Bayesian Analysis

Why Isn't Everyone Bayesian?

Dr. Iris Ivy M. Gauran

School of Statistics
University of the Philippines

Go to `www.menti.com` and use the code 41 32 49 0

It is common knowledge that with a fair coin, "the probability of flipping Heads is 0.5." How do you interpret this probability?

A. If I flip this coin over and over and over and over, roughly 50% of the flips will be Heads.

B. If I flip this coin, Heads or Tails are equally plausible.

C. Both of the above make sense.

Go to `www.menti.com` and use the code 41 32 49 0

It is common knowledge that with a fair coin, "the probability of flipping Heads is 0.5." How do you interpret this probability?

A. If I flip this coin over and over and over and over, roughly 50% of the flips will be Heads.

B. If I flip this coin, Heads or Tails are equally plausible.

C. Both of the above make sense.

A. 1 point B. 3 points C. 2 points

Go to `www.menti.com` and use the code 41 32 49 0

A civil engineer warns that "there is a 0.5 probability of the local bridge collapsing within 1 year." How do you interpret this probability?

A. If the engineer observes this year over and over and over and over, the bridge will collapse in roughly 50% of these years.

B. The engineer's calculation indicates that the bridge is equally likely to collapse or not collapse this year.

C. The engineer's calculation is wrong. The bridge will either collapse or it won't. Thus the probability of collapse can only be 0 or 1 (not 0.5).

Go to www.menti.com and use the code 41 32 49 0

A civil engineer warns that "there is a 0.5 probability of the local bridge collapsing within 1 year." How do you interpret this probability?

A. If the engineer observes this year over and over and over and over, the bridge will collapse in roughly 50% of these years.

B. The engineer's calculation indicates that the bridge is equally likely to collapse or not collapse this year.

C. The engineer's calculation is wrong. The bridge will either collapse or it won't. Thus the probability of collapse can only be 0 or 1 (not 0.5).

A. 1 point B. 3 points C. 1 points

Go to www.menti.com and use the code 88 77 44 7

In a survey of 10 UP students, two experiments are performed:

- **I ask each student if they are in favor of the death penalty. 10 out of 10 say "yes".**
- **I give each student a sample of coffee and a sample of chai tea. 10 out of 10 correctly identify which sample is which.**

In light of these experiments, what do you conclude?

A. You're more confident that UP students can distinguish between coffee & chai than UP students favoring the death penalty.

B. The evidence in favor of UP students' intention to favor death penalty is just as strong as the evidence in favor of UP students' ability to identify between coffee and chai.

Go to www.menti.com and use the code 88 77 44 7

In a survey of 10 UP students, two experiments are performed:

- **I ask each student if they are in favor of the death penalty. 10 out of 10 say "yes".**
- **I give each student a sample of coffee and a sample of chai tea. 10 out of 10 correctly identify which sample is which.**

In light of these experiments, what do you conclude?

A. You're more confident that UP students can distinguish between coffee & chai than UP students favoring the death penalty.

B. The evidence in favor of UP students' intention to favor death penalty is just as strong as the evidence in favor of UP students' ability to identify between coffee and chai.

A. 3 points B. 1 point

Go to `www.menti.com` and use the code 88 77 44 7

Suppose that during a recent doctor's visit, you tested positive for SARS-CoV-2 virus. If you only get to ask the doctor one question, which would it be?

A. What are the chances that I actually have the disease?

B. If in fact I do not have the disease, what are the chances that I would've gotten this positive test result?

Go to www.menti.com and use the code 88 77 44 7

Suppose that during a recent doctor's visit, you tested positive for SARS-CoV-2 virus. If you only get to ask the doctor one question, which would it be?

A. What are the chances that I actually have the disease?

B. If in fact I do not have the disease, what are the chances that I would've gotten this positive test result?

A. 3 points B. 1 point

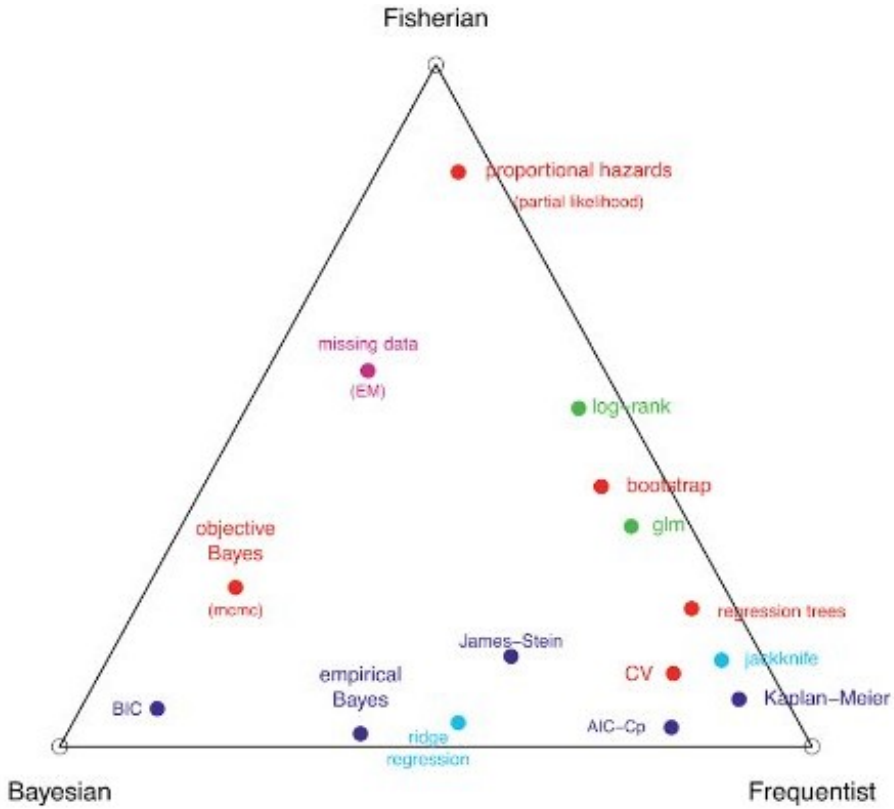
Brief History

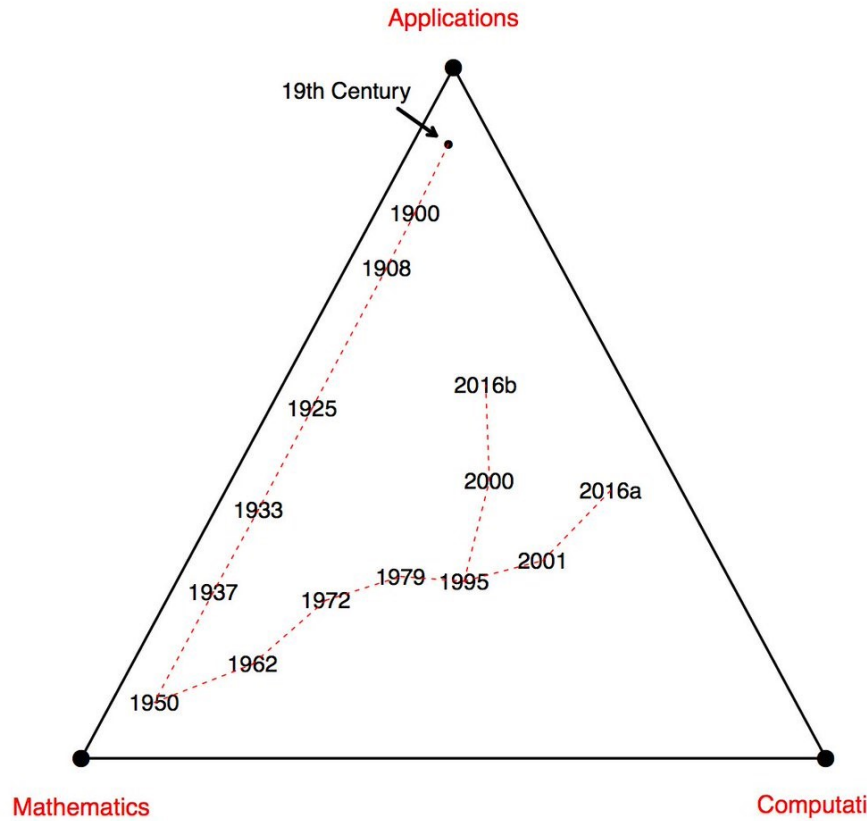
Thomas Bayes (1701 – 1761) was an English statistician, philosopher and Presbyterian minister who is known for formulating a specific case of the theorem that bears his name: **Bayes' theorem**.



Bayes never published what would become his most famous accomplishment; his notes were edited and published after his death by **Richard Price**.

Computer Age Statistical Inference





Illustration



Contains 4 marbles

Possible contents:

(1) ○○○○

(2) ●○○○

(3) ●●○○

(4) ●●●○

(5) ●●●●

Observe:

●○●

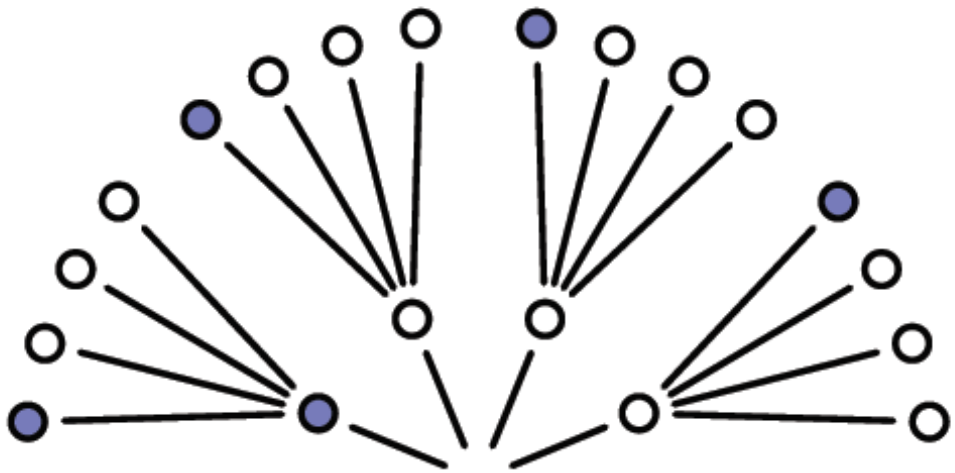
Conjecture: 

Data: 

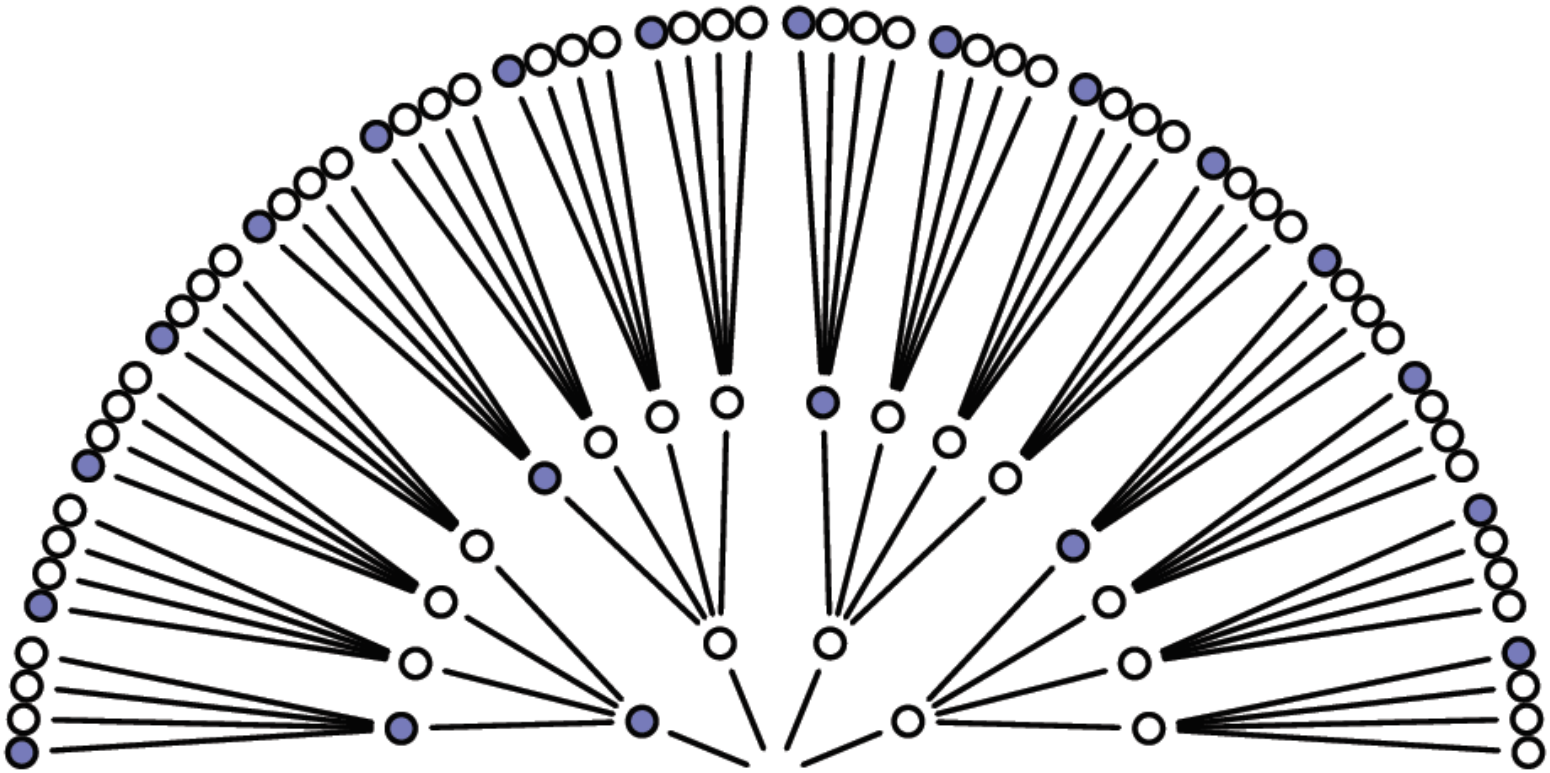


Conjecture: 

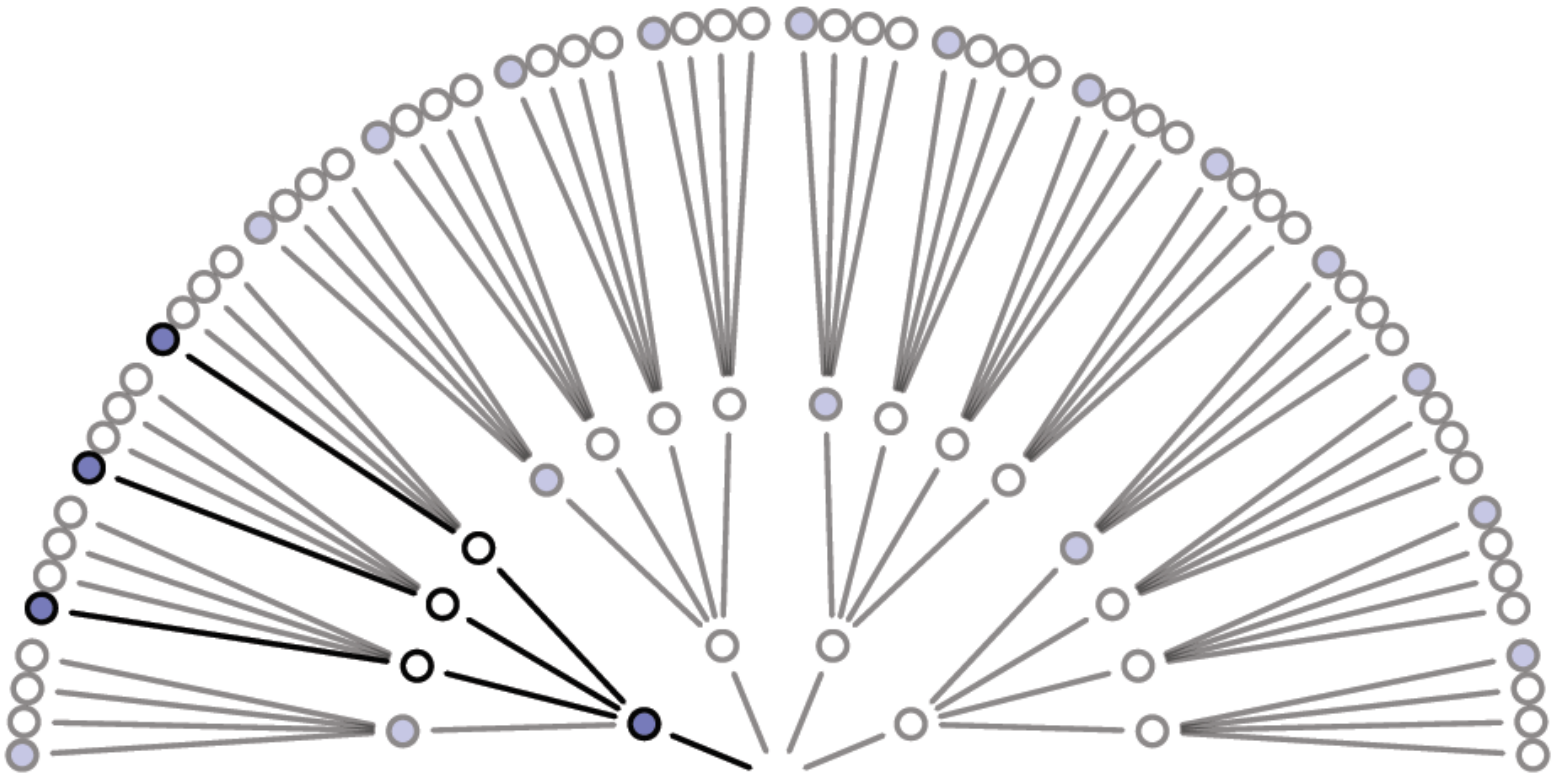
Data: 



