TEWA 1: Advanced Data Analysis

Lecture 06

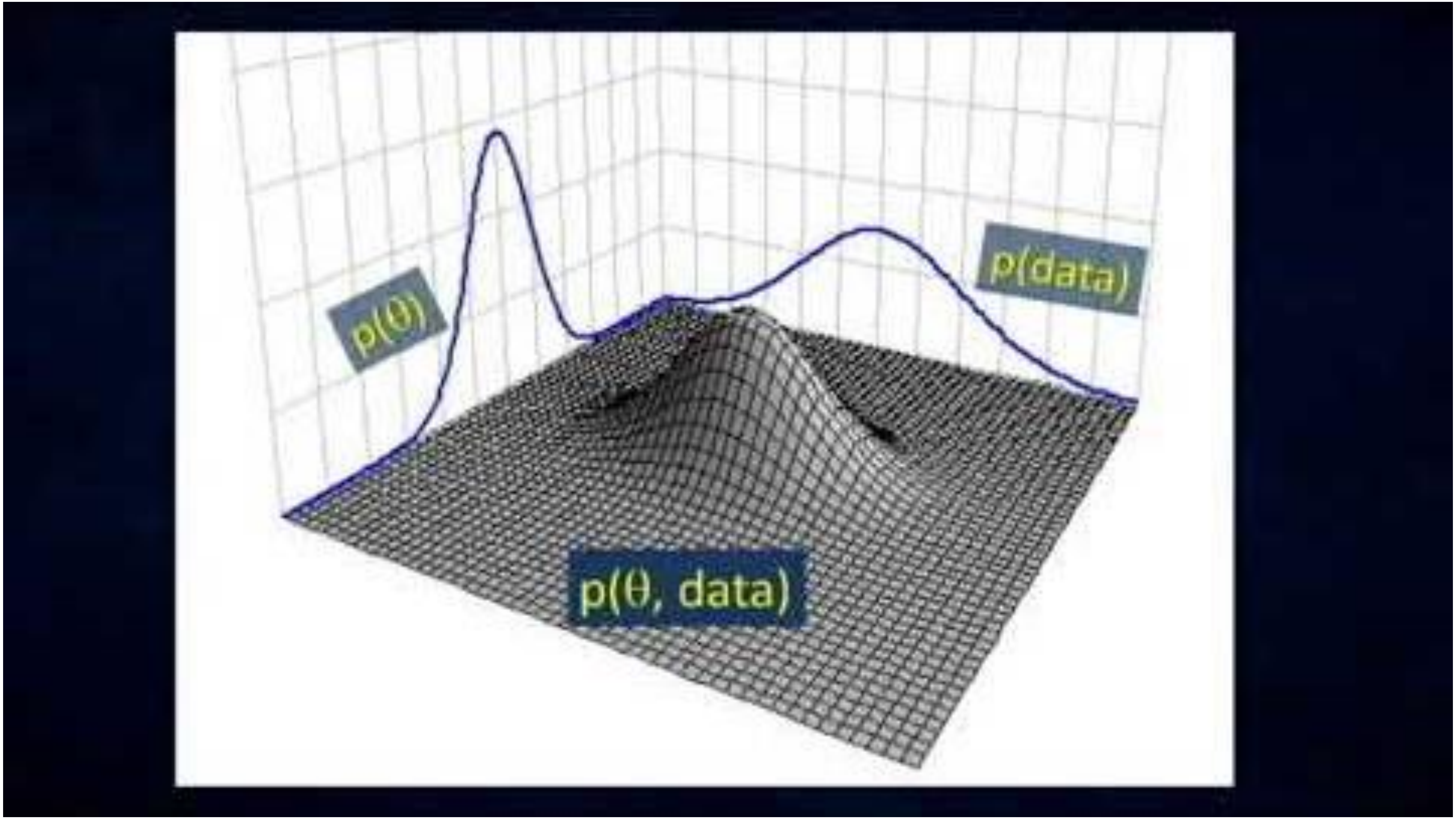
Lei Zhang
Social, Cognitive and Affective Neuroscience Unit (SCAN-Unit)
Department of Cognition, Emotion, and Methods in Psychology

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

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

universität wien
Fakultät für Psychologie

Bayesian warm-up?

