AARHUS UNIVERSITY

Methods 4 - 2

Chris Mathys

BSc Programme in Cognitive Science

Spring 2024

Recap: Bayes' rule

- The product rule of probability states that

$$p(A|B)p(B) = p(B|A)p(A)$$

- If we divide by $p(B)$, we get **Bayes' rule:**

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)p(A)}{\sum_a p(B|a)p(a)}$$

- The last equality comes from unpacking $p(B)$ according to the product and sum rules:

$$p(B) = \sum_a p(B, a) = \sum_a p(B|a)p(a)$$

Bayes' rule: what problem does it solve?

- Why is Bayes' rule important?
- It allows us to invert conditional probabilities, ie to pass from $p(B|A)$ to $p(A|B)$:

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

- In other words, it allows us to update our belief about A in light of observation B

Bayes' rule: the chocolate example

In our example, it is immediately clear that $P(Nobel|chocolate)$ is very different from $P(chocolate|Nobel)$. While the first is hopeless to determine directly, the second is much easier to find out: ask Nobel laureates how much chocolate they eat. Once we know that, we can use Bayes' rule:

likelihood
model

$$p(Nobel|chocolate) = \frac{p(chocolate|Nobel)P(Nobel)}{p(chocolate)}$$

prior

posterior

evidence

However: *note that no amount of statistical analysis will tell you anything about the causal mechanism behind this if you don't have* ***a hypothesis about that mechanism and a causal scientific model of it****!*

Flipped classroom ☺

The procedure in brief:

- You’ve done the readings (hopefully…)
- You may also have watched the videos

Now we go through the content again and explain it to each other.

Bayesian data analysis

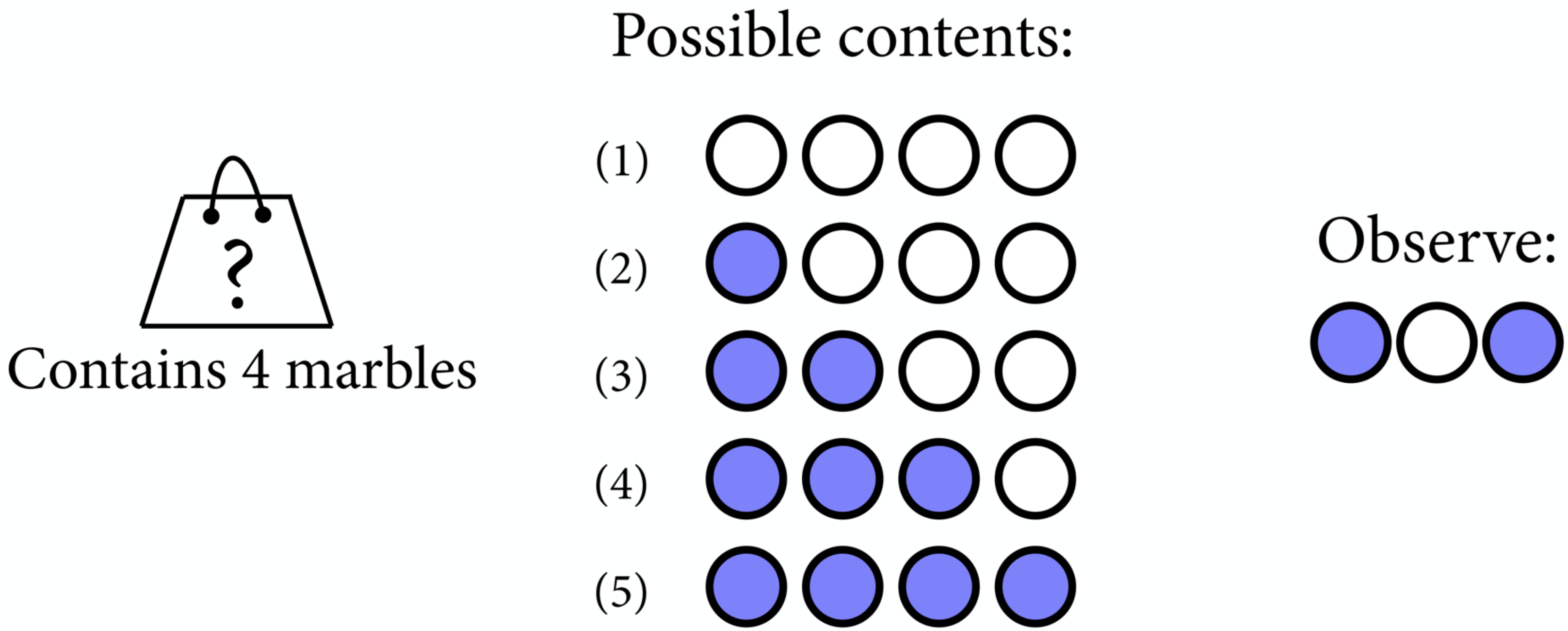
For each possible explanation of the data,

Count all the ways data can happen.

Explanations with more ways to produce the data are more plausible.

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

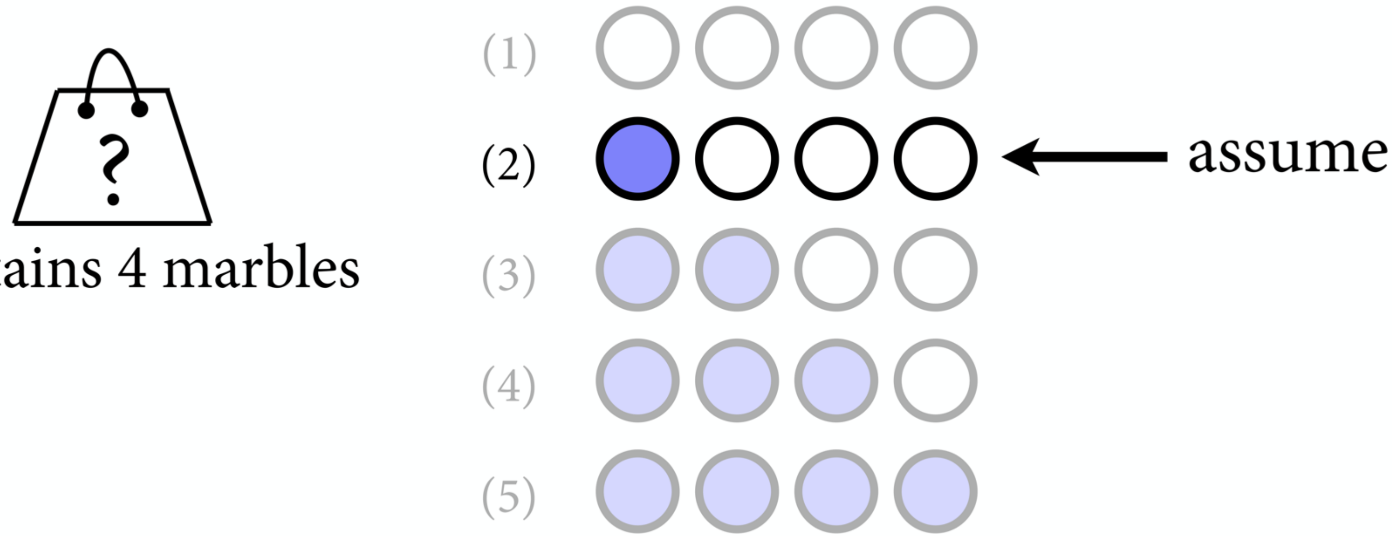
Garden of Forking Data



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Garden of Forking Data

Possible contents:



Contains 4 marbles

How many ways to observe ?

3 Ways to see

if the bag contains

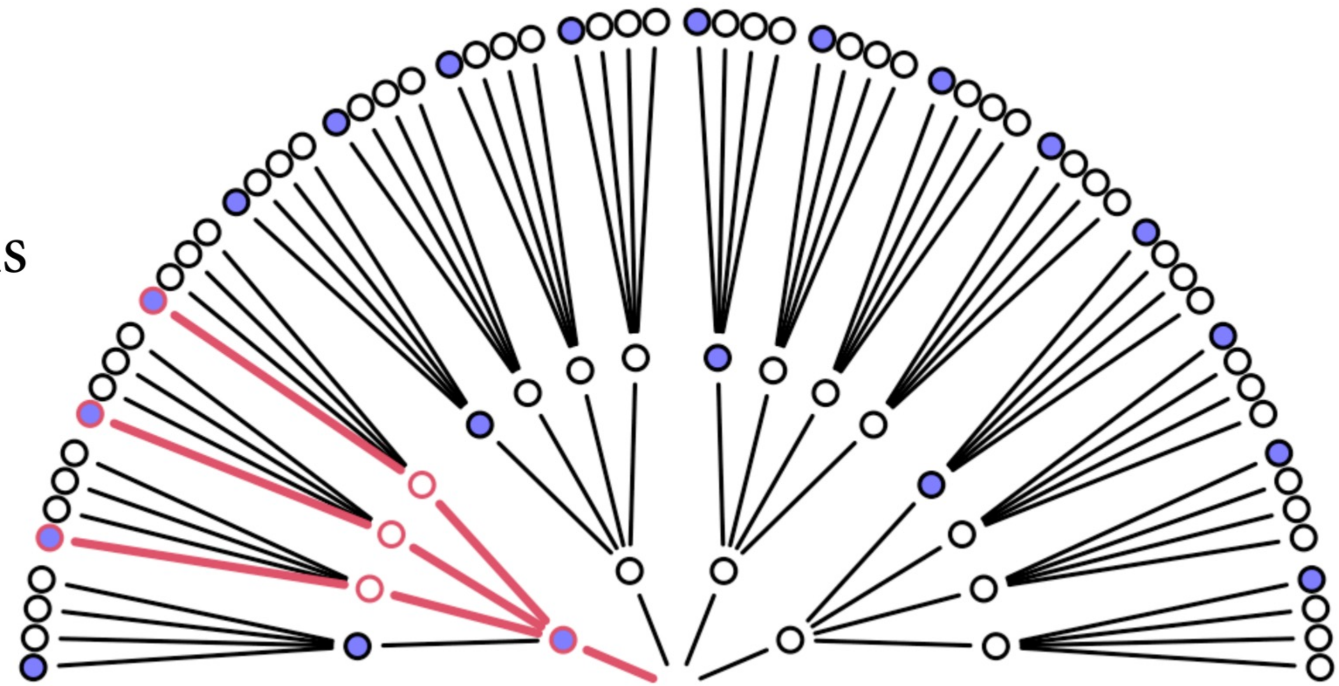
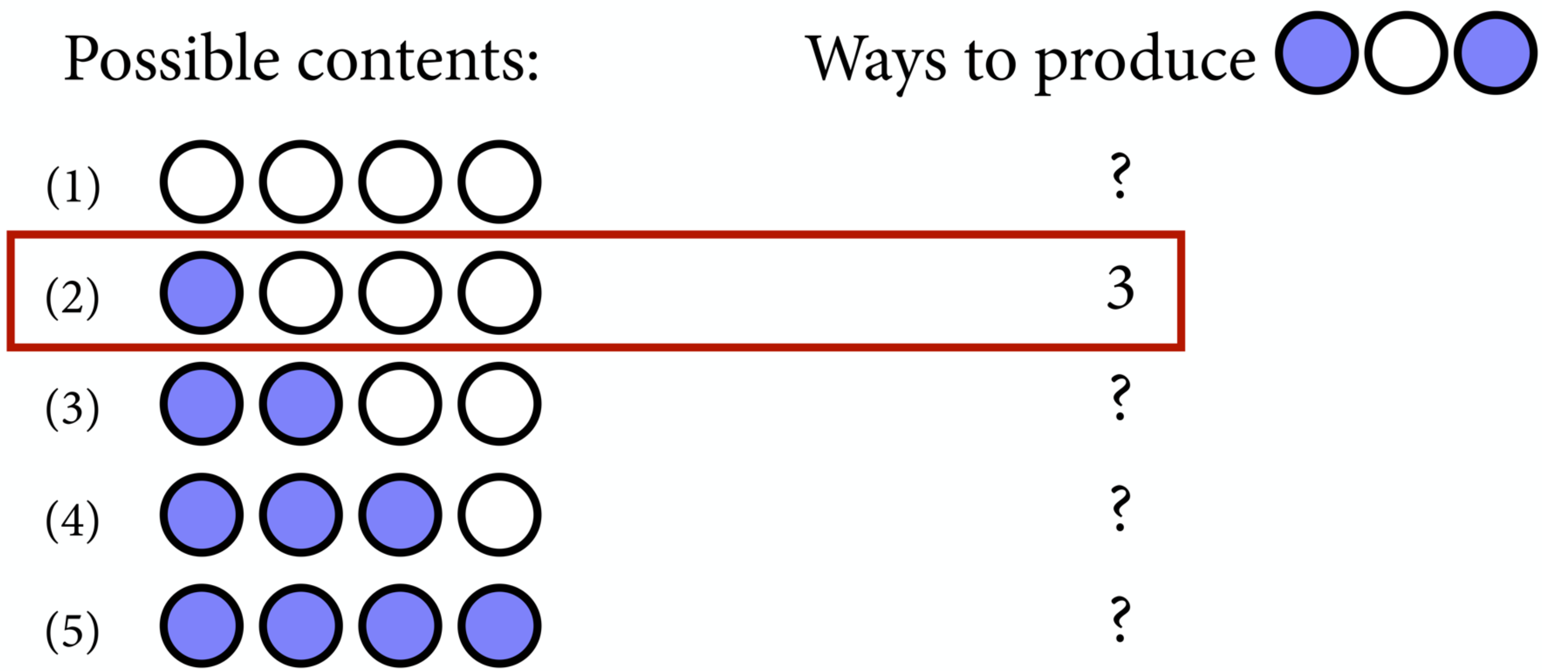