Conjecture: ●○○○ Data: ●○●



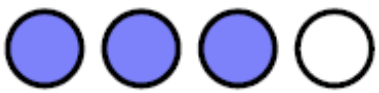
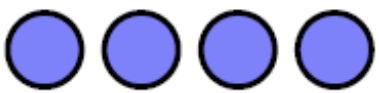
Conjecture: ●○○○ Data: ●○●

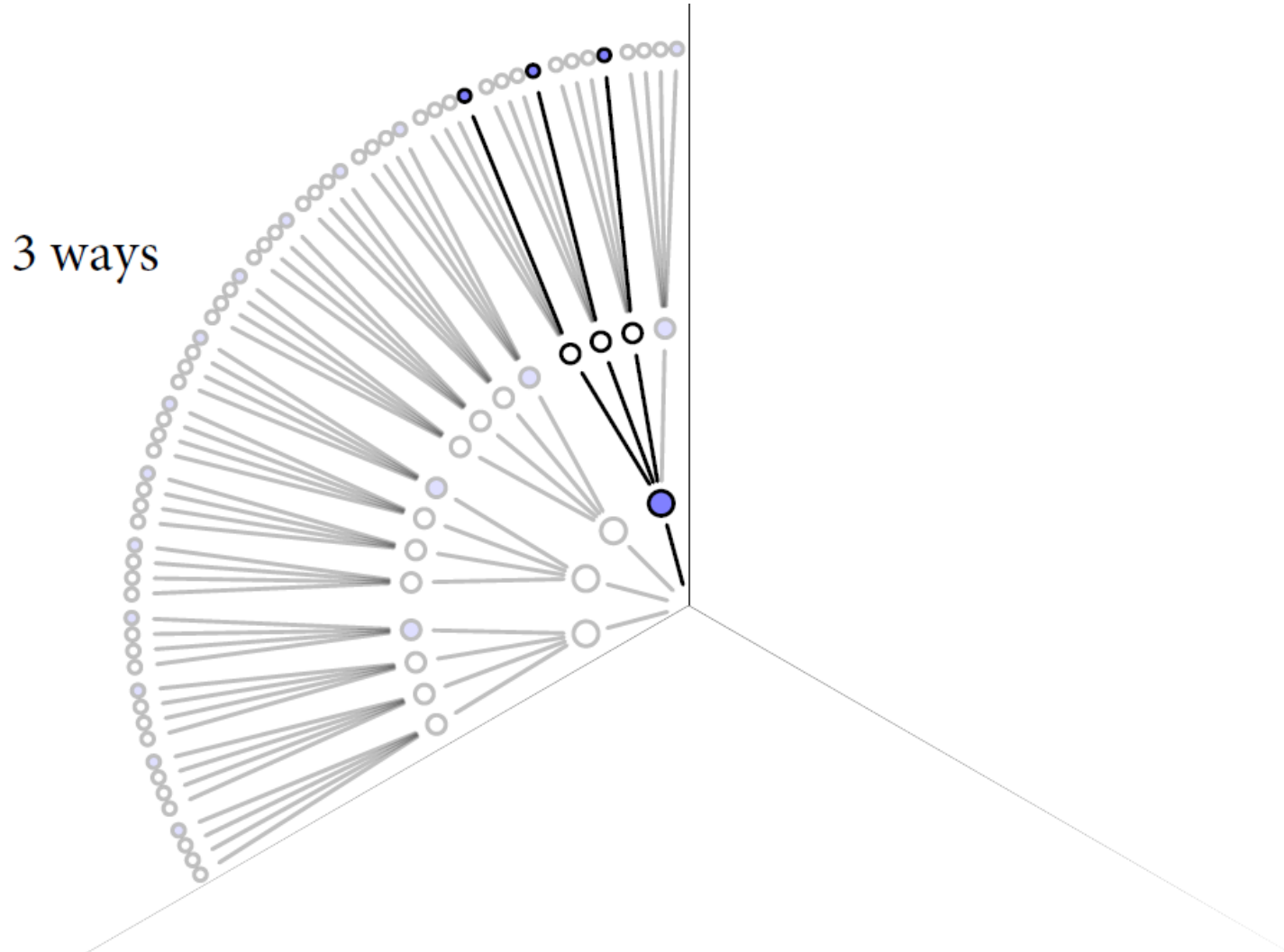


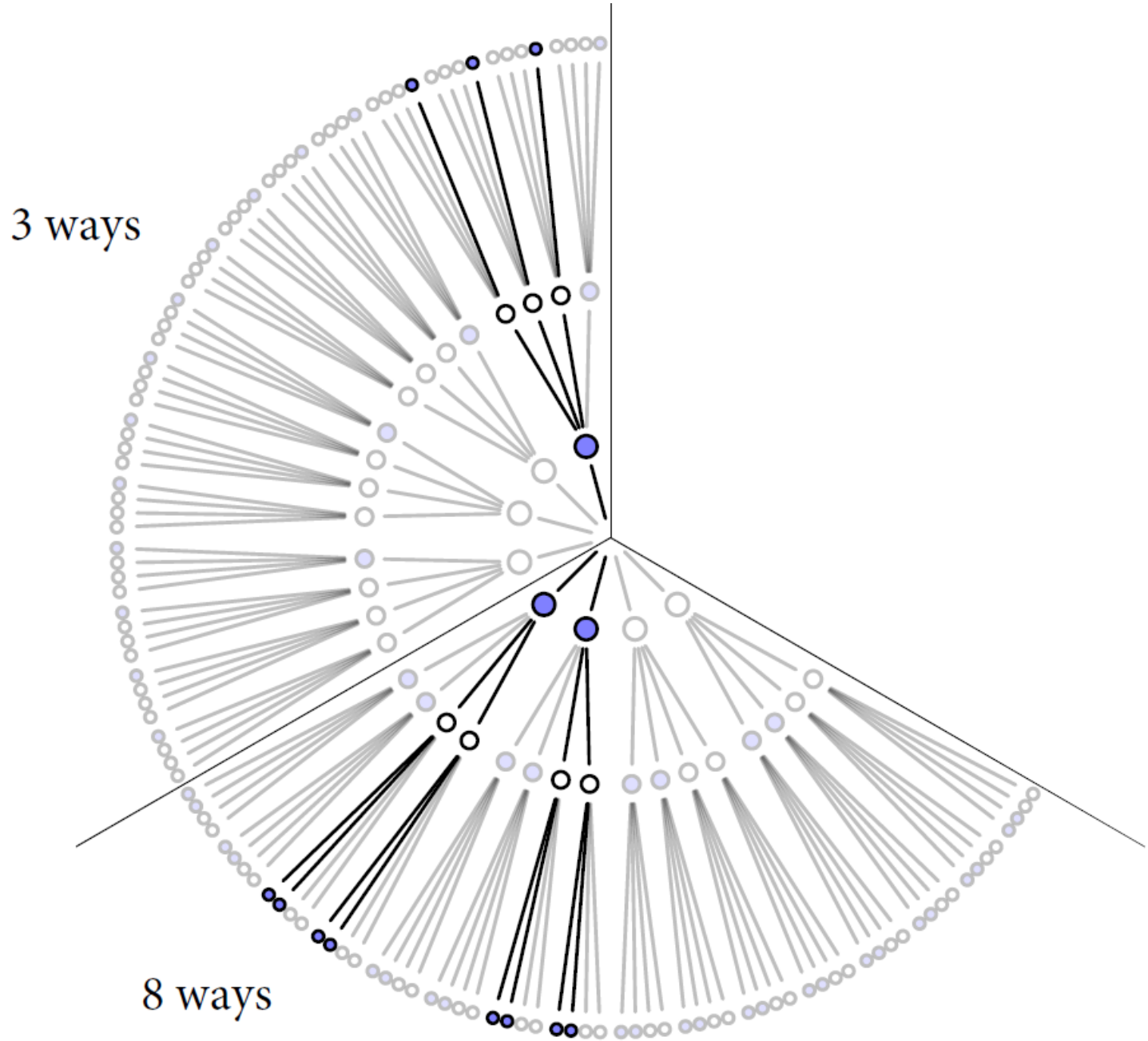
3 paths consistent with data

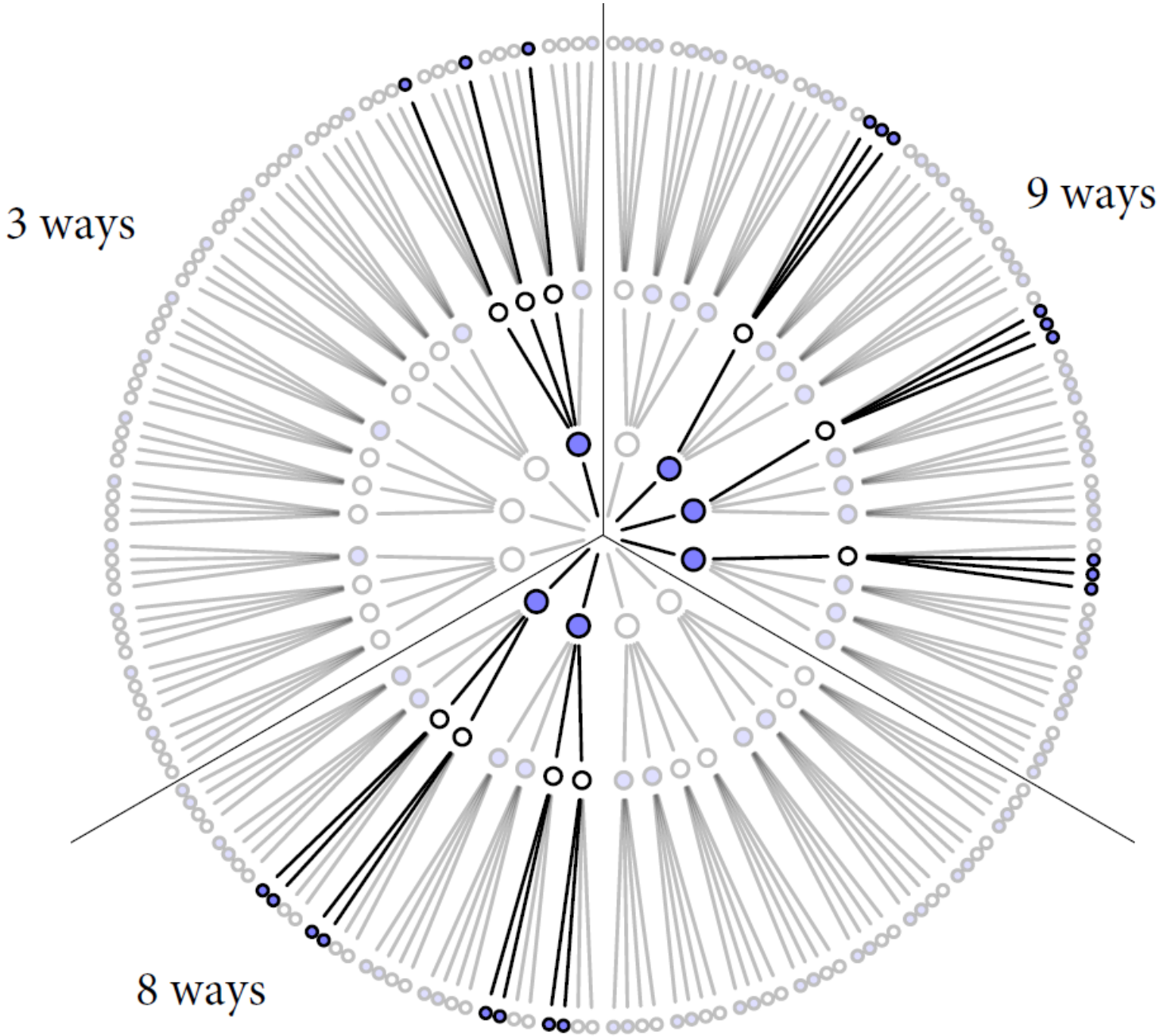
Illustration

Possible contents:	Ways to produce ●○●
(1) ○○○○	?
(2) ●○○○	3
(3) ●●○○	?
(4) ●●●○	?
(5) ●●●●	?

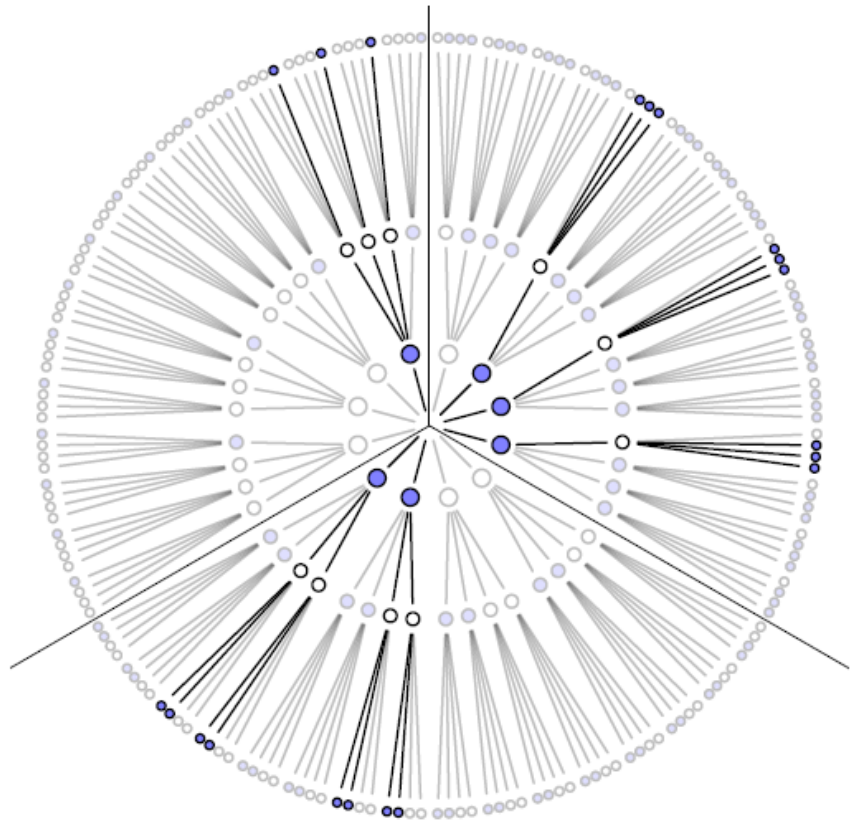
Possible contents:	Ways to produce 
(1) 	0
(2) 	3
(3) 	?
(4) 	?
(5) 	0


3 ways


3 ways
8 ways


3 ways
9 ways
8 ways

Conjecture	Ways to produce ●○●
[○○○○]	$0 \times 4 \times 0 = 0$
[●○○○]	$1 \times 3 \times 1 = 3$
[●●○○]	$2 \times 2 \times 2 = 8$
[●●●○]	$3 \times 1 \times 3 = 9$
[●●●●]	$4 \times 0 \times 4 = 0$



Updating

Another draw from the bag: ●

Conjecture	Ways to produce ●	Previous counts	New count
[○○○○]	0	0	$0 \times 0 = 0$
[●○○○]	1	3	$3 \times 1 = 3$
[●●○○]	2	8	$8 \times 2 = 16$
[●●●○]	3	9	$9 \times 3 = 27$
[●●●●]	4	0	$0 \times 4 = 0$

Using other Information

Factory says: ● marbles rare, but every bag contains at least one.