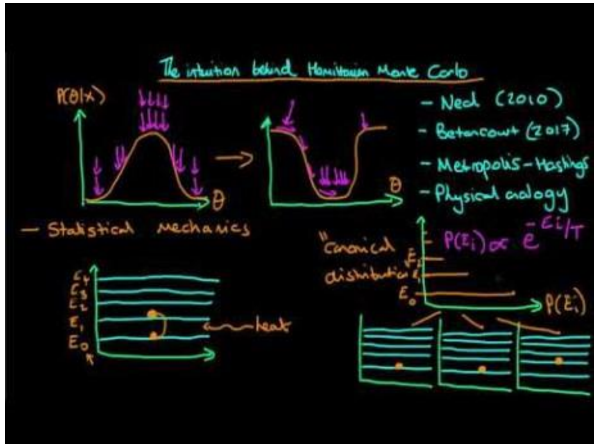

p(θ)
p(data)
p(θ, data)

MCMC Sampling Algorithms

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

Stan!

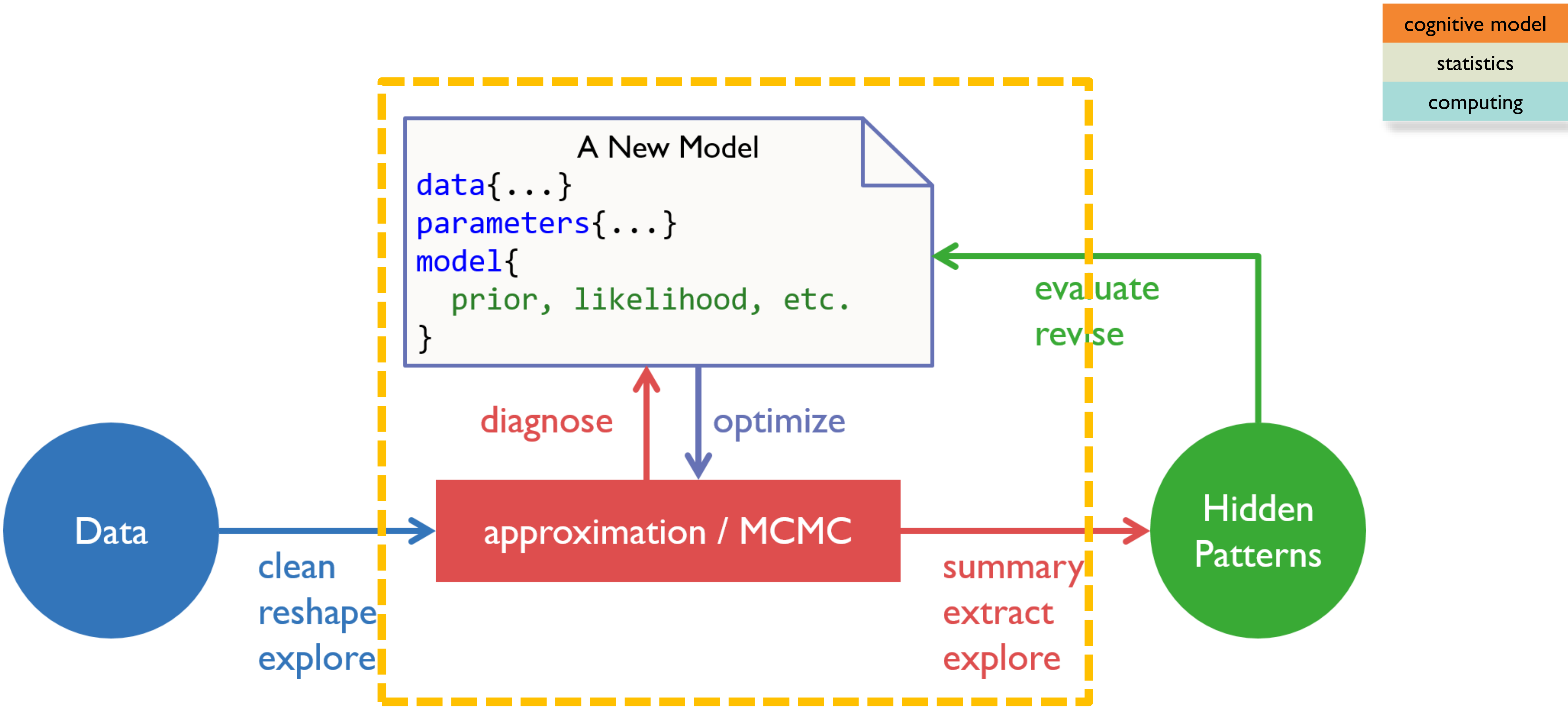
Optional homework



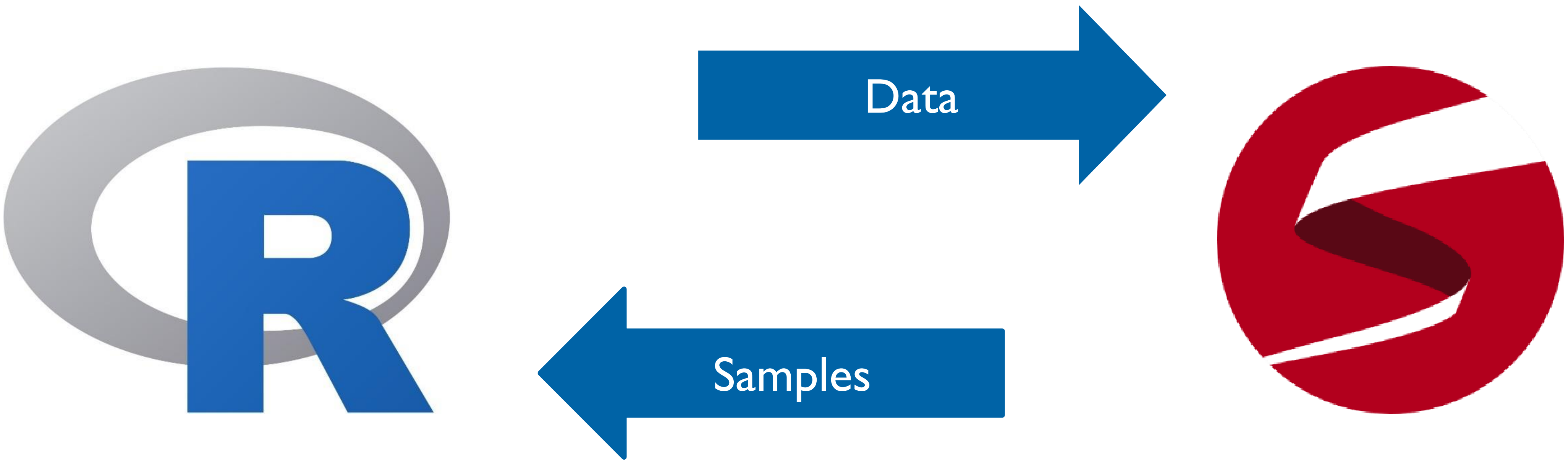
https://youtu.be/a-wydhEuAm0

*: also PyMC3

STAN
PROGRAMMING
LANGUAGE I


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

Stan and RStan



Steps of Bayesian Modeling, with Stan

A data story	Think about how the data might arise. It can be *descriptive* or even *causal*. Write a Stan program (*.stan).
Update	Educate your model by feeding it the data. Bayesian Update: update the prior, in light of data, to produce posterior. Run Stan using RStan (PyStan, MatlabStan etc.)
Evaluate	Compare model with reality. Revise your model. Evaluate in RStan and ShinyStan.

McElreath (2016)

Steps of Using Stan

1. Stan program read into memory
2. Source-to-source transformation into C++
3. C++ compiled and linked (takes a while)
4. Run Stan program
5. Posterior analysis / interface

```
data {
  int<lower=0> N;
  int<lower=0,upper=1> y[N];
}
parameters {
  real<lower=0,upper=1> theta;
}
model {
  y ~ bernoulli(theta);
}
```

Stan Language

model blocks

```
data {
 //... read in external data...
}
```

```
transformed data {
 //... pre-processing of data ...
}
```

```
parameters {
 //... parameters to be sampled by HMC ...
}
```

```
transformed parameters {
 //... pre-processing of parameters ...
}
```

```
model {
 //... statistical/cognitive model ...
}
```

```
generated quantities {
 //... post-processing of the model ...
}
```

REVISIT BINOMIAL MODEL

Binomial Model

cognitive model
statistics
computing

W L W W W L W L W

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

$$w \sim \text{Binomial}(N, \theta)$$

reads as:
w is distributed as a binomial distribution, with number of trials *N*, and success rate ϑ.

