TEWA 1: Advanced Data Analysis

Lecture 04

Lei Zhang
Social, Cognitive and Affective Neuroscience Unit (SCAN-Unit)
Department of Cognition, Emotion, and Methods in Psychology

lei.zhang@univie.ac.at
lei-zhang.net
@lei_zhang_lz

https://github.com/lei-zhang/tewa1_univie

universität wien
Fakultät für Psychologie

Bayesian warm-up?

Example I: discrete

cognitive model
statistics
computing

Joint probability :

$$P(X=0, Y=1) =$$

$$\sum_{x,y} P(X=x, Y=y) = 1$$

rain

		X	
		1	0
cold Y	1	0.5	0.1
	0	0.1	0.3

Marginal probability :

$$P(Y=1) =$$

$$P(X=0) =$$

$$P(X=x) = \sum_{y} P(X=x, Y=y)$$

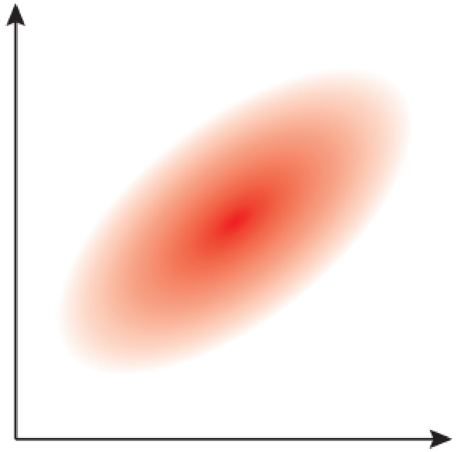
Conditional probability :

$$P(X=1|Y=1) =$$

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{P(X=x, Y=y)}{\sum_x P(X=x, Y=y)}$$

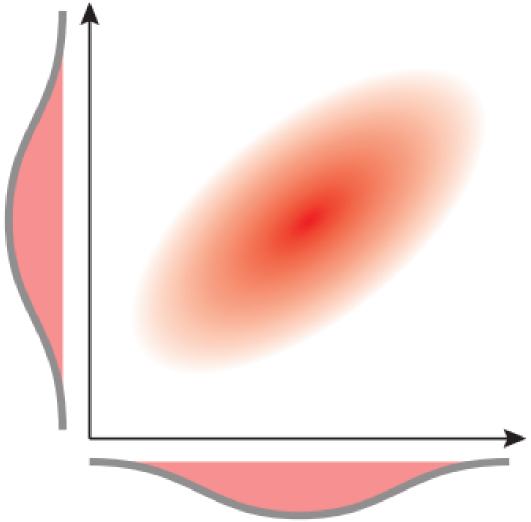
Example I: continuous

joint distribution



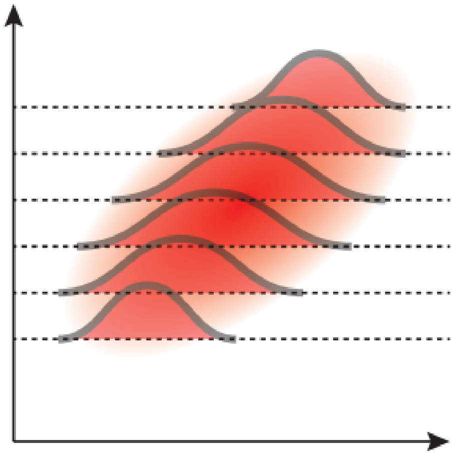
The "co-distribution" of x and y.

mariginal distribution



The density of x- (or y-) values, without knowing the other's value.

conditional distribution



The probability distribution of x, given that we know the value of y.

Credit: [Eric Ma]