Conjecture	Factory count
[○○○○]	0
[●○○○]	3
[●●○○]	2
[●●●○]	1
[●●●●]	0

Using other Information

Factory says: ● marbles rare.

Conjecture	Prior ways	Factory count	New count
[○○○○]	0	0	$0 \times 0 = 0$
[●○○○]	3	3	$3 \times 3 = 9$
[●●○○]	16	2	$16 \times 2 = 32$
[●●●○]	27	1	$27 \times 1 = 27$
[●●●●]	0	0	$0 \times 0 = 0$

Counts to Plausibility

Unglamorous Basis of Probability:

Things that can happen more ways are more plausible.

Possible composition	p	ways to produce data	plausibility
[○○○○]	0	0	0
[●○○○]	0.25	3	0.15
[●●○○]	0.5	8	0.40
[●●●○]	0.75	9	0.45
[●●●●]	1	0	0

Counts to Plausibility

Possible composition	p	ways to produce data	plausibility
[○○○○]	0	0	0
[●○○○]	0.25	3	0.15
[●●○○]	0.5	8	0.40
[●●●○]	0.75	9	0.45
[●●●●]	1	0	0

Plausibility is Probability: Set of non-negative real numbers that sum up to one.

Probability Theory provides us with a set of shortcuts for counting possibilities.

Key Idea in Bayesian Analysis

The chance that something is TRUE changes as new information becomes available.

Monty Hall Problem

- At the beginning, the car can be behind any of the three doors with equal probability. That is,

$$P(D_1) = P(D_2) = P(D_3) = 1/3$$

- Let's say we choose door number 1, and Monty open door number 2.

- Now let's write down the conditional probability of opening, OD_2, given the three possibilities (i.e., D_1, D_2, and D_3):

$$\begin{aligned} P(OD_2|D_1) &= 1/2 \\ P(OD_2|D_2) &= 0 \\ P(OD_2|D_3) &= 1 \end{aligned}$$

- Now using the law of total probability we can find the marginal probability for opening door number 2:

$$P(OD_2) = 1/2 \times 1/3 + 1/2 \times 0 + 1/2 \times 1/3 = 1/2$$

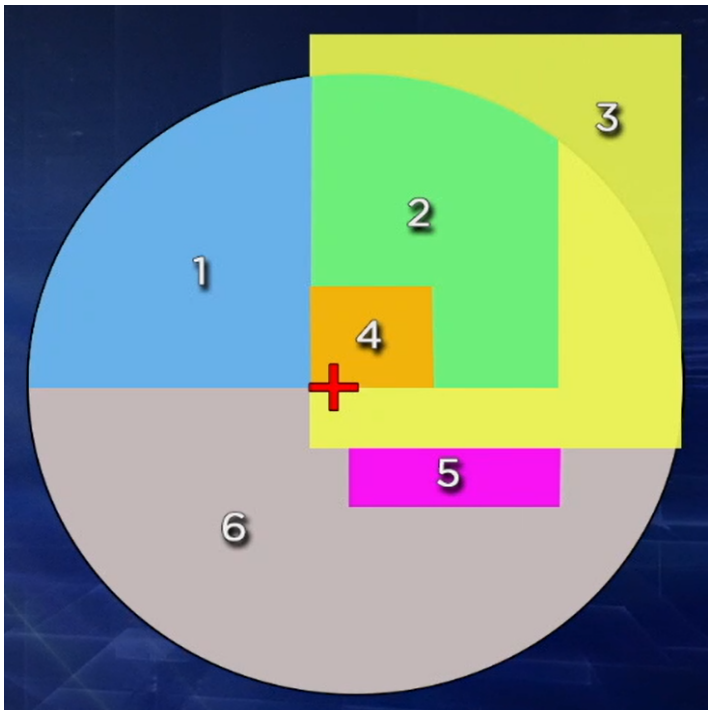
Monty Hall Problem

- Using Bayes' theorem, we have:

$$
\begin{aligned}
P(D_3|OD_2) &= \frac{p(D_3)P(OD_2|D_3)}{P(OD_2)} \\
&= \frac{1/3 \times 1}{1/2} = 2/3 \\
P(D_1|OD_2) &= \frac{P(D_1)P(openD_2|D_1)}{P(OD_2)} \\
&= \frac{1/3 \times 1/2}{1/2} = 1/3
\end{aligned}
$$

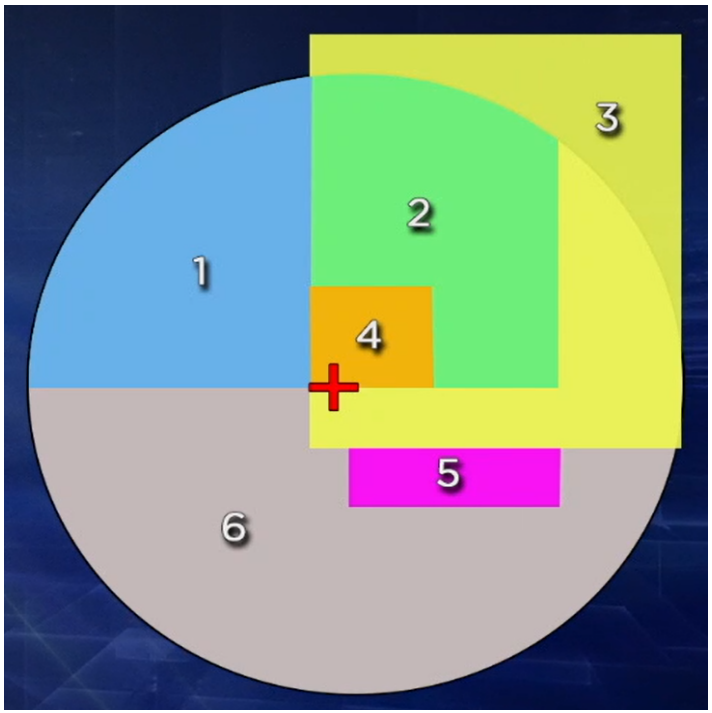
- Therefore, probability of winning doubles if we switch.

Simplified Search for AF 447



- **Stage 1:** Looked for pingers in the rectangular region including Zones 2, 3, and 4
- **Stage 2:** Looked for pingers in Zone 5
- **Stage 3:** Looked for pingers in the wedge including Zones 1, 2
- Zone 6 was never searched

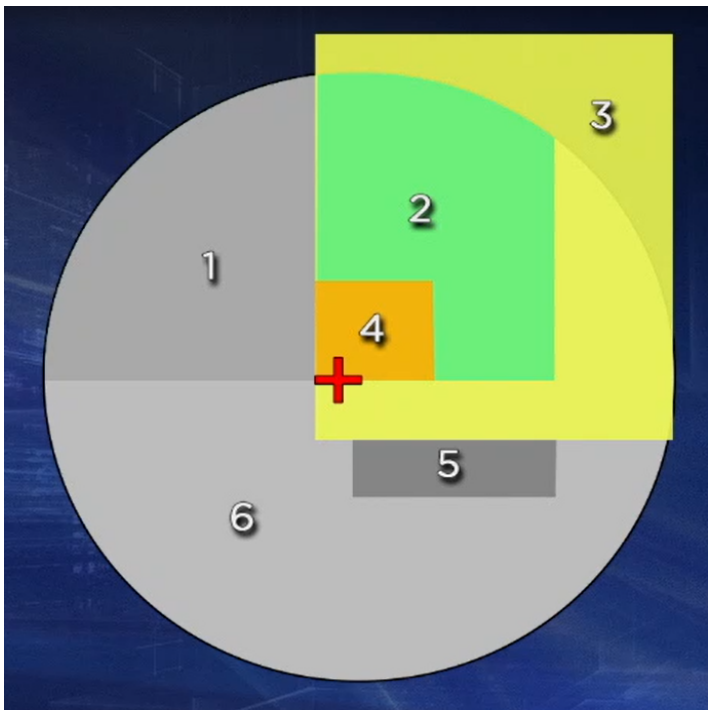
Simplified Search for AF 447



Zone	Probability
1	0.10
2	0.20
3	0.25
4	0.20
5	0.15
6	0.10

What do we know so far?

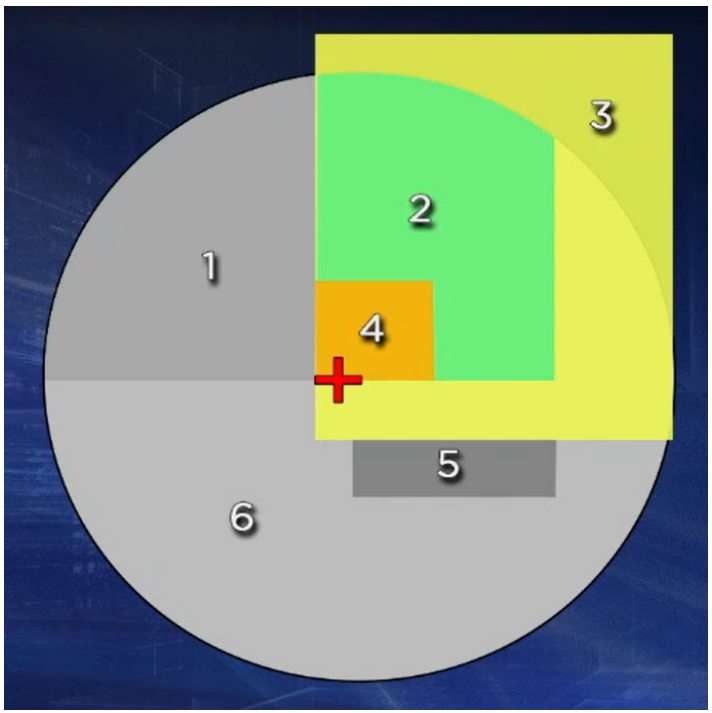
Stage 1: Looked for pingers in the rectangular region including Zones 2, 3, and 4



Given that you got this search result, how likely is it that the wreck is actually there?

What do we know so far?

Stage 1: Looked for pingers in the rectangular region including Zones 2, 3, and 4



Given that you got this search result, how likely is it that the wreck is actually there?

P(wreck is in Zone 2 | wreck is not found in Zones 2, 3, 4)

P(wreck is in Zone 2 | search fails)

search fails = wreck is not found in Zones 2, 3, 4

$$P(\text{wreck is in Zone 2 and search fails}) = P(\text{search fails} \mid \text{Zone 2})P(\text{Zone 2})$$

P(wreck is in Zone 2 | search fails)

search fails = wreck is not found in Zones 2, 3, 4

$$P(\text{wreck is in Zone 2 and search fails}) = P(\text{search fails} \mid \text{Zone 2})P(\text{Zone 2})$$

$$P(\text{wreck is in Zone 1 and search fails}) = P(\text{search fails} \mid \text{Zone 1})P(\text{Zone 1})$$

P(wreck is in Zone 2 | search fails)

search fails = wreck is not found in Zones 2, 3, 4

$$P(\text{wreck is in Zone 2 and search fails}) = P(\text{search fails} \mid \text{Zone 2})P(\text{Zone 2})$$

$$P(\text{wreck is in Zone 1 and search fails}) = P(\text{search fails} \mid \text{Zone 1})P(\text{Zone 1})$$

$$P(\text{wreck is in Zone 3 and search fails}) = P(\text{search fails} \mid \text{Zone 3})P(\text{Zone 3})$$

P(wreck is in Zone 2 | search fails)

search fails = wreck is not found in Zones 2, 3, 4

$$P(\text{wreck is in Zone 2 and search fails}) = P(\text{search fails} \mid \text{Zone 2})P(\text{Zone 2})$$

$$P(\text{wreck is in Zone 1 and search fails}) = P(\text{search fails} \mid \text{Zone 1})P(\text{Zone 1})$$

$$P(\text{wreck is in Zone 3 and search fails}) = P(\text{search fails} \mid \text{Zone 3})P(\text{Zone 3})$$

$$P(\text{wreck is in Zone 4 and search fails}) = P(\text{search fails} \mid \text{Zone 4})P(\text{Zone 4})$$

P(wreck is in Zone 2 | search fails)

search fails = wreck is not found in Zones 2, 3, 4

$$P(\text{wreck is in Zone 2 and search fails}) = P(\text{search fails} \mid \text{Zone 2})P(\text{Zone 2})$$

$$P(\text{wreck is in Zone 1 and search fails}) = P(\text{search fails} \mid \text{Zone 1})P(\text{Zone 1})$$

$$P(\text{wreck is in Zone 3 and search fails}) = P(\text{search fails} \mid \text{Zone 3})P(\text{Zone 3})$$

$$P(\text{wreck is in Zone 4 and search fails}) = P(\text{search fails} \mid \text{Zone 4})P(\text{Zone 4})$$

$$P(\text{wreck is in Zone 5 and search fails}) = P(\text{search fails} \mid \text{Zone 5})P(\text{Zone 5})$$

P(wreck is in Zone 2 | search fails)

search fails = wreck is not found in Zones 2, 3, 4

$$P(\text{wreck is in Zone 2 and search fails}) = P(\text{search fails} \mid \text{Zone 2})P(\text{Zone 2})$$

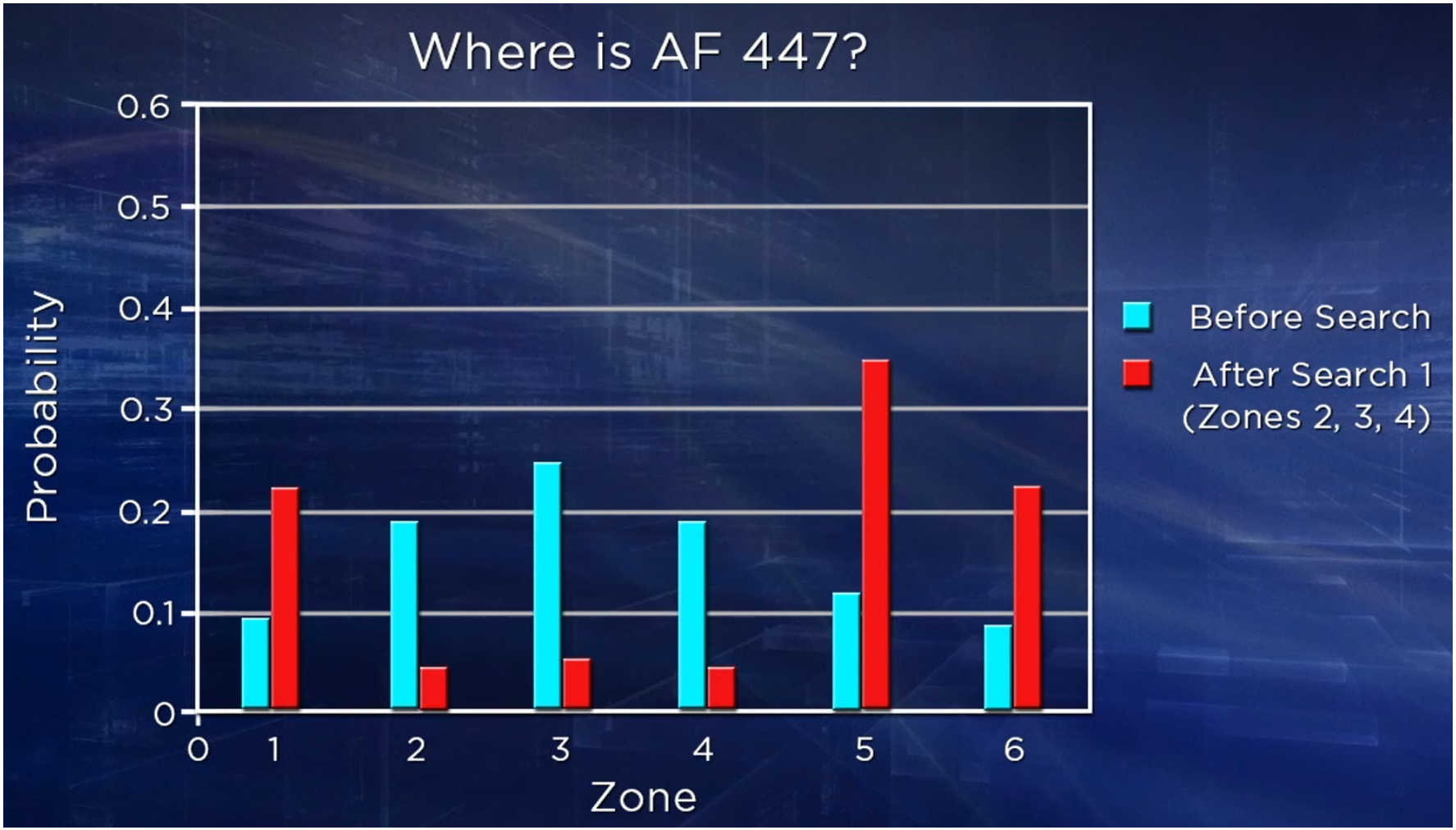
$$P(\text{wreck is in Zone 1 and search fails}) = P(\text{search fails} \mid \text{Zone 1})P(\text{Zone 1})$$