McElreath (2016)

Graphical Model Notations

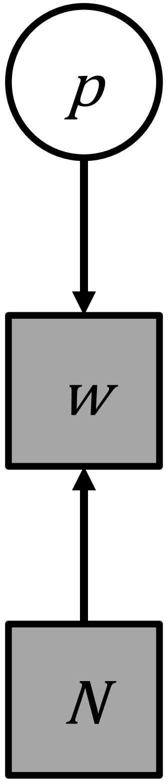
	continuous	discrete
unobserved	θ	δ
observed	y	N

Lee & Wagenmakers (2013)

Binomial Model

W L W W W L W L W

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



$$\theta \sim \text{Uniform}(0, 1)$$

$$w \sim \text{Binomial}(N, \theta)$$

	continuous	discrete
unobserved	θ	δ
observed	y	N

McElreath (2016)

Binomial Model

W L W W W L W L W

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

```
data {
    int<lower=0> w;
    int<lower=0> N;
}

parameters {
    real<lower=0,upper=1> theta;
}

model {
    w ~ binomial(N, theta);
}
```

McElreath (2016)

Running Binomial Model with Stan


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

```
> R.version
R version 3.5.1 (2018-07-02)

> stan_version()
[1] "2.18.0"
```

Model Summary

cognitive model
statistics
computing

```
> print(fit_globe)
Inference for Stan model: binomial_globe_model.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

       mean se_mean   sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
theta  0.64    0.00 0.14  0.35  0.54  0.65  0.74  0.87  1278    1
lp__  -7.72    0.02 0.69 -9.77 -7.89 -7.46 -7.27 -7.21  1824    1

Samples were drawn using NUTS(diag_e) at Tue Apr 09 12:44:04 2019.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Gelman-Rubin convergence diagnostic
(Gelman & Rubin, 1992)

Binomial Model

W L W W W L W L W

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