Bayes' theorem

$$p(A,B) = p(B,A)$$

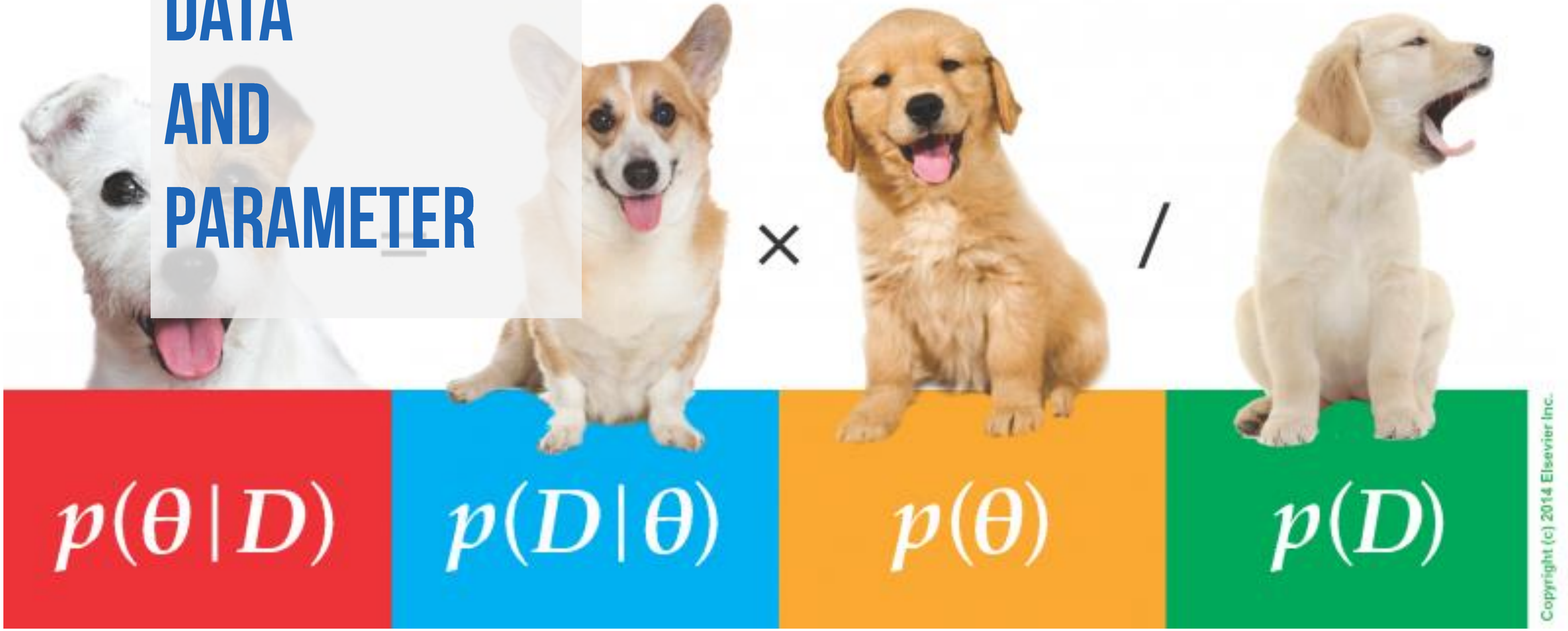
$$p(A,B) = p(A|B)\,p(B)$$

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

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

$$\boxed{p(A \mid B) = \frac{p(B \mid A)\,p(A)}{p(B)}}$$

LINKING DATA AND PARAMETER



Linking Data and Parameter

$$\theta \qquad\qquad D$$

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

Linking Data and Parameter

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Linking Data and Parameter

Likelihood
How plausible is the data given our parameter is true?

Prior
How plausible is our parameter before observing the data?

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Posterior
How plausible is our parameter given the observed data?

Evidence
How plausible is the data under all possible parameters?

What is $p(\text{Data}|\,\vartheta)$

L(theta | Data)

- This is the "Model"
- Data is fixed, ϑ varies
- Not a probability distribution
  - the sum is not "one"

Probability of coin landing heads up, θ	Number of heads, X 0	1	2	Total
0.0	1.00	0.00	0.00	1.00
0.2	0.64	0.32	0.04	1.00
0.4	0.36	0.48	0.16	1.00
0.6	0.16	0.48	0.36	1.00
0.8	0.04	0.32	0.64	1.00
1.0	0.00	0.00	1.00	1.00
Total	2.20	1.60	2.20	

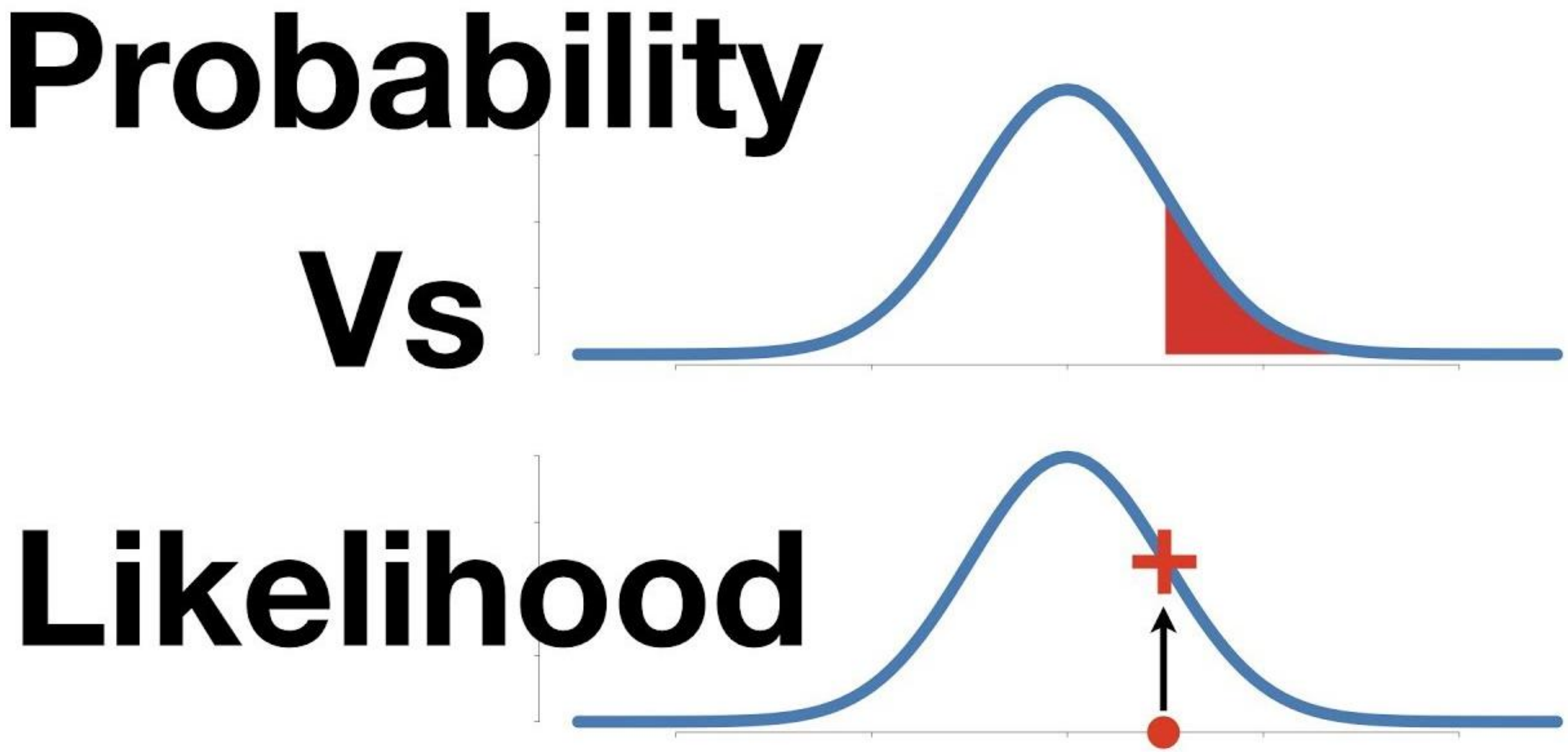
$$Pr(X=0\,|\,\theta)=Pr(T,T\,|\,\theta)=Pr(T\,|\,\theta)\times Pr(T\,|\,\theta)=(1-\theta)^2$$