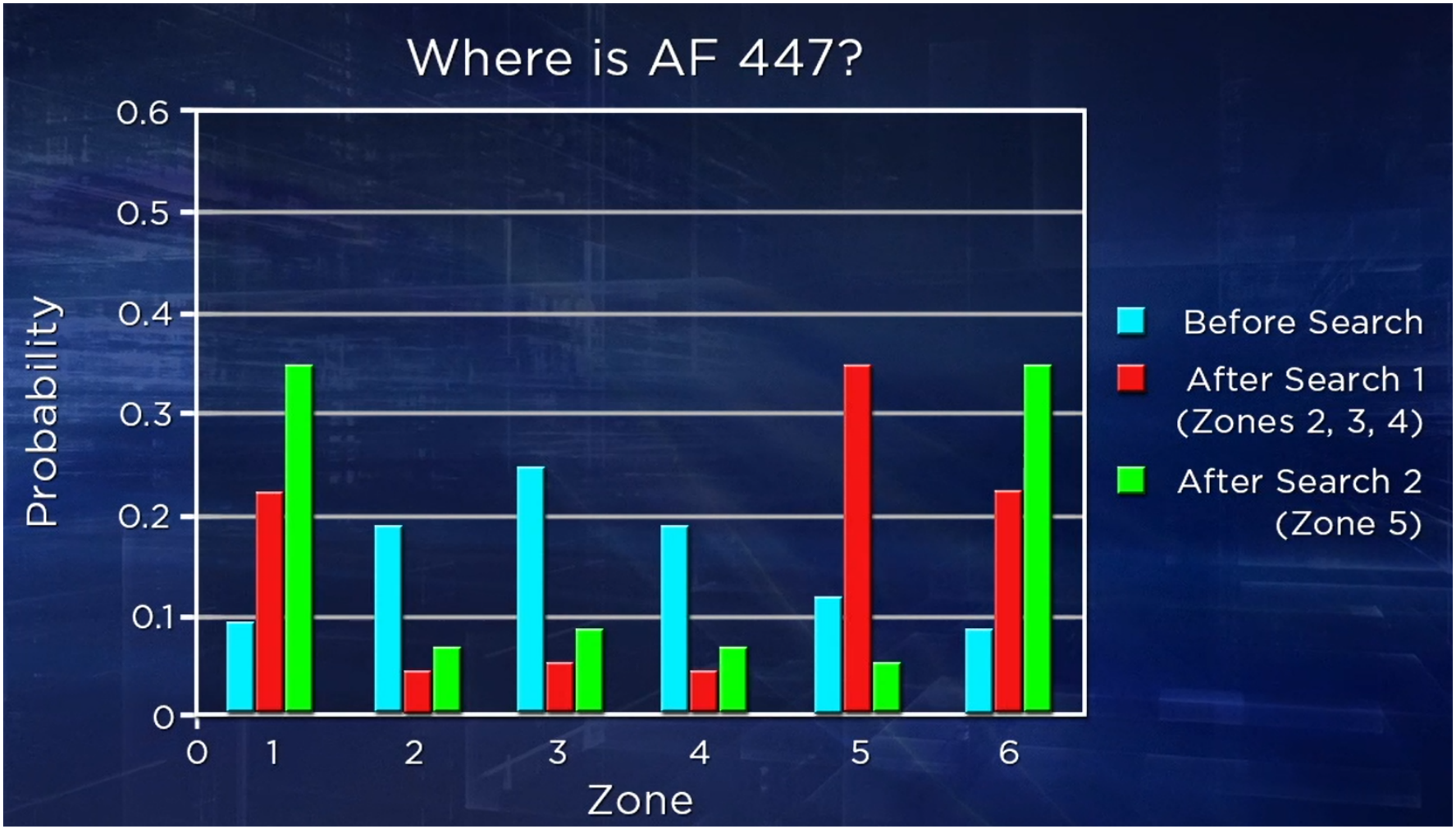
$$P(\text{wreck is in Zone 3 and search fails}) = P(\text{search fails} \mid \text{Zone 3})P(\text{Zone 3})$$

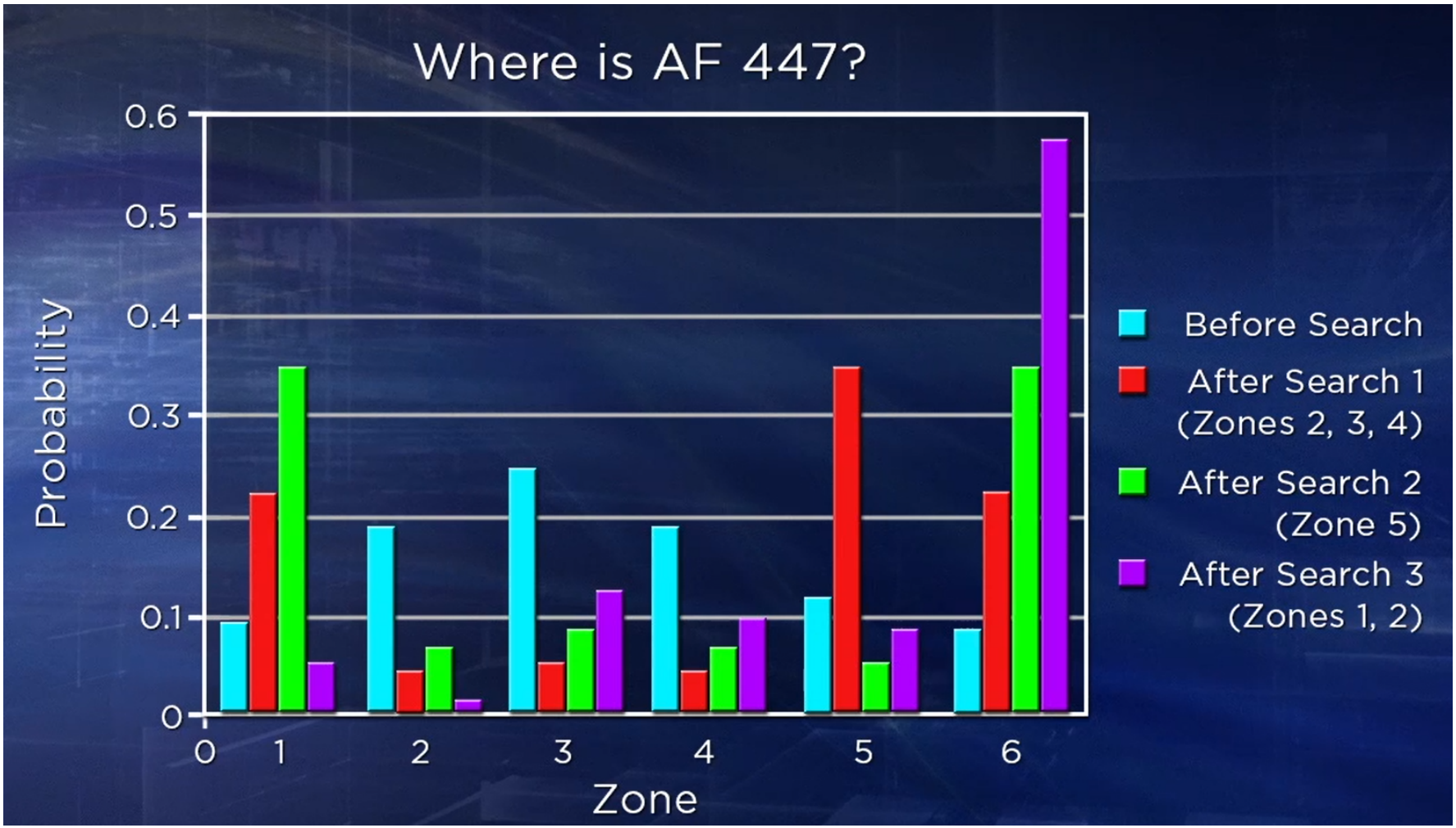
$$P(\text{wreck is in Zone 4 and search fails}) = P(\text{search fails} \mid \text{Zone 4})P(\text{Zone 4})$$

$$P(\text{wreck is in Zone 5 and search fails}) = P(\text{search fails} \mid \text{Zone 5})P(\text{Zone 5})$$

$$P(\text{wreck is in Zone 6 and search fails}) = P(\text{search fails} \mid \text{Zone 6})P(\text{Zone 6})$$


Where is AF 447?
Probability
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
Zone
Before Search
After Search 1
(Zones 2, 3, 4)


Where is AF 447?
Probability
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
Zone
Before Search
After Search 1
(Zones 2, 3, 4)
After Search 2
(Zone 5)


Where is AF 447?
Probability
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
Zone
Before Search
After Search 1
(Zones 2, 3, 4)
After Search 2
(Zone 5)
After Search 3
(Zones 1, 2)

Questions to Think About

1. How does Bayesian probability relate to **drug testing**? What procedures do you think should be followed when someone tests positive for an illegal substance?

2. Think of several examples where people commonly make the error of reversing conditional probabilities — that is, they confound the chance of A when B is true and the chance of B when A is true.

A Bayesian Statistician

Let

- $\mathbf{y}$ be the data we will collect from an experiment,
- K be everything we know for certain about the world (aside from y), and
- θ be anything we don't know for certain.

A Bayesian Statistician

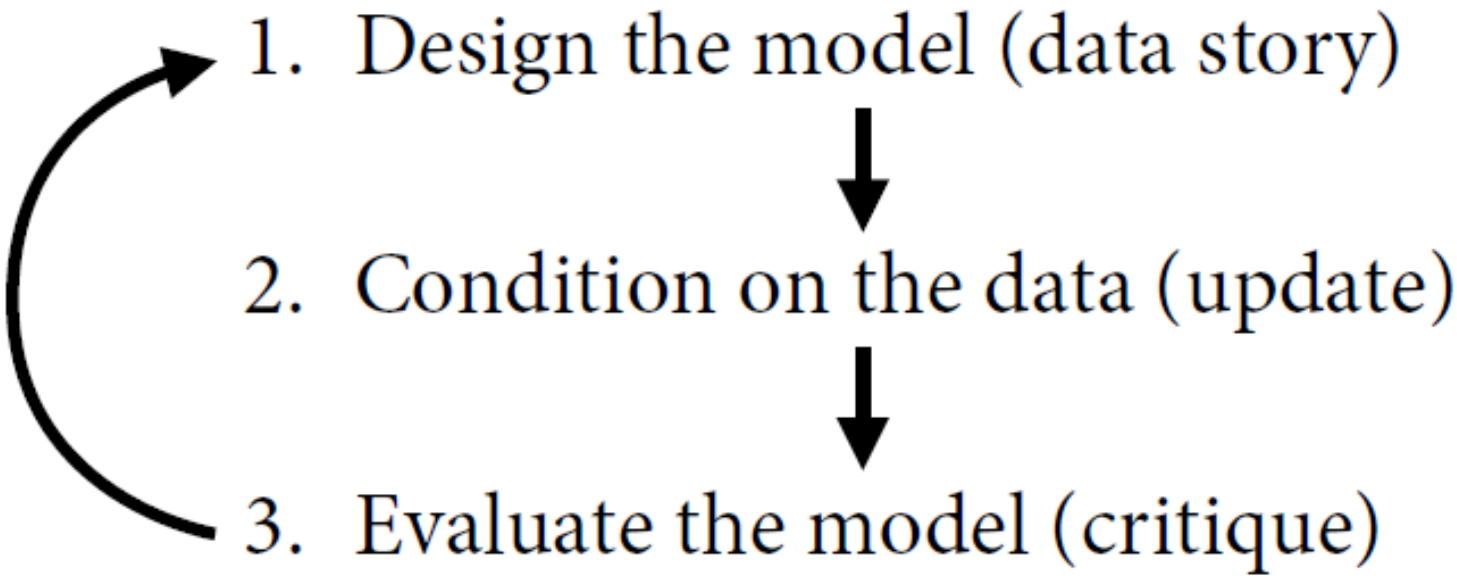
Let

- $\mathbf{y}$ be the data we will collect from an experiment,
- K be everything we know for certain about the world (aside from y), and
- θ be anything we don't know for certain.

A Bayesian statistician is an individual who makes decisions based on the probability distribution of those things we don't know conditional on what we know, i.e.

$$p(\theta \mid \mathbf{y}, K).$$

How to use probability to do statistical modeling?



Data Story



Nine tosses of the globe:

W L W W W L W L W

Design > Condition > Evaluate

- Data story motivates the model

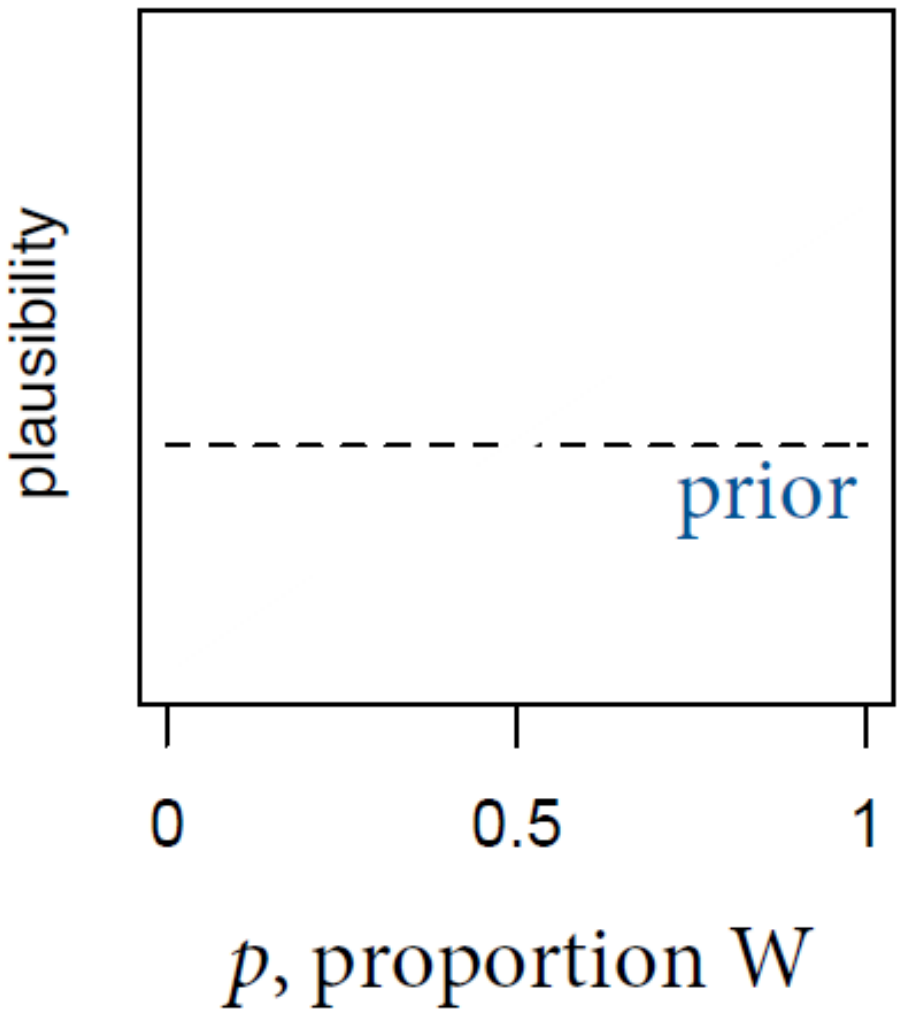
How do the data arise?