Figure 2.2

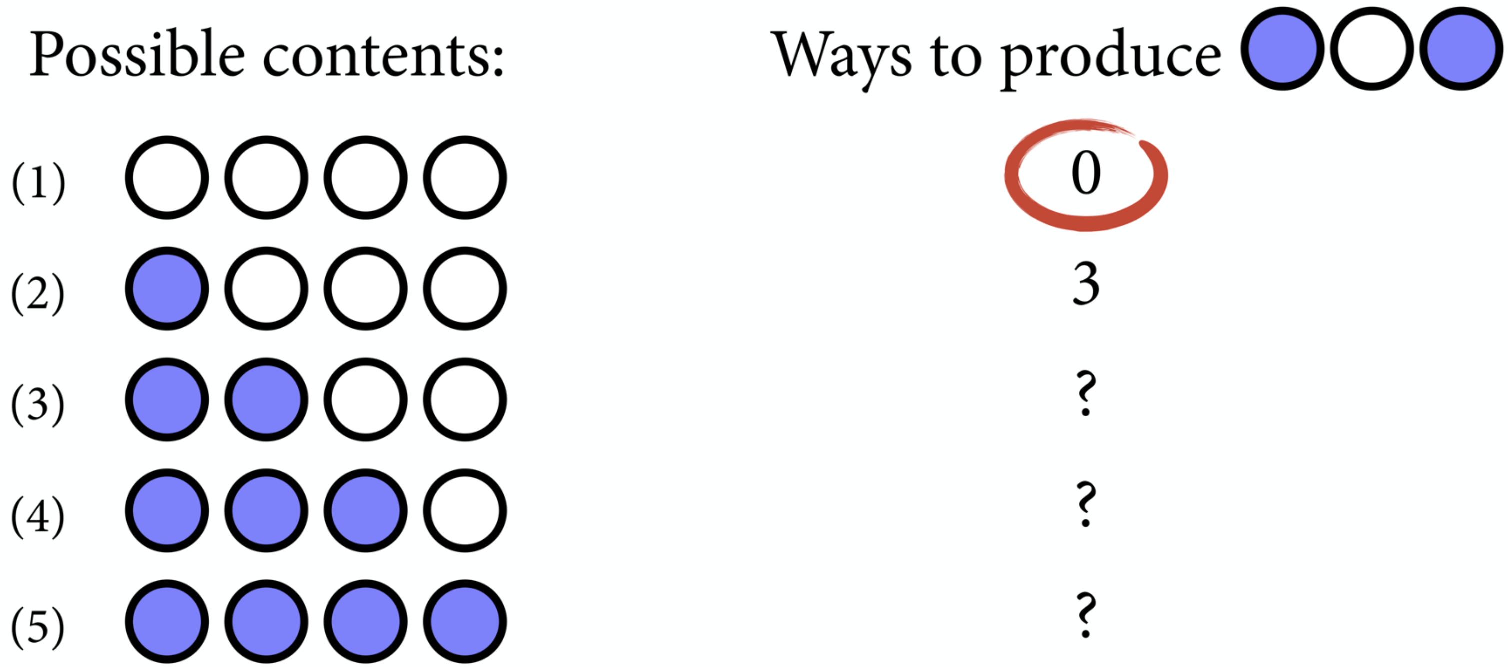
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Garden of Forking Data



Source: Richard McElreath, https://github.com/rmcelreath/stat_rethinking_2022

Garden of Forking Data

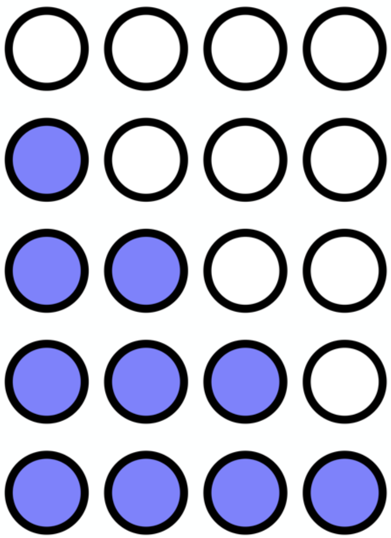


Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Garden of Forking Data

Possible contents:

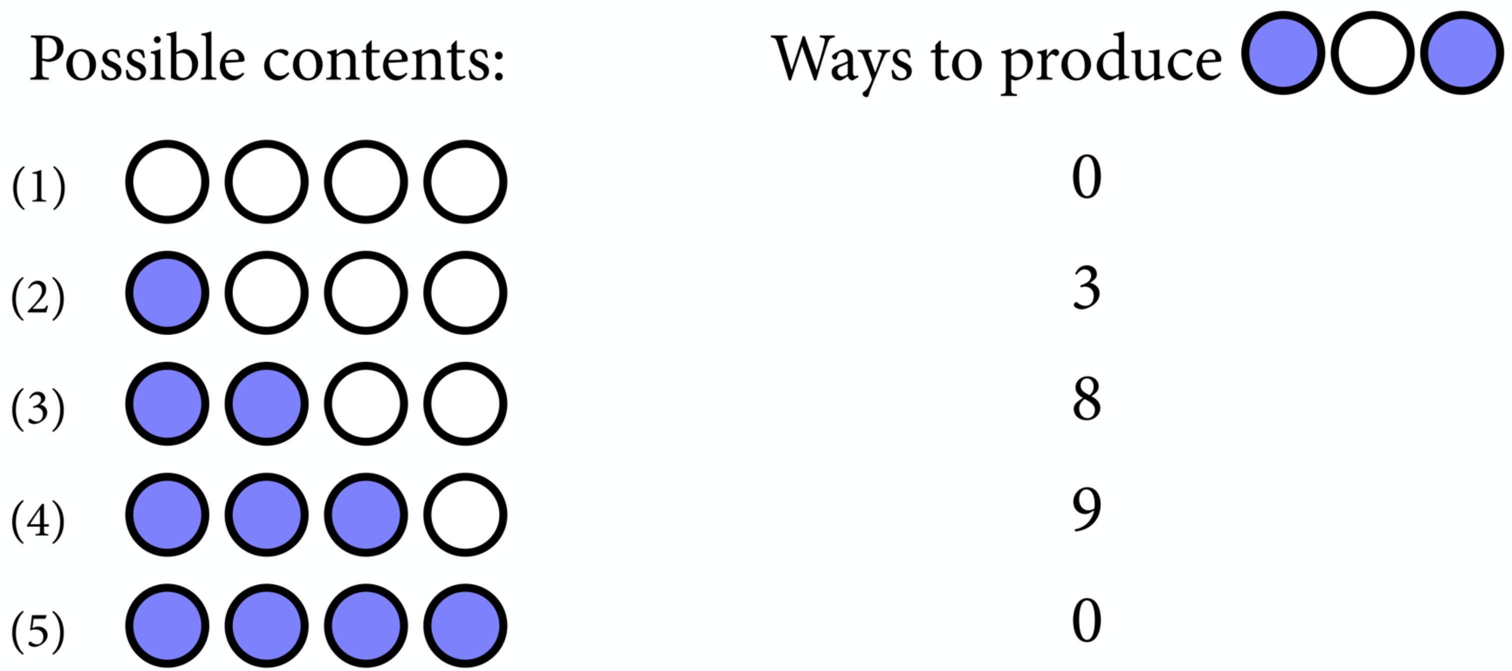
Ways to produce 



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

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Garden of Forking Data

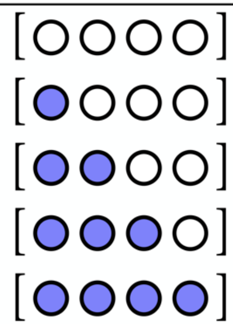


Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Counts to plausibility

Unglamorous basis of applied probability:
Things that can happen more ways are more plausible.