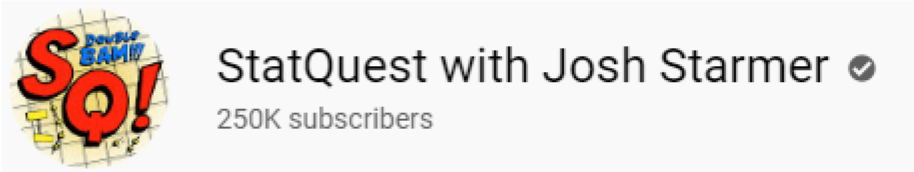
$$Pr(X=1\,|\,\theta)=Pr(H,T\,|\,\theta)+Pr(T,H\,|\,\theta)=2\times Pr(T\,|\,\theta)\times Pr(H\,|\,\theta)=2\theta(1-\theta)$$

$$Pr(X=2\,|\,\theta)=Pr(H,H\,|\,\theta)=Pr(H\,|\,\theta)\times Pr(H\,|\,\theta)=\theta^2.$$

Lambert (2018)

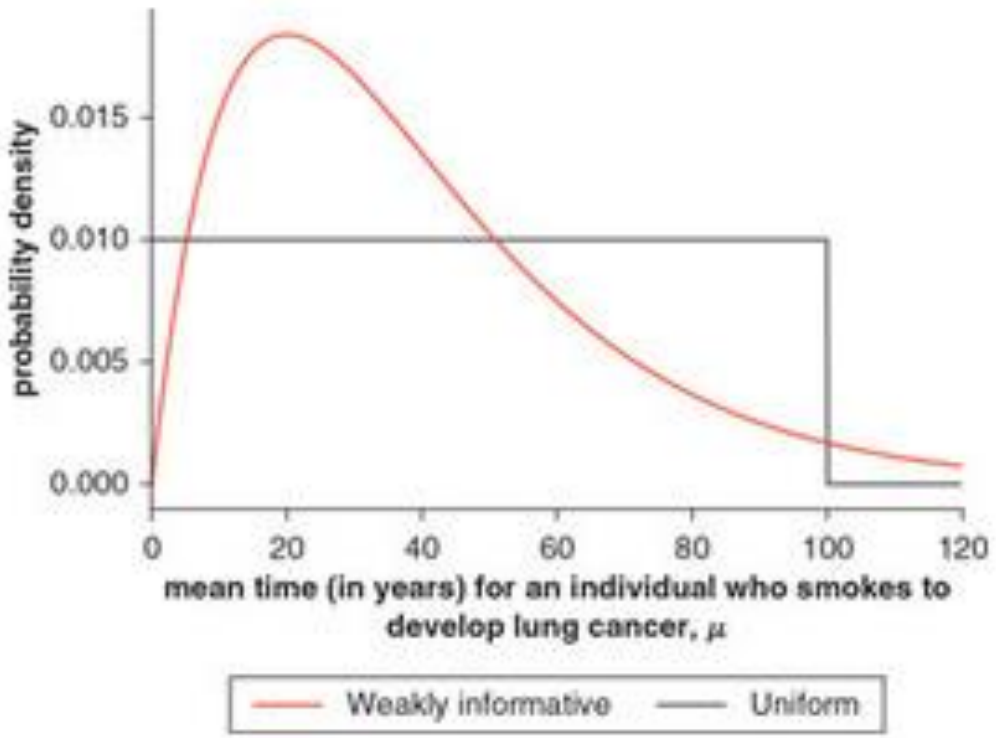
Watch this video!





https://youtu.be/pYxNSUDSFH4

What is $p(\vartheta)$?



Lambert (2018)

What is *p*(*Data*)?

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

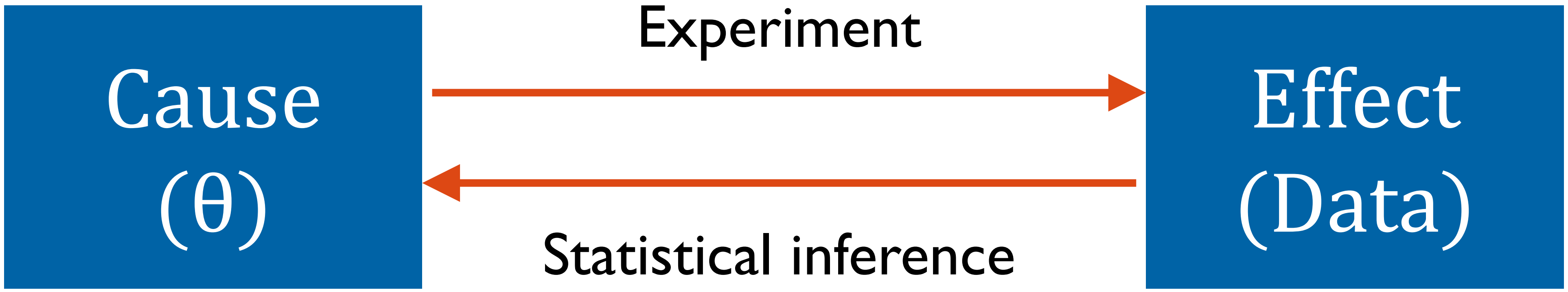
discrete parameters

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\boxed{\sum_{\theta^*} p(D \mid \theta^*)\, p(\theta^*)}}$$

