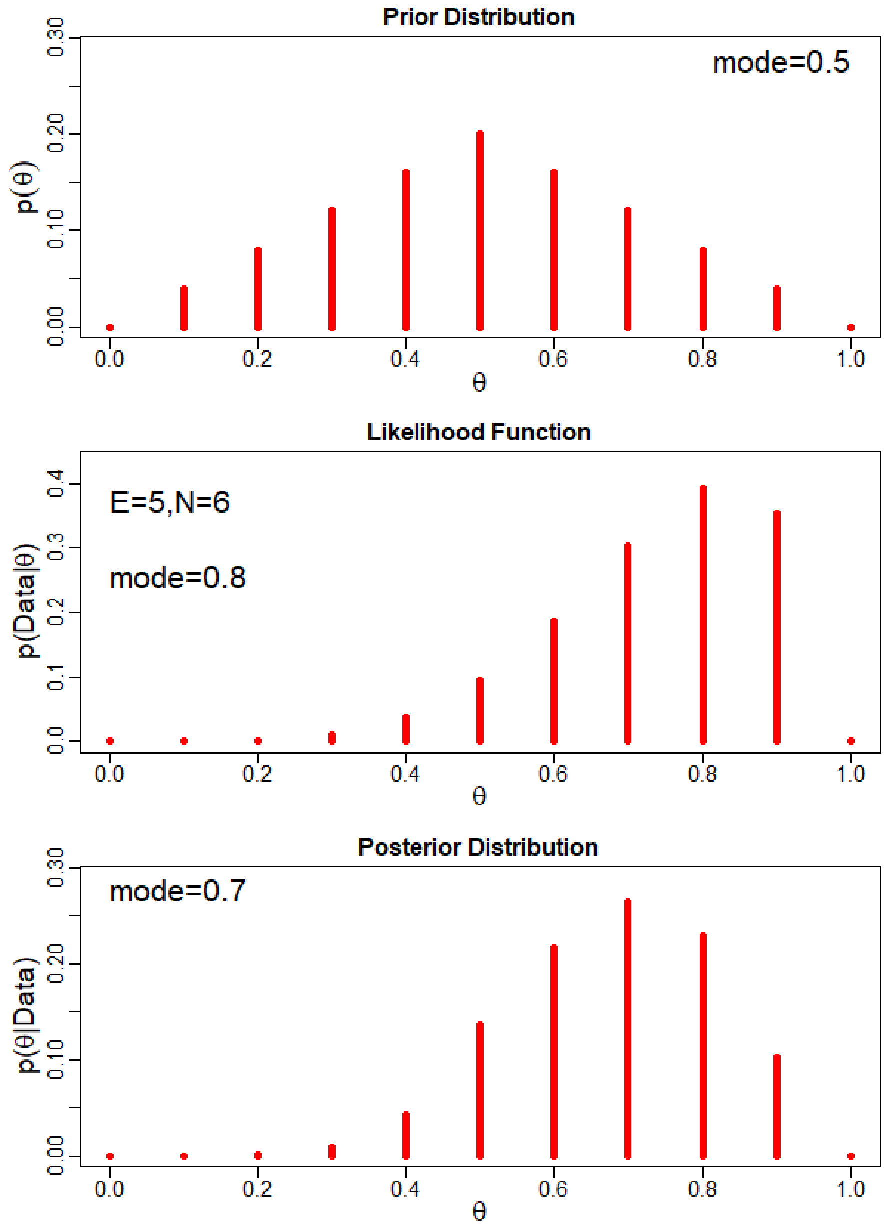

Prior Distribution
mode=0.6
p(θ)
0.00
0.10
0.20
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood Function
E=1,N=1
mode=1
p(Data|θ)
0.0
0.2
0.4
0.6
0.8
1.0
θ
Posterior Distribution
mode=0.7
p(θ|Data)
0.00
0.10
0.20
θ

O_6=Y



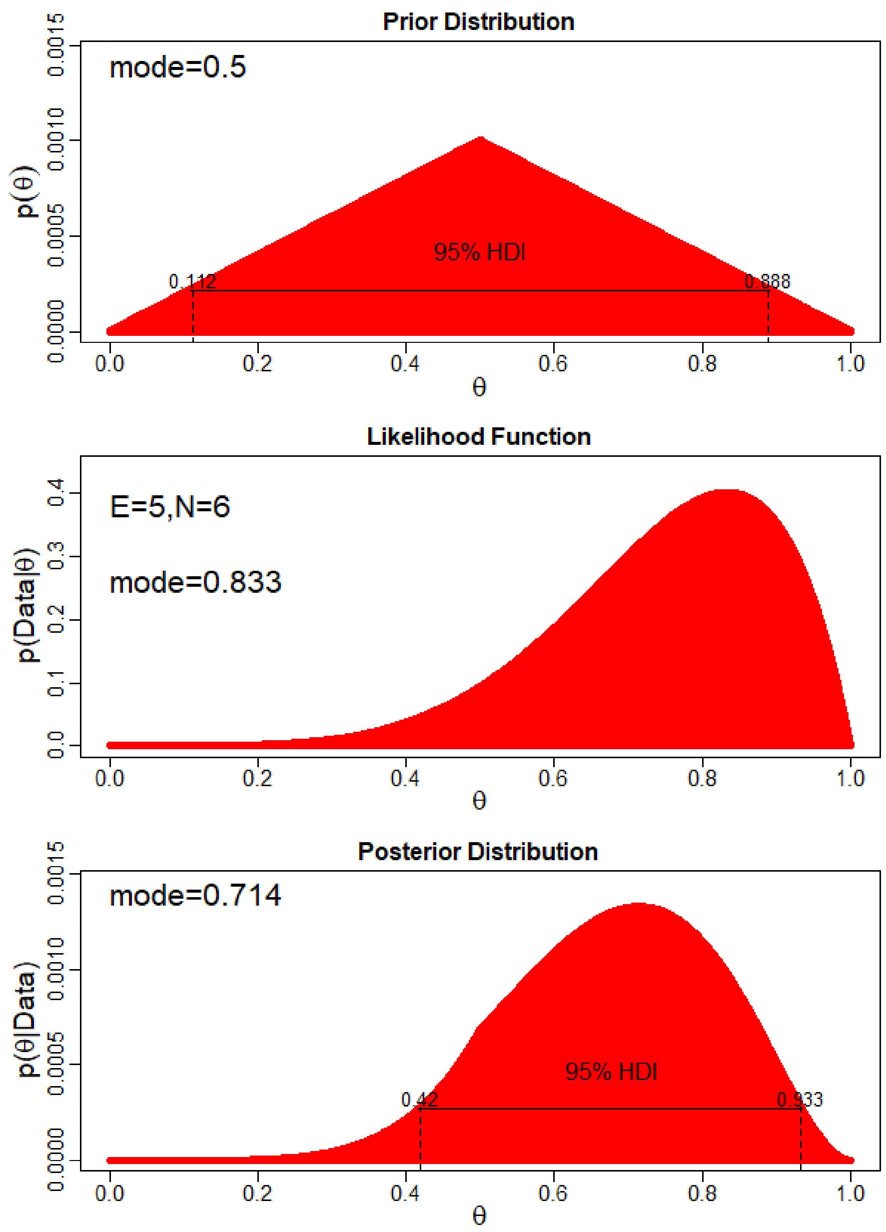
Prior Distribution
mode=0.7
p(θ)
0.00
0.10
0.20
0.30
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood Function
E=1,N=1
mode=1
p(Data|θ)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
θ
Posterior Distribution
mode=0.7
p(θ|Data)
0.00
0.10
0.20
0.30
0.0
0.2
0.4
0.6
0.8
1.0
θ

$O_{1..6}$=YYNYYY



Prior Distribution
mode=0.5
p(θ)
θ
Likelihood Function
E=5,N=6
mode=0.8
p(Data|θ)
θ
Posterior Distribution
mode=0.7
p(θ|Data)
θ

Higher resolution



Conjugate Prior

- We need a distribution to describe our prior belief such that posterior has a closed form distribution
- Beta distribution is an excellent option for parameters in the range [0,1]
  - It is the conjugate prior for binomial distribution.
  - Beta prior + binomial = Beta posterior