Possible composition



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Counts to plausibility

Unglamorous basis of applied 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

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

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

R code 2.1

```
ways <- c( 3 , 8 , 9 )
ways/sum(ways)
```

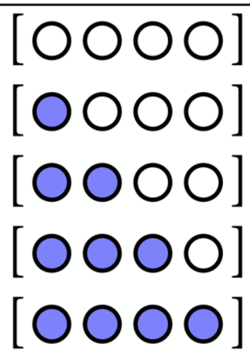
```
[1] 0.15 0.40 0.45
```

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Updating

Another draw from the bag: 

Conjecture



Updating

Another draw from the bag: ●

Conjecture	Ways to produce ●
[○○○○]	0
[●○○○]	1
[●●○○]	2
[●●●○]	3
[●●●●]	4

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Updating

Another draw from the bag: ●

Conjecture	Ways to produce ●	Previous counts
[○○○○]	0	0
[●○○○]	1	3
[●●○○]	2	8
[●●●○]	3	9
[●●●●]	4	0

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

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	$0 \times 4 = 0$

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Bayesian updating

The rules:

1. State a causal model for how the observations arise, given each possible explanation

2. Count ways data could arise for each explanation

3. Relative plausibility is relative value from (2)

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Globe of Forking Water

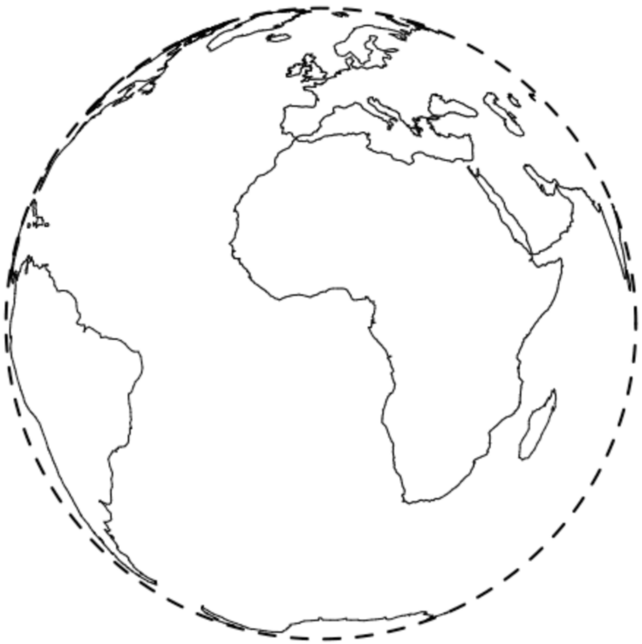
For each possible proportion of water,

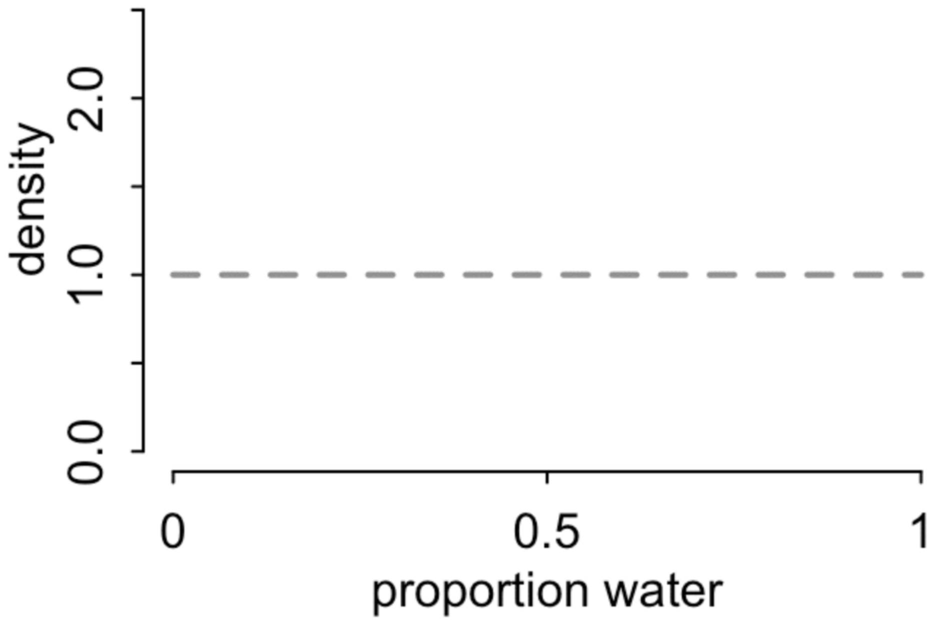
Count number of ways data could happen.

Must state how observations are generated

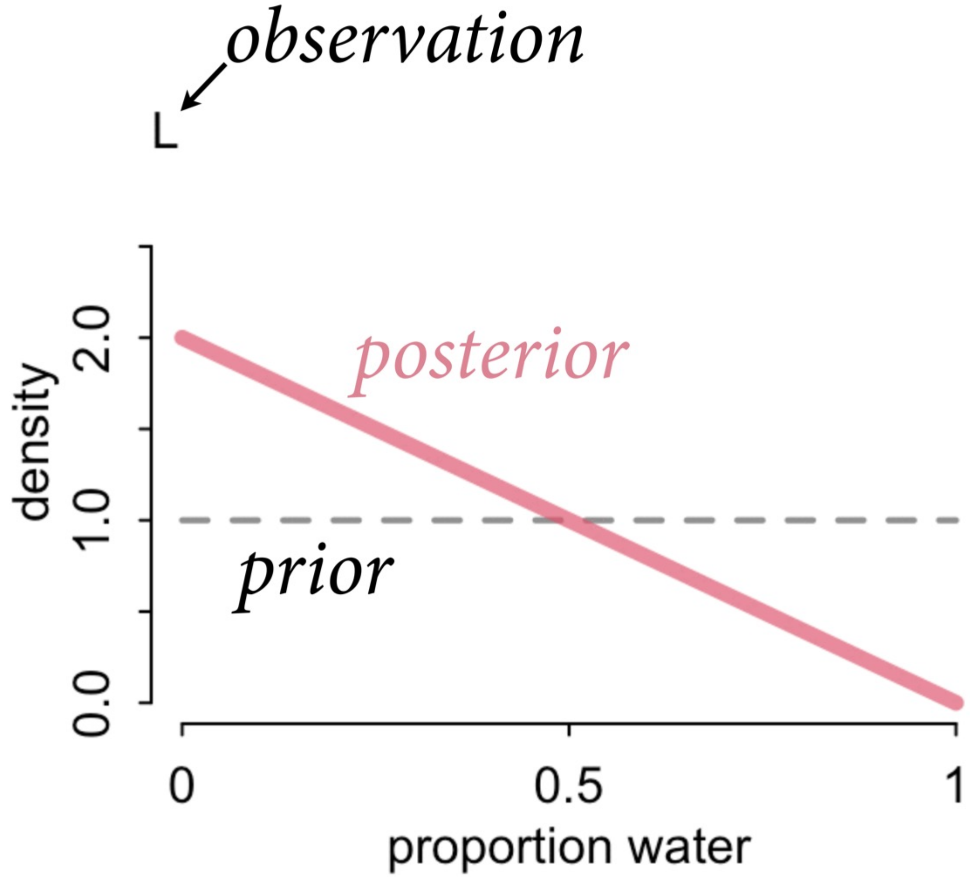
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Toss The First

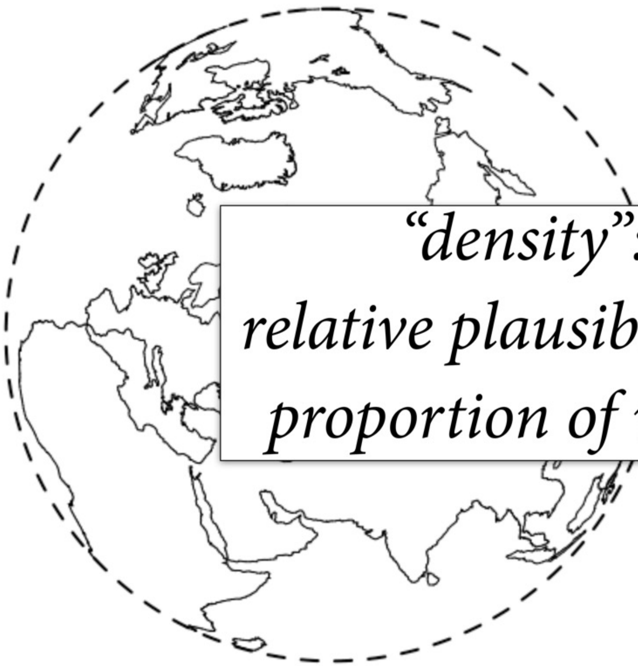




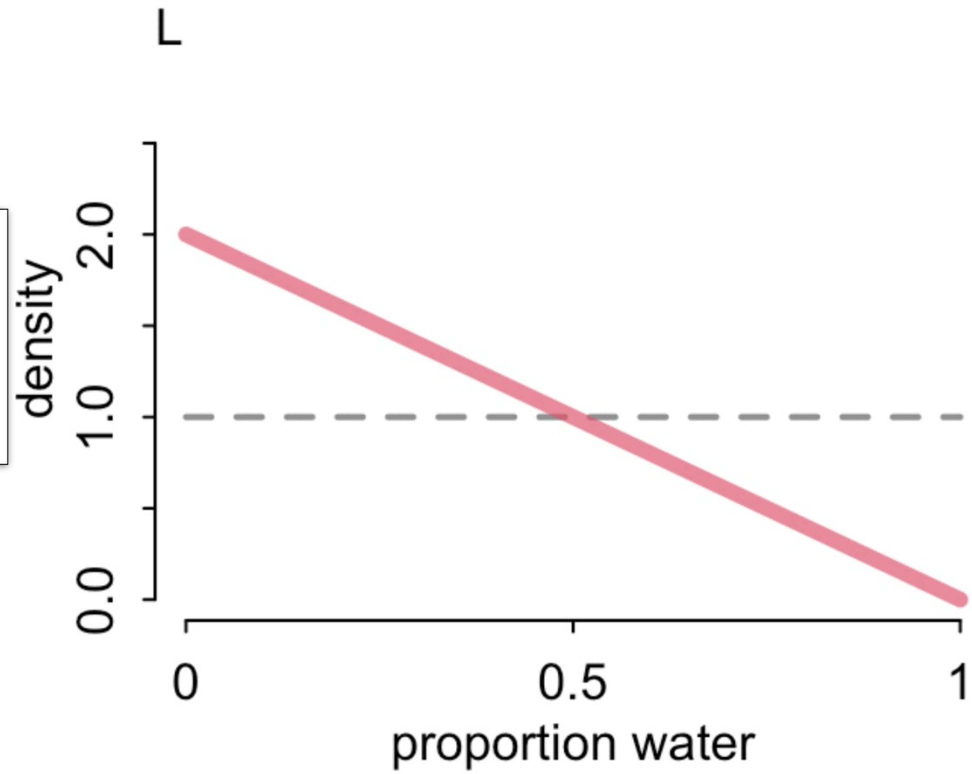
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