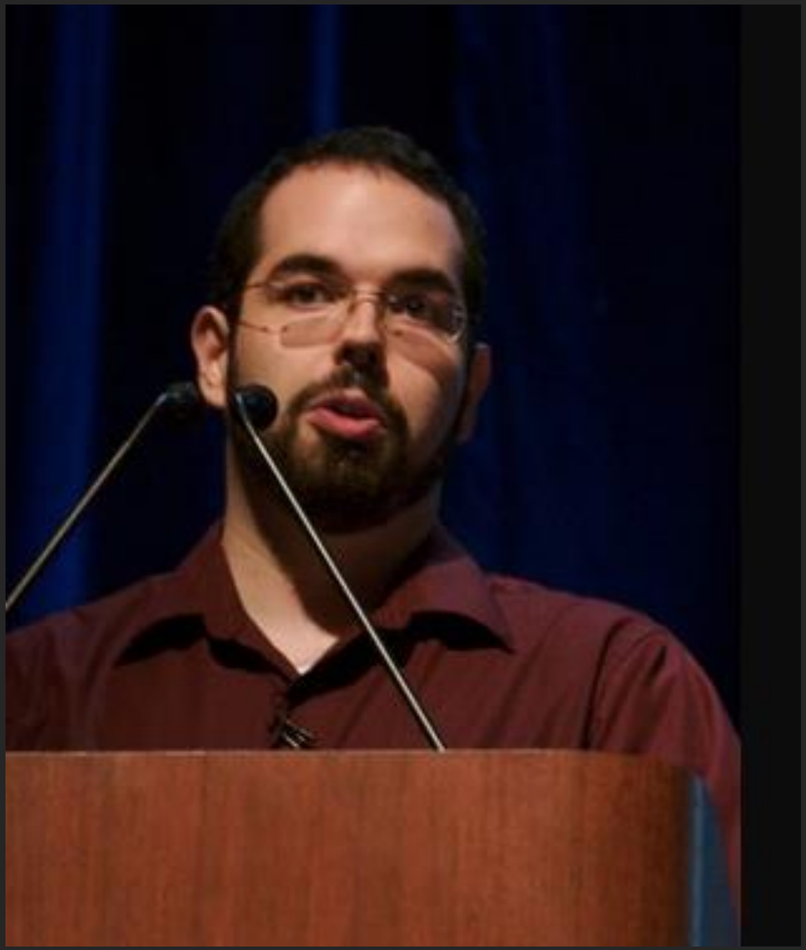
continuous parameters

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\boxed{\int p(D \mid \theta^*)\, p(\theta^*)\, d\theta^*}}$$

Why the Bayes' theorem is important?



$$\boxed{p(\theta|D)} = \frac{p(D|\theta)p(\theta)}{p(D)}$$



"Probability is orderly opinion and <u>inference</u> from data <u>is</u> nothing other than <u>the revision of</u> such <u>opinion</u> in the light of relevant new information."

Eliezer S. Yudkowsky

BINOMIAL MODEL

Binomial Model

cognitive model
statistics
computing

- You are curious how much of the surface is covered in water.
- You will toss the globe up in the air.
- You will record whether or not the surface under your right index finger is water (W) or land (L).
- You might observe: W L W W W L W L W
- → 6/9 = 0.666667?
- Is it right? If not, what to do next?

McElreath (2016)

Steps of (Bayesian) Modeling?

A data story

Think about how the data might arise.
It can be *descriptive* or even *causal*.

Update

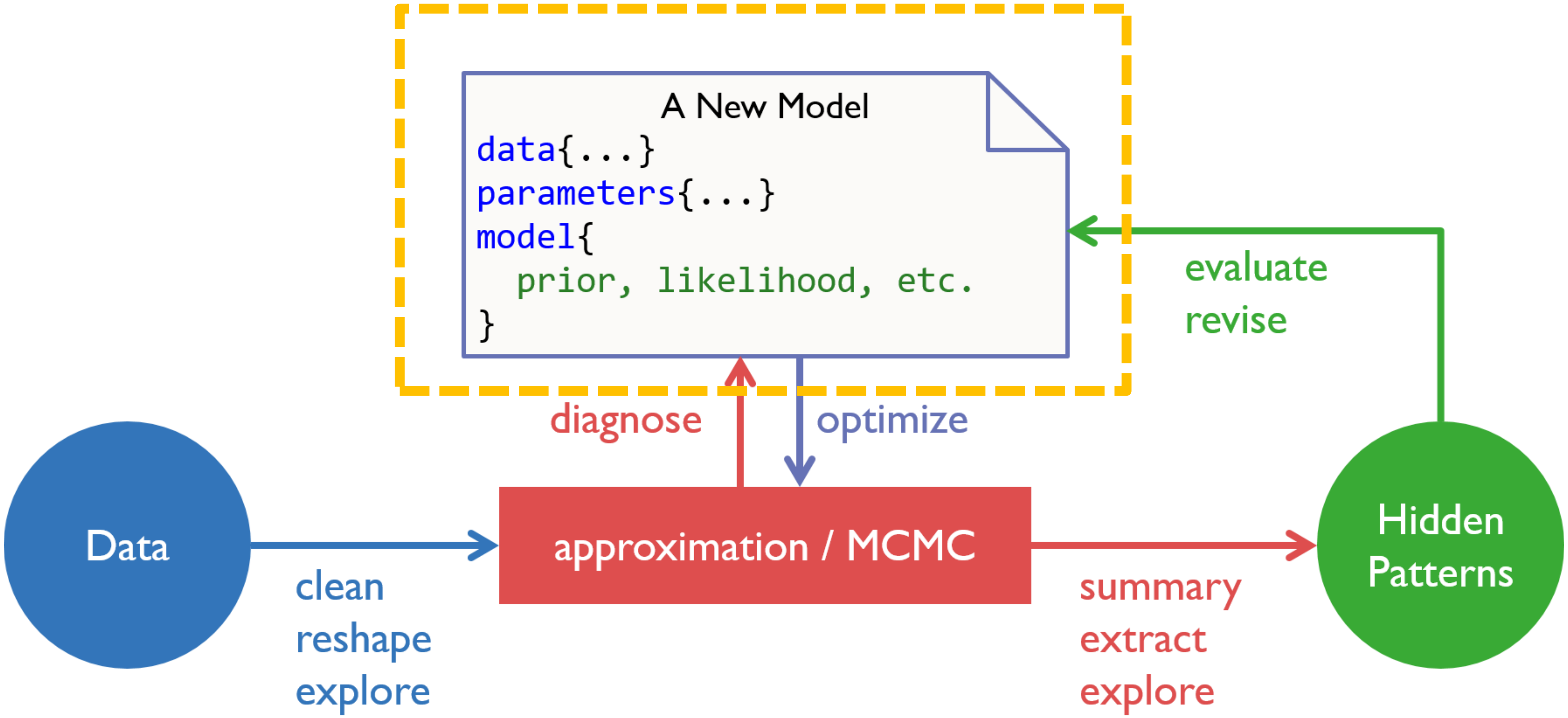
Educate your model by feeding it with data.
Bayesian Update:

