Bayesian Statistics and Hierarchical Bayesian Modeling for Psychological Science

Lecture 06

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

universität wien

Fakultät für Psychologie

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

Bayesian warm-up?

Binomial Model

- 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?

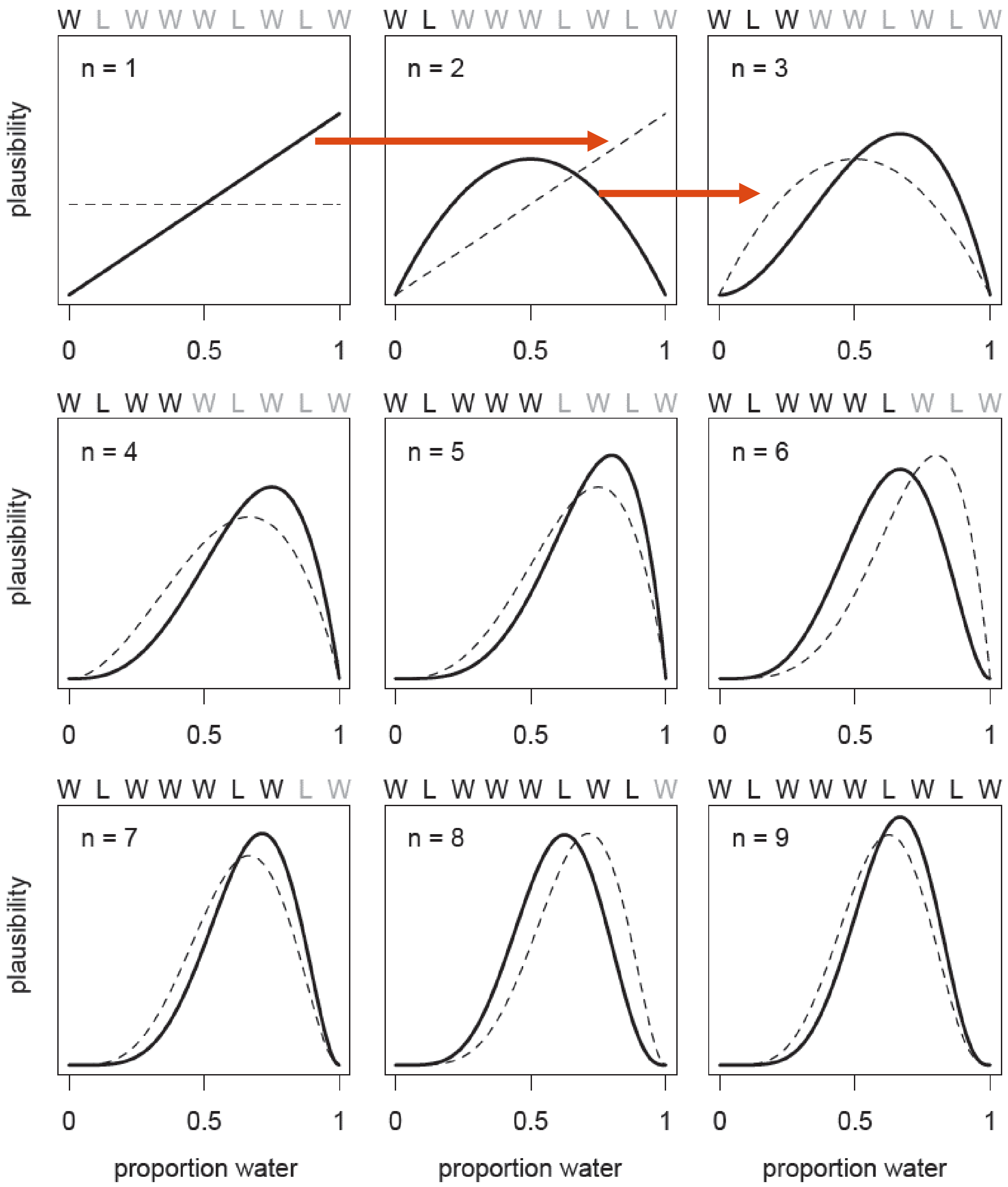
McElreath (2016)

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)

Update



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

McElreath (2016)

Solve it by Grid Approximation

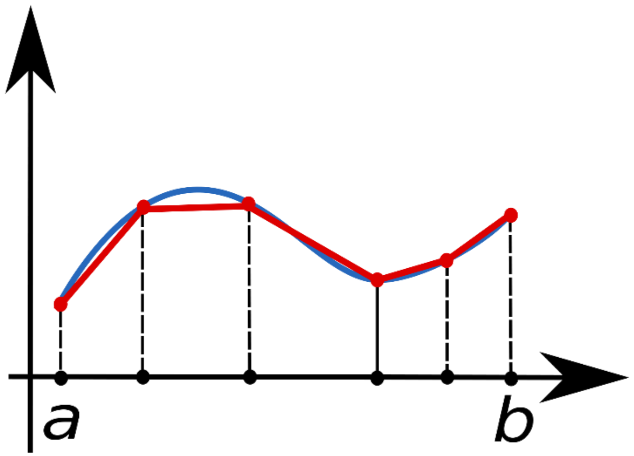
cognitive model
statistics
computing

discrete parameters

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

continuous parameters

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

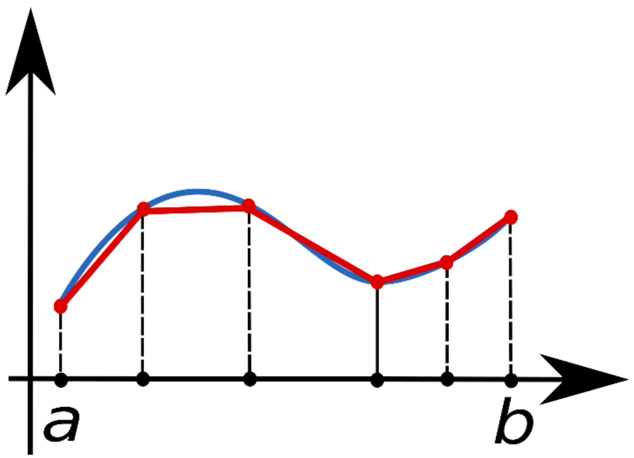


Binomial Model – Grid Approximation

```
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)
```