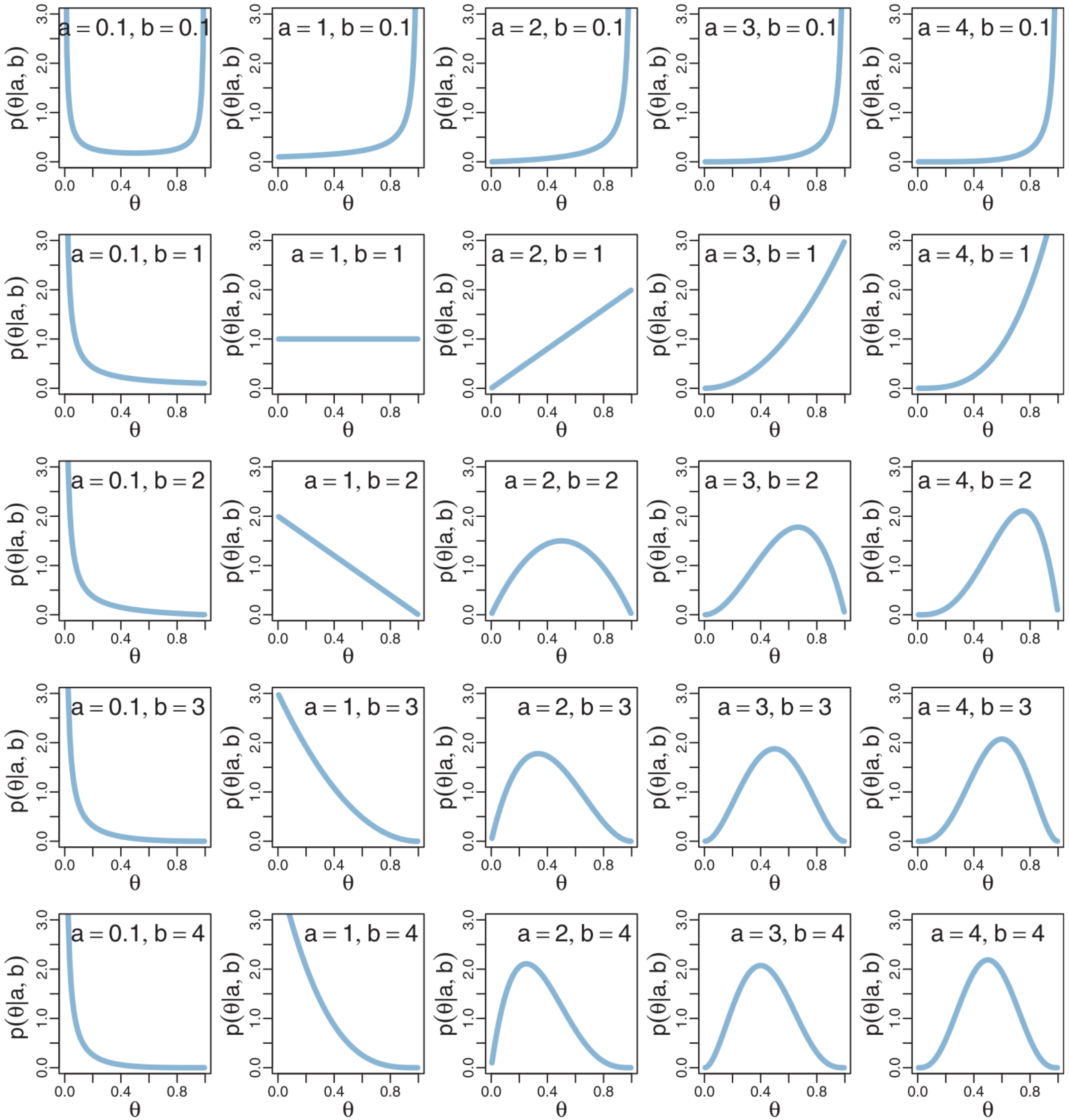
$$p(\theta|a,b) = \mathrm{beta(a,b)} \propto \theta^{(a-1)}(1-\theta)^{(b-1)}, 0 \le \theta \le 1$$

- Mean=$\frac{a}{a+b}$
- Mode=$\frac{a-1}{a+b-2}$
- N samples and E events

$$p(data|\theta) = \mathrm{beta(a+E, b+N-E)}$$

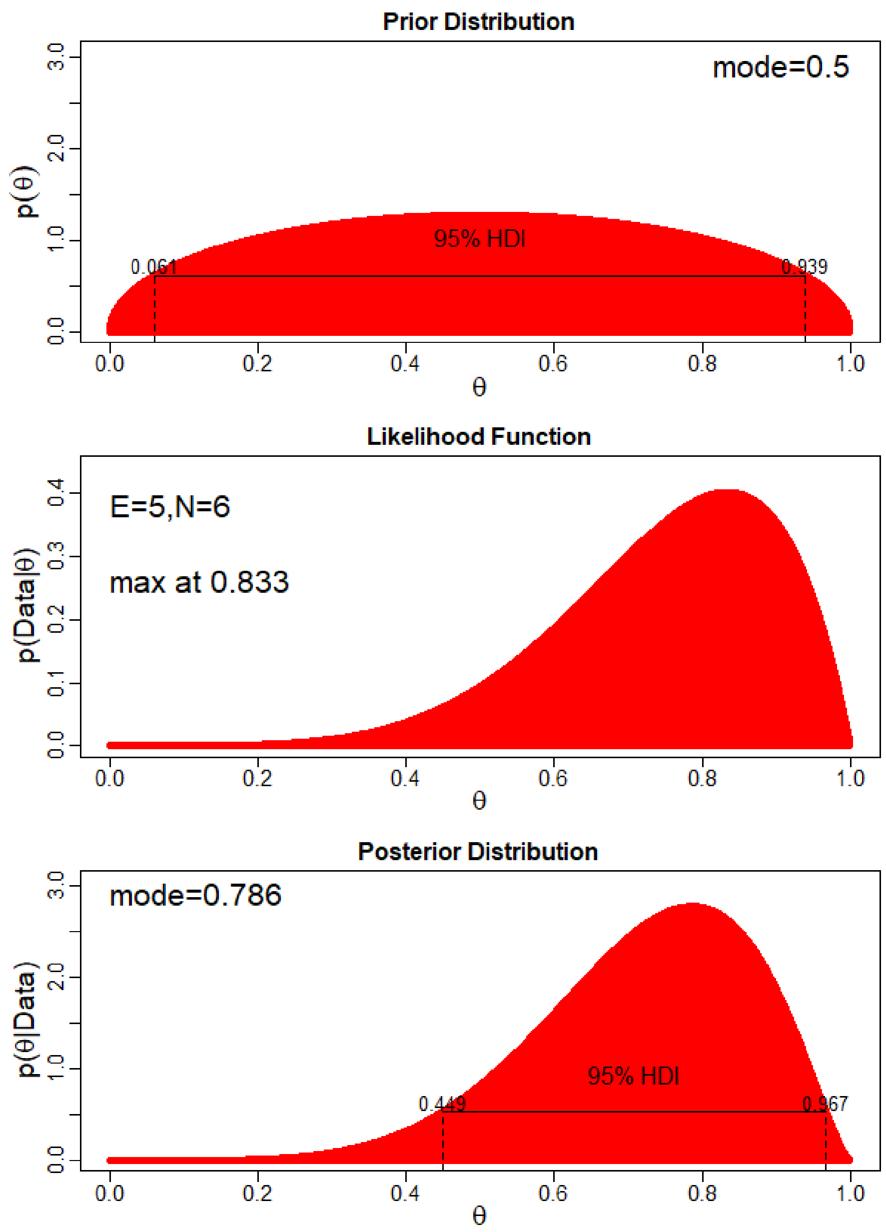
- $a + b$ is effective sample size of prior

