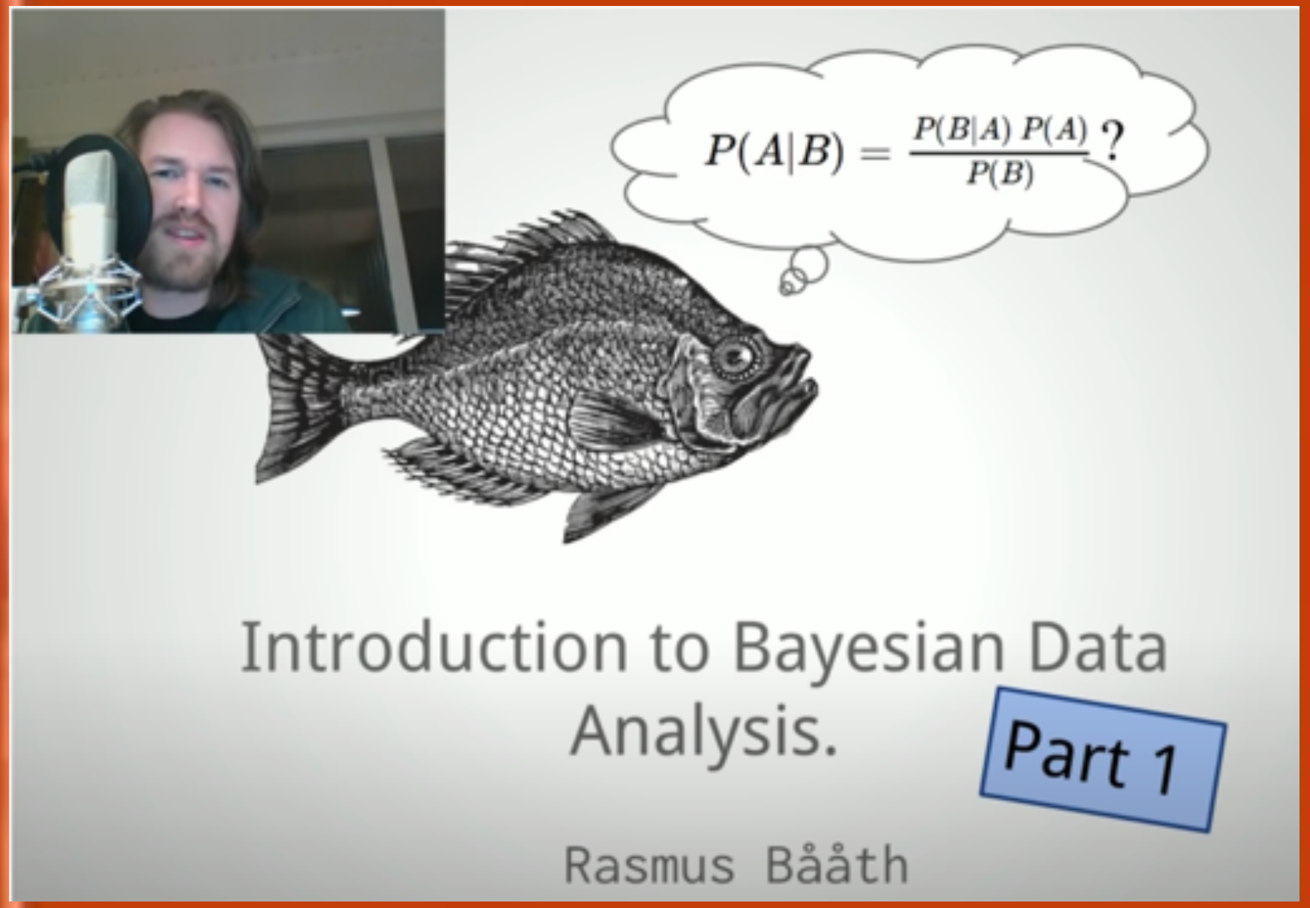
What we talk about When we talk about Bayes