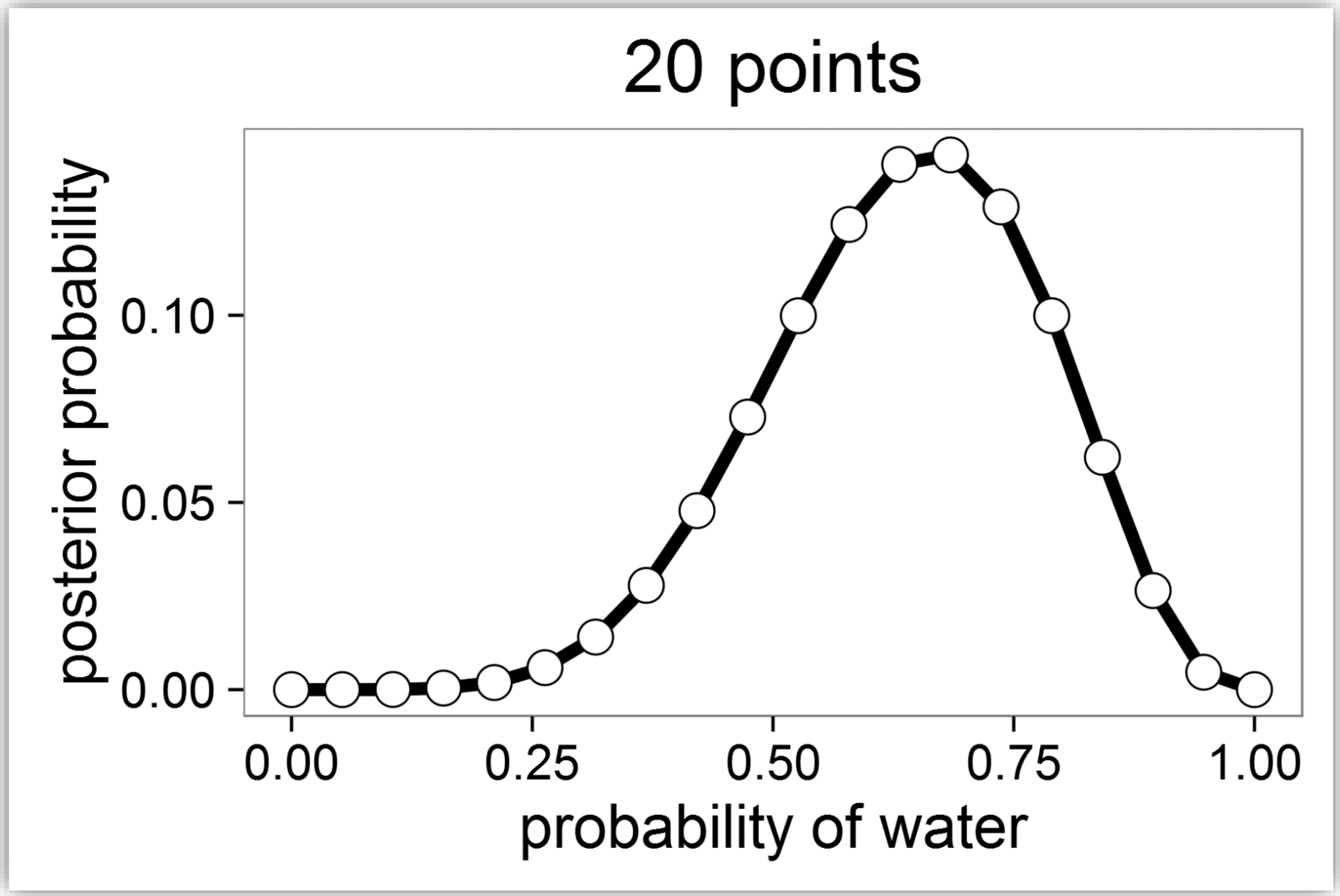
$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\int p(D \mid \theta^*)\, p(\theta^*)\, d\theta^*}$$