Beta Distribution



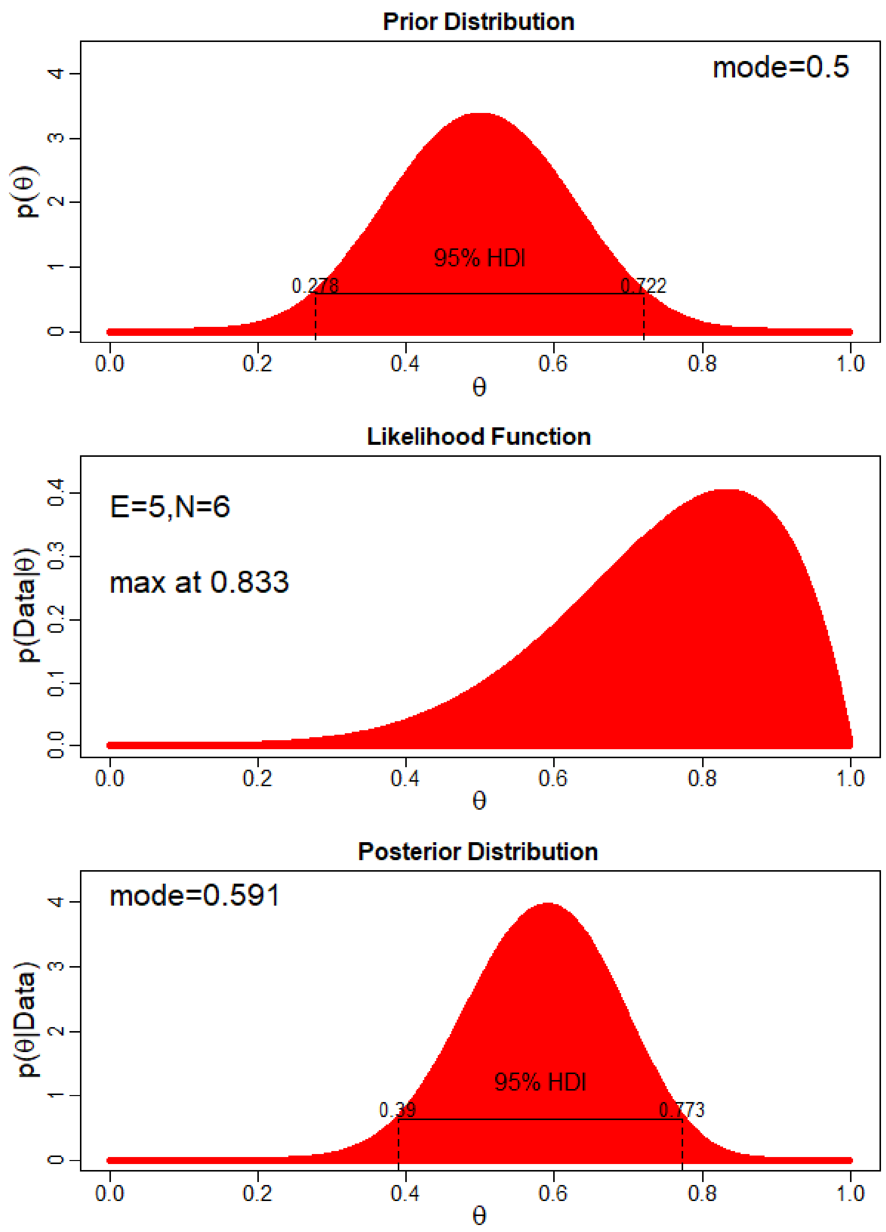
Kruschke, John K. (Ed.) (2014): Doing Bayesian data analysis

$$beta(1.5, 1.5)$$



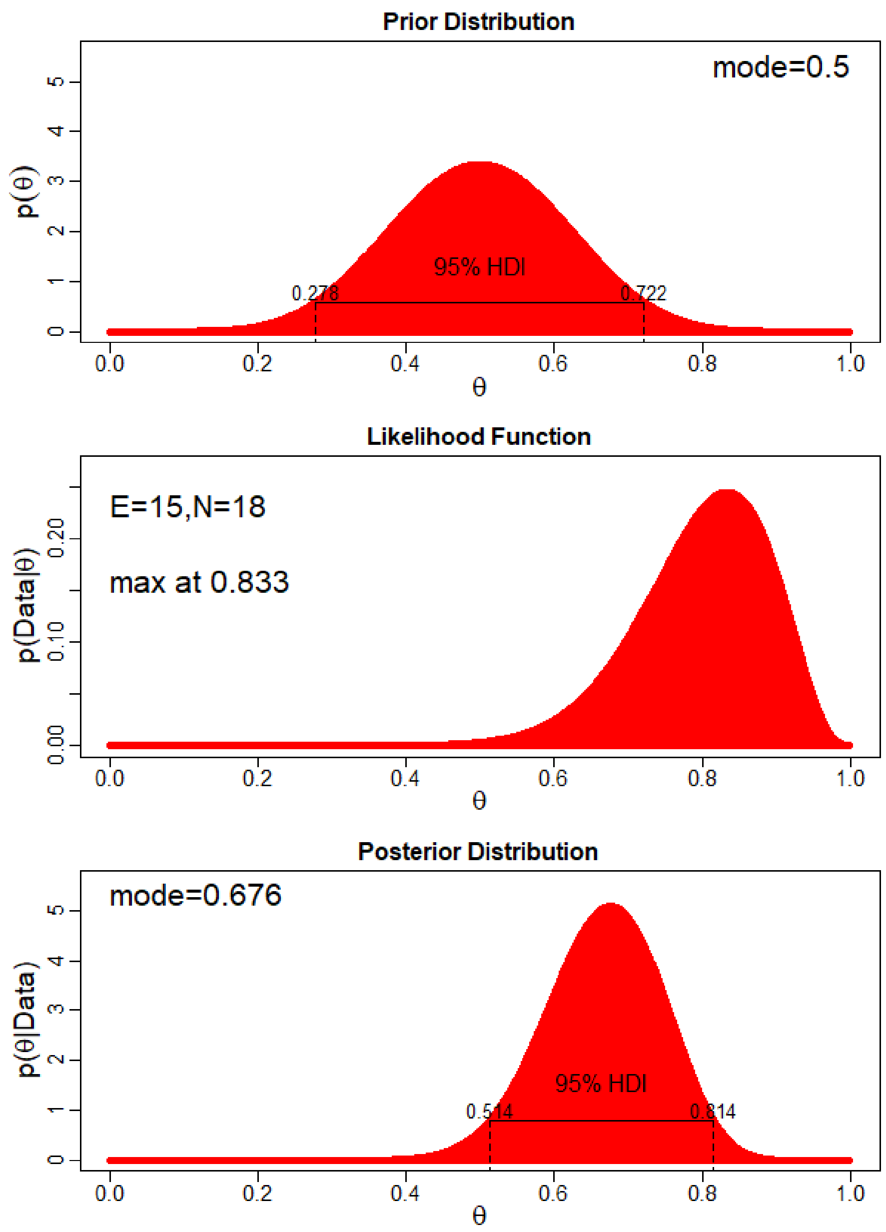
Prior Distribution
mode=0.5
95% HDI
0.061
0.939
p(θ)
θ
Likelihood Function
E=5,N=6
max at 0.833
p(Data|θ)
θ
Posterior Distribution
mode=0.786
95% HDI
0.449
0.967
p(θ|Data)
θ