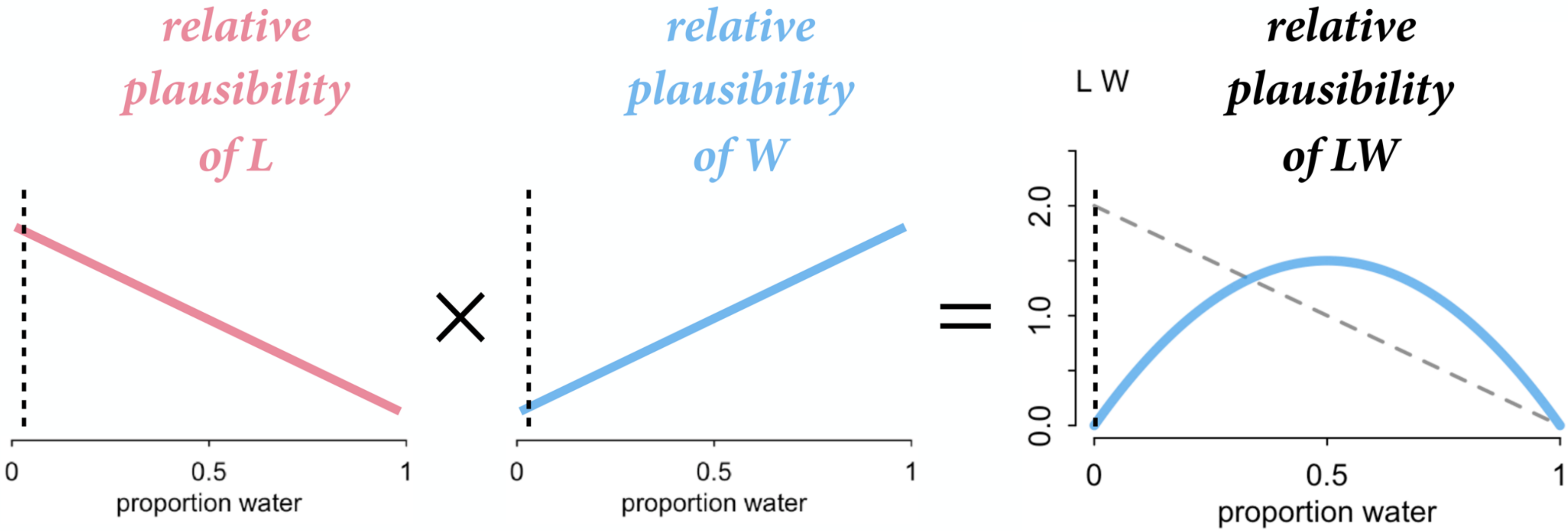


"density":
relative plausibility of proportion of water



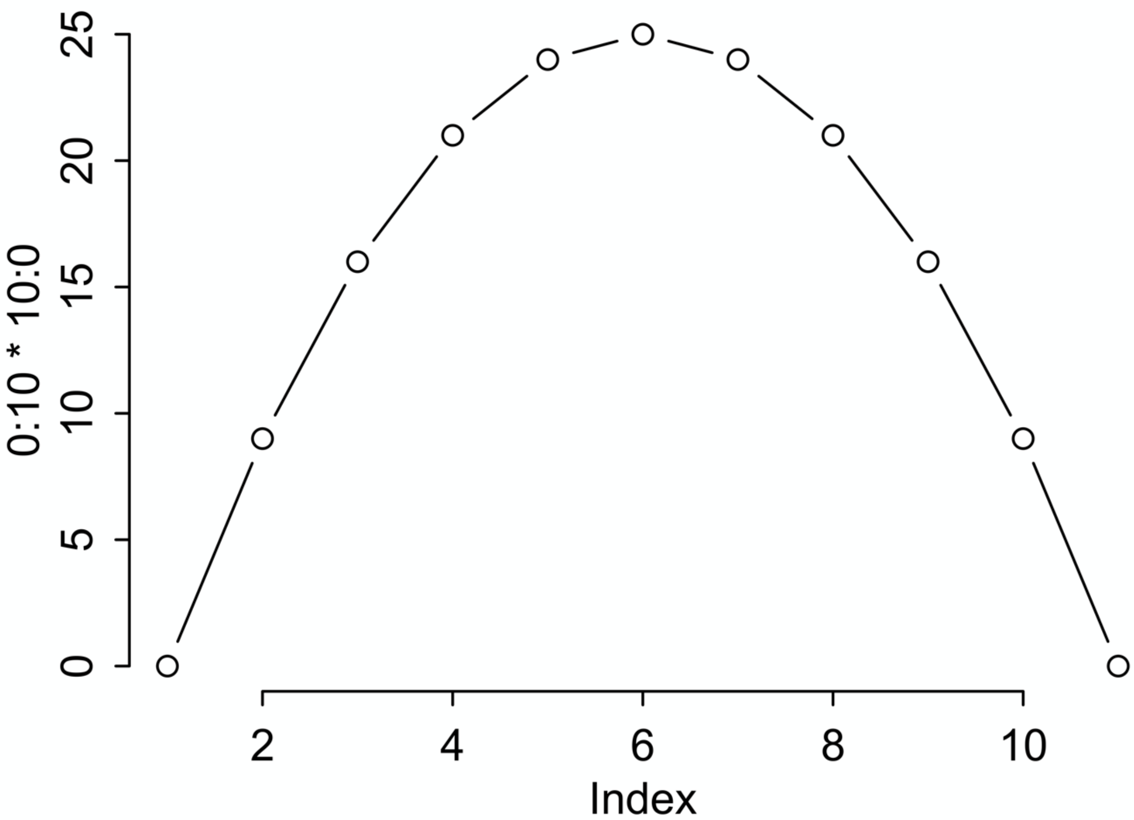
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Toss The Second



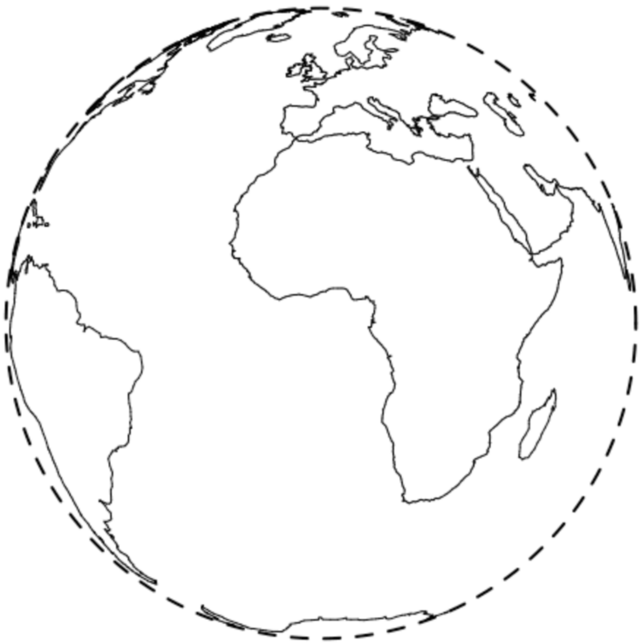
Source: Richard McElreath, https://github.com/rmcelreath/stat_rethinking_2022

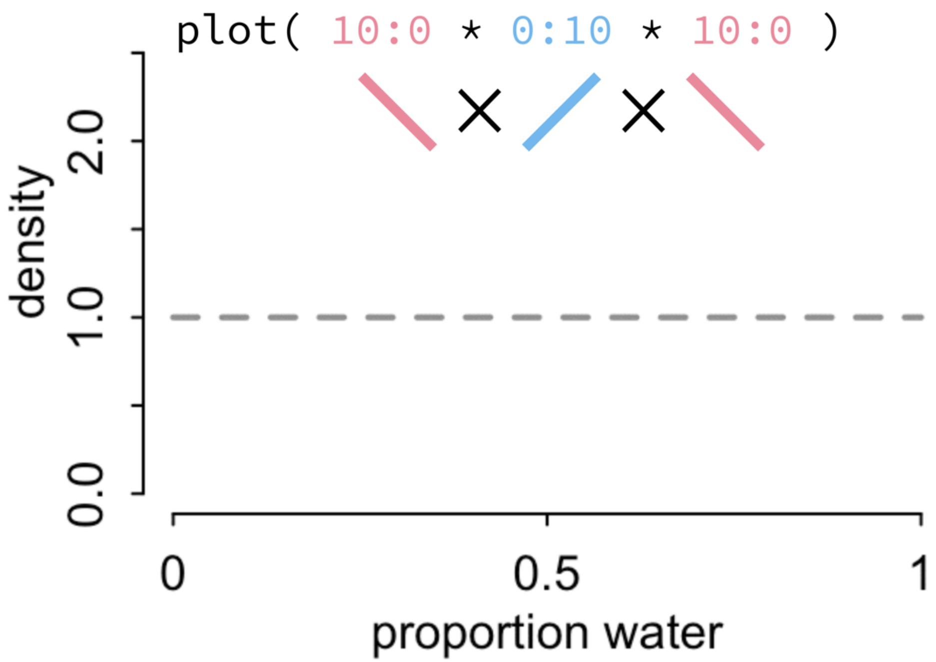
```
plot( 0:10 * 10:0 )
```



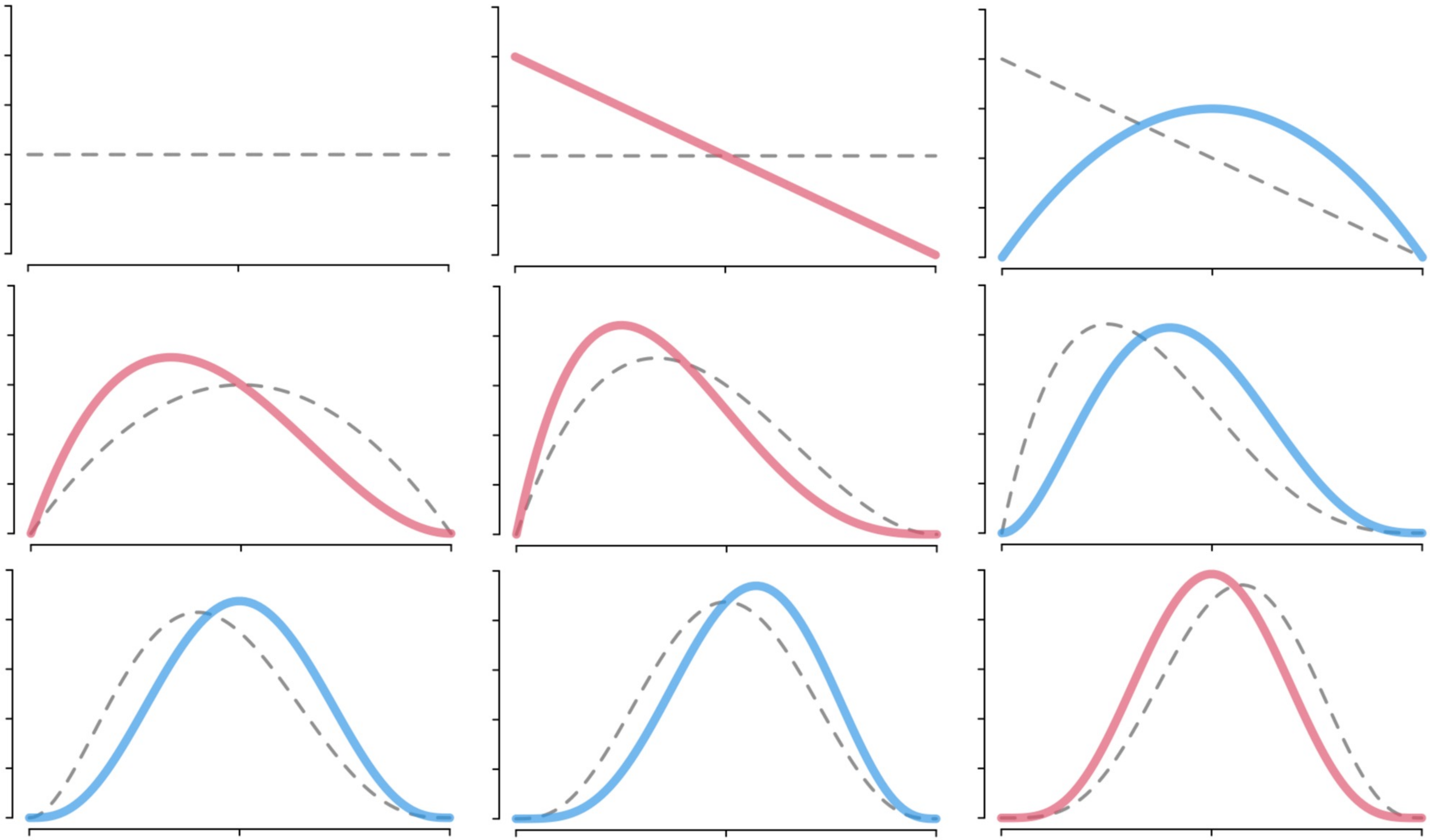
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Toss The Third