```
data {
    int<lower=0> w;
    int<lower=0> N;
}

parameters {
    real<lower=0,upper=1> theta;
}

model {
    w ~ binomial(N, theta);
}
```

McElreath (2016)

Model Summary

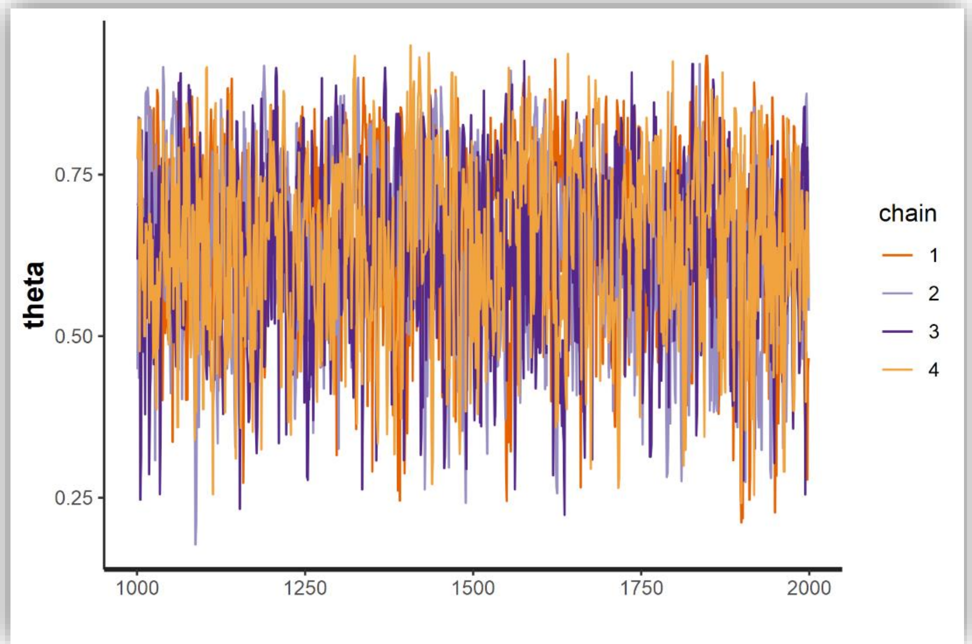
cognitive model
statistics
computing

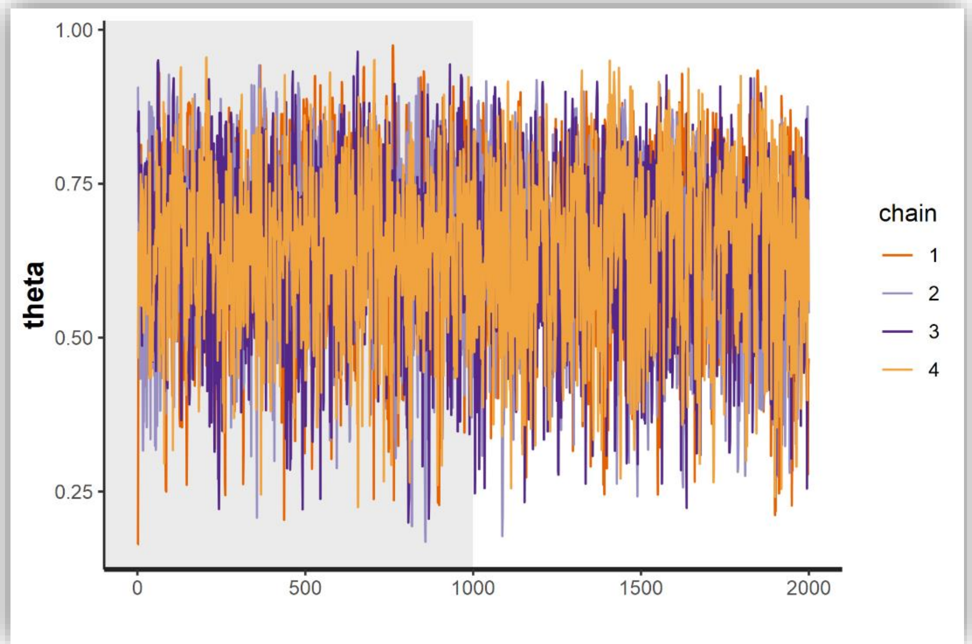
```
> print(fit_globe)
Inference for Stan model: binomial_globe_model.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

       mean se_mean   sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
theta  0.64    0.00 0.14  0.35  0.54  0.65  0.74  0.87  1278    1
lp__  -7.72    0.02 0.69 -9.77 -7.89 -7.46 -7.27 -7.21  1824    1

Samples were drawn using NUTS(diag_e) at Tue Apr 09 12:44:04 2019.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Gelman-Rubin convergence diagnostic
(Gelman & Rubin, 1992)