$$beta(9,9)$$



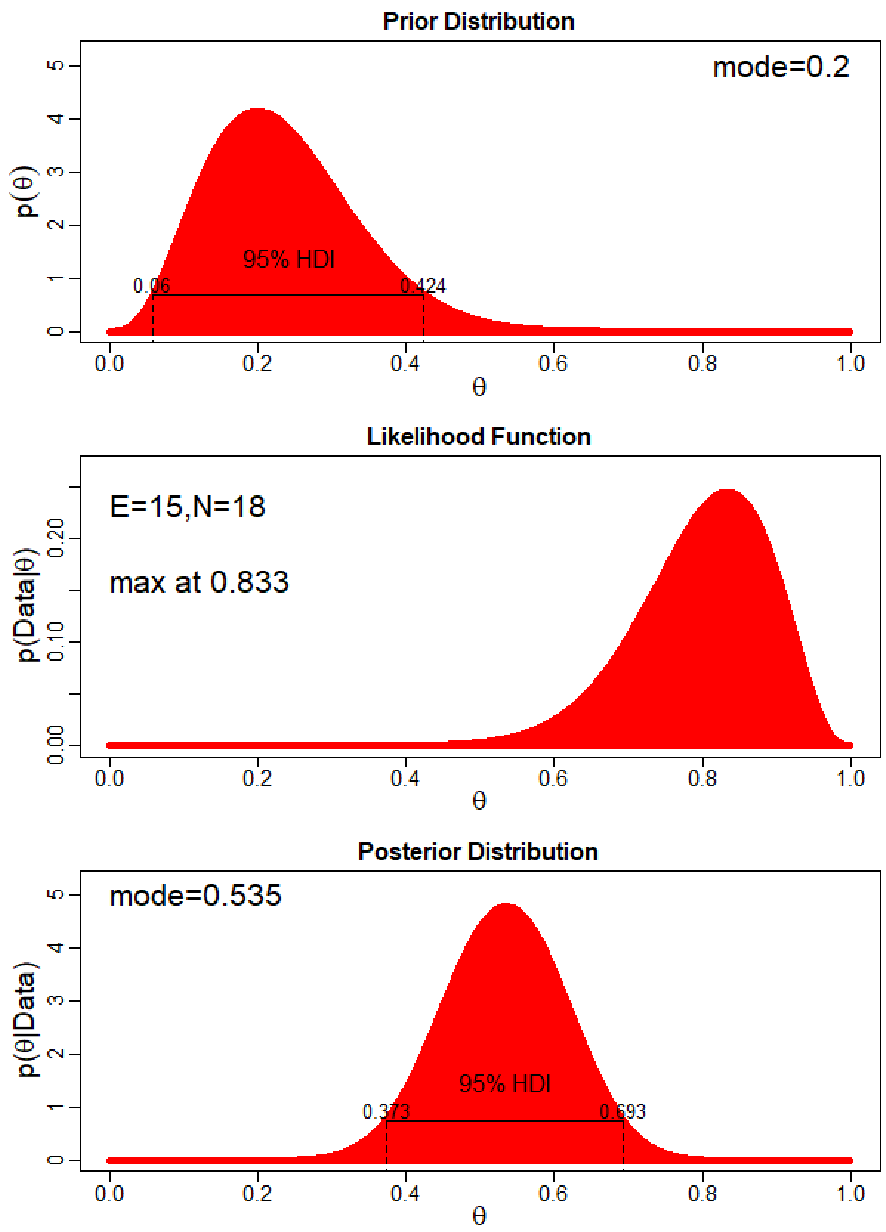
Prior Distribution
mode=0.5
95% HDI
0.278
0.722
p(θ)
θ
Likelihood Function
E=5,N=6
max at 0.833
p(Data|θ)
θ
Posterior Distribution
mode=0.591
95% HDI
0.39
0.773
p(θ|Data)
θ

$$beta(9, 9)$$



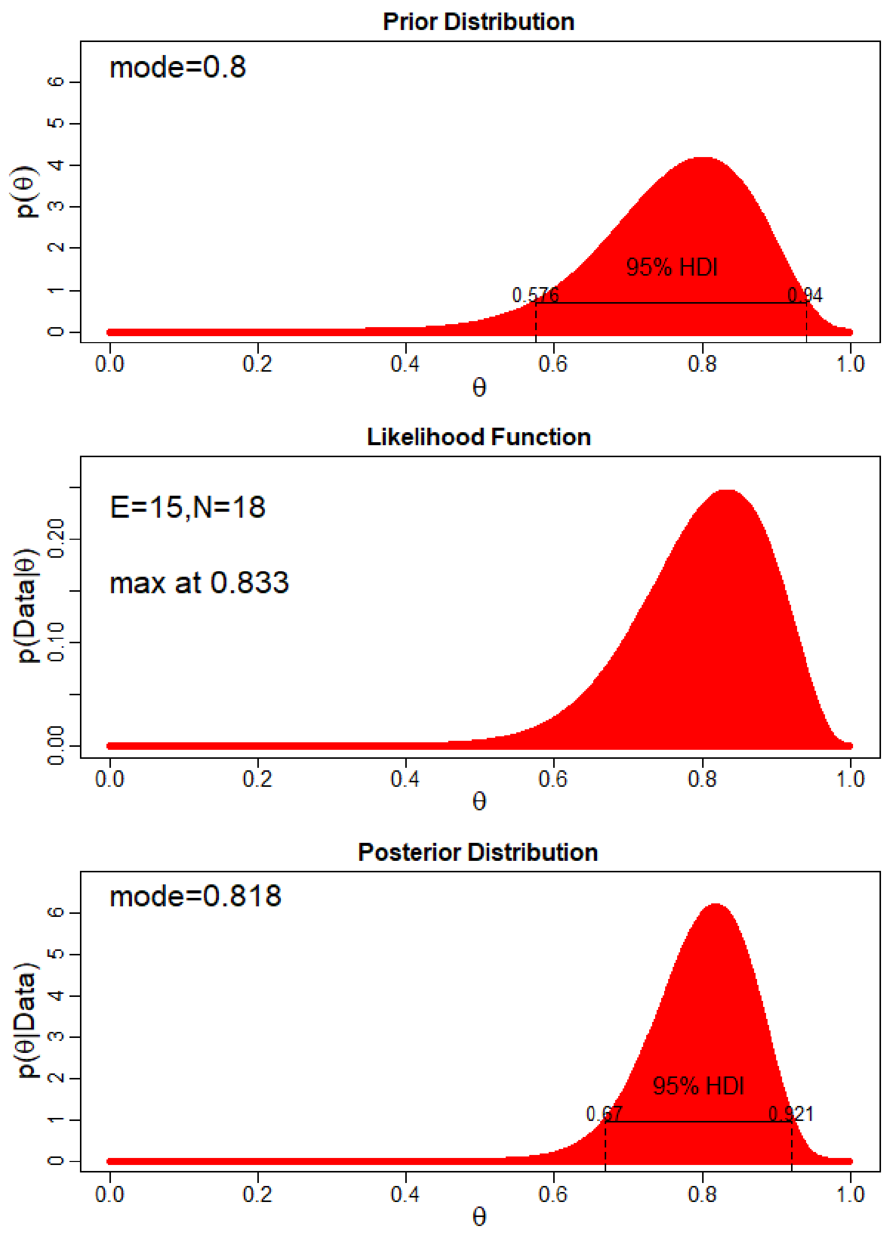
Prior Distribution
mode=0.5
95% HDI
0.278
0.722
p(θ)
θ
Likelihood Function
E=15,N=18
max at 0.833
p(Data|θ)
θ
Posterior Distribution
mode=0.676
95% HDI
0.514
0.814
p(θ|Data)
θ

$$beta(4.2, 13.8)$$



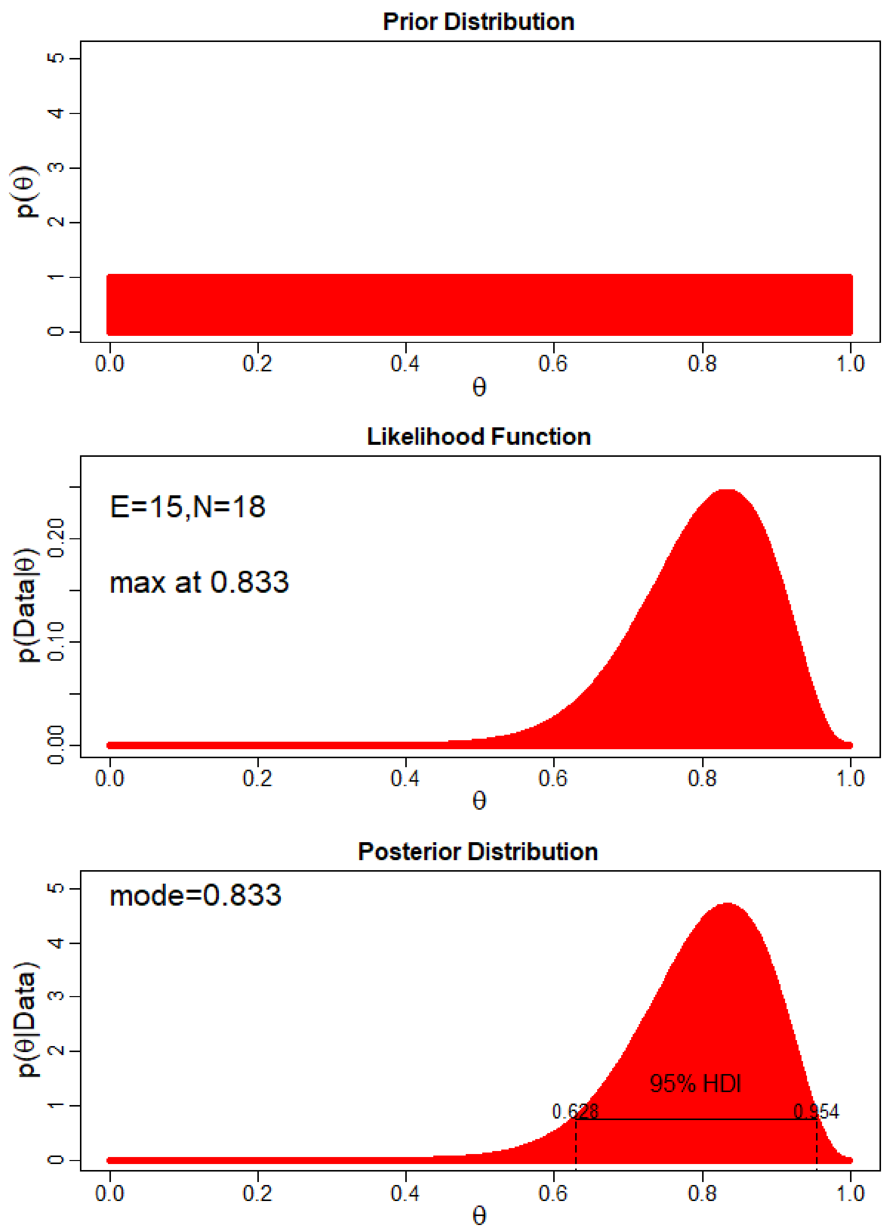
Prior Distribution
mode=0.2
95% HDI
0.06
0.424
p(θ)
θ
Likelihood Function
E=15,N=18
max at 0.833
p(Data|θ)
θ
Posterior Distribution
mode=0.535
95% HDI
0.373
0.693
p(θ|Data)
θ