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

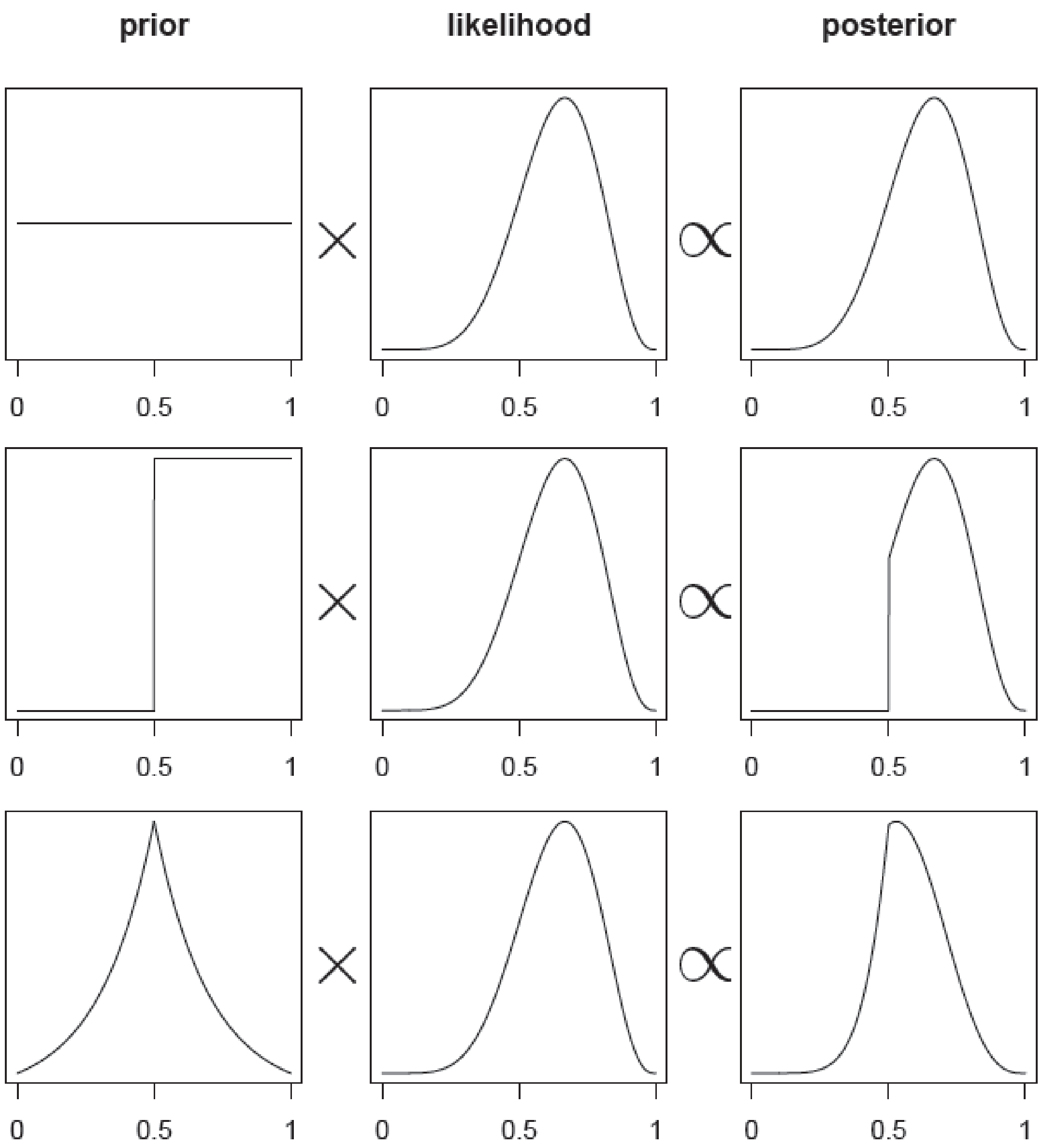


Binomial Model – Grid Approximation

cognitive model
statistics
computing



Impact of Prior



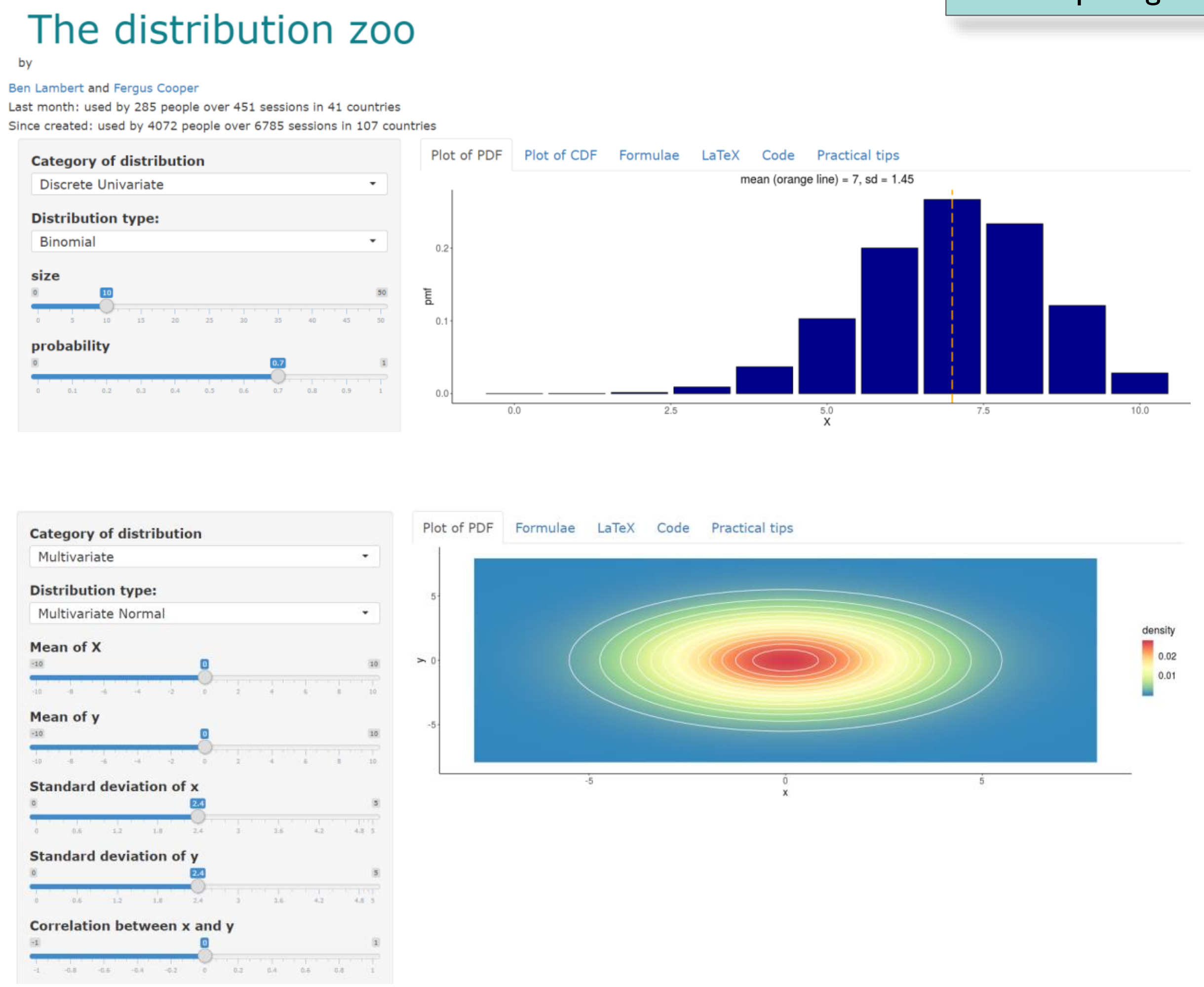
McElreath (2016)

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?