- update the prior, in light of data, to produce posterior
- the updated posterior then becomes the prior of next update

Evaluate

Compare model with reality.
Revise your model.

McElreath (2016)


A New Model
data{...}
parameters{...}
model{
prior, likelihood, etc.
}
evaluate
revise
diagnose
optimize
Data
clean
reshape
explore
approximation / MCMC
summary
extract
explore
Hidden
Patterns

A Data Story of the Globe

- The true proportion of water covering the globe is ϑ.
- A single toss of the globe has a probability ϑ of producing a water (W) observation.
- It has a probability $(1 - \vartheta)$ of producing a land (L) observation.
- Each toss of the globe is independent of the others.

McElreath (2016)

Components of a Model

think about the likelihood function (of Binomial):

$$p(\theta \mid D) = \frac{\boxed{p(D \mid \theta)}\, p(\theta)}{\int p(D \mid \theta^*)\, p(\theta^*)\, d\theta^*}$$

$$p(w \mid N, \theta) = \binom{N}{w} \theta^w (1-\theta)^{N-w}$$

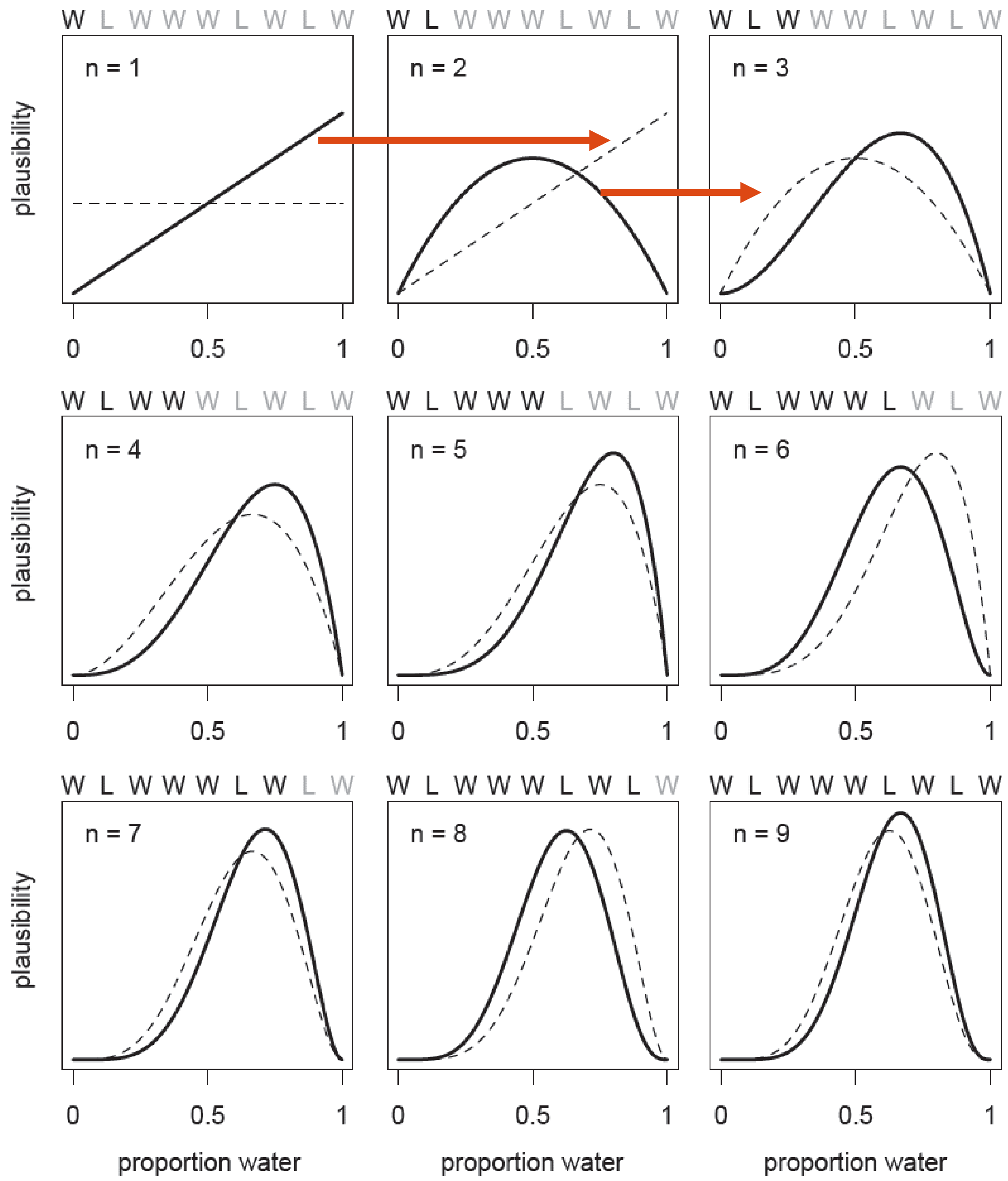
N: total number of observations
w: number of water

known (data)