- For W L W W W L W L W:
  - Some true proportion of water, p
  - Toss globe, probability p of observing W, $1 - p$ of L
  - Each toss therefore independent of other tosses

- Translate data story into probability statements

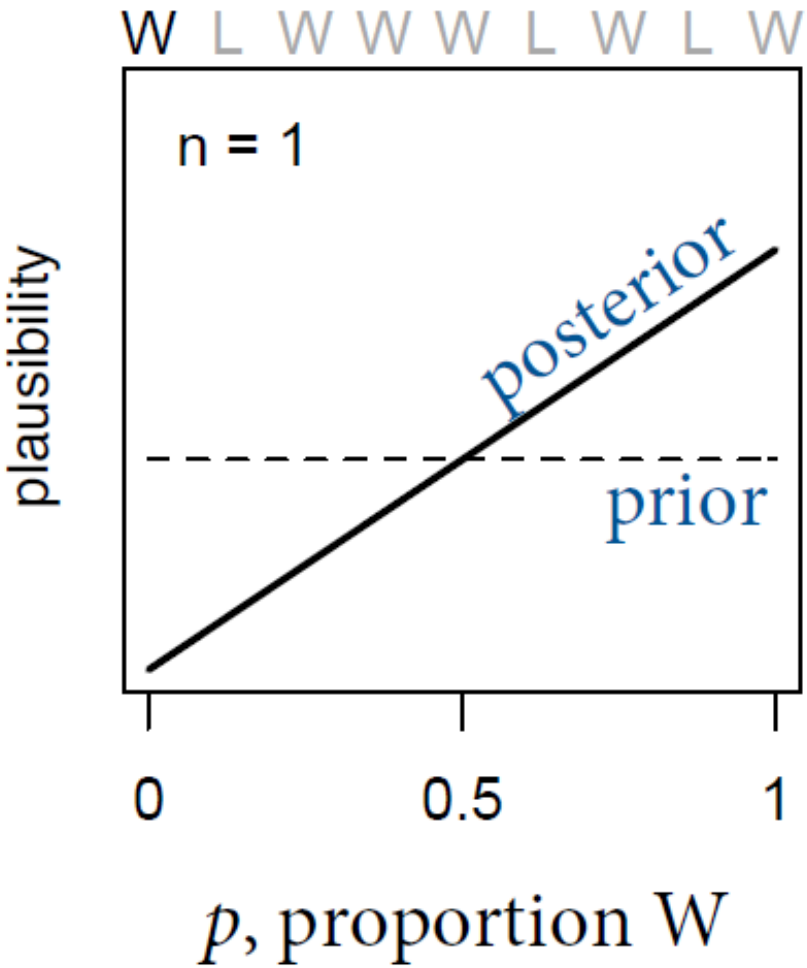
Design > **Condition** > Evaluate

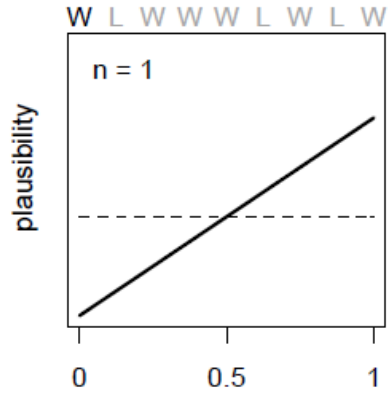
Bayesian updating converts prior into posterior

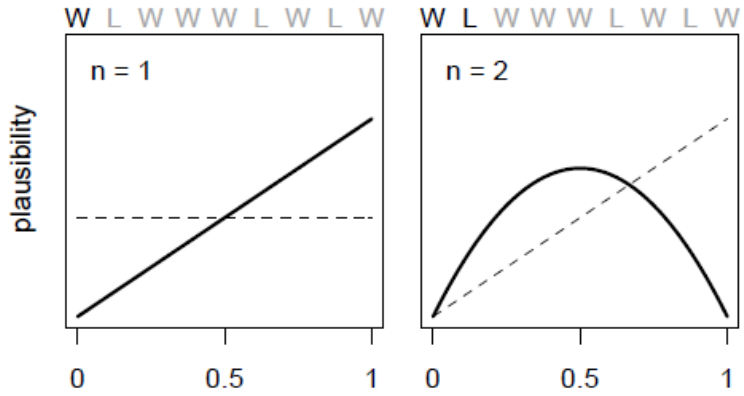


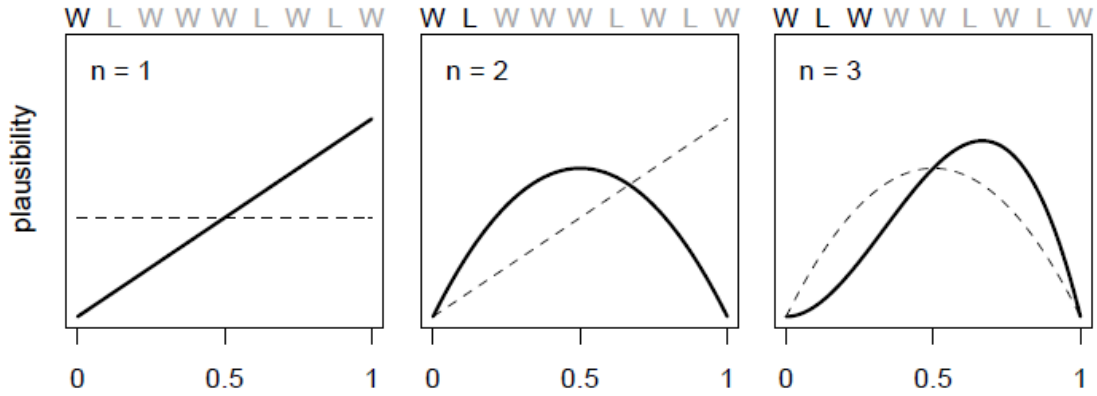
Design > **Condition** > Evaluate

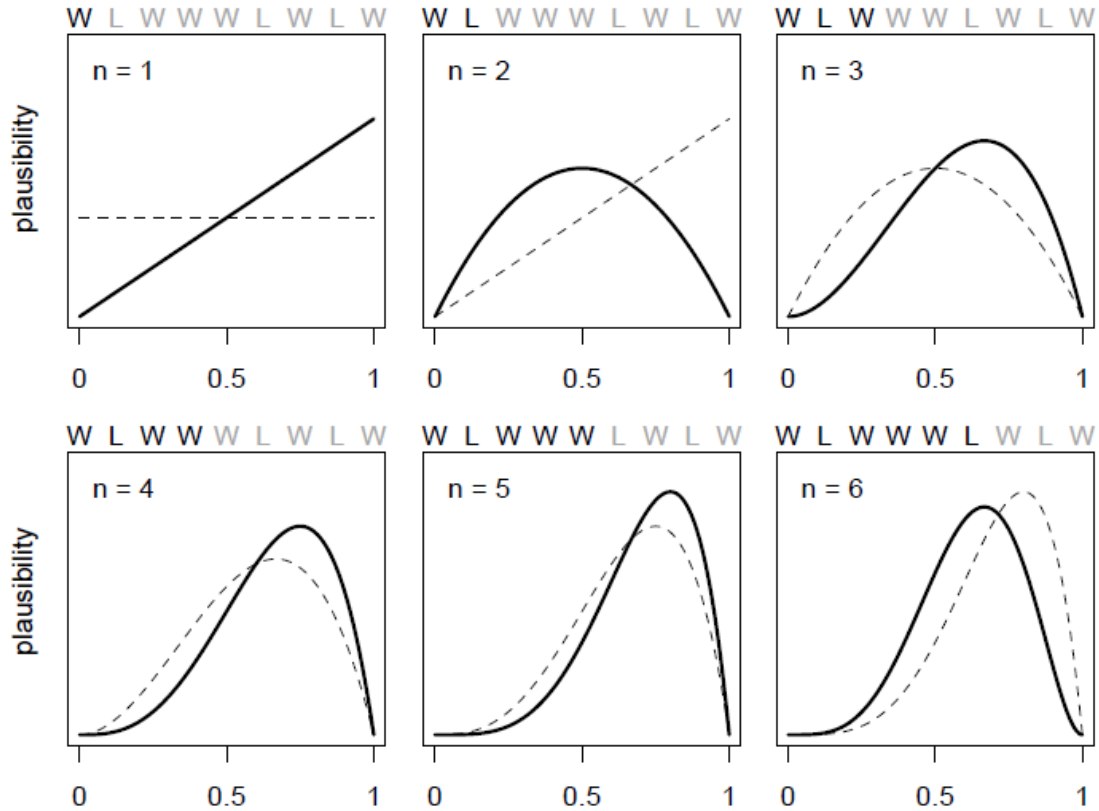
- **Condition on data to update information state:** New confidence in each value of p, conditional on data

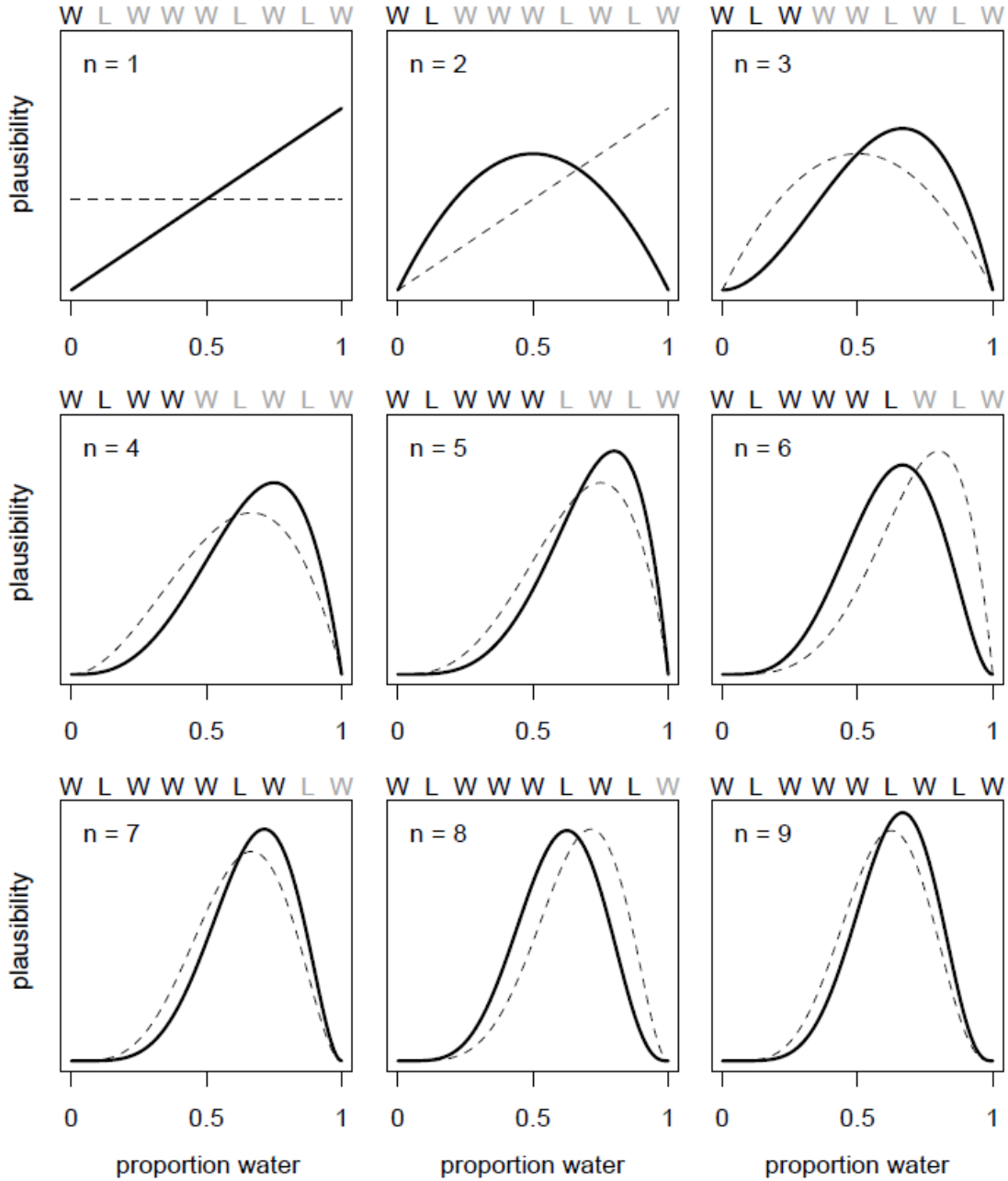



W L W W W L W L W
n = 1
plausibility
0
0.5
1


W L W W W L W L W
n = 1
plausibility
0
0.5
1
W L W W W L W L W
n = 2
0
0.5
1


W L W W W L W L W
n = 1
plausibility
0
0.5
1
W L W W W L W L W
n = 2
0
0.5
1
W L W W W L W L W
n = 3
0
0.5
1


W L W W W L W L W
n = 1
n = 2
n = 3
n = 4
n = 5
n = 6
plausibility
plausibility
0
0.5
1


W L W W W L W L W
n = 1
n = 2
n = 3
n = 4
n = 5
n = 6
n = 7
n = 8
n = 9
plausibility
0
0.5
1
proportion water

Bayesian Updating

We can use our previous posterior as the **new prior** and derive a **new posterior**:

Bayesian Updating

We can use our previous posterior as the **new prior** and derive a **new posterior**:

$$p(\theta|\boldsymbol{x}_1, \boldsymbol{x}_2) \quad \propto \quad p(\theta \mid \boldsymbol{x}_1) L(\theta \mid \boldsymbol{x}_2)$$

Bayesian Updating

We can use our previous posterior as the **new prior** and derive a **new posterior**:

$$\begin{aligned} p(\theta|\boldsymbol{x}_1, \boldsymbol{x}_2) &\propto p(\theta \mid \boldsymbol{x}_1) L(\theta \mid \boldsymbol{x}_2) \\ &\propto p(\theta) L(\theta \mid \boldsymbol{x}_1) L(\theta \mid \boldsymbol{x}_2) \end{aligned}$$

Bayesian Updating

We can use our previous posterior as the **new prior** and derive a **new posterior**:

$$\begin{aligned} p(\theta|\boldsymbol{x}_1, \boldsymbol{x}_2) &\propto p(\theta \mid \boldsymbol{x}_1) L(\theta \mid \boldsymbol{x}_2) \\ &\propto p(\theta) L(\theta \mid \boldsymbol{x}_1) L(\theta \mid \boldsymbol{x}_2) \\ &\propto p(\theta) L(\theta \mid \boldsymbol{x}_1, \boldsymbol{x}_2) \end{aligned}$$

(since $\boldsymbol{x}_1, \boldsymbol{x}_2$ independent)

Note this is the same posterior we would have obtained had $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ arrived at the same time!

Bayesian Updating

We can use our previous posterior as the **new prior** and derive a **new posterior**:

$$\begin{aligned} p(\theta|\boldsymbol{x}_1, \boldsymbol{x}_2) &\propto p(\theta \mid \boldsymbol{x}_1) L(\theta \mid \boldsymbol{x}_2) \\ &\propto p(\theta) L(\theta \mid \boldsymbol{x}_1) L(\theta \mid \boldsymbol{x}_2) \\ &\propto p(\theta) L(\theta \mid \boldsymbol{x}_1, \boldsymbol{x}_2) \end{aligned}$$

(since $\boldsymbol{x}_1, \boldsymbol{x}_2$ independent)

Note this is the same posterior we would have obtained had $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ arrived at the same time!