Lambert (2018)

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

What if I have multiple parameters?

cognitive model
statistics
computing

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

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

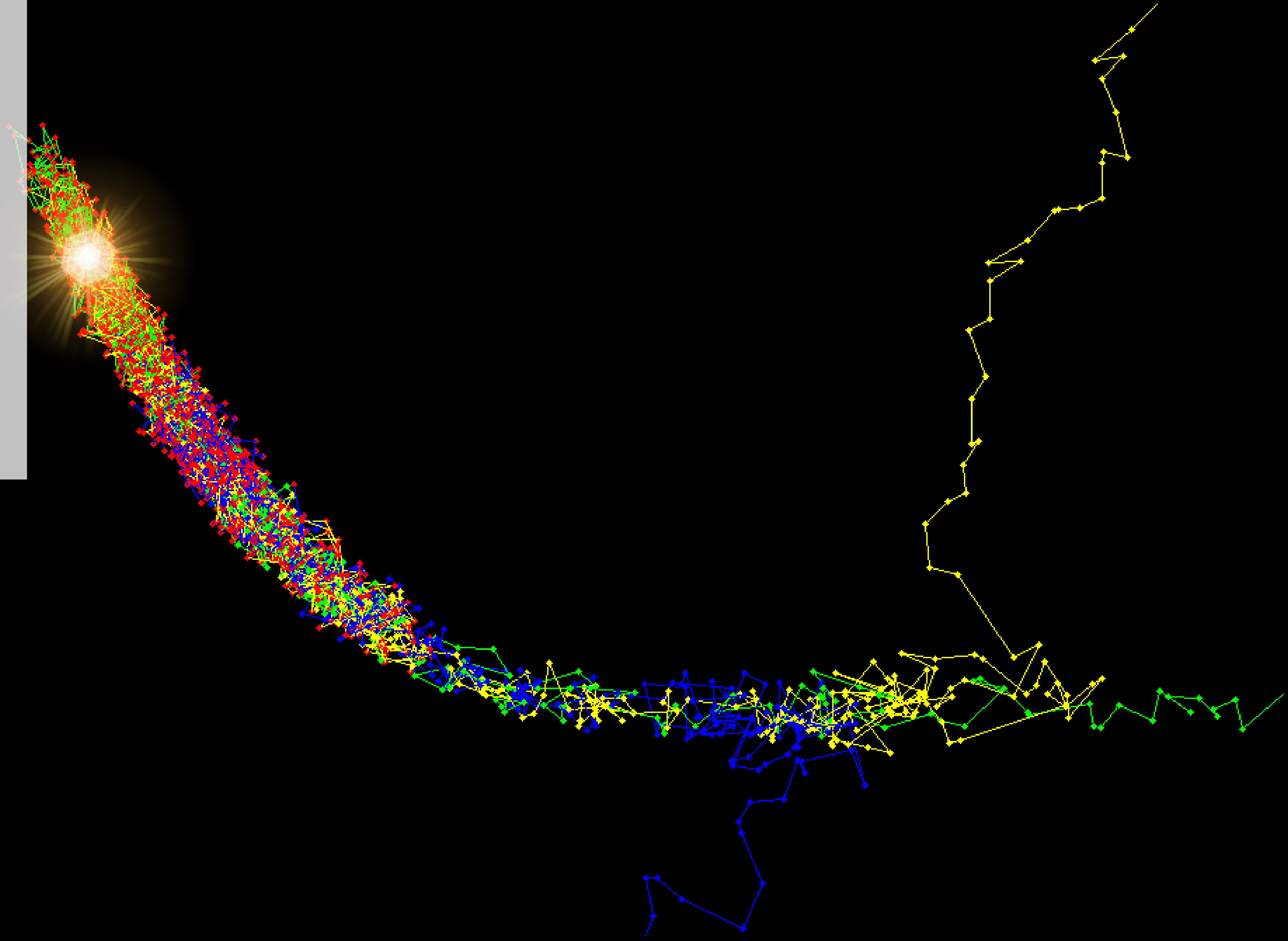
$$p(data) = \int_{\text{All}\,\theta_1} \int_{\text{All}\,\theta_2} p(data, \theta_1, \theta_2)\, \mathrm{d}\theta_1 \mathrm{d}\theta_2$$

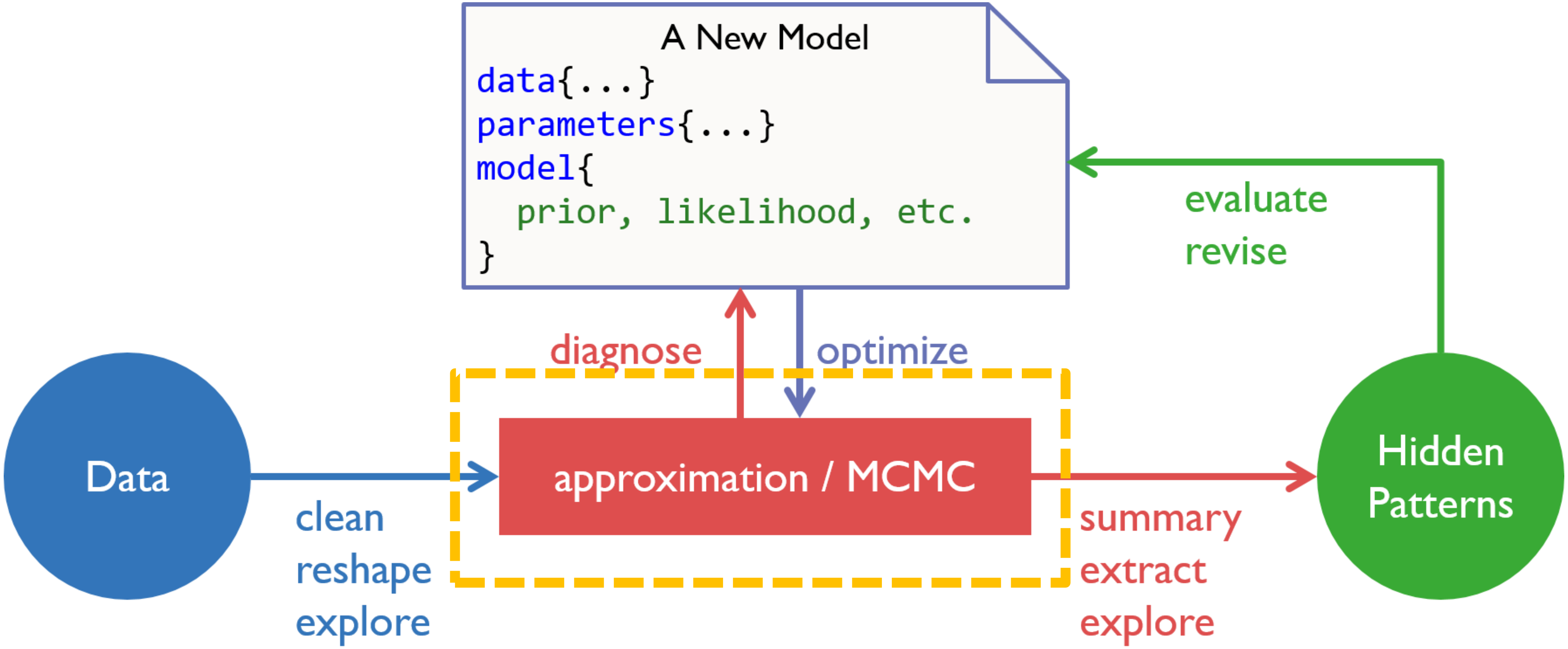
$$p(data) = \int_{\mu_1} \int_{\sigma_1} \cdots \int_{\mu_{100}} \int_{\sigma_{100}} \underbrace{p(data \mid \mu_1, \sigma_1, \ldots, \mu_{100}, \sigma_{100})}_{\text{likelihood}} \times \underbrace{p(\mu_1, \sigma_1, \ldots, \mu_{100}, \sigma_{100})}_{\text{prior}} \, \mathrm{d}\mu_1 \mathrm{d}\sigma_1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