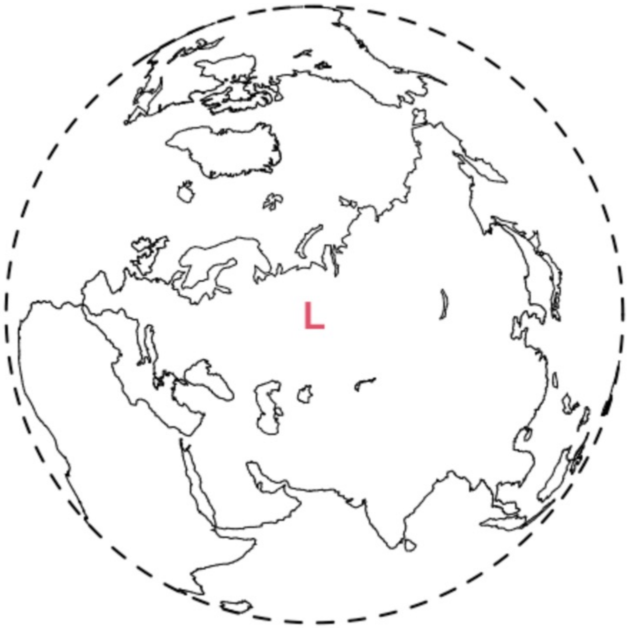


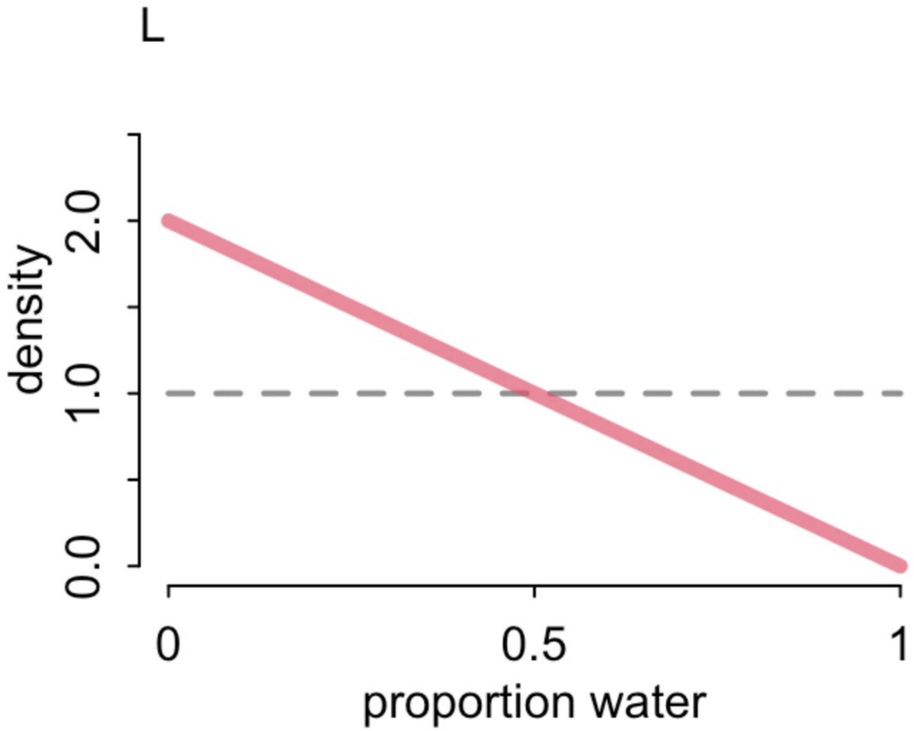
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

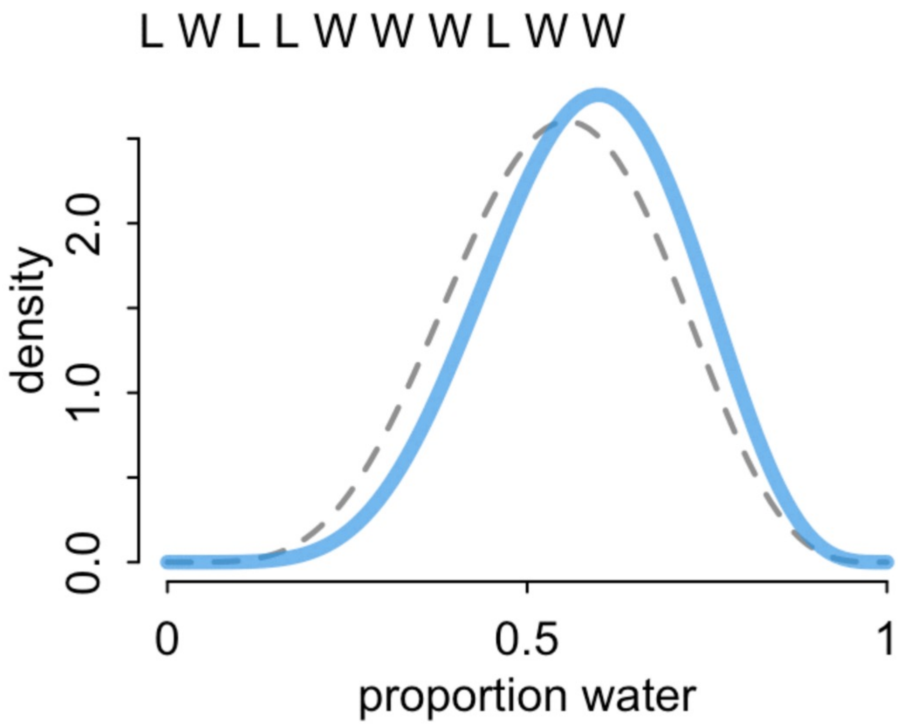
(1) No minimum sample size





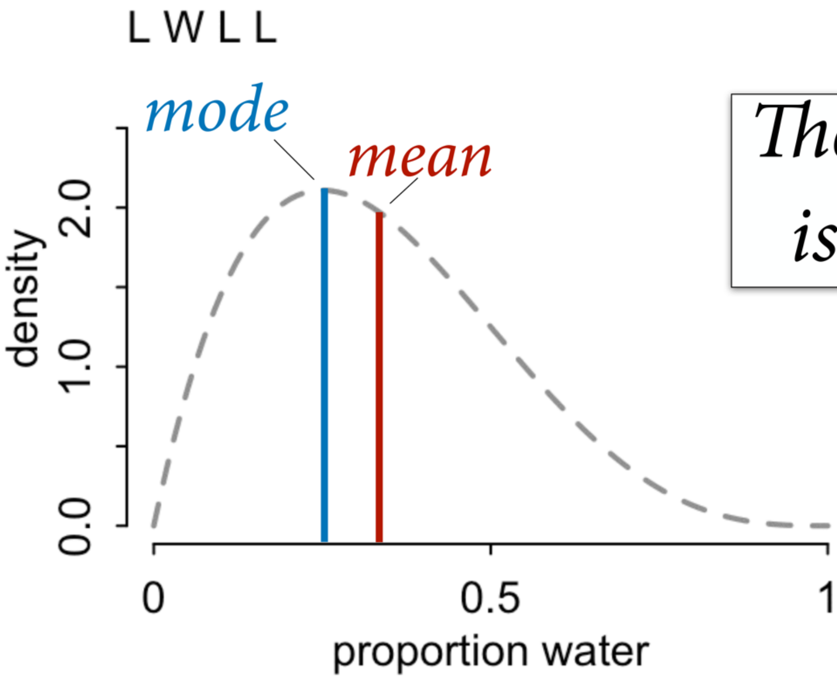
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