ϑ : proportion of water

unknown (parameter)

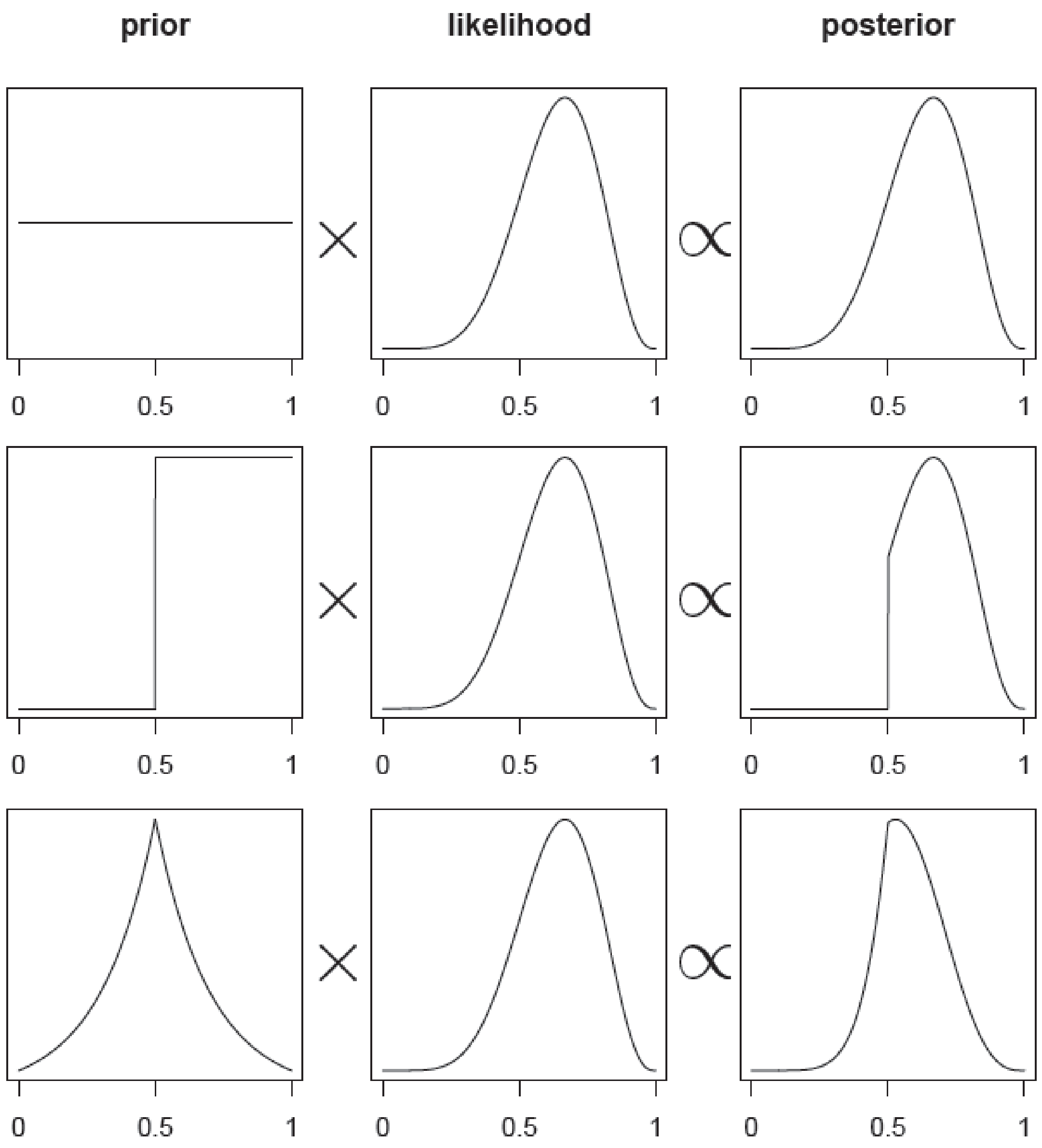
Update



- order doesn't matter
- 2/3 is most likely
- others are not ruled out

McElreath (2016)

Impact of Prior



McElreath (2016)

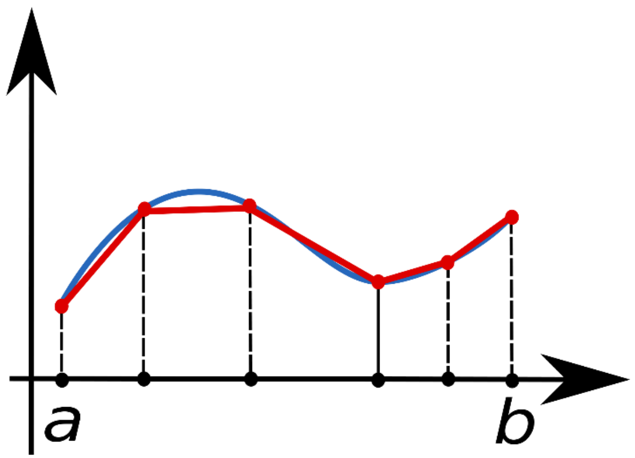
Solve it by Grid Approximation

discrete parameters

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\boxed{\sum_{\theta^*} p(D \mid \theta^*)\, p(\theta^*)}}$$

continuous parameters

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\boxed{\int p(D \mid \theta^*)\, p(\theta^*)\, d\theta^*}}$$