$beta(13.8, 4.2)$



Prior Distribution
mode=0.8
95% HDI
0.576
0.94
p(θ)
θ
Likelihood Function
E=15,N=18
max at 0.833
p(Data|θ)
θ
Posterior Distribution
mode=0.818
95% HDI
0.67
0.921
p(θ|Data)
θ

$$beta(1.0, 1.0)$$

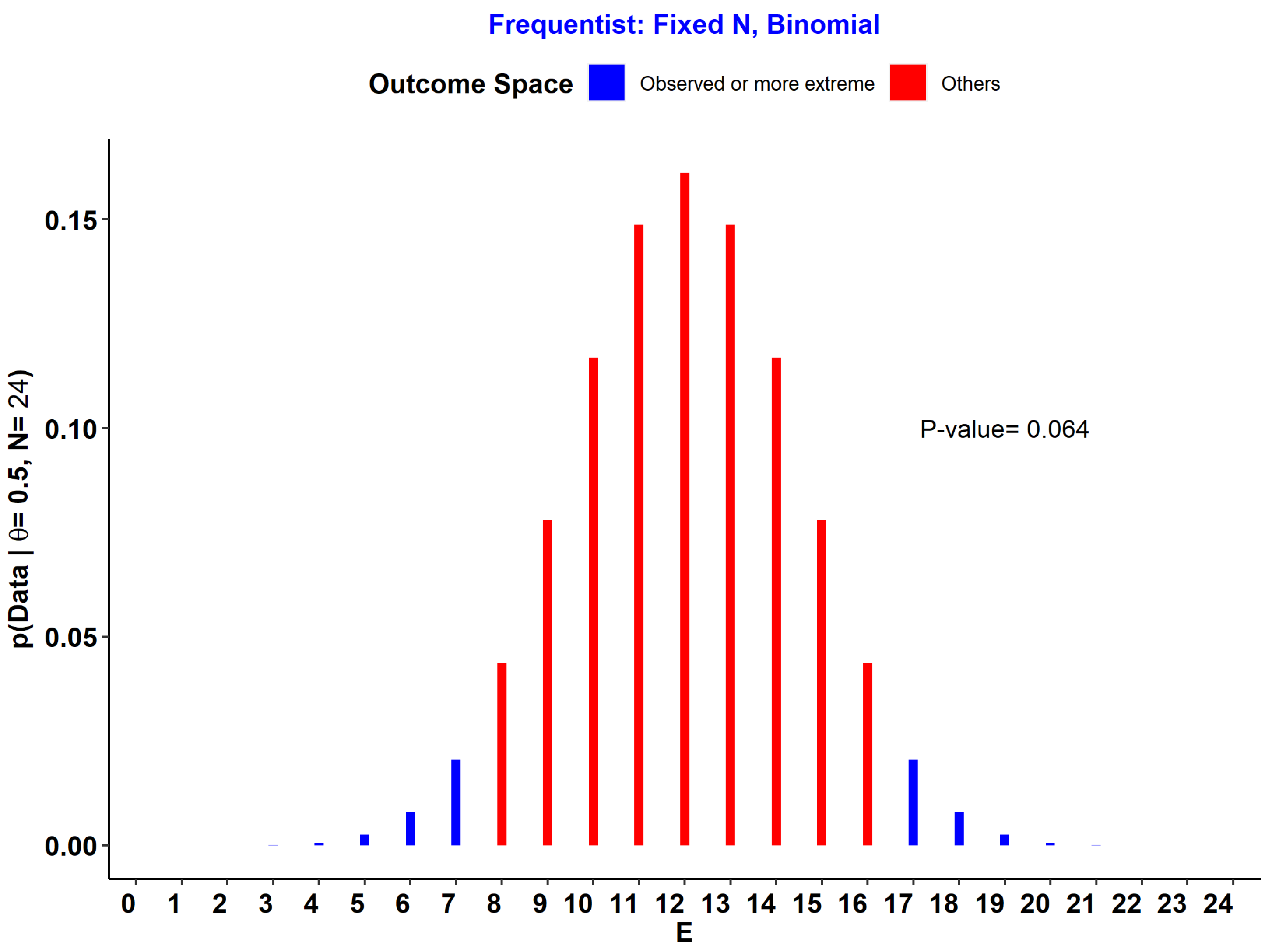

Prior Distribution
p(θ)
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood Function
E=15,N=18
max at 0.833
p(Data|θ)
0.00
0.10
0.20
0.0
0.2
0.4
0.6
0.8
1.0
θ
Posterior Distribution
mode=0.833
95% HDI
0.628
0.954
p(θ|Data)
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
θ

Example: Stopping Rules

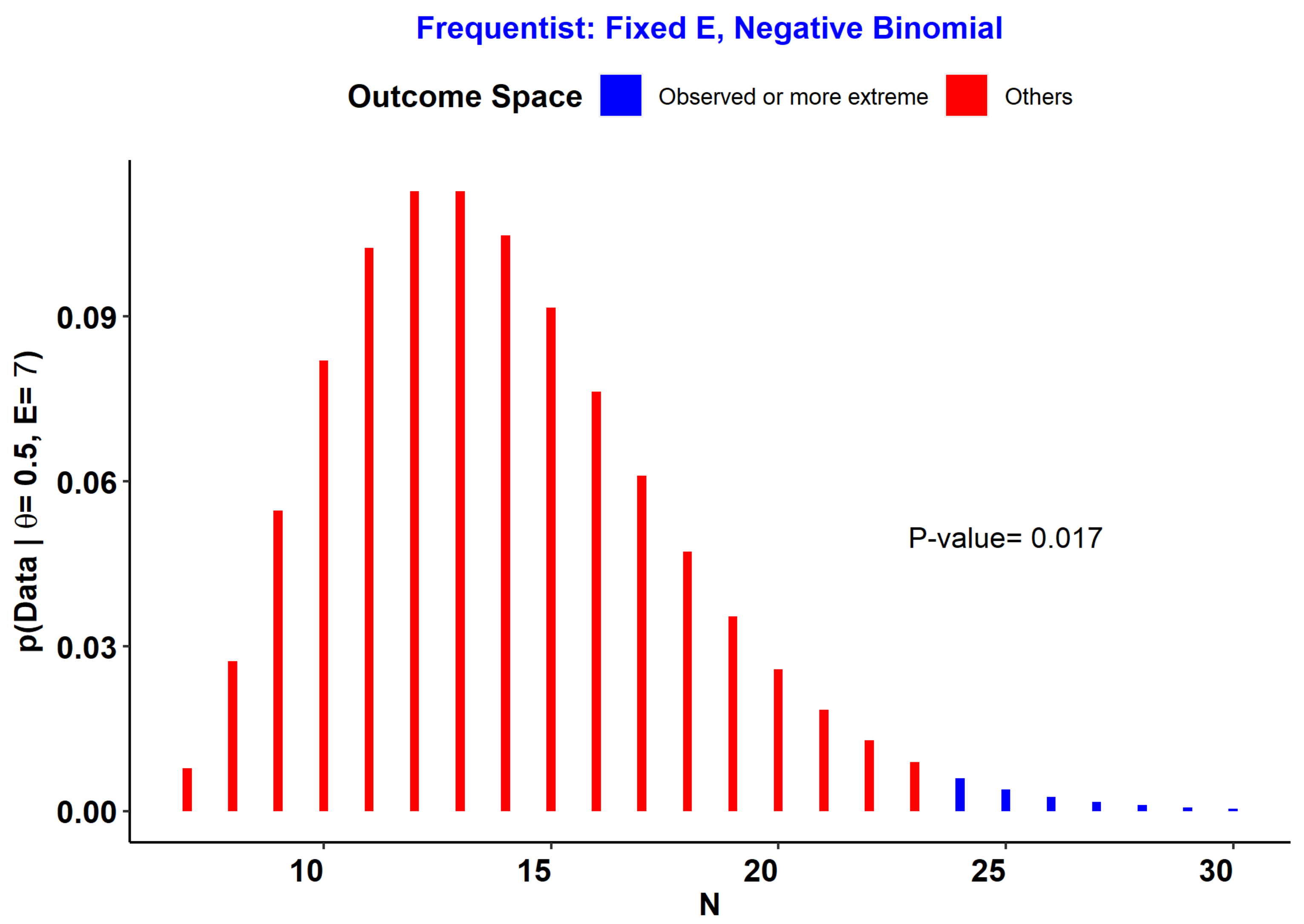
+ N=24 and E=7.