(2) Shape embodies sample size



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

(3) No point estimate

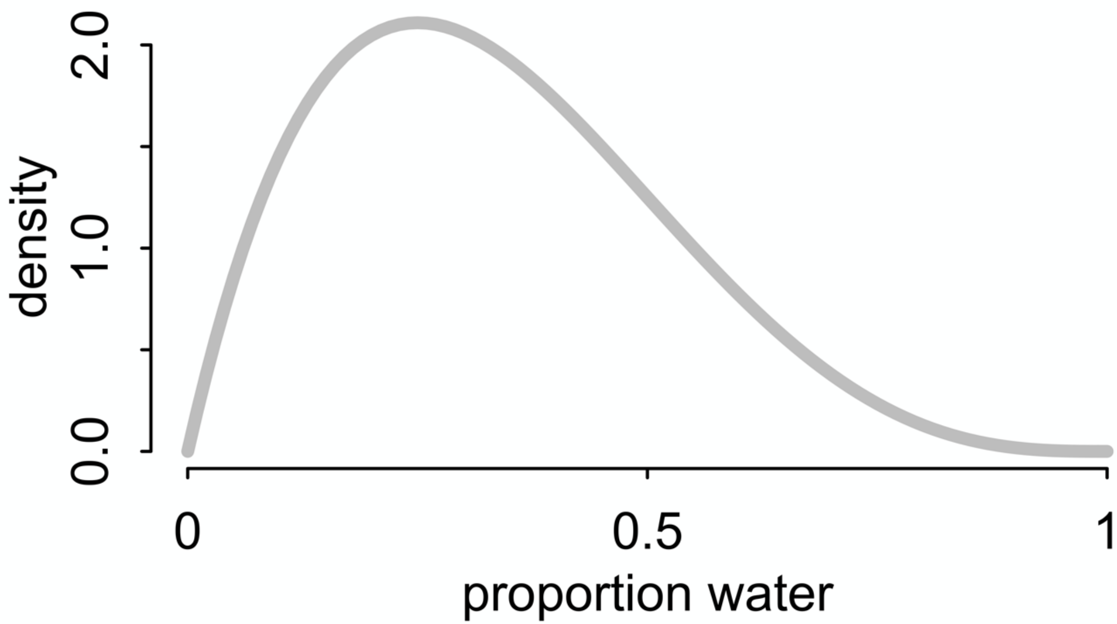


*The **distribution** is the estimate*

Always use the entire distribution

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

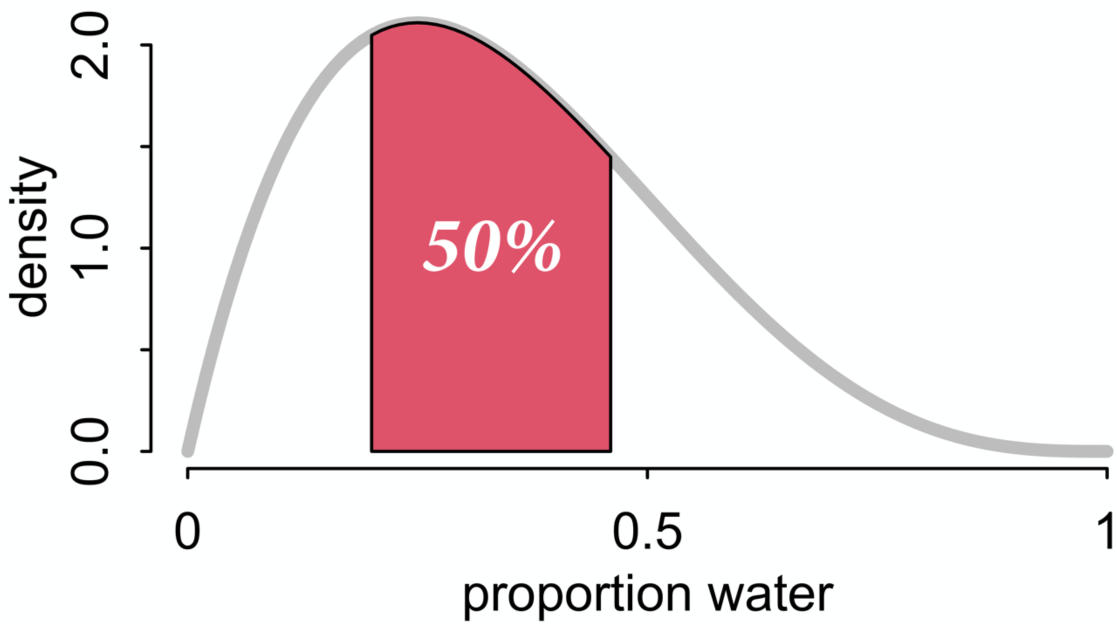
(4) No one true interval



Intervals communicate shape of posterior

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

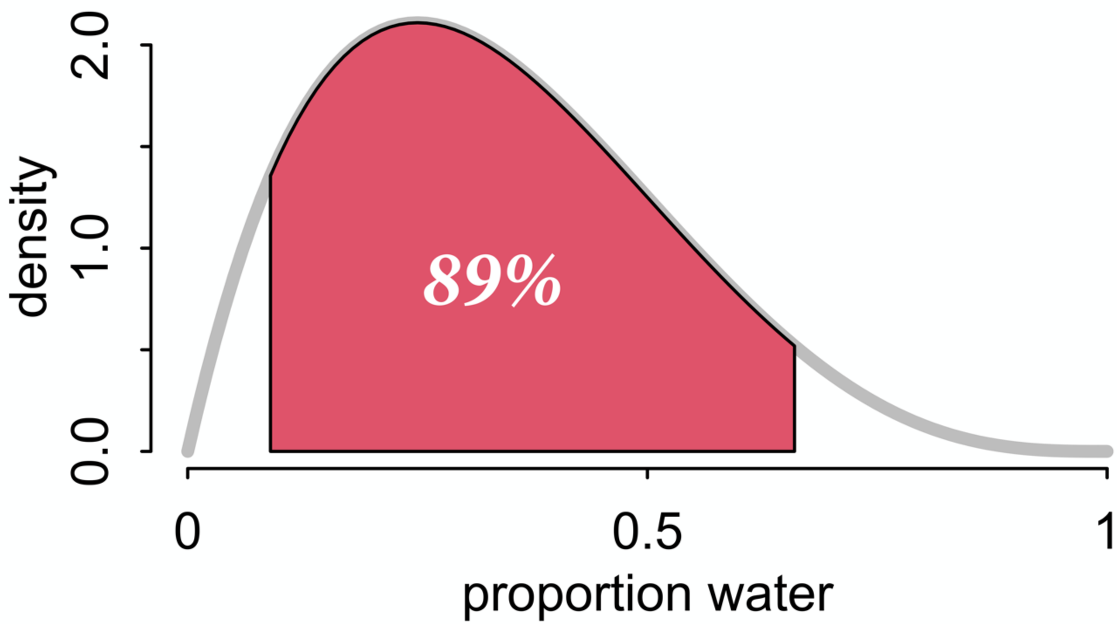
(4) No one true interval



Intervals communicate shape of posterior

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

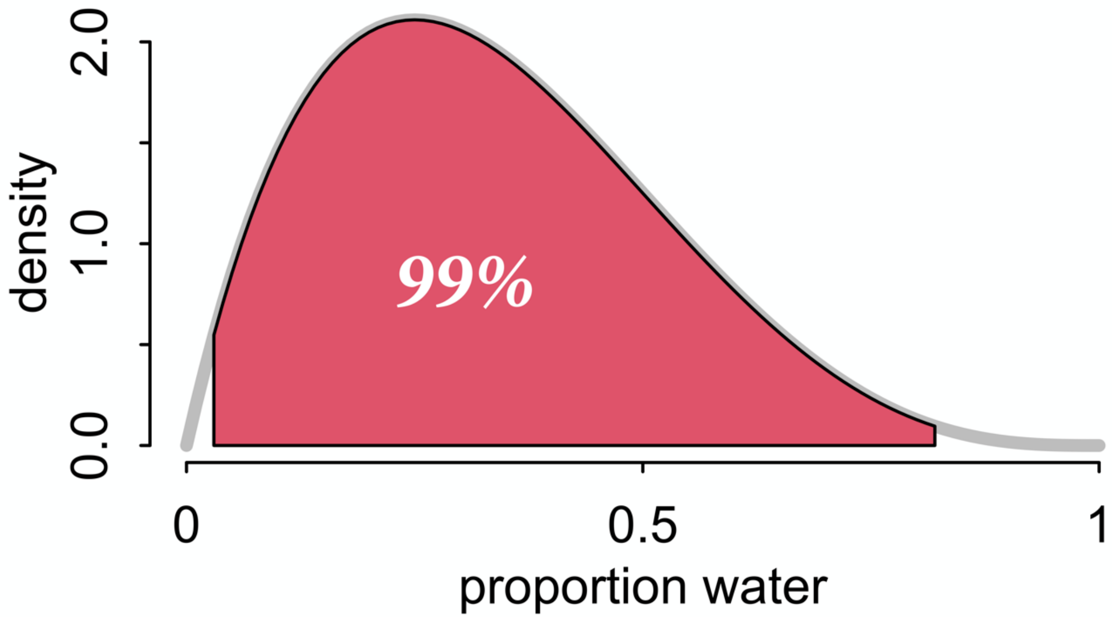
(4) No one true interval



Intervals communicate shape of posterior

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

(4) No one true interval



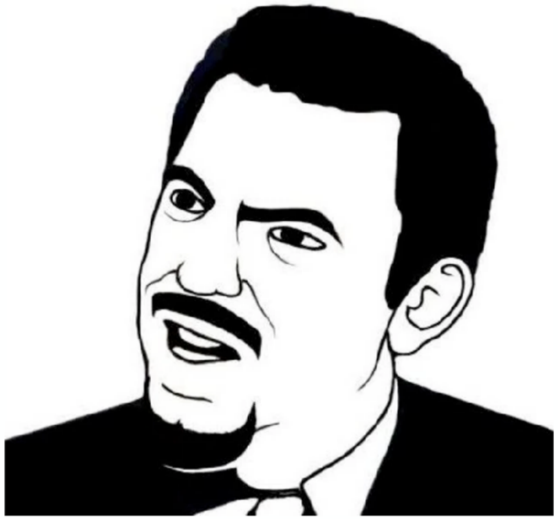
Intervals communicate shape of posterior

95% is obvious superstition. Nothing magical happens at the boundary.

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Letters From My Reviewers

"The author uses these cute 89% intervals, but we need to see the 95% intervals so we can tell whether any of the effects are robust."



That an arbitrary interval contains an arbitrary value is not meaningful. Use the whole distribution.

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

The Formalities

Data: *W* and *L*, the number of water and land observations

$$\Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$$

The number of ways to realize W,L given p

Binomial probability function

```
dbinom( W , W+L , p )
```

```
> dbinom( 6 , 9 , 0.7 )
[1] 0.2668279
>
```

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

The Formalities

Data: W and L, the number of water and land observations

$$\Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$$

The number of ways to realize W,L given p

Parameters: p, the proportion of water on the globe

$$\Pr(p) = \frac{1}{1-0} = 1.$$

Relative plausibility of each possible p

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

The Formalities

$$\Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$$

$$\Pr(p) = \frac{1}{1-0} = 1.$$

Posterior is (normalized) product:

$$\Pr(p|W, L) = \frac{\Pr(W, L|p) \Pr(p)}{\Pr(W, L)}$$

Relative plausibility of each possible p, after learning W,L

We multiply because that's how the garden counts!

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

With Numbers

Ignore the mathematics for the moment and just draw the owl with numbers

1. For each possible value of p

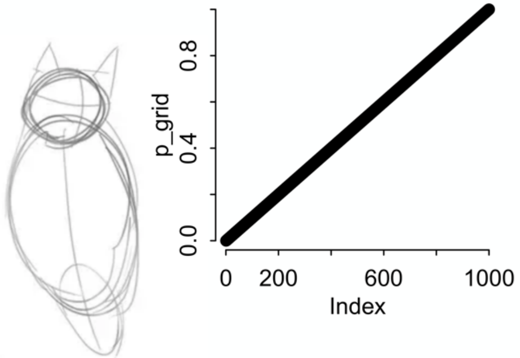
2. Compute product $\Pr(W,L|p)\Pr(p)$

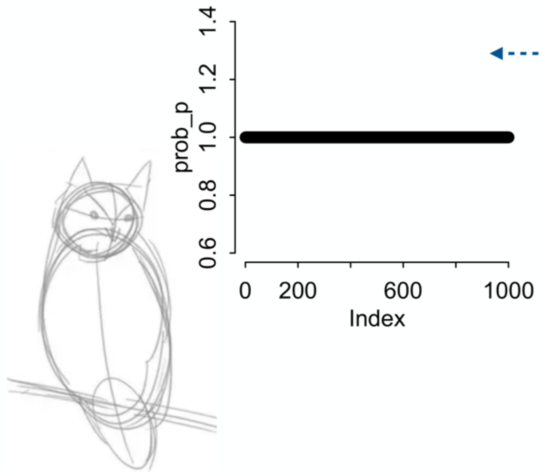
3. Relative sizes of products in (2) are posterior probabilities