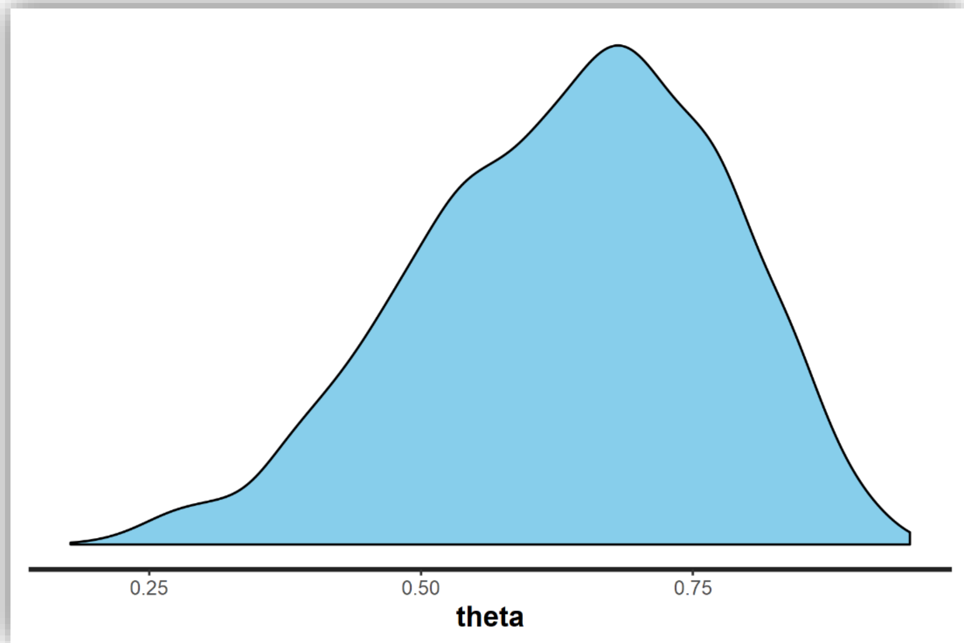
Diagnostics - traceplot

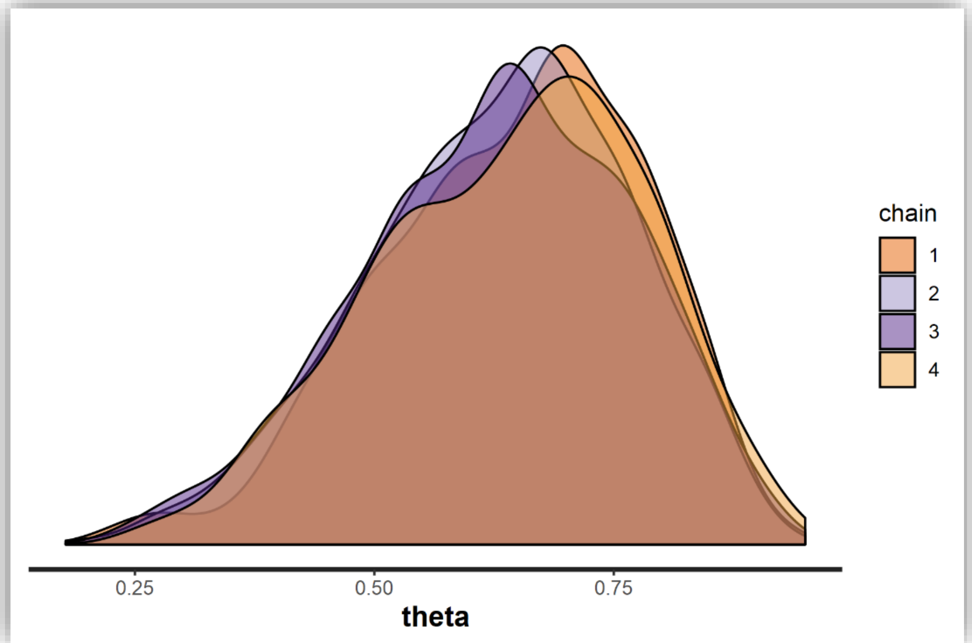
cognitive model
statistics
computing





Diagnostics - density

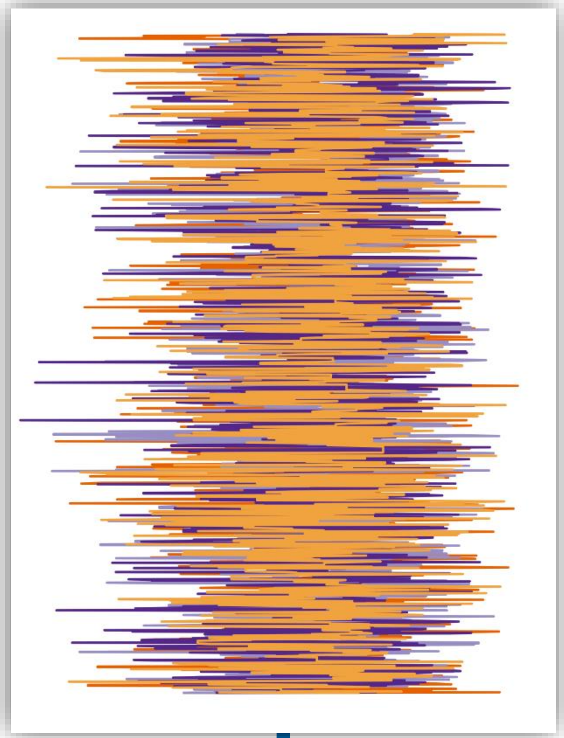


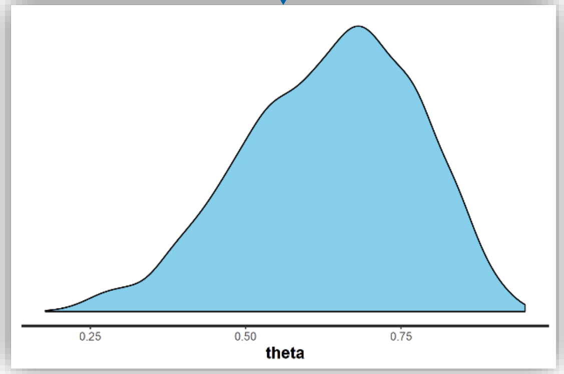