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

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

MARKOV CHAIN MONTE CARLO




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

Solving the Problem by Approximation

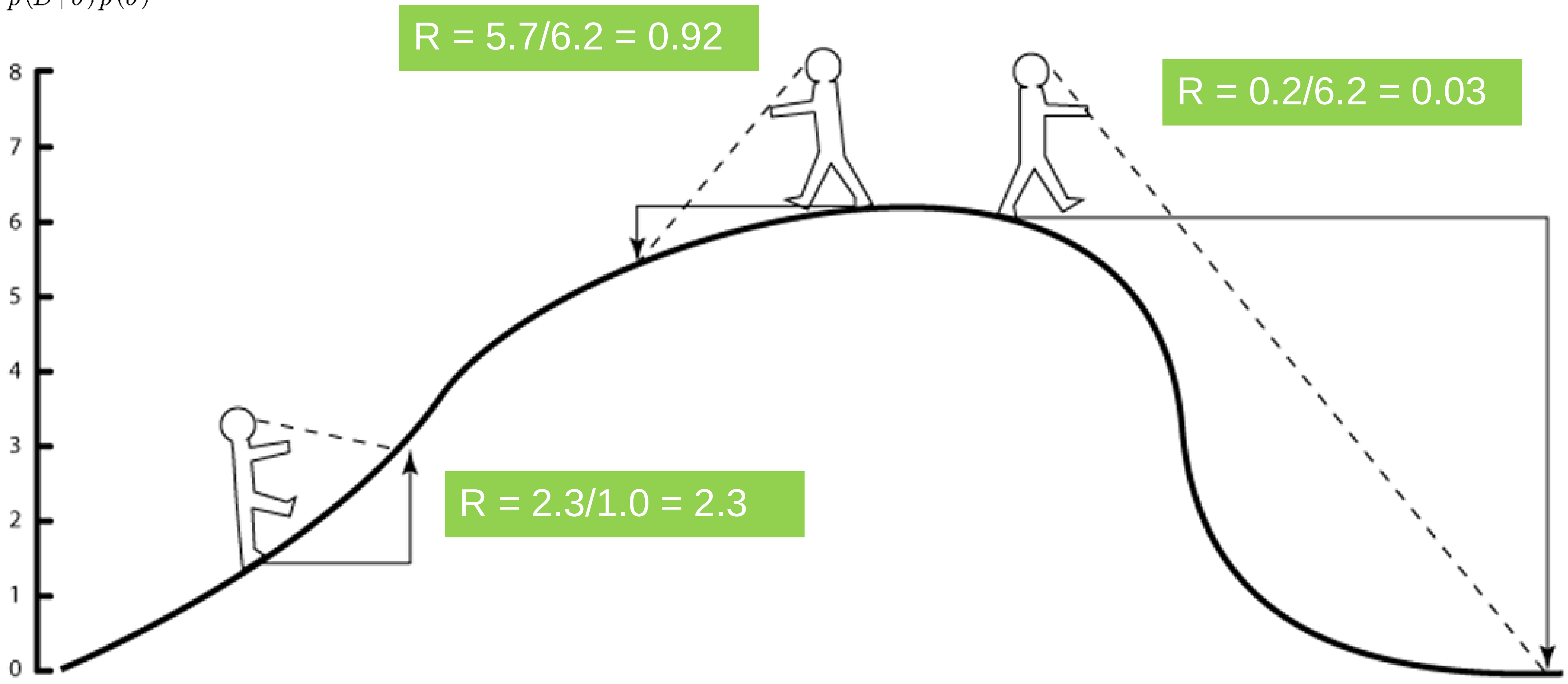
$$p(\theta \mid D) \propto p(D \mid \theta)\, p(\theta)$$

Deterministic Approximation	Stochastic Approximation
→ Variational Bayes	→ Sampling Methods