This "sequential updating" process can continue indefinitely in the Bayesian setup.

- Bayesian inference: Logical answer to a question in the form of a model

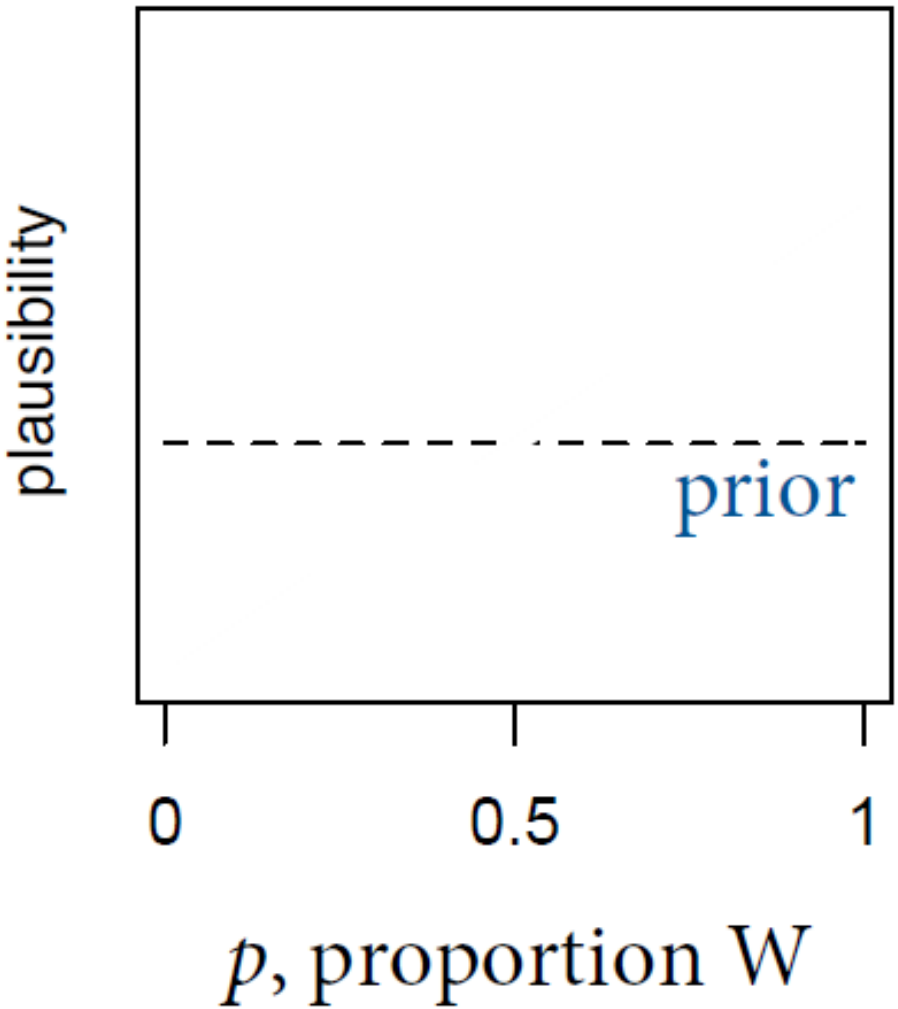
"How plausible is each proportion of water, given these data?"

- Model must be evaluated
  - Did it "malfunction"?
  - Does the answer make sense?
  - Does the question make sense?
  - Check sensitivity of answer to changes in assumptions

Likelihood



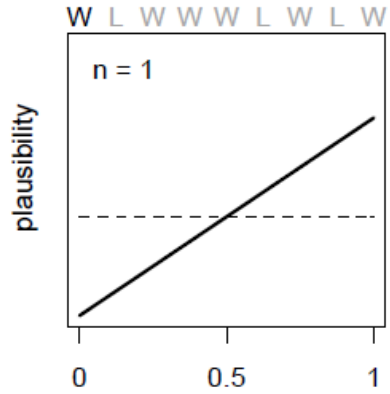
Prior Probability

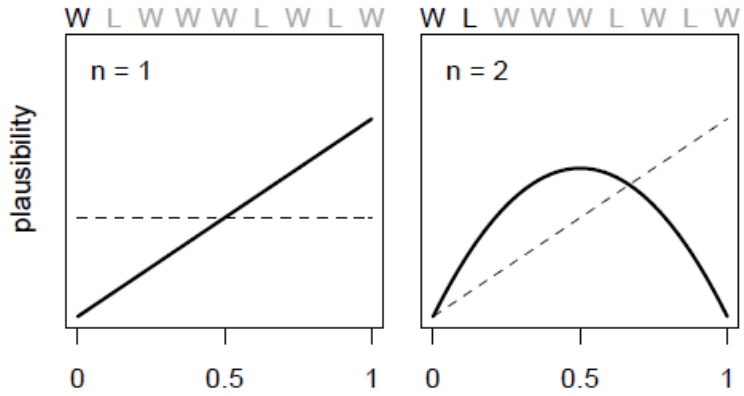


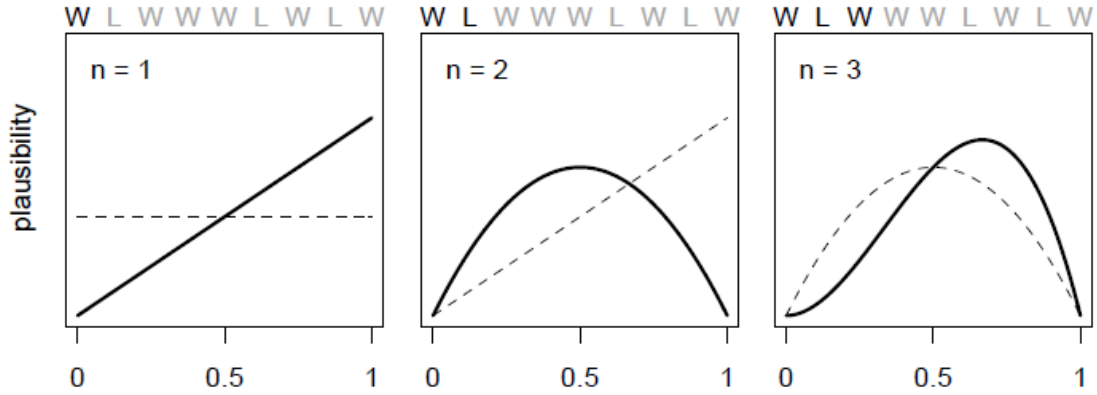
Joint Model

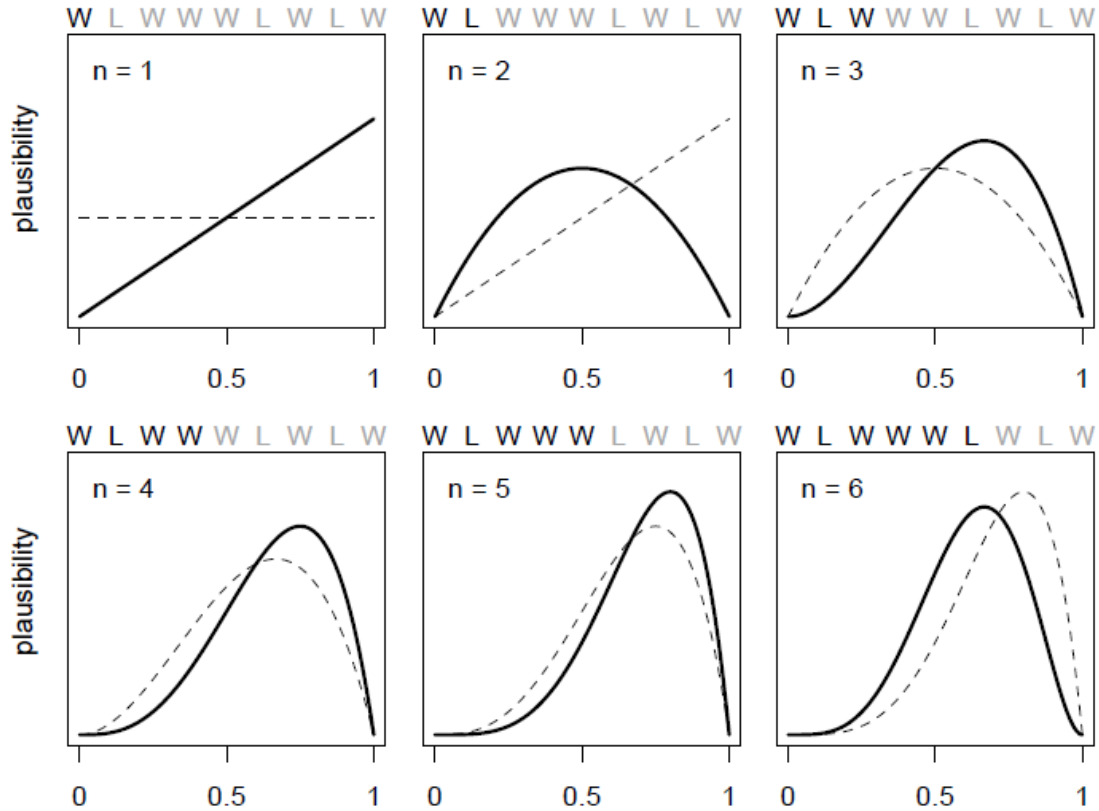
$$
\begin{aligned}
W \mid n, p &\sim \text{Binomial}(n, p) \\
p &\sim \text{Uniform}(0, 1)
\end{aligned}
$$

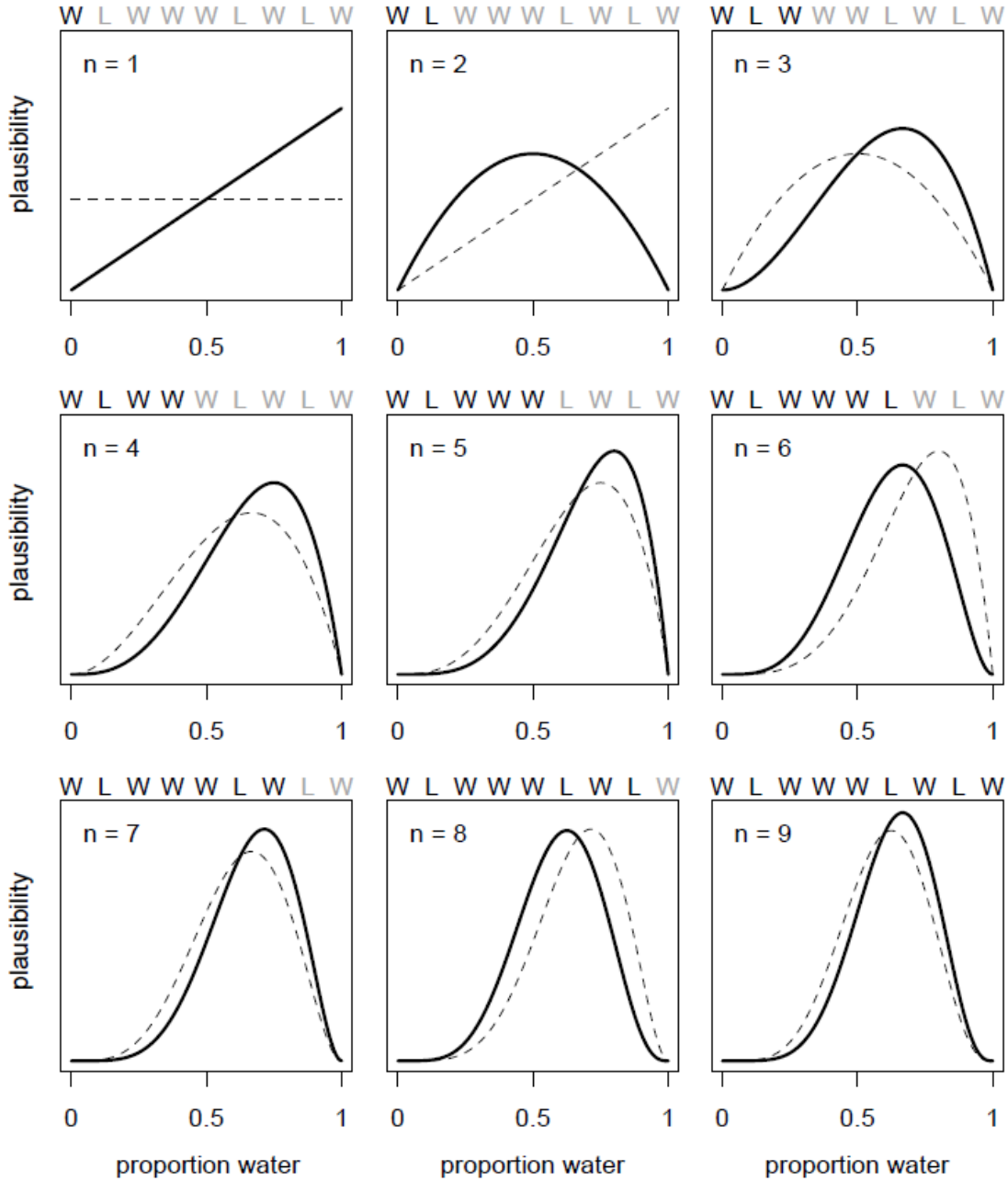
$$
p \mid n, W \sim ?
$$

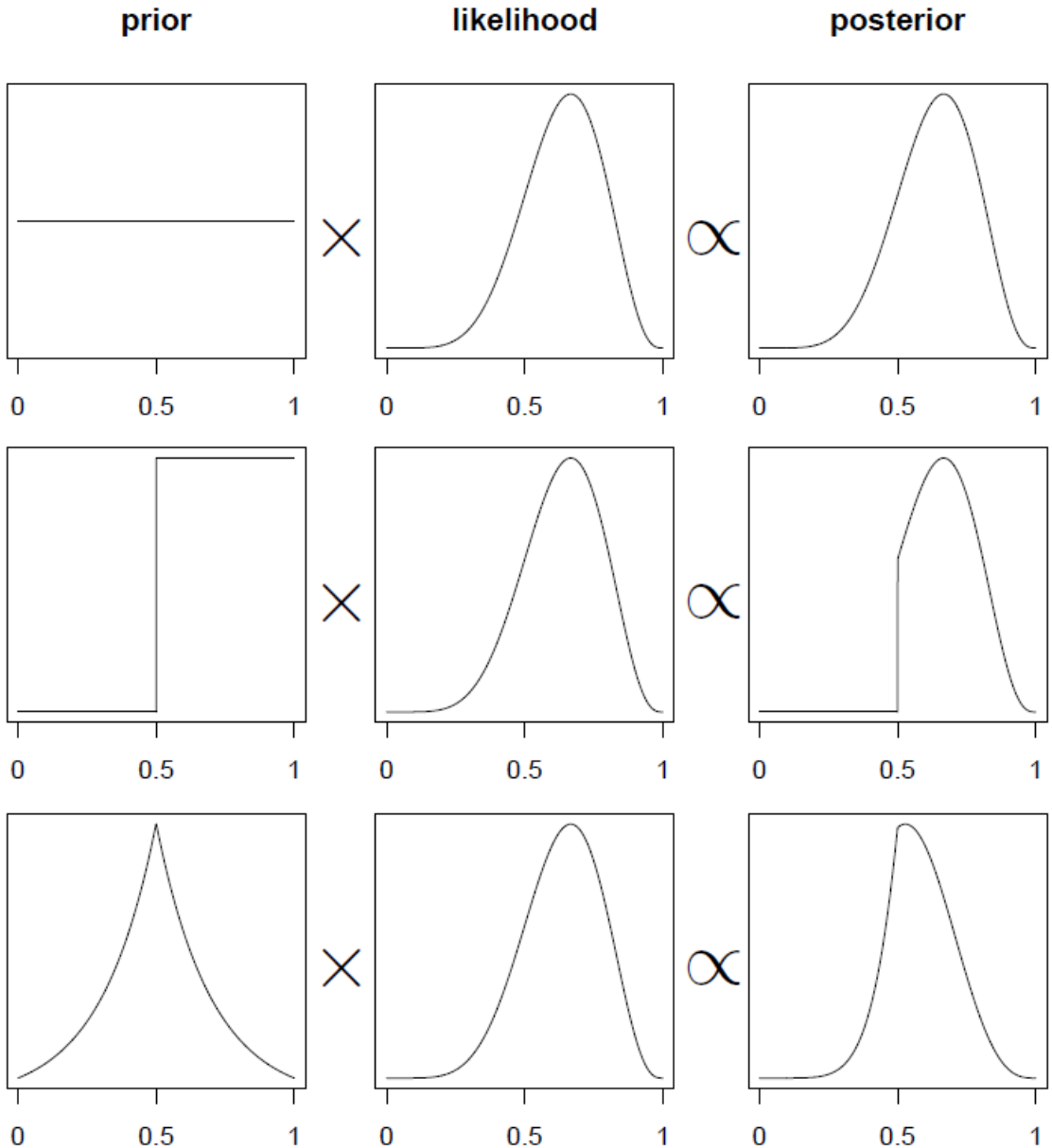

W L W W W L W L W
n = 1
plausibility
0
0.5
1


W L W W W L W L W
n = 1
plausibility
0
0.5
1
W L W W W L W L W
n = 2
0
0.5
1


W L W W W L W L W
n = 1
plausibility
0
0.5
1
W L W W W L W L W
n = 2
0
0.5
1
W L W W W L W L W
n = 3
0
0.5
1