Diagnostics

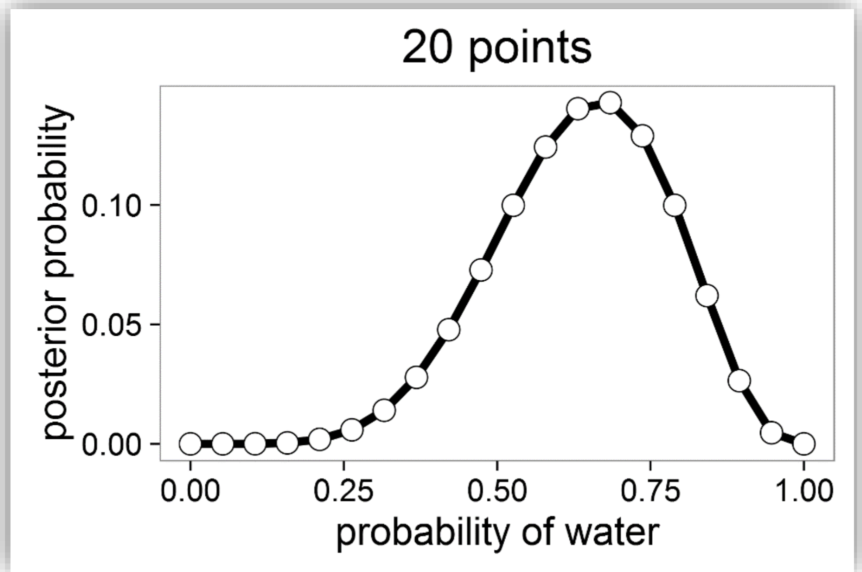
cognitive model
statistics
computing

MCMC



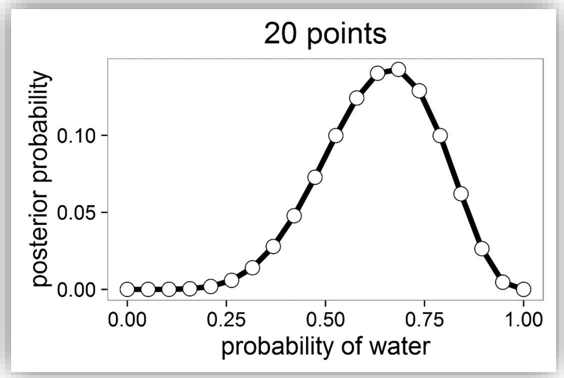


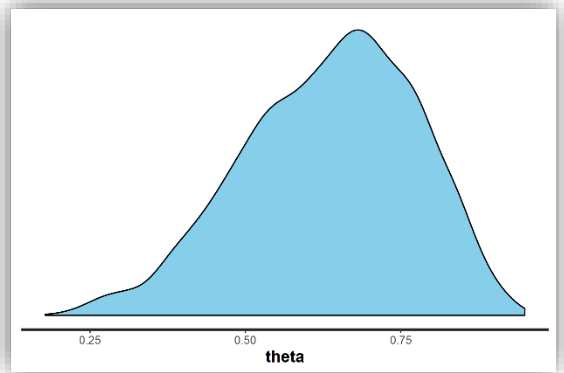
Grid Approximation



Draw a Conclusion?