1. N is fixed. **Binomial**

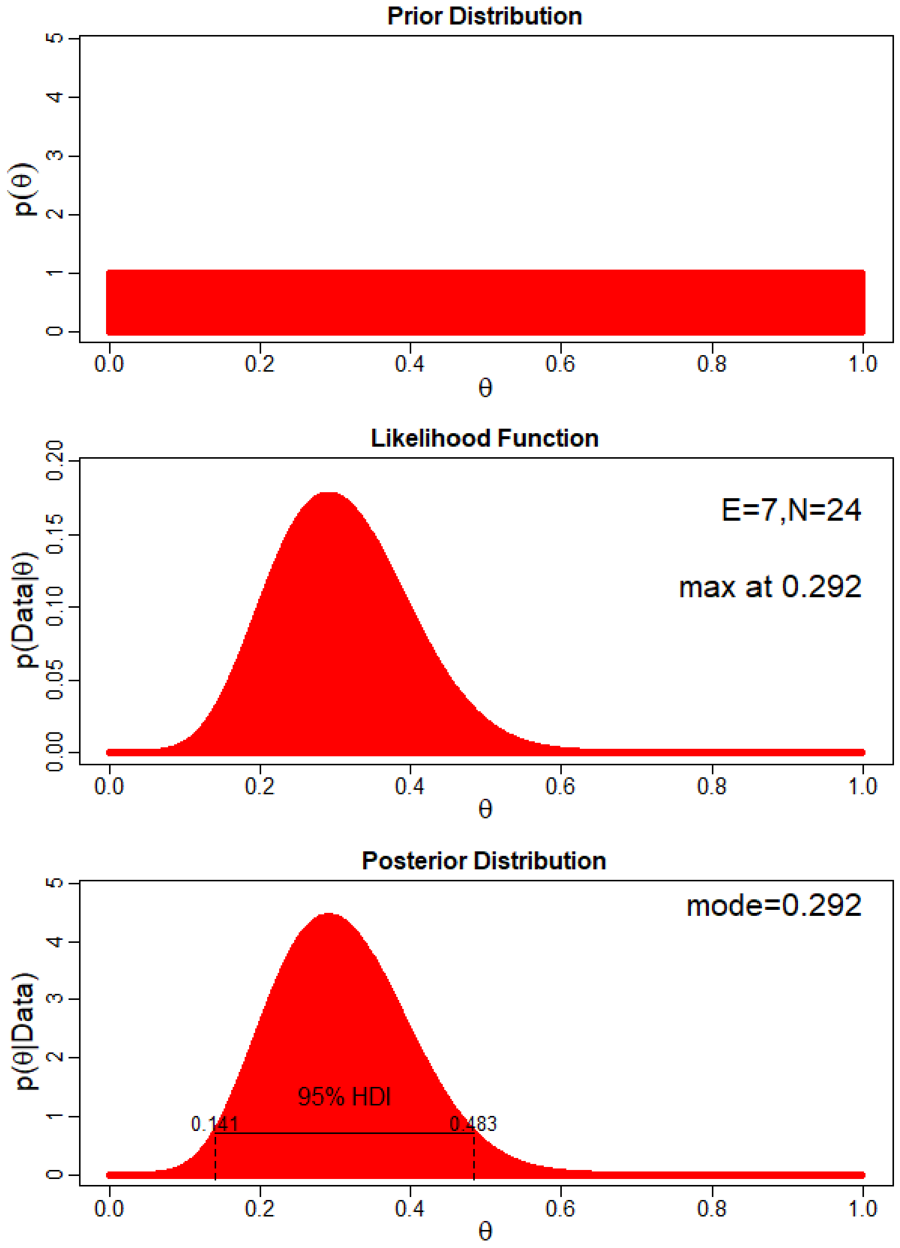
2. E is fixed. **Negative Binomial**


Frequentist: Fixed N, Binomial
Outcome Space
Observed or more extreme
Others
p(Data | θ= 0.5, N= 24)
0.15
0.10
0.05
0.00
P-value= 0.064
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
E

Fixed N


Frequentist: Fixed E, Negative Binomial
Outcome Space
Observed or more extreme
Others
p(Data | θ= 0.5, E= 7)
0.09
0.06
0.03
0.00
P-value= 0.017
10
15
20
25
30
N

Fixed E



Uniform Prior

Problems with Bayesian Inference

Subjectivity

- Most serious objection to Bayesian statistics.
- Two observers/researchers can arrive at different conclusions
  - Same statistical model
  - Different priors

Denominator is hard to calculate

- In some cases, we can use conjugate priors
  - But in many cases, we cannot
- If the number of parameters are small, we can use grid approximation
- However, even when we have moderate number of parameters, it is not practical to use grid approximation.

Problems with Bayesian Inference

$$P(\theta|Data) = \frac{P(Data|\theta)P(\theta)}{P(Data)} = \frac{P(Data|\theta)P(\theta)}{\int_{\theta'} P(Data|\theta')p(\theta')d\theta'}$$

Subjectivity

- Most serious objection to Bayesian statistics.
- Two observers/researchers can arrive at different conclusions
  - Same statistical model
  - Different priors

Denominator is hard to calculate

- In some cases, we can use conjugate priors
  - But in many cases, we cannot
- If the number of parameters are small, we can use grid approximation
- However, even when we have moderate number of parameters, it is not practical to use grid approximation.

Sampling from Posterior

Markov chain Monte Carlo (MCMC)

- Metropolis–Hastings
  - Gibbs sampling
- JAGS, BUGS

Hamiltonian Monte Carlo (HMC)