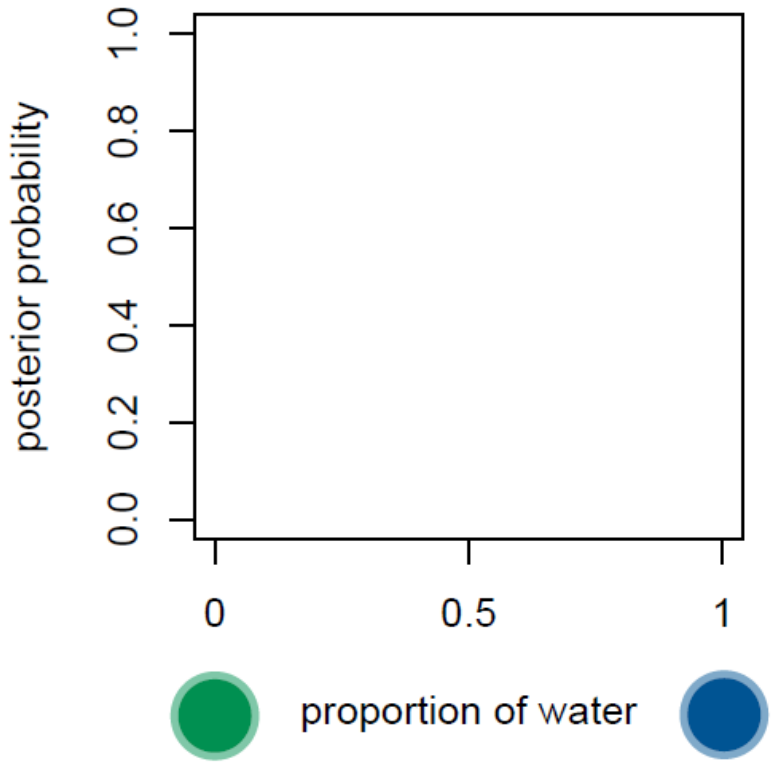

W L W W W L W L W
n = 1
n = 2
n = 3
n = 4
n = 5
n = 6
plausibility
plausibility
0
0.5
1


W L W W W L W L W
n = 1
n = 2
n = 3
n = 4
n = 5
n = 6
n = 7
n = 8
n = 9
plausibility
0
0.5
1
proportion water

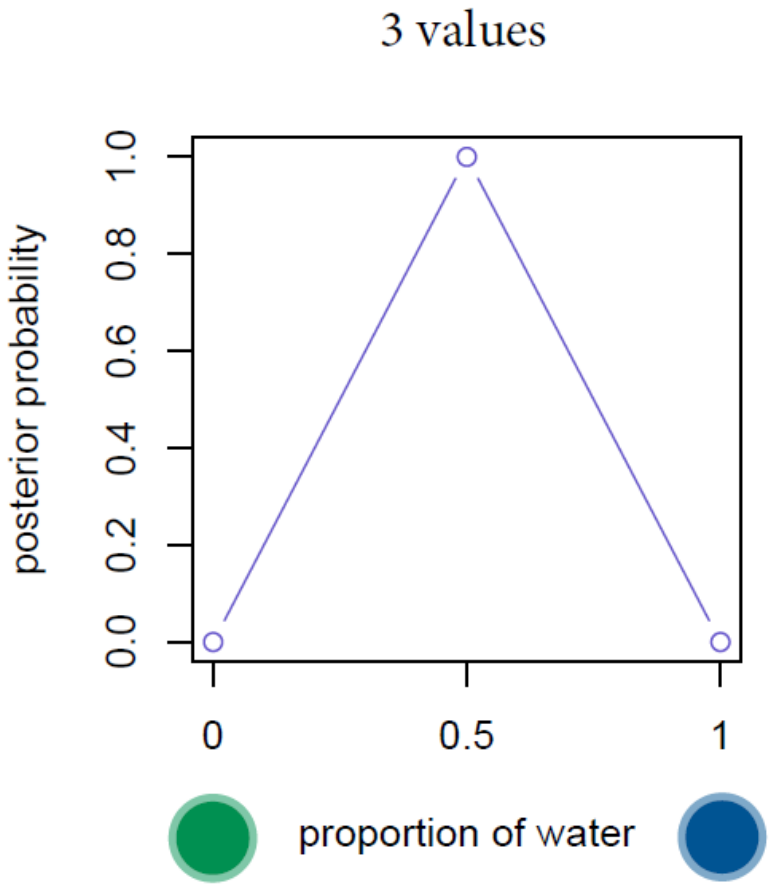

prior
likelihood
posterior
0
0.5
1

Computing the Posterior

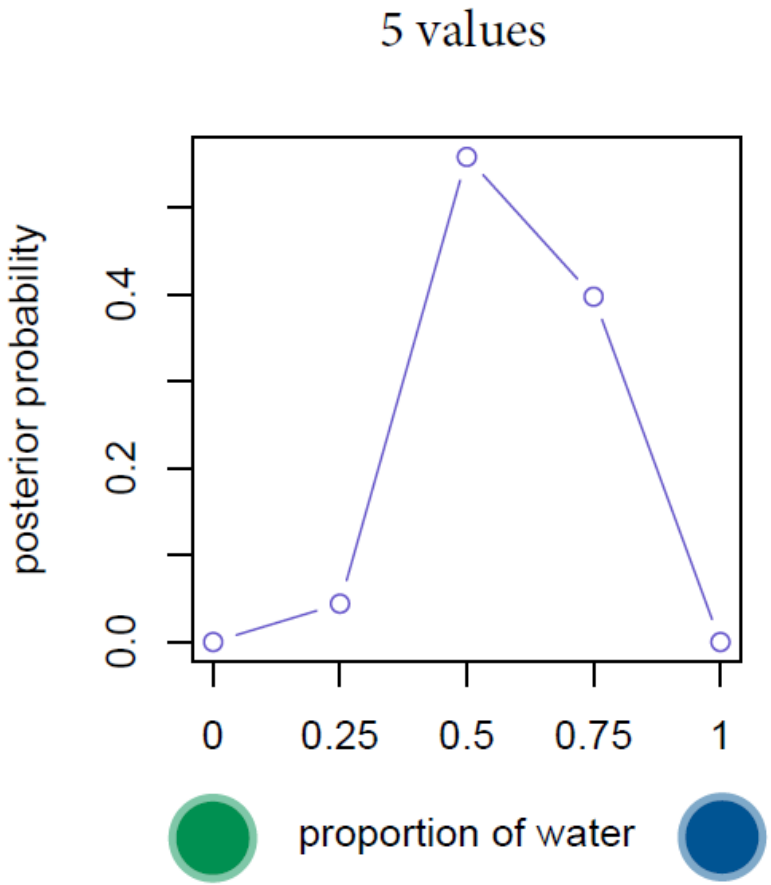
1. Analytical approach (often impossible)
2. Grid approximation (very intensive)
3. Quadratic approximation (limited)
4. Markov chain Monte Carlo (intensive)


posterior probability
0.0
0.2
0.4
0.6
0.8
1.0
0
0.5
1
proportion of water

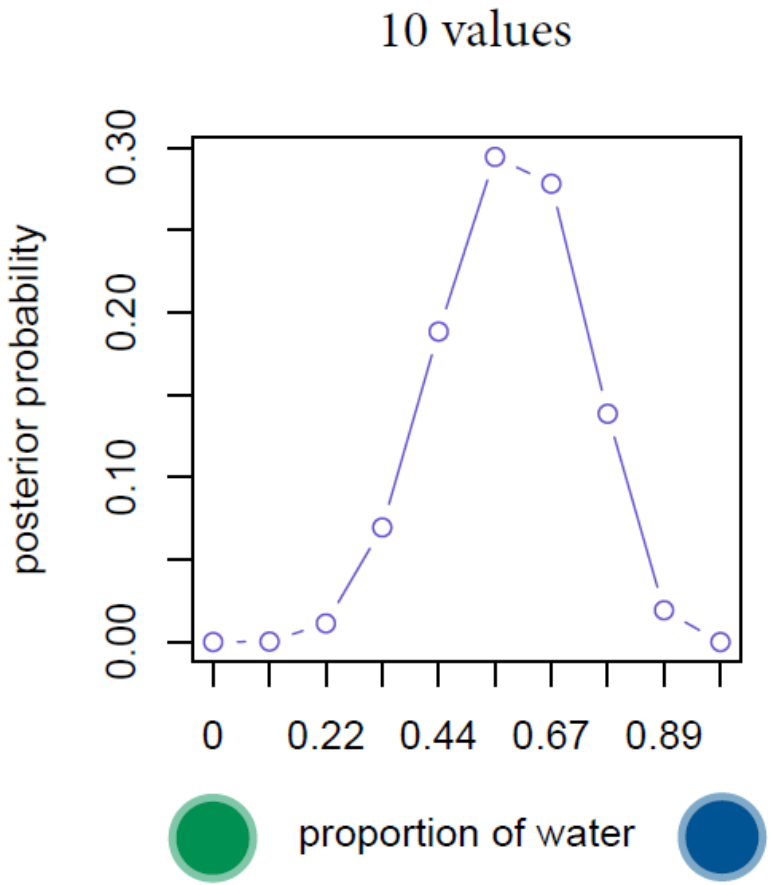
Grid Approximation



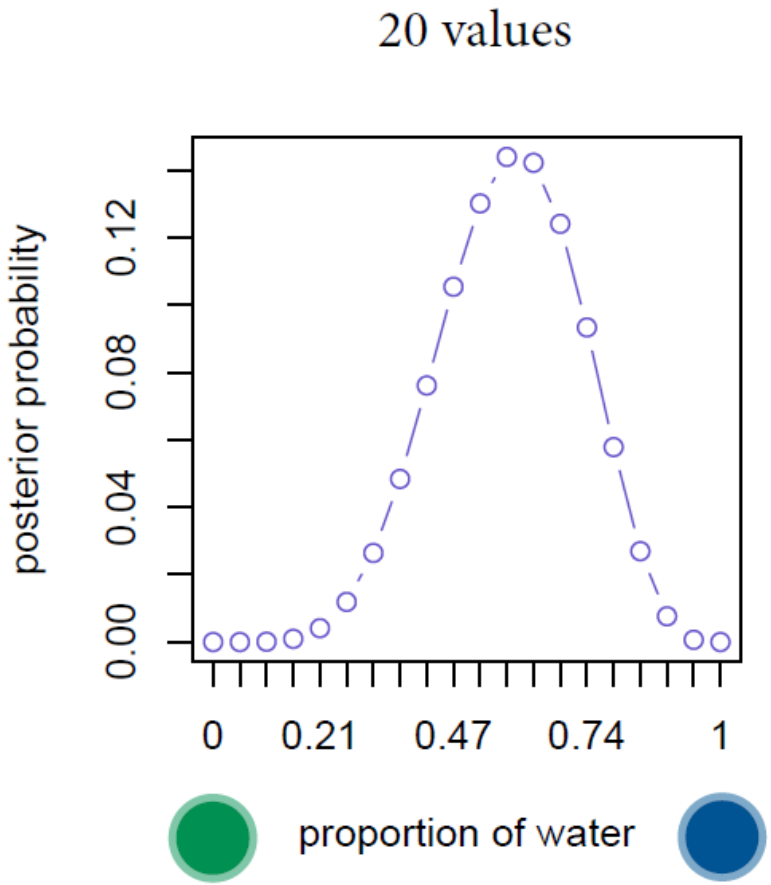
Grid Approximation

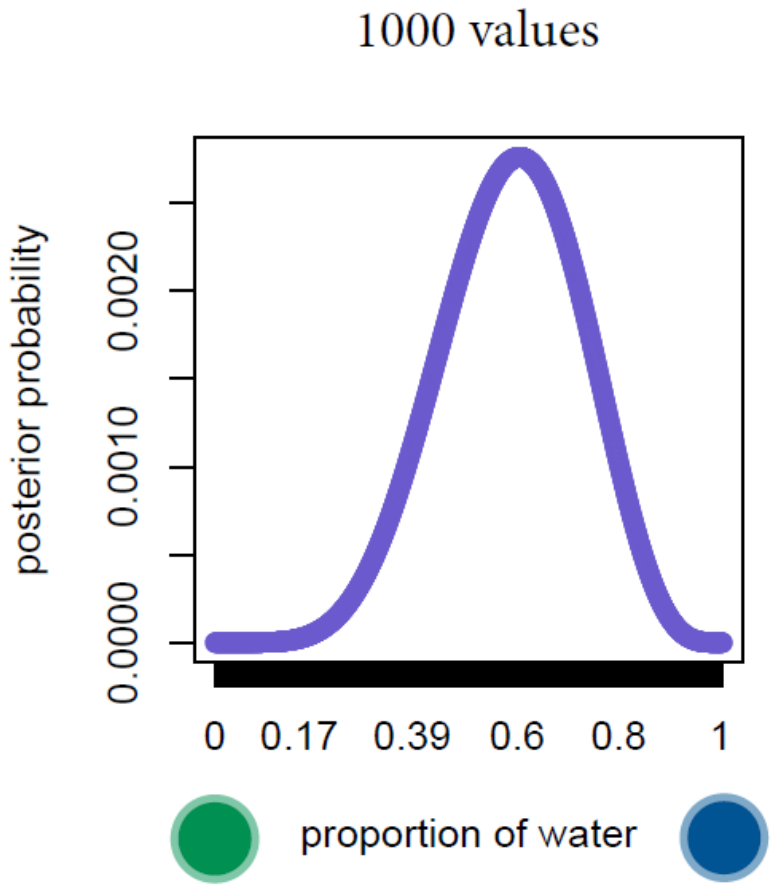


Grid Approximation

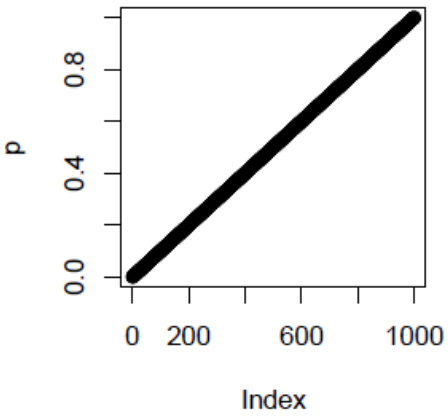


Grid Approximation

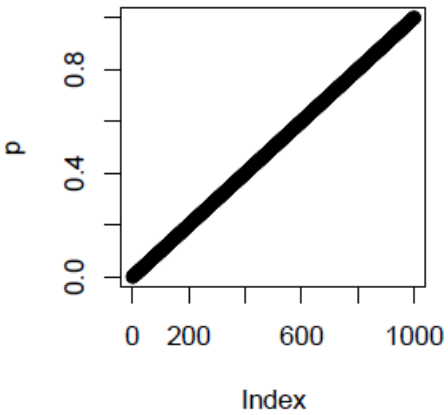



1000 values
posterior probability
0.0000
0.0010
0.0020
0
0.17
0.39
0.6
0.8
1
proportion of water