- W = 6 out of N = 9
- uncertainty (relative plausibility) of all ϑ values
- the relative plausibility of ϑ = 0.64 is the highest, but it never rules out the possibility of ϑ being other values, e.g., 0.5, 0.75
- → when ϑ = 0.5, you may still observe 6W / 9 trials





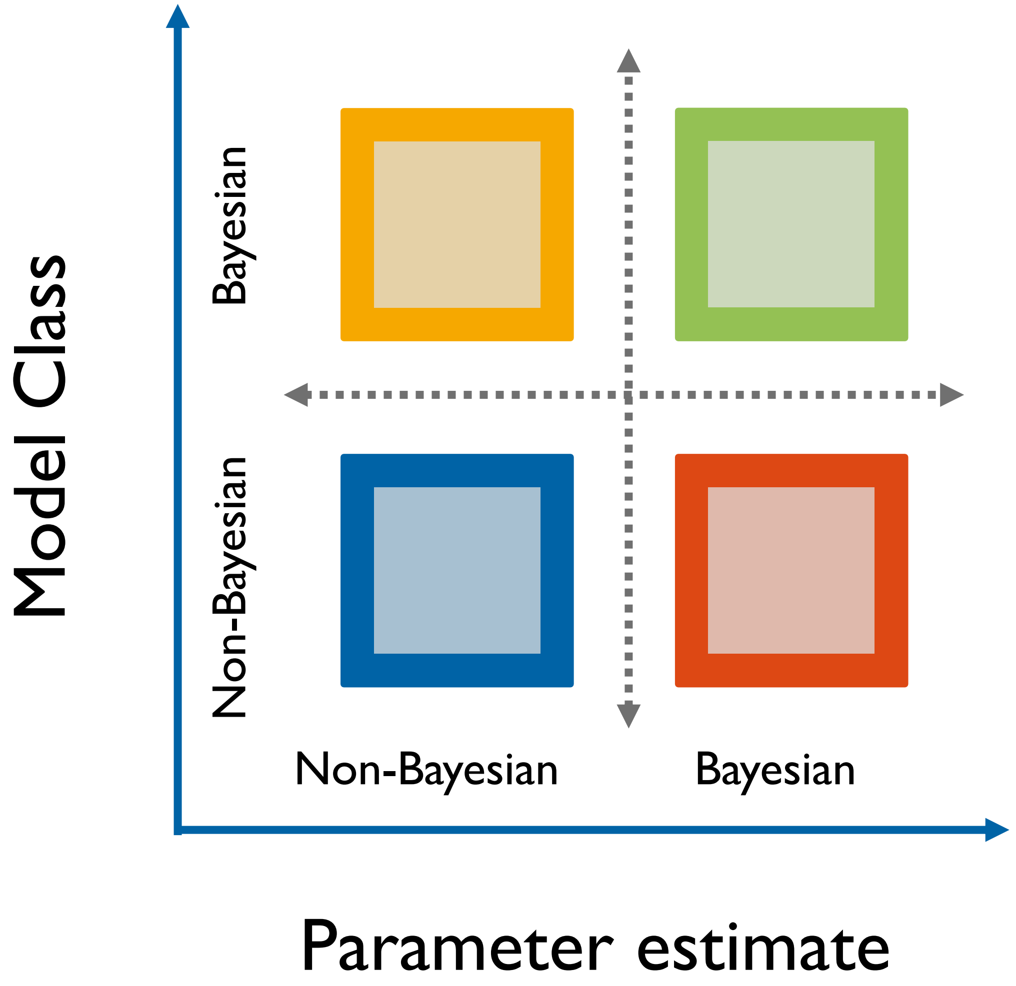
Is Anything Missing? – NO

cognitive model
statistics
computing

```
data {
    int<lower=0> w;
    int<lower=0> N;
}

parameters {
    real<lower=0,upper=1> theta;
}

model {
    theta ~ uniform(0,1);
    w ~ binomial(N, theta);
}
```

```
data {
    int<lower=0> w;
    int<lower=0> N;
}

parameters {
    real<lower=0,upper=1> theta;
}

model {
    w ~ binomial(N, theta);
}
```

What We Talk About When We Talk About "Bayesian" Models




ANY
QUESTIONS
?

Happy Computing!