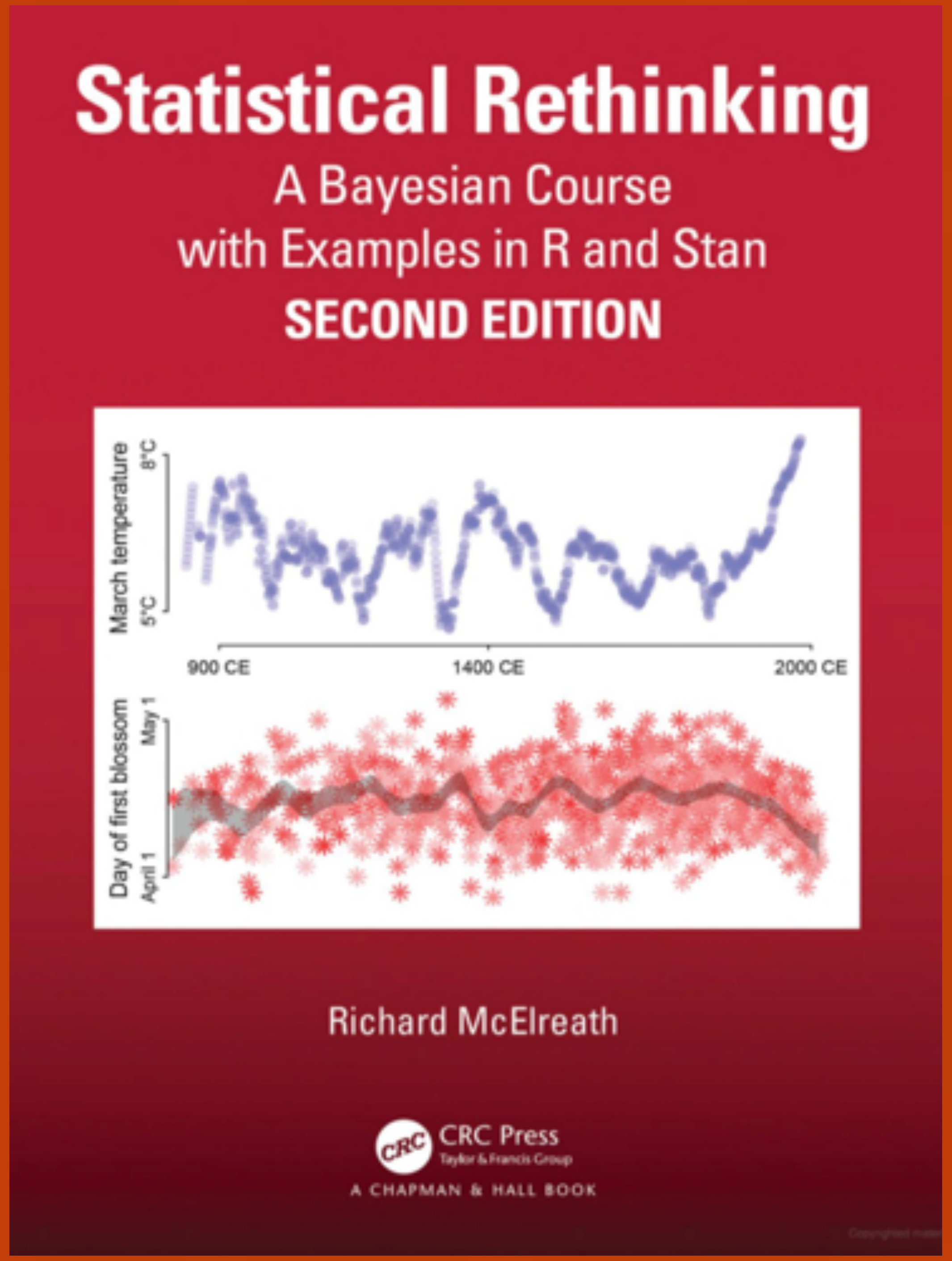
浅谈贝叶斯

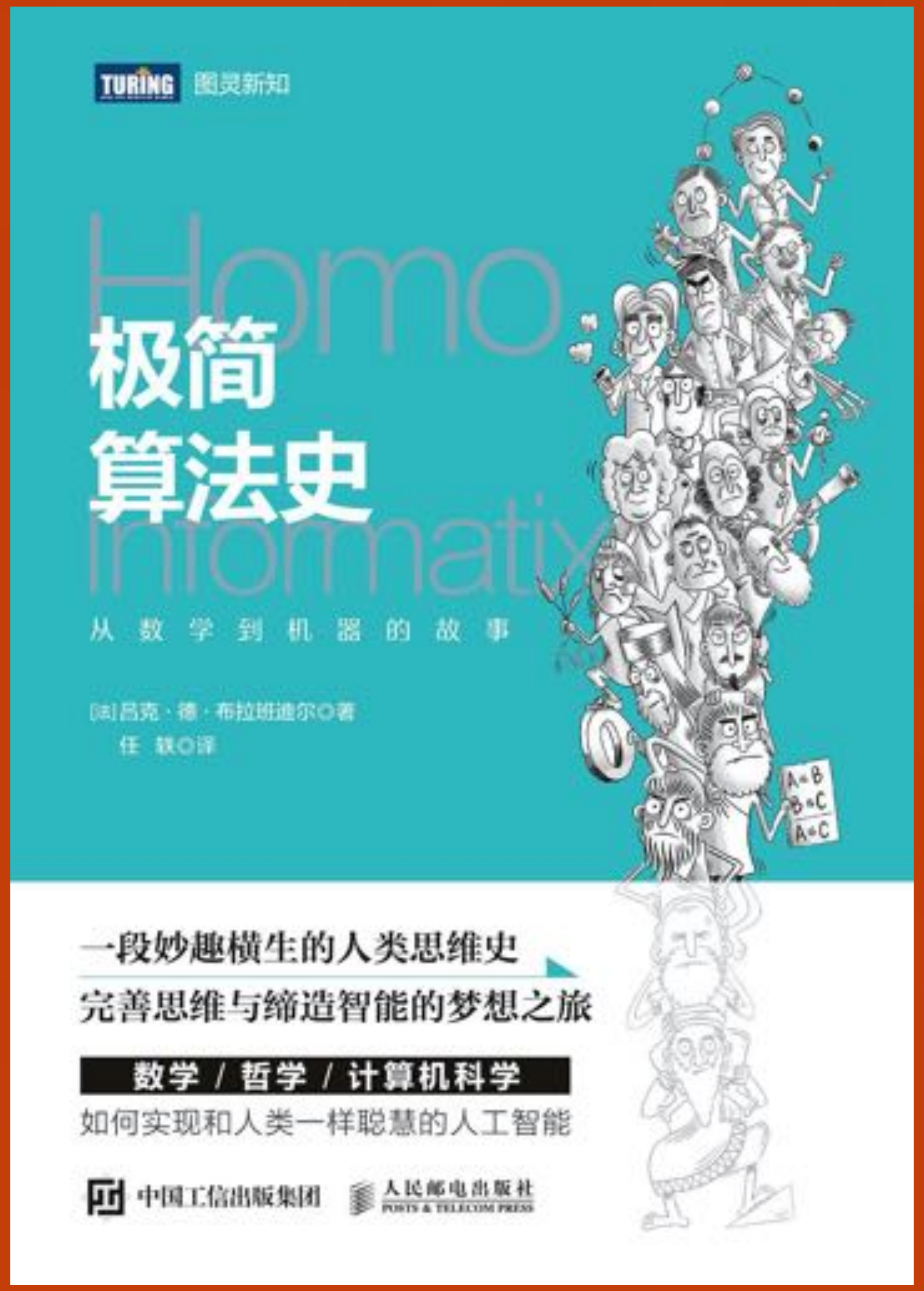
Huan FAN
CEC
XTBG
June 2021

Acknowledgement



混合效应模型的贝叶斯实现





Term Logic 亚里士多德的三段论

A is B	All men are mortal.
B is C	All Greeks are men.
Therefore A is C.	Therefore all Greeks are mortal.