An MCMC Robot

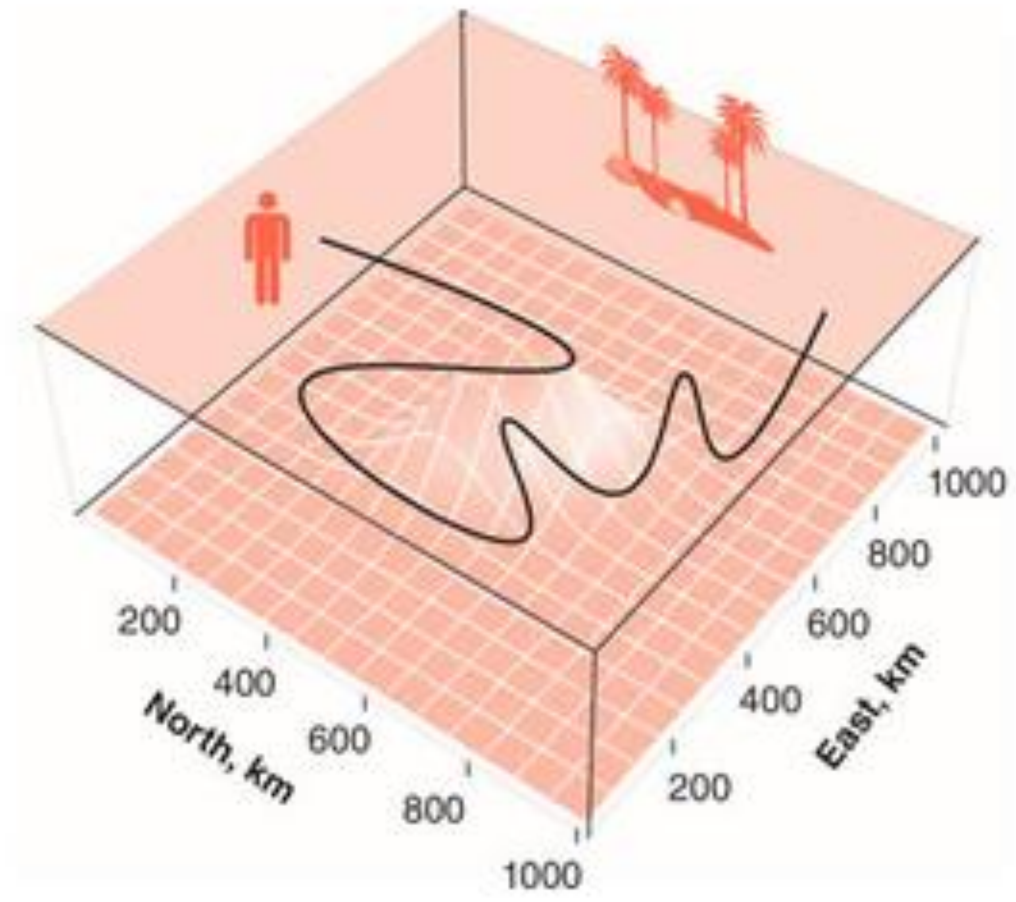
cognitive model
statistics
computing

$p(\theta \mid D) \propto p(D \mid \theta) p(\theta)$



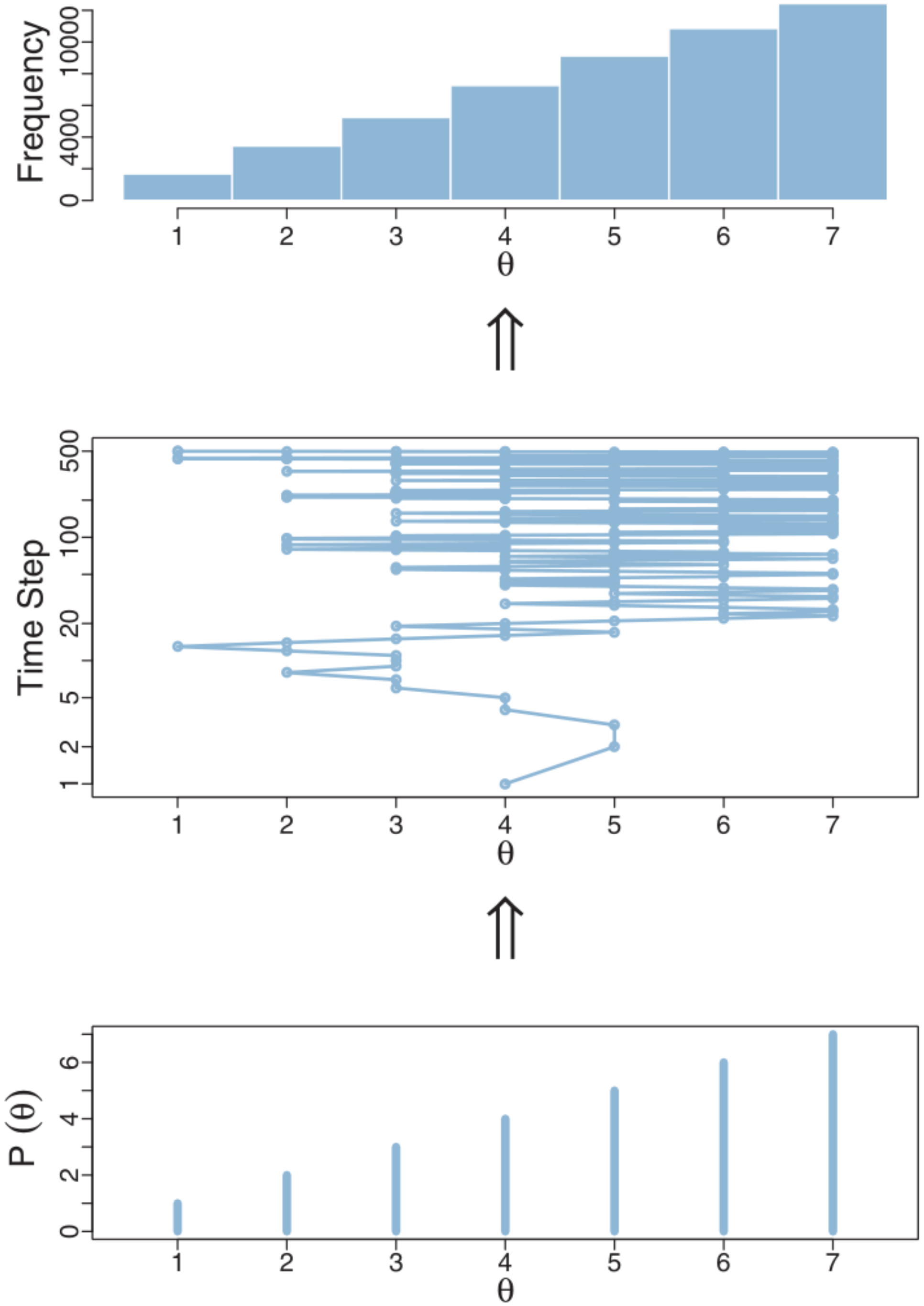
http://marple.eeb.uconn.edu/mcmcrobot/

An MCMC Robert in 3D



Lambert (2018)