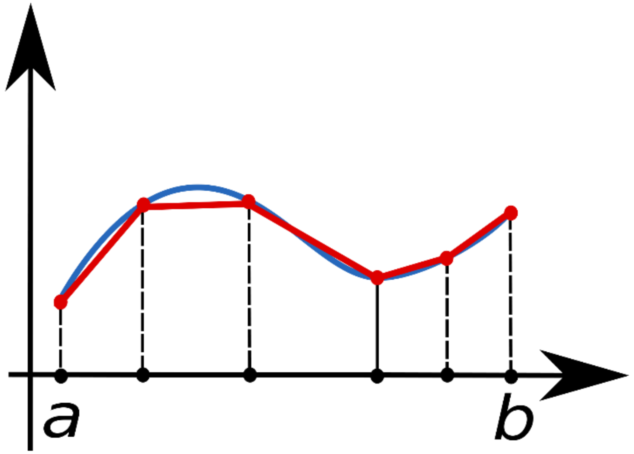


Binomial Model – Grid Approximation

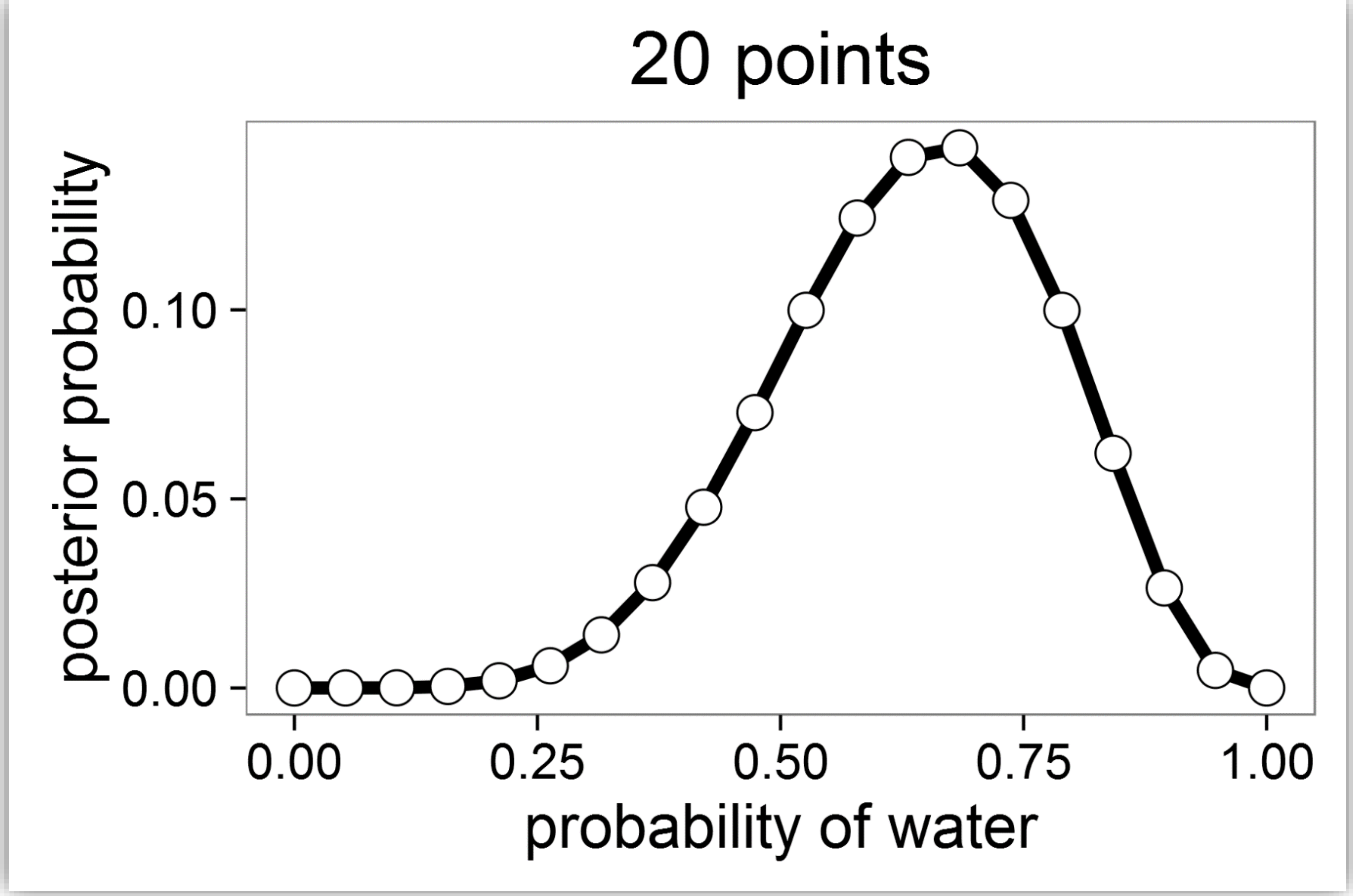

```
theta_start <- 0; theta_end <- 1; n_grid <- 20
w <- 6; N <- 9

# define grid
theta_grid <- seq(from = theta_start, to = theta_end,
                  length.out = n_grid)

# define prior
prior <- rep(1 , n_grid)

# compute likelihood at each value in grid
likelihood <- dbinom(w, size = N, prob = theta_grid)

# compute product of likelihood and prior
unstd.posterior <- likelihood * prior

# standardize the posterior, so it sums to 1
posterior <- unstd.posterior / sum(unstd.posterior)
```