Three ways of reasoning
三大推理方式

Deductive 演绎推理

All men are mortal.

All Greeks are men.

Therefore all Greeks are mortal.

Inductive 归纳推理

All Greeks are men.

All Greeks are mortal.

Therefore all man are mortal.

Abductive 溯因推理

Greeks are mortal.

All men are mortal.

Greeks are men.

Three ways of reasoning
三大推理方式

Deductive 演绎推理	Inductive 归纳推理	Abductive 溯因推理
All men are mortal.	All Greeks are men.	Greeks are mortal.
	sample from a population	
All Greeks are men.	All Greeks are mortal.	All men are mortal.
	get conclusion from the sample	
Therefore all Greeks are mortal.	Therefore all man are mortal.	Greeks are men.
	interpolate the conclusion to the population	

Three ways of reasoning
三大推理方式

Deductive 演绎推理

All men are mortal.

All Greeks are men.

Therefore all Greeks are mortal.

Inductive 归纳推理

All Greeks are men.

All Greeks are mortal.

Therefore all man are mortal.

Abductive 溯因推理

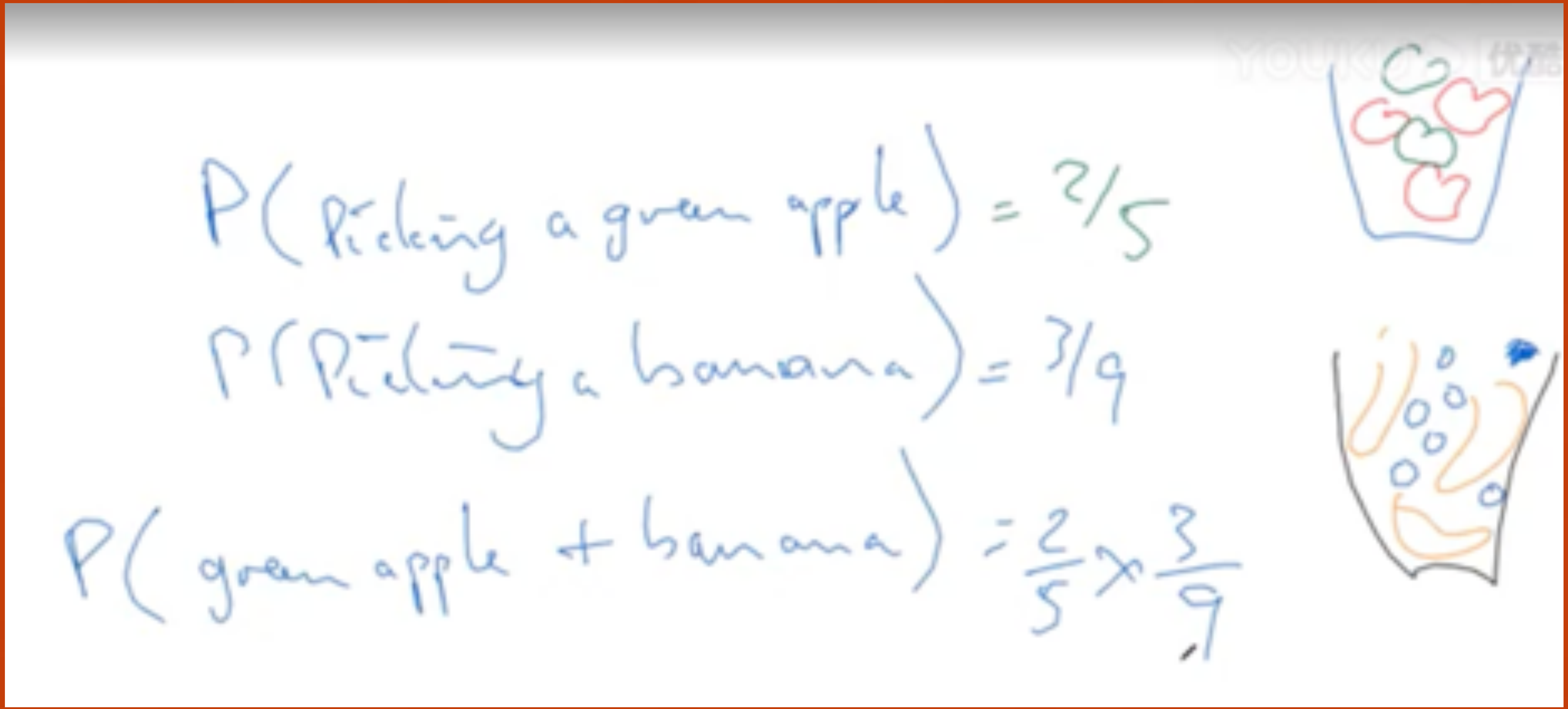
All greeks are mortal.