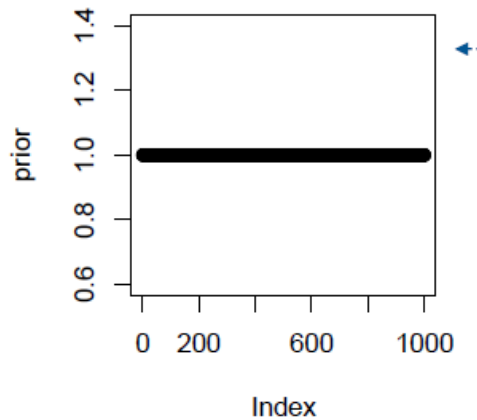
```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```

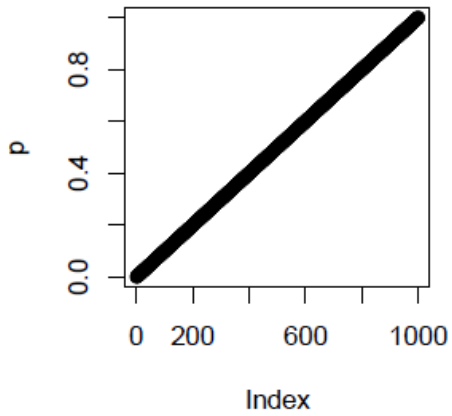



```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```

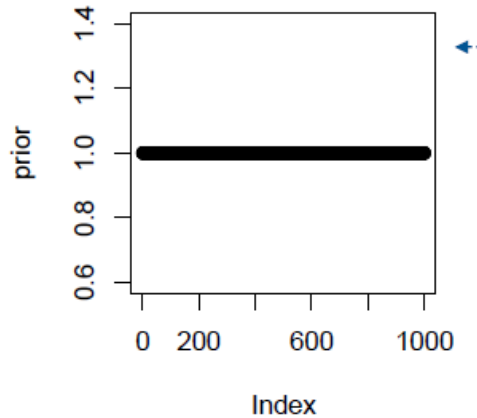


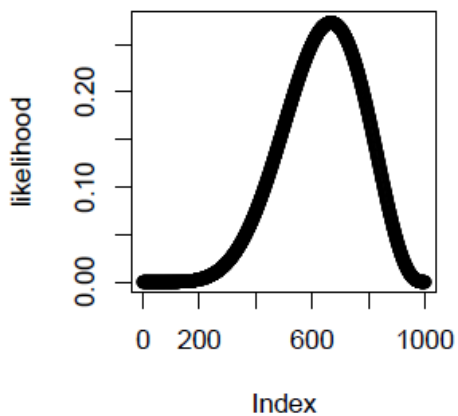

```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```

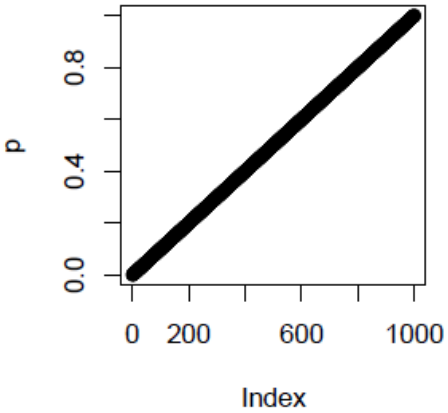




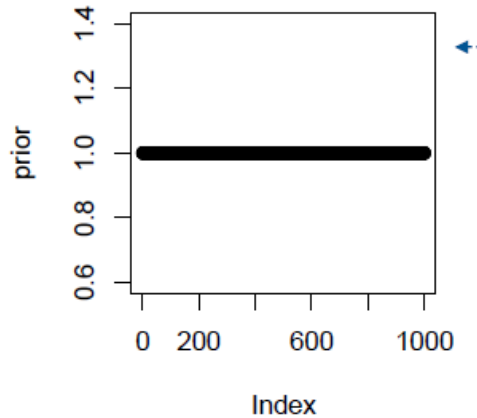

```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```

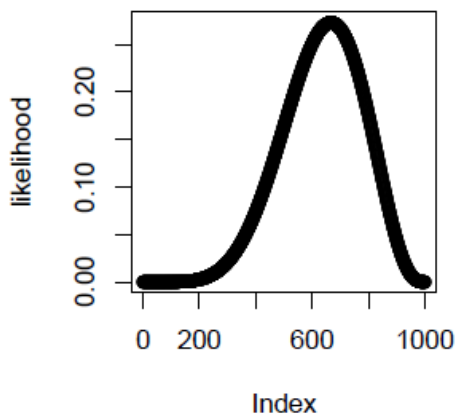


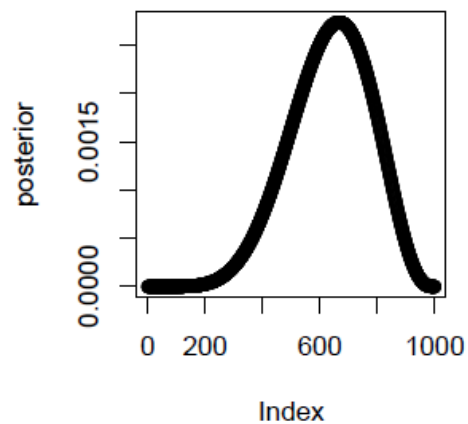




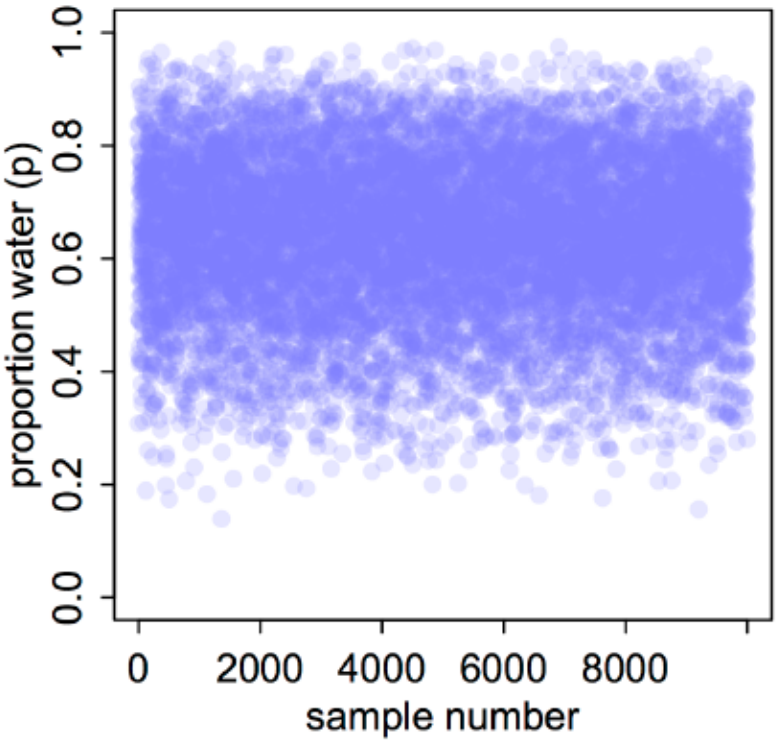

```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```

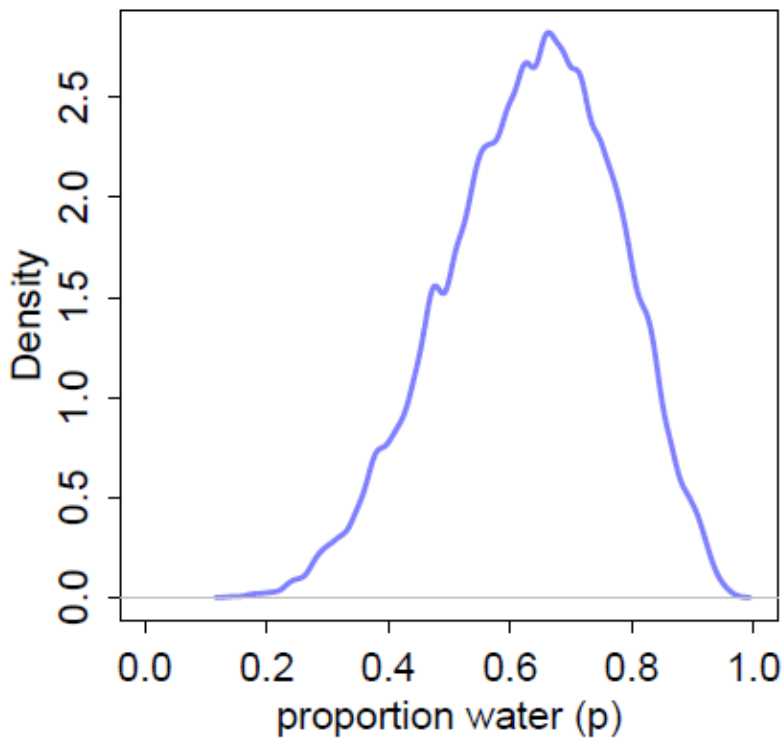






Sampling from the Posterior

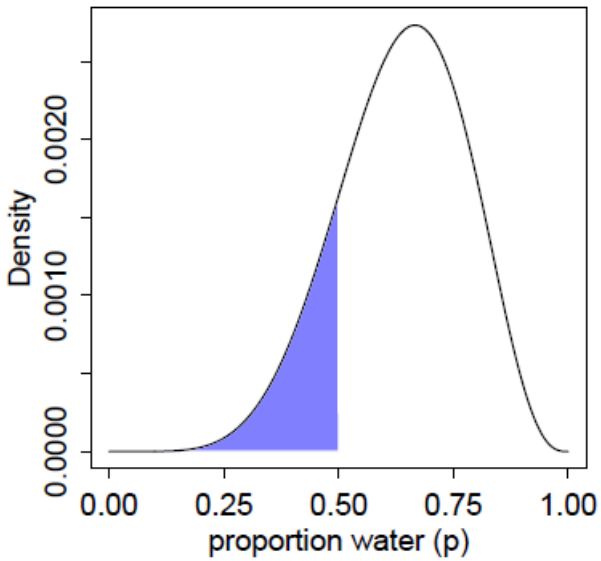


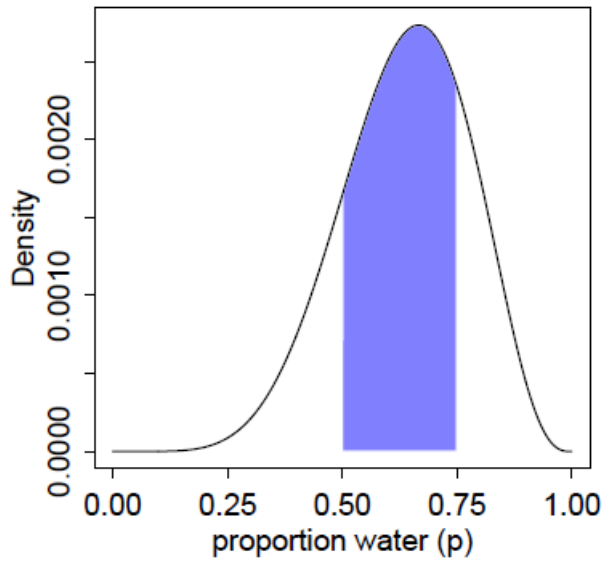



```
samples <- sample( p , prob=posterior , size=1e4 , replace=TRUE )
```

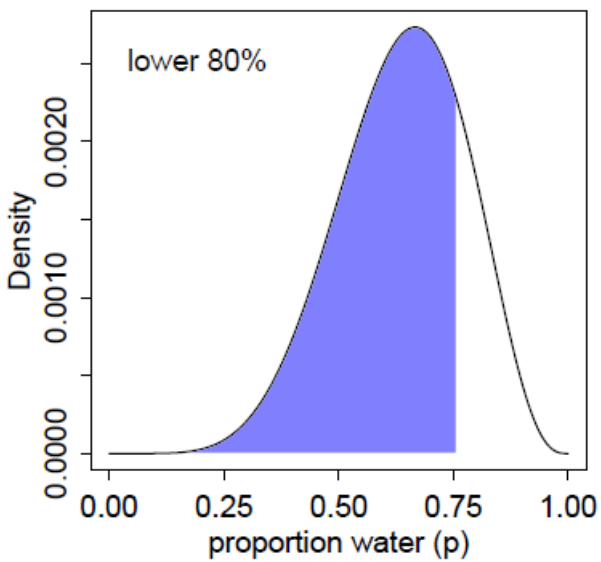
Compute what we want

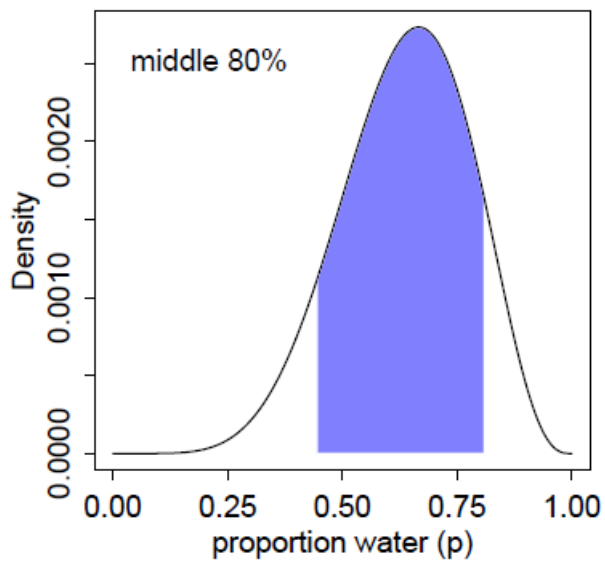
Intervals of defined boundary ask *how much mass?*





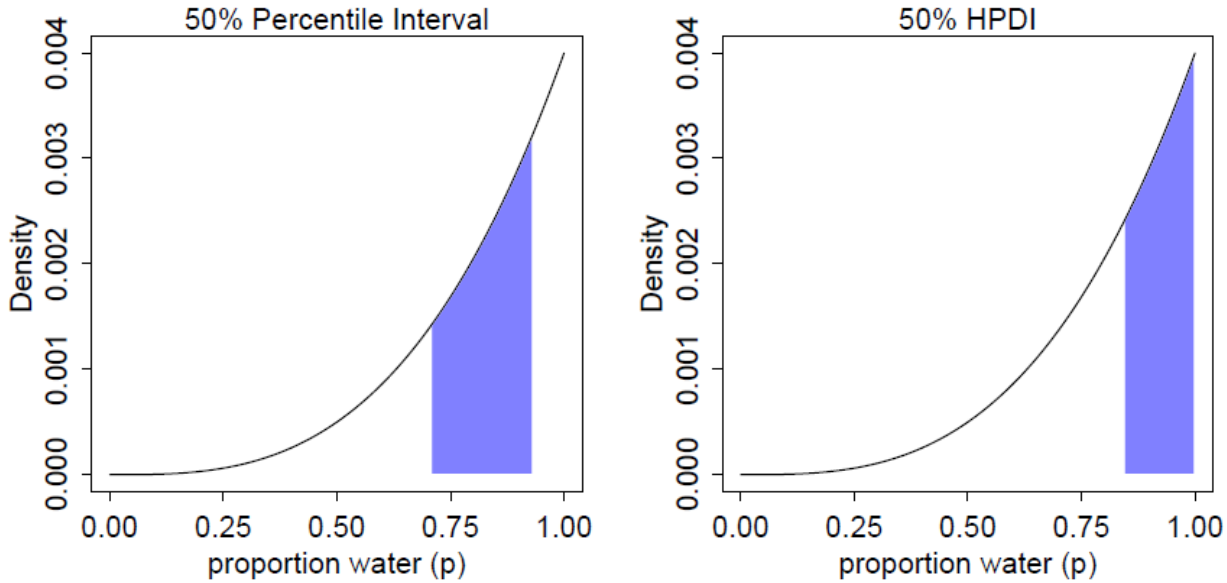
Intervals of defined mass ask *which values?*





Questions to Think About

What could be wrong about getting confidence intervals?



- Percentile Intervals (PI): equal area in each tail
- Highest Posterior Density Intervals (HPDI): narrowest interval containing mass