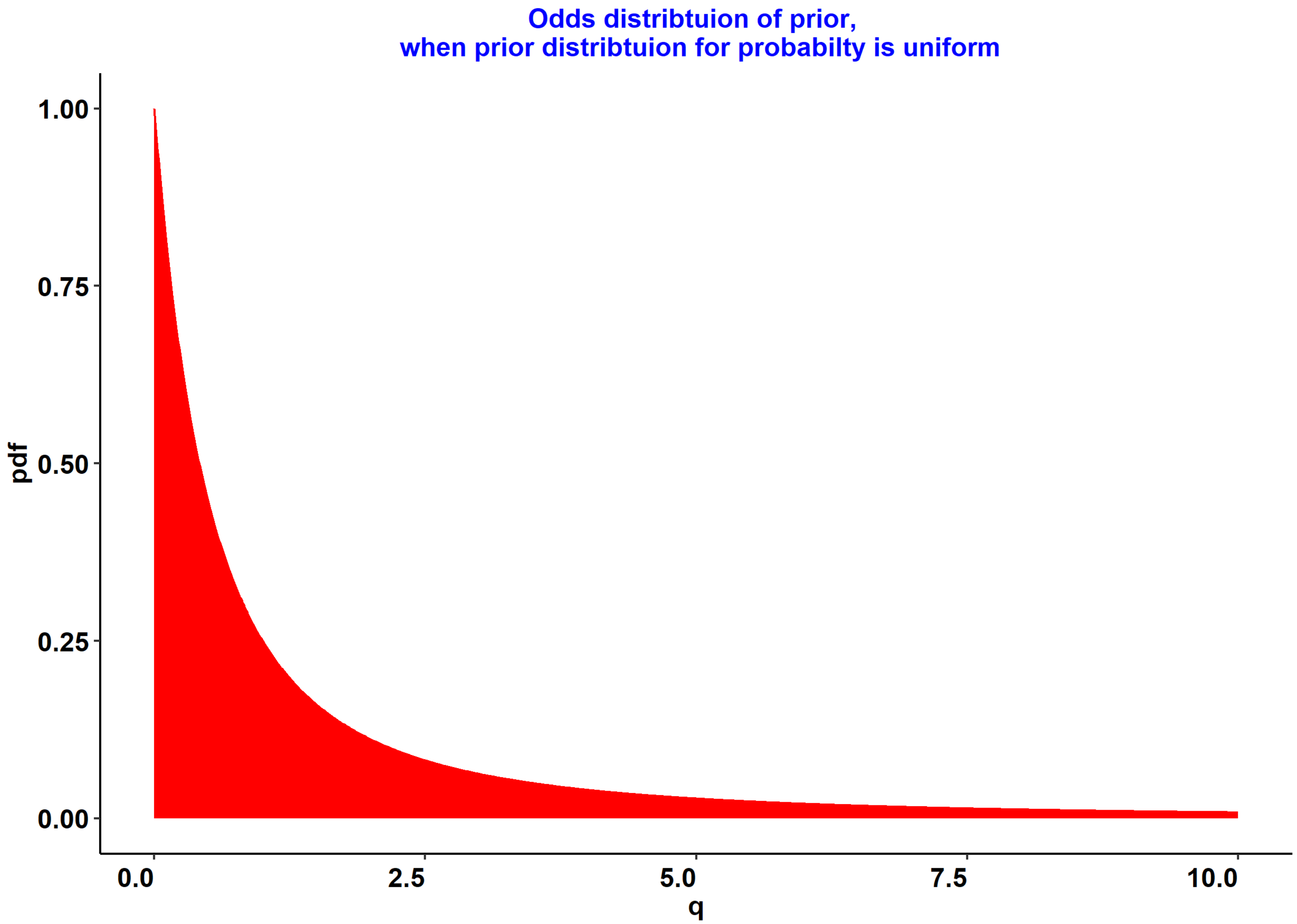
- STAN

Principle of Indifference

- *"If you are completely ignorant about which of a set of exclusive and exhaustive propositions is true, that you should assign them equal probabilities that sum to one."*

SOBER, ELLIOTT (2008): Evidence and evolution. The logic behind the science. Cambridge University Press.

Bayesian Inference Violates Principle of indifference



- Uniform prior.
  - We believe all $0 \leq \theta \leq 1$ have the same prior probability.
  - We might think that this prior is "*uninformative*".
- Change the θ which is the probability metric to odds

$$q = \frac{\theta}{1-\theta}$$

How to set the prior?