All men are mortal.

All greeks are men.

Bayesian

Kyle's fruit basket

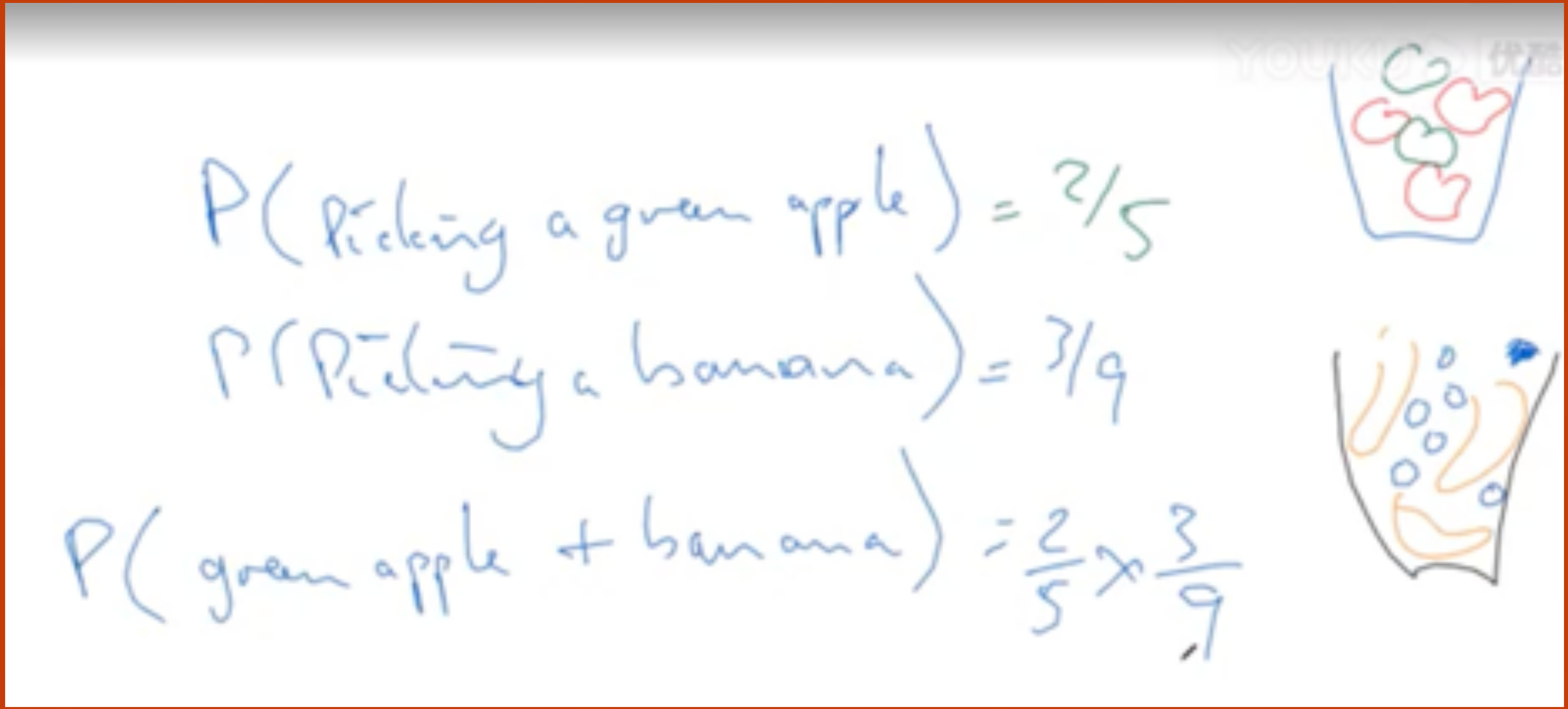


Joint probability 联合概率

The likelihood that two events happen at the same time

Kyle’s fruit basket

沿用Kyle的水果篮子例子

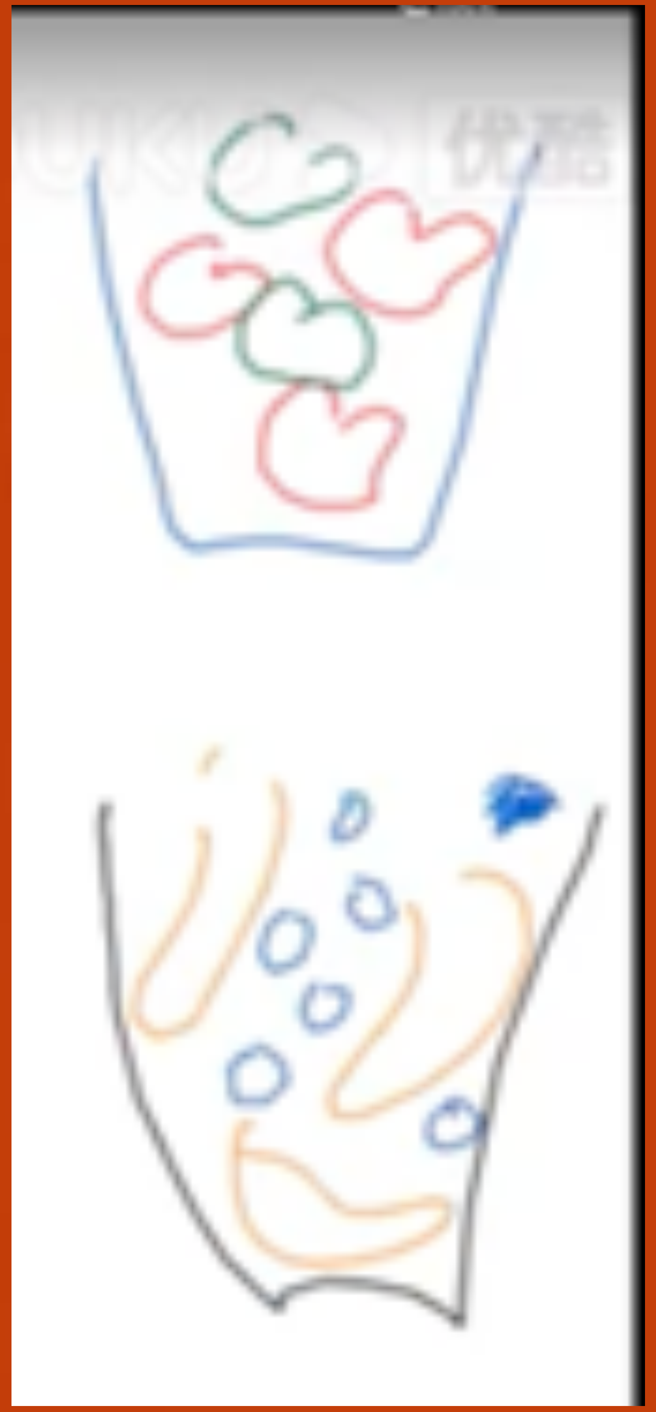


joint probability of independent events

独立事件的联合概率

Kyle's fruit basket

沿用Kyle的水果篮子例子



Chance of drawing a banana

1. P (pick a basket)
2. P (draw a banana)

joint probability of dependent events

非独立事件的联合概率

Bayesian theorem

贝叶斯定理

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A : pick the 2nd basket

B : pick a banana

Joint Probability (A , B) = P (B I A) * P (A)