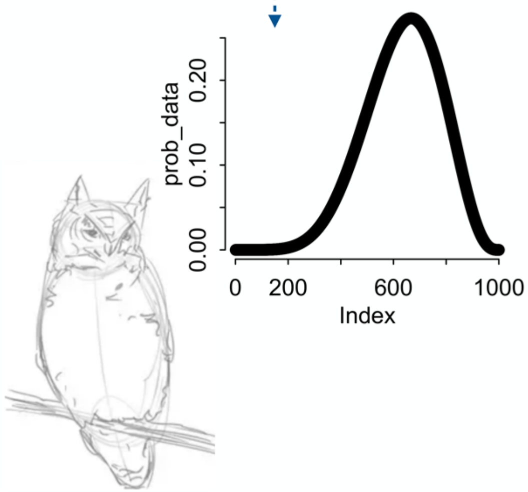
Bayesian owl

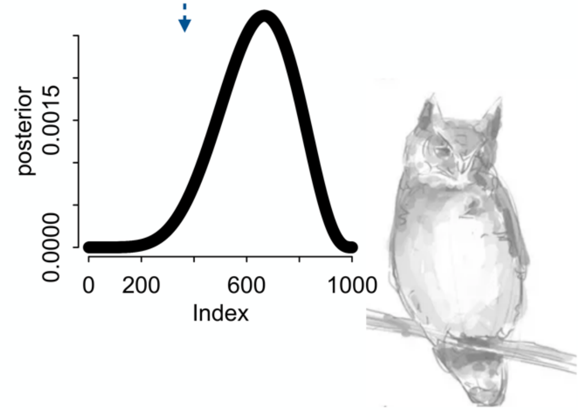
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

```
p_grid ← seq( from=0 , to=1 , len=1000 )
prob_p ← rep( 1 , 1000 )
prob_data ← dbinom( 6 , 9 , prob=p_grid )
posterior ← prob_data * prob_p
posterior ← posterior / sum(posterior)
```



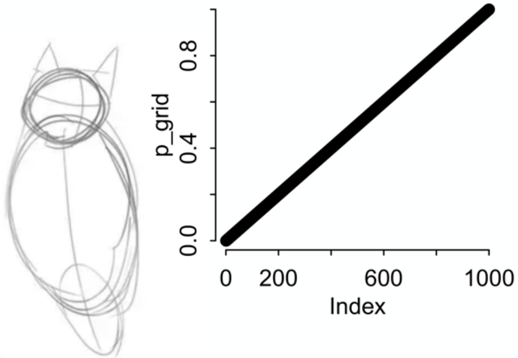


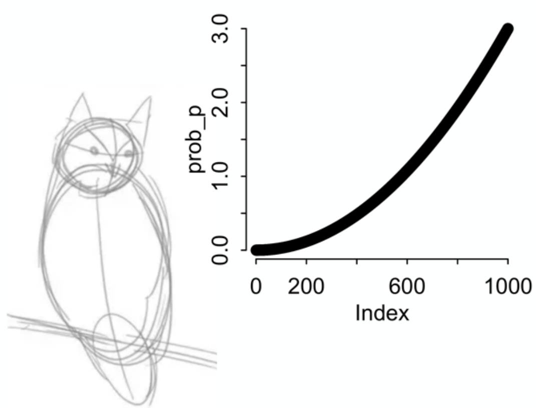


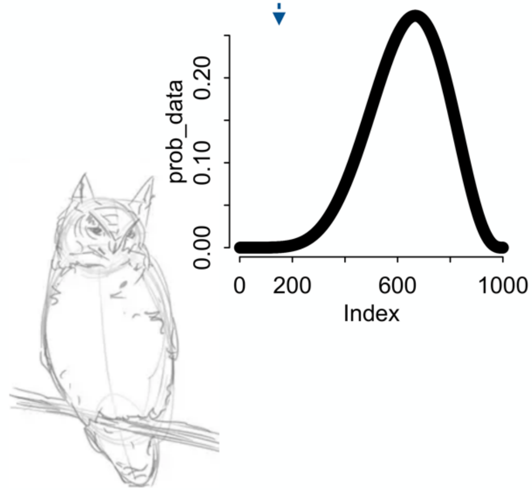


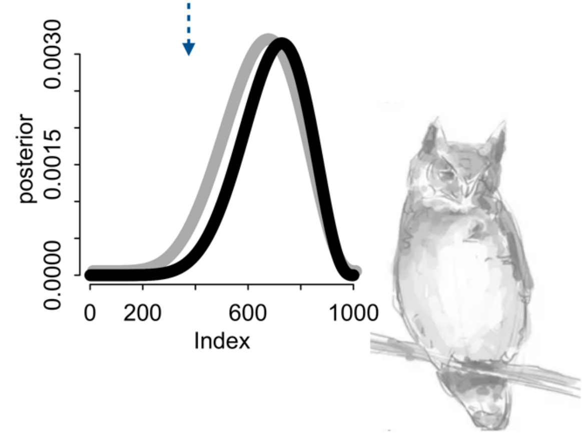
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

```r
p_grid <- seq( from=0 , to=1 , len=1000 )
prob_p <- dbeta( p_grid , 3 , 1 )
prob_data <- dbinom( 6 , 9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```









Source: Richard McElreath, https://github.com/rmcelreath/stat_rethinking_2022

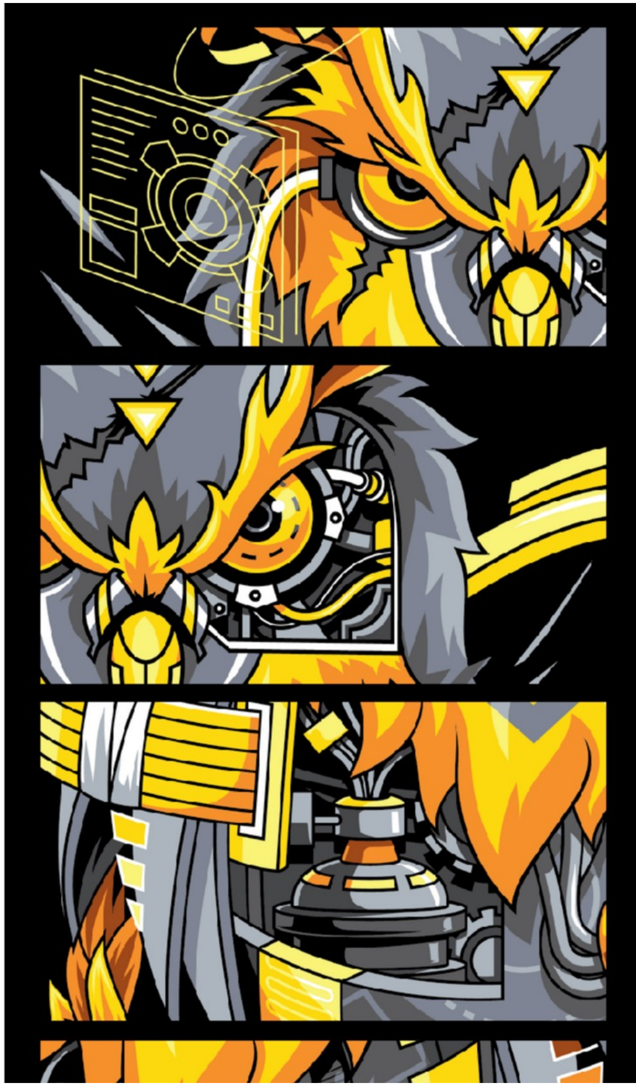
Many Ways to Count

Grid Approximation inefficient

Other methods:

Quadratic approximation

Markov chain Monte Carlo (MCMC)



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

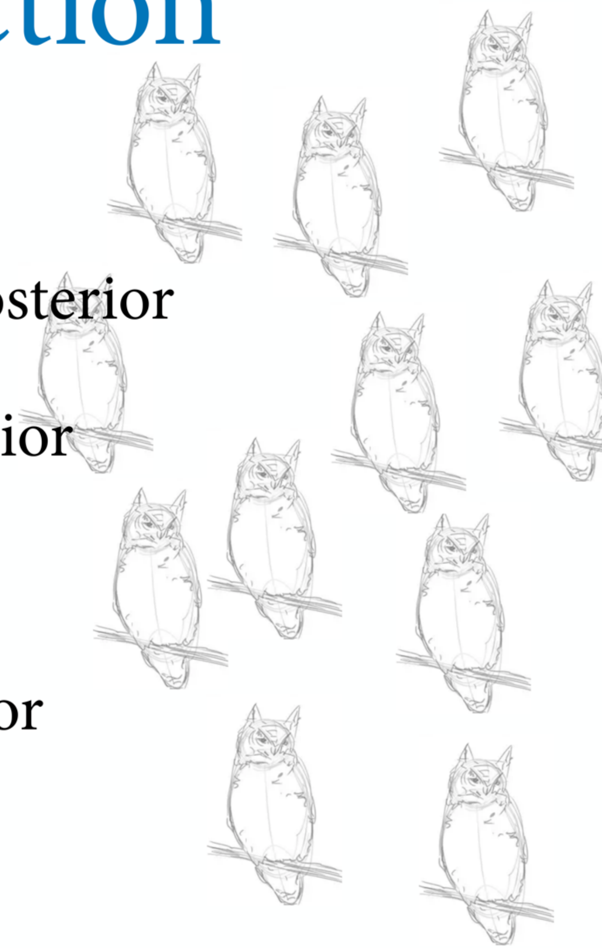
From Posterior to Prediction

Implications of model depend upon **entire** posterior

Must average any inference over entire posterior

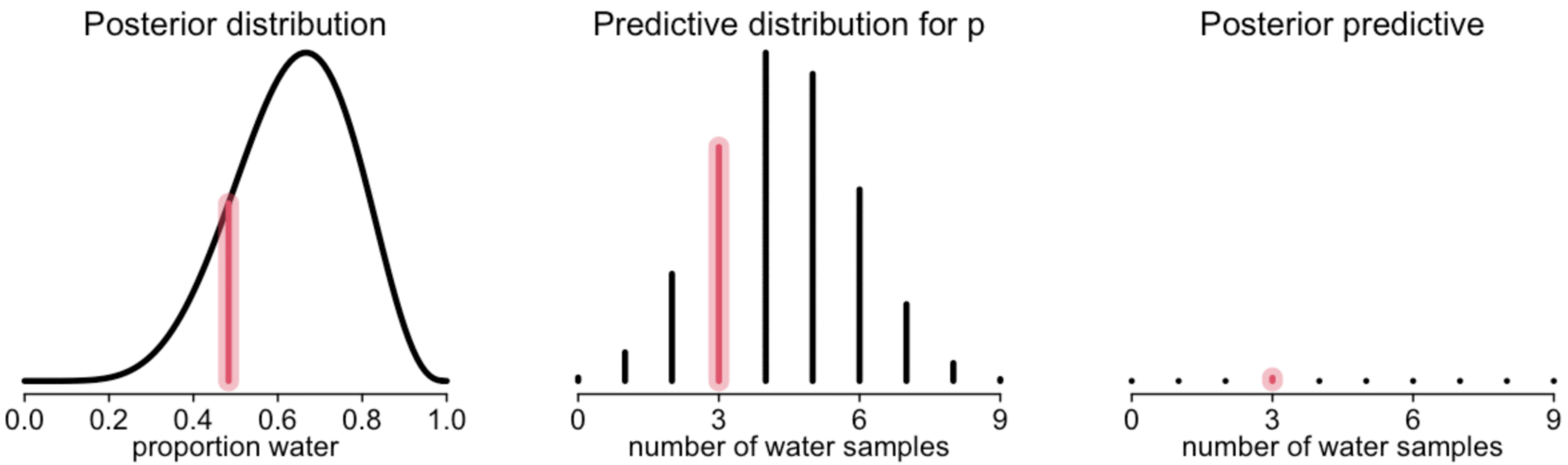
This usually requires integral calculus

OR we can just take samples from the posterior



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Uncertainty ⇨ Causal model ⇨ Implications



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Sample from posterior

R code 3.2