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\int p(D \mid \theta^*)\, p(\theta^*)\, d\theta^*}$$

$$p(w \mid N, \theta) = \binom{N}{w} \theta^w (1-\theta)^{N-w}$$



Binomial Model – Grid Approximation



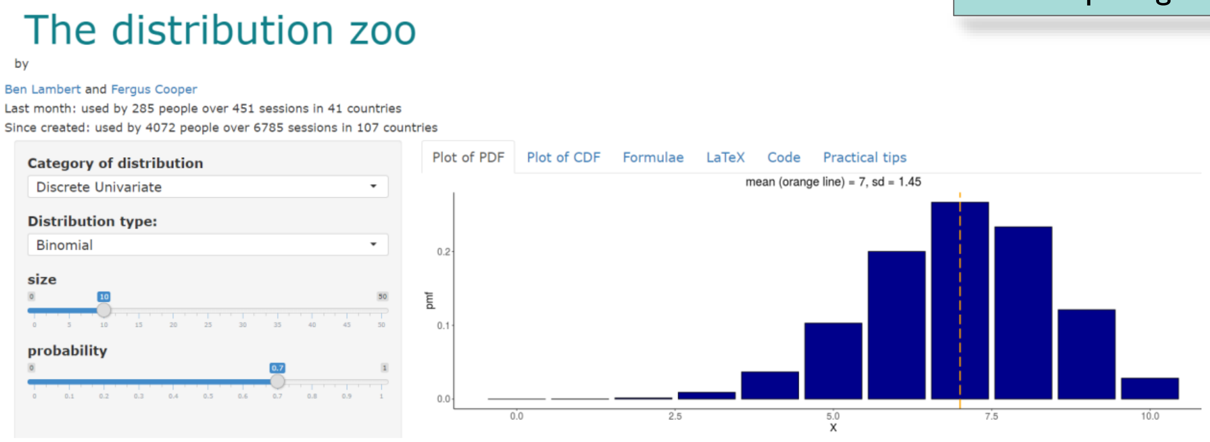
Exercise VII

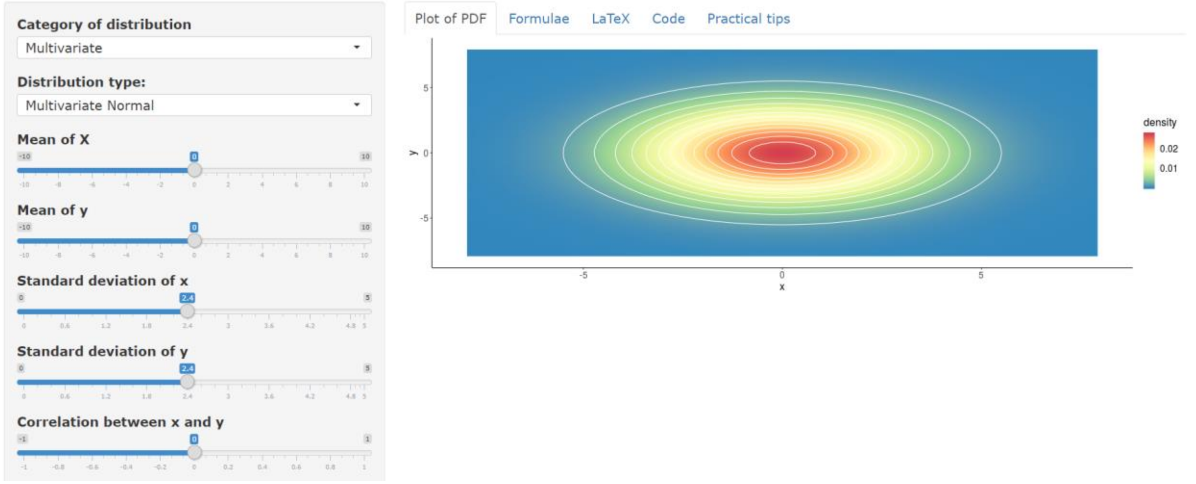

```
…/BayesCog/02.binomial_globe/_scripts/binomial_globe_grid.R
```

TASK: run a grid approximation with `grid_size = 50`

How do I know which likelihood to use?

cognitive model
statistics
computing





Lambert (2018)

https://ben18785.shinyapps.io/distribution-zoo/

What if I have multiple parameters?

grid approximation for
2 parameters?
5 parameters?
10 parameters?

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\boxed{\int p(D \mid \theta^*)\, p(\theta^*)\, d\theta^*}}$$

$$p(data) = \int_{\text{All}\,\theta_1} \int_{\text{All}\,\theta_2} p(data, \theta_1, \theta_2)\, \mathrm{d}\theta_1 \mathrm{d}\theta_2$$

$$p(data) = \int_{\mu_1} \int_{\sigma_1} \cdots \int_{\mu_{100}} \int_{\sigma_{100}} \underbrace{p(data \mid \mu_1, \sigma_1, \ldots, \mu_{100}, \sigma_{100})}_{\text{likelihood}} \times \underbrace{p(\mu_1, \sigma_1, \ldots, \mu_{100}, \sigma_{100})}_{\text{prior}} \, \mathrm{d}\mu_1 \mathrm{d}\sigma_1 \ldots \mathrm{d}\mu_{100} \mathrm{d}\sigma_{100},$$

- Analytical solutions (often does not exist)
- Grid approximation (takes too long)
- Markov Chain Monte Carlo

$$\boxed{p(\theta \mid D) \propto p(D \mid \theta)\, p(\theta)}$$


ANY
QUESTIONS
?

Happy Computing!