Weakly Informative

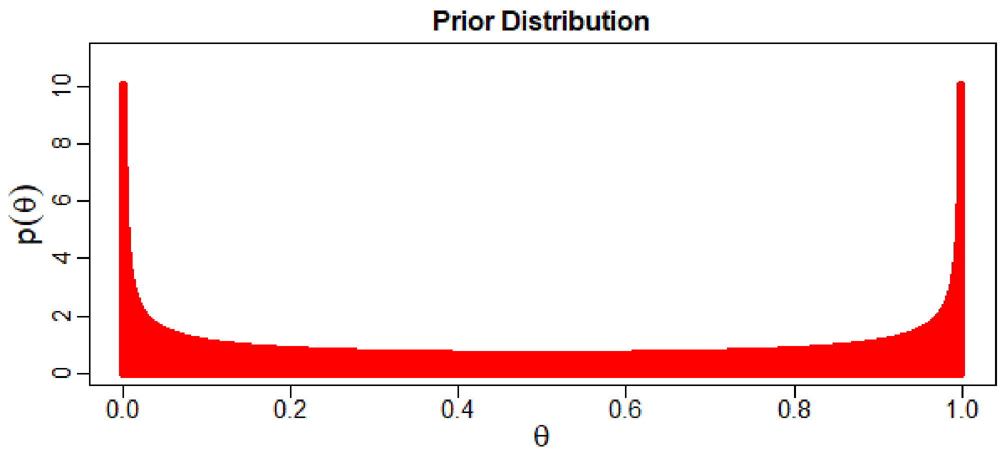
Informative priors

- Prior Studies
  - Moment-Matching
- Expert Knowledge

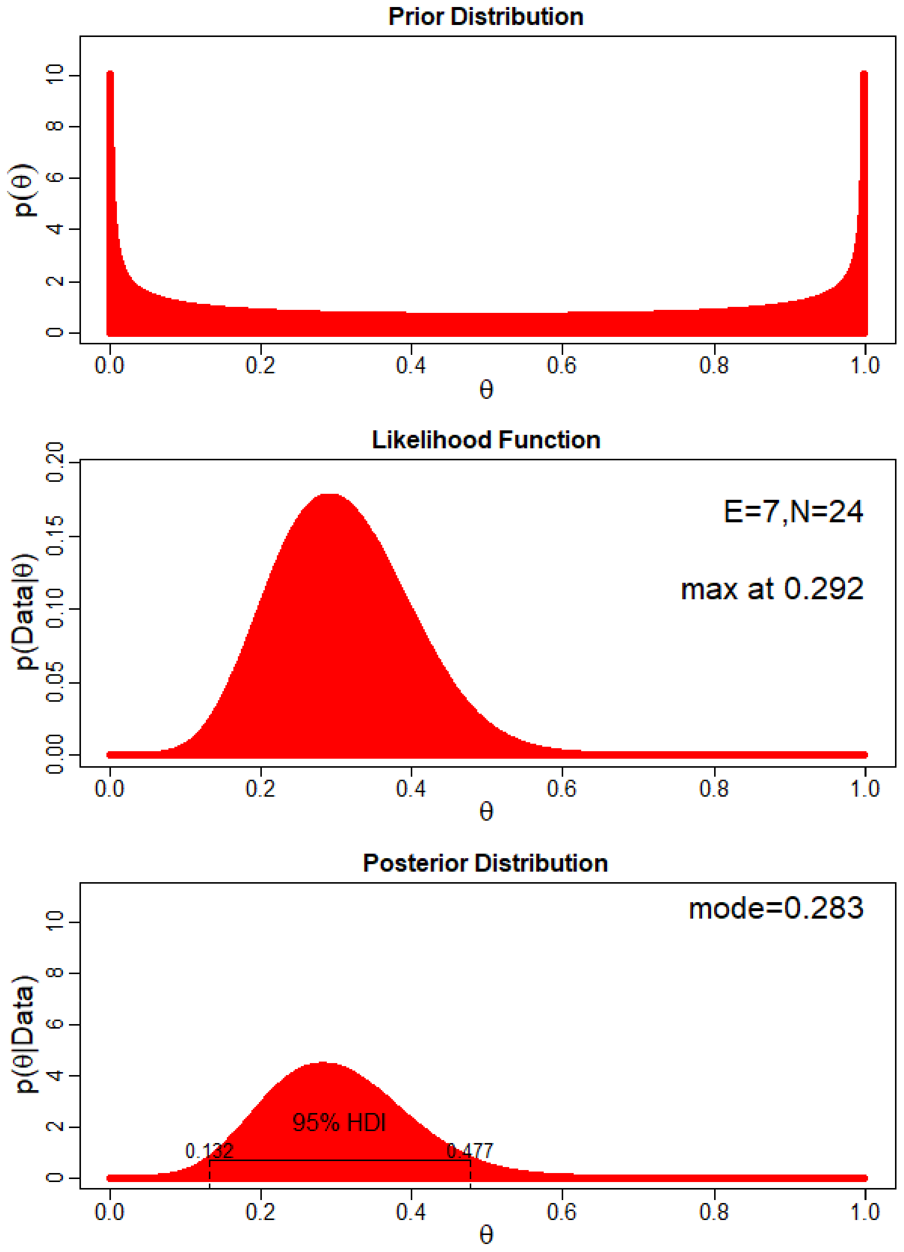
Objective Priors

- Jeffreys Prior
- Reference Prior

Jeffreys Prior



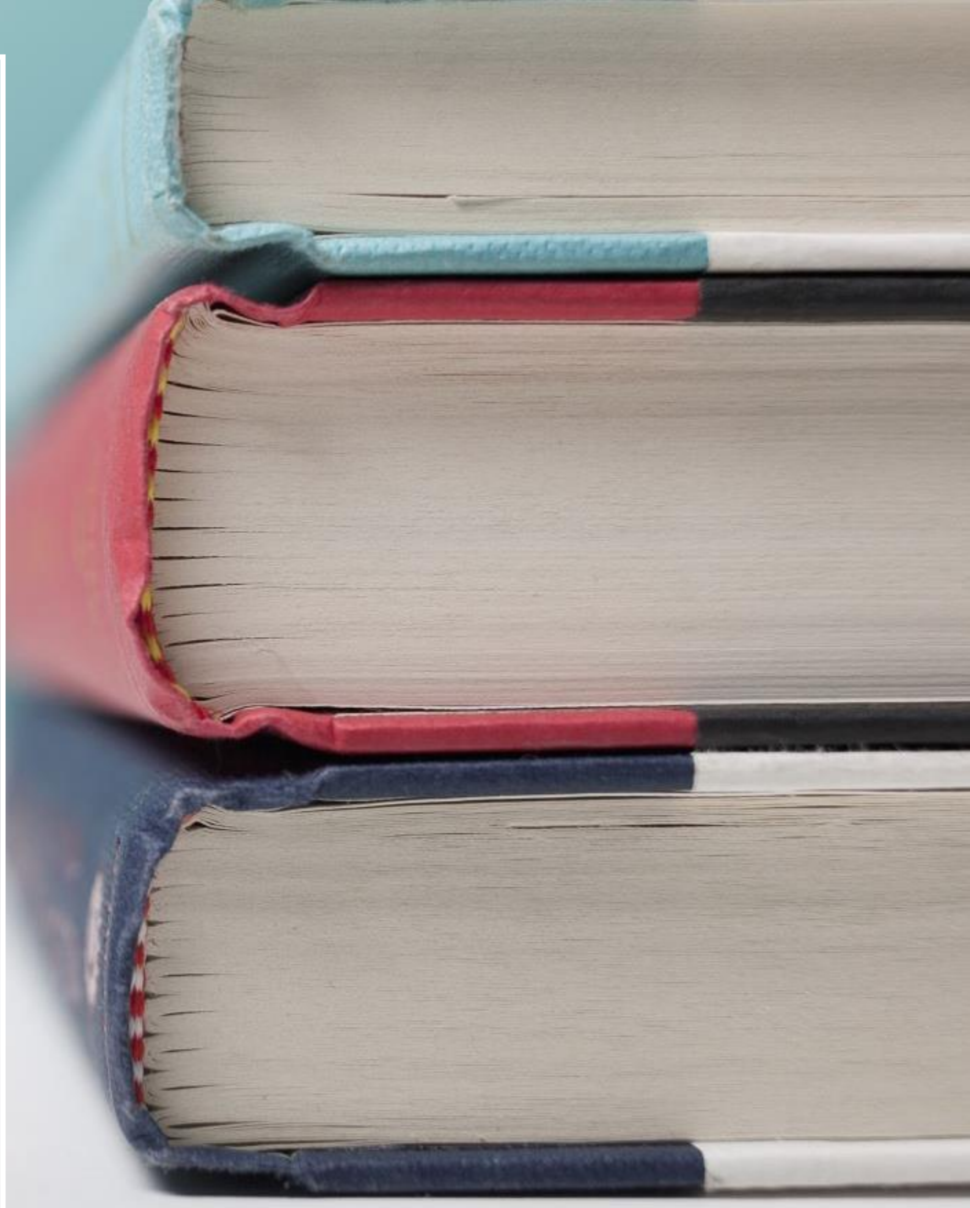
- Jeffreys proposed an “objective” prior that is invariant under monotone transformations of the parameter.
  - Based on Fisher information
  - It is not uninformative
- For example, for binomial distribution, Jeffreys Prior is beta(0.5,0.5).



Jeffreys prior

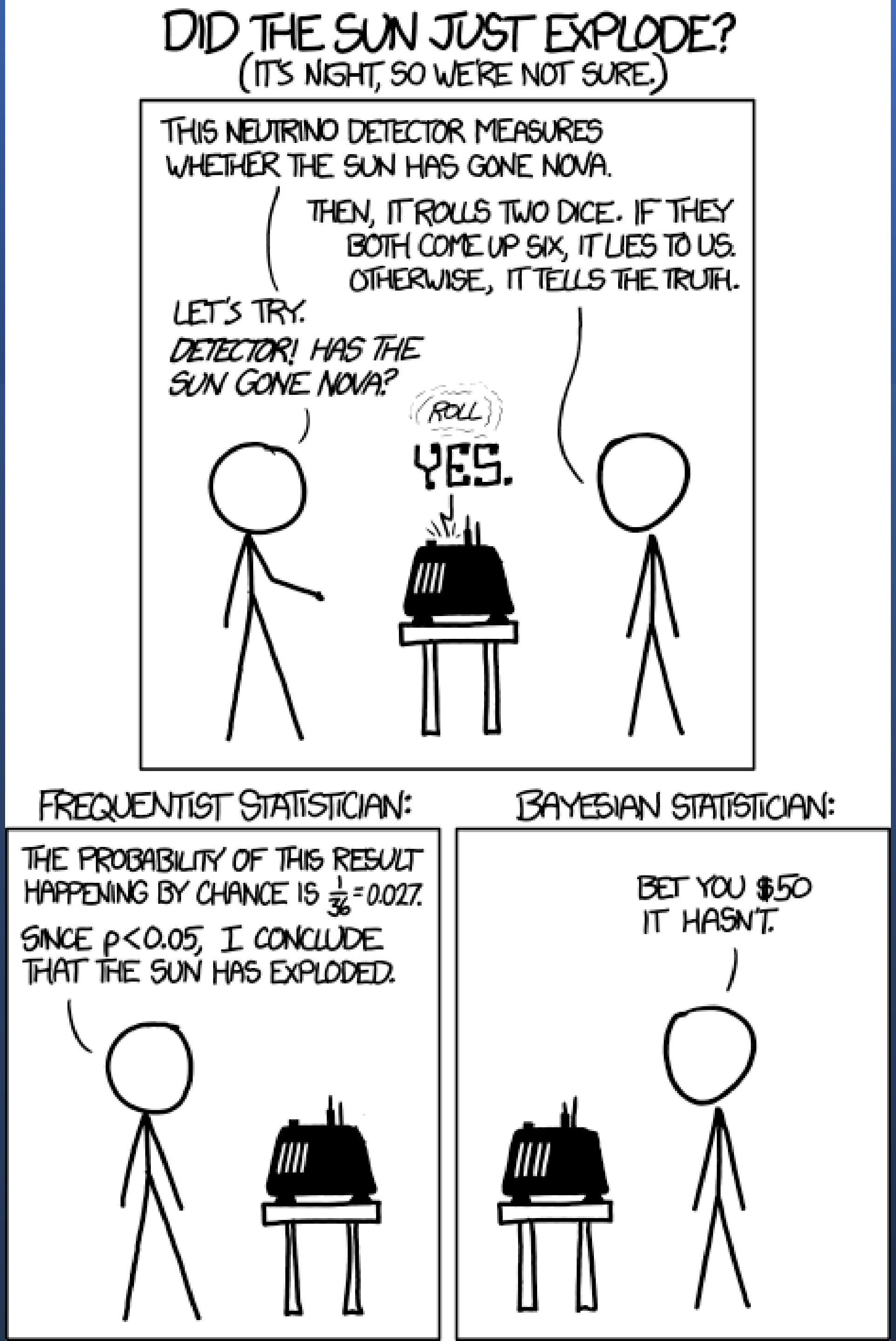
Reading Suggestions

- Kruschke, John K. (Ed.) (2014): Doing Bayesian data analysis. A tutorial with R, JAGS, and Stan. Academic Press.
  - Some of the simulations was based on the codes from this book.
- Lambert, Ben (2018): A student's guide to Bayesian statistics. 1st. Los Angeles: SAGE.
- McElreath, Richard (2020): Statistical rethinking. A Bayesian course with examples in R and Stan. Taylor and Francis CRC Press.
- SOBER, ELLIOTT (2008): Evidence and evolution. The logic behind the science. Cambridge, UK: Cambridge University Press.



Conclusions

- Bayesian Statistics is a very flexible approach
  - Update our belief after observing data
  - Natural statement about the parameters
- Bayesian Inference Violates Principle of indifference



https://xkcd.com/1132/

Thank you!