Sampling Example: Discrete



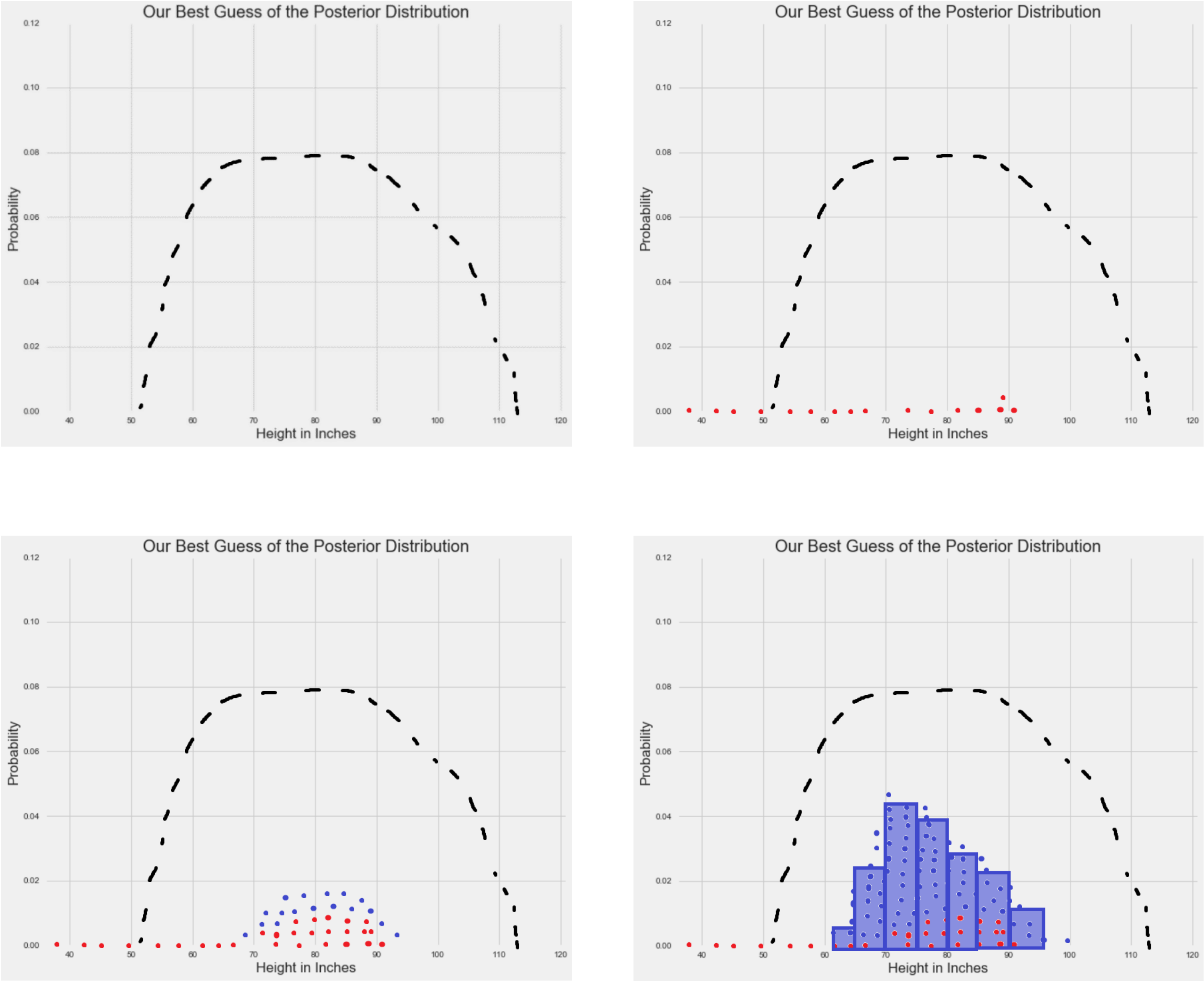
MCMC summary

MCMC trace

True distribution

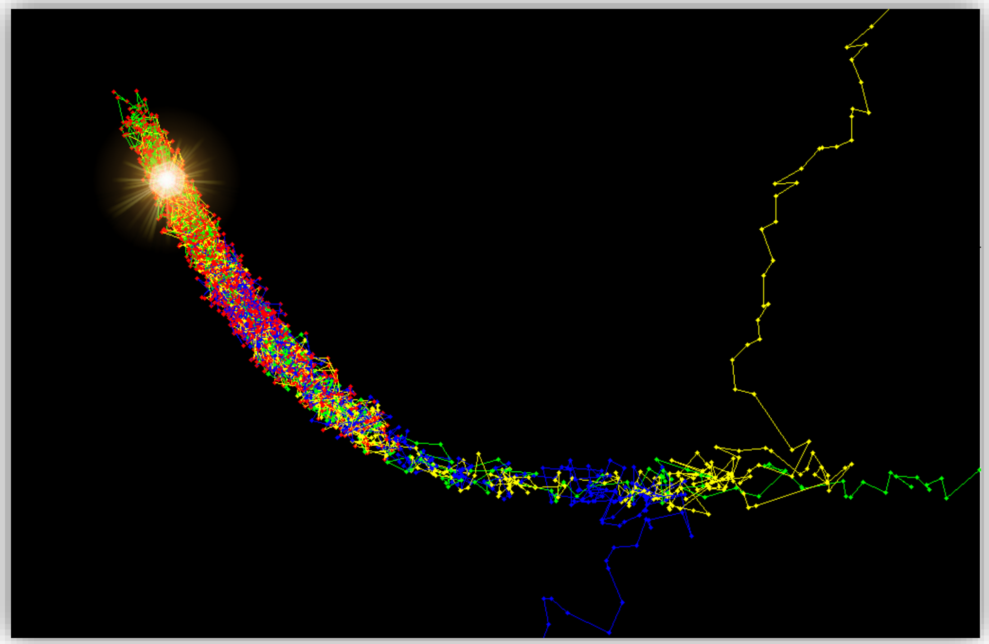
Kruschke (2015)