```
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)
```

R code 3.3

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

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Sample from posterior

R code 3.3

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

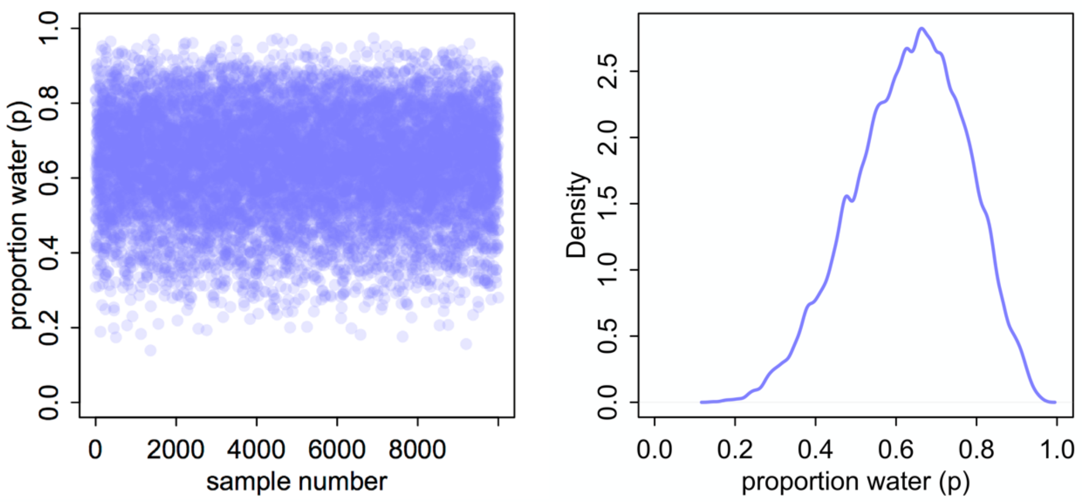


Figure 3.1

Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

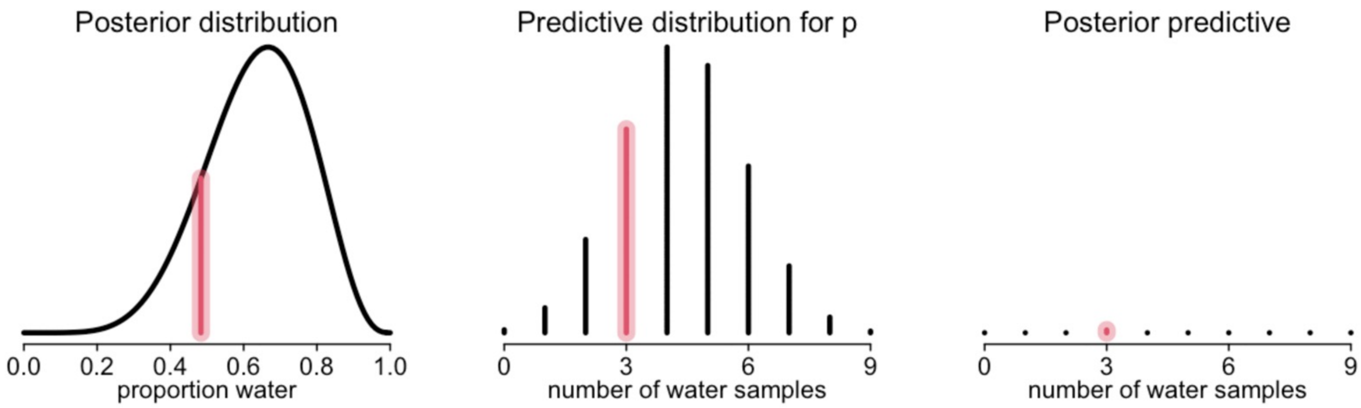
Sample predictions

R code 3.3

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

R code 3.26

```
w <- rbinom( 1e4 , size=9 , prob=samples )
```



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

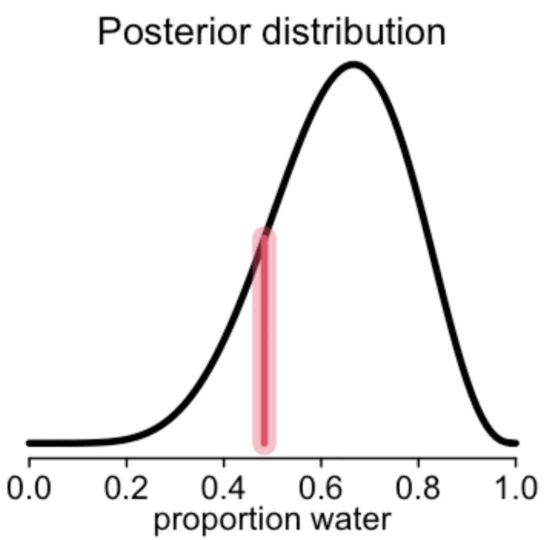
Sampling is Fun & Easy

Sample from posterior, compute desired quantity for each sample, profit

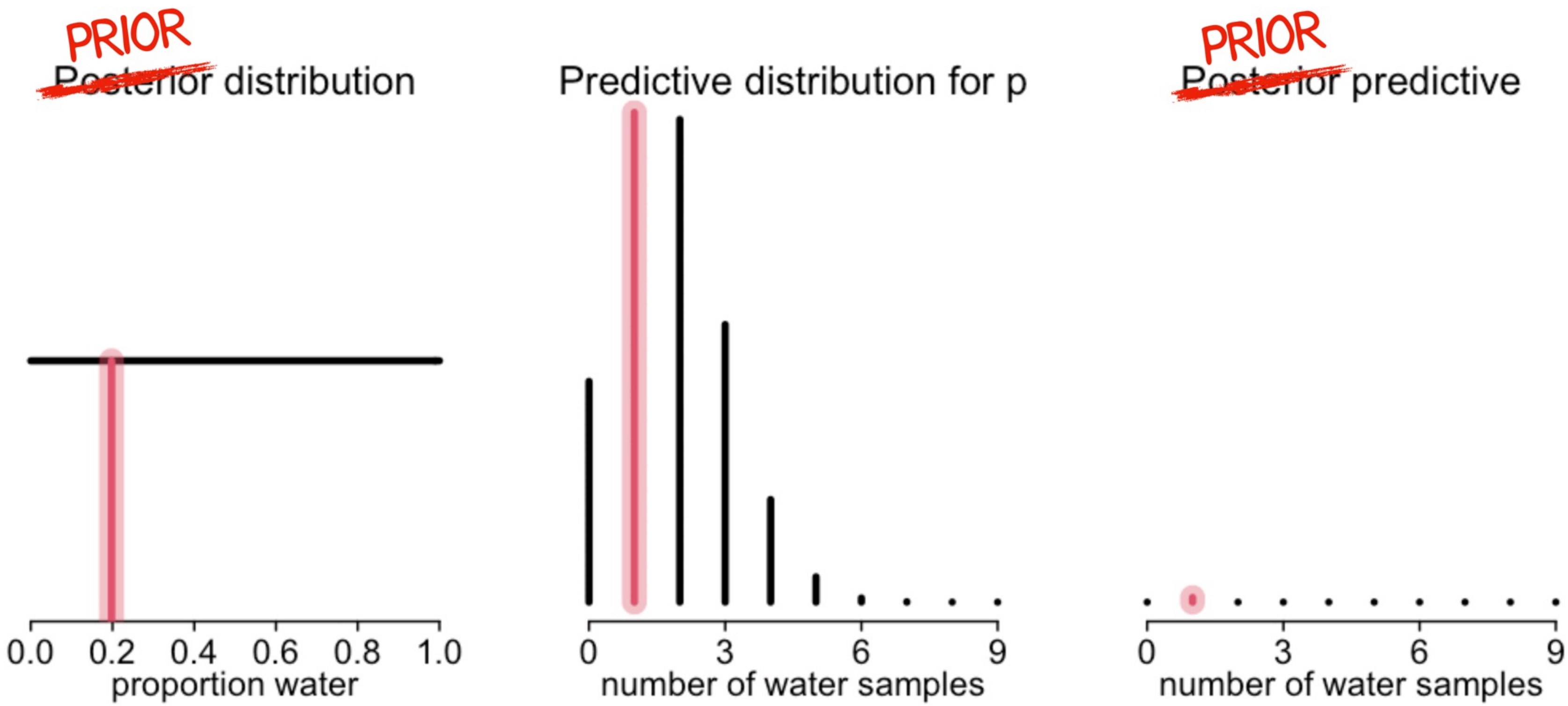
Much easier than doing integrals

Turn a **calculus problem** into a **data summary problem**

MCMC produces only samples anyway



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022



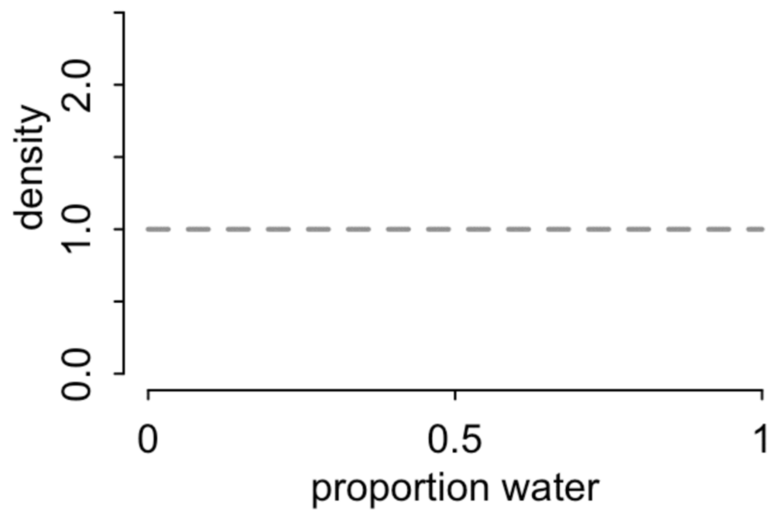
Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022

Bayesian modesty

No guarantees except ***logical***

Probability theory is a method of logically deducing ***implications of data*** *under assumptions that you must choose*

Any framework selling you more is hiding assumptions



Source: Richard McElraeth, https://github.com/rmcelreath/stat_rethinking_2022