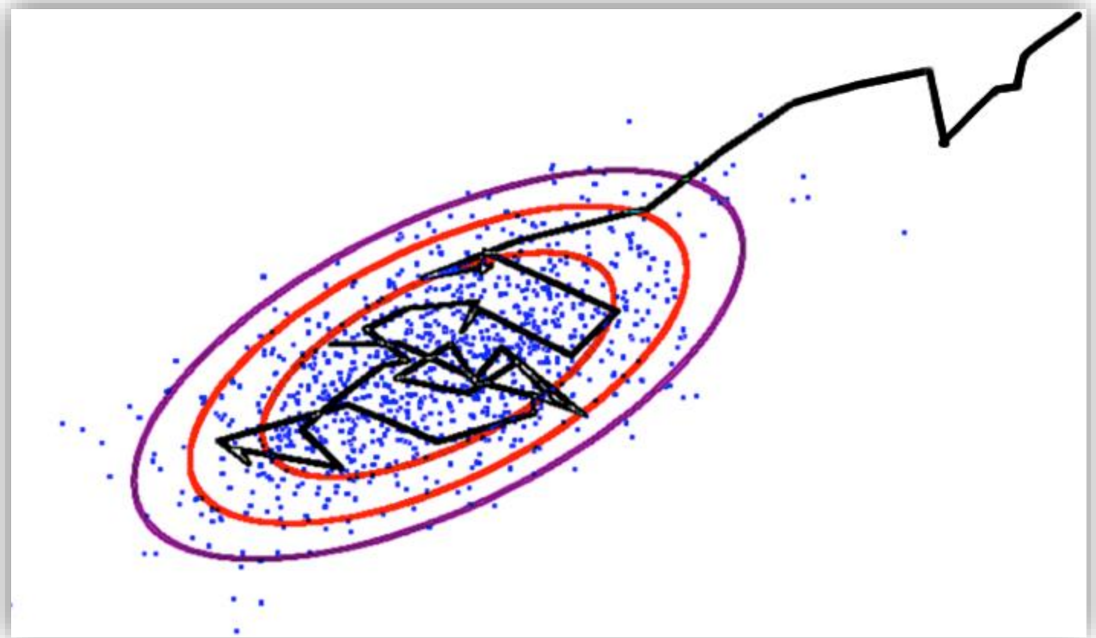
Sampling Example: Continuous



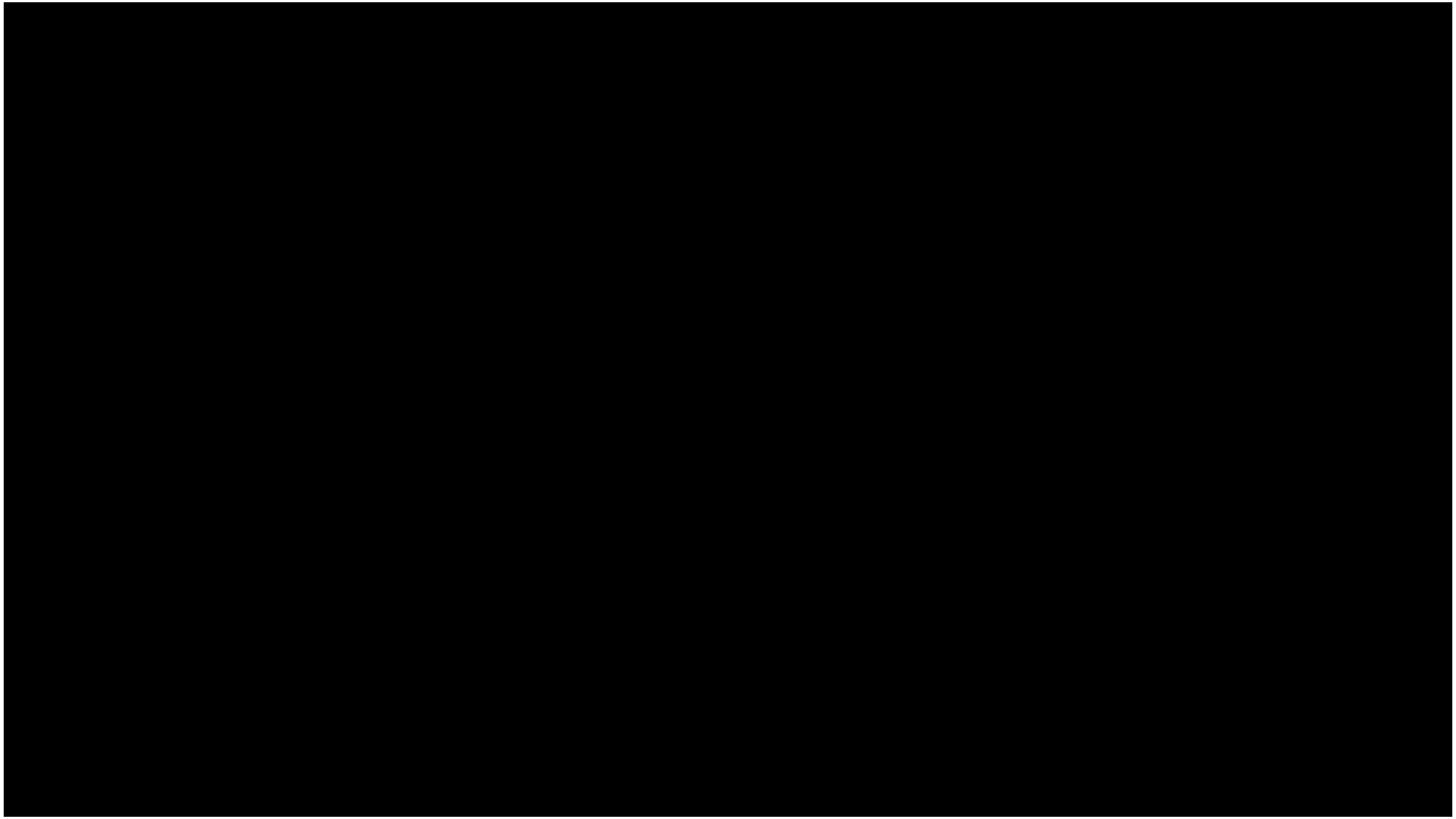
[A Zero-Math Introduction to Markov Chain Monte Carlo Methods]()

Visual Example





Let’s watch a video!



Credit: Thomas Wiecki

MCMC Sampling Algorithms

- Rejection sampling
- Importance sampling
- Metropolis algorithm
- Gibbs sampling → JAGS
- HMC sampling*

Stan!

*: also PyMC3

PyMC3


ANY
QUESTIONS
?

Happy Computing!