= P (A I B) * P (B)

JP(A,B): picked a banana from the 2nd basket

P(A) = p of picking the 2nd basket

P(B) = p of picking a banana

P(B I A): if we picked the second basket, the p of picking a banana

P(A I B): if we picked a banana, what's the probability that it is from the second basket

A more complex case

贝叶斯定理

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

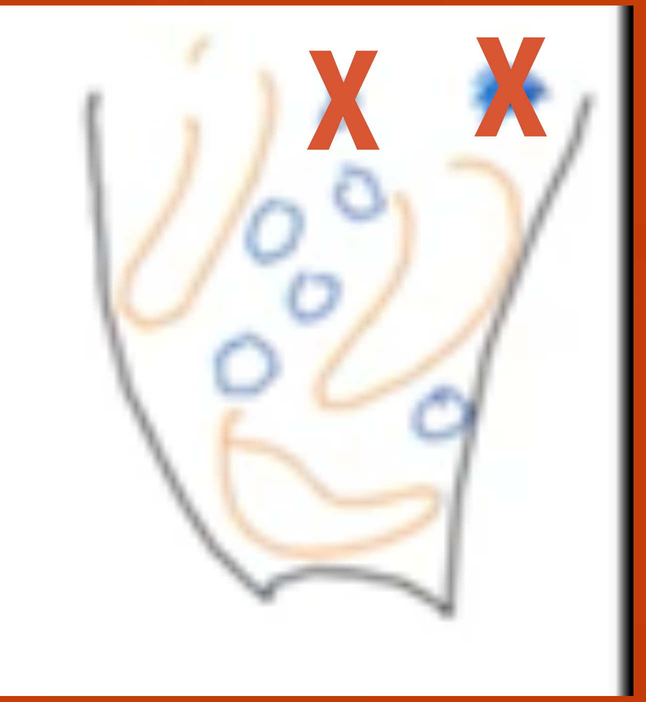
A : pick the 2nd basket

B : pick a banana

Joint Probability (A , B) = P (A | B) * P (B)

= P (B | A) * P (A)

JP(A,B): picked a banana from the 2nd basket



P(A) = p of picking the 2nd basket 1/2

P(B) = p of picking a banana 6/18

P(B | A): if we picked the second basket, the p of picking a banana 3/8

P(A | B): if we picked a banana, what's the probability that it is from the second basket

And that is very helpful!

Customer Profiling

用户画像

P(A) = Being a young mother	1%
P(B) = Bought diaper	5%
P(B I A): probability of a young mother buying diaper	50%
P(A I B): M bought diaper, what's the probability of A being a mother?	10%

Three ways of reasoning
三大推理方式

Deductive 演绎推理

All men are mortal.

All Greeks are men.

Therefore all Greeks are mortal.

Inductive 归纳推理

All Greeks are men.

All Greeks are mortal.

Therefore all man are mortal.

Abductive 溯因推理

M buys diaper

All greeks are mortal.

Young mothers buy diaper

All men are mortal.

M is a young mother

All greeks are men.

Bayesian

Targeted marketing

P(A) = M Being a young mother 10%

P(B) = Bought baby formula 10%

P(B | A): probability of a young mother buying baby formula 50%

P(A | B): M bought baby formula,
what’s the probability of A being a mother? 50%

Now let's show her ads about baby toys and clothes!

Bayesian Thinking
贝叶斯逻辑

1. Estimates the statistical probability of something being true

2. updates that probability as new evidence appears

3. approaching the truth without achieving absolute certainty

人：什么是机器学习?
机器：问我一个问题
人：3乘以7等于多少?
机器：7
人：连三七二十一都不知道
机器：21

Prior and Posterior
先验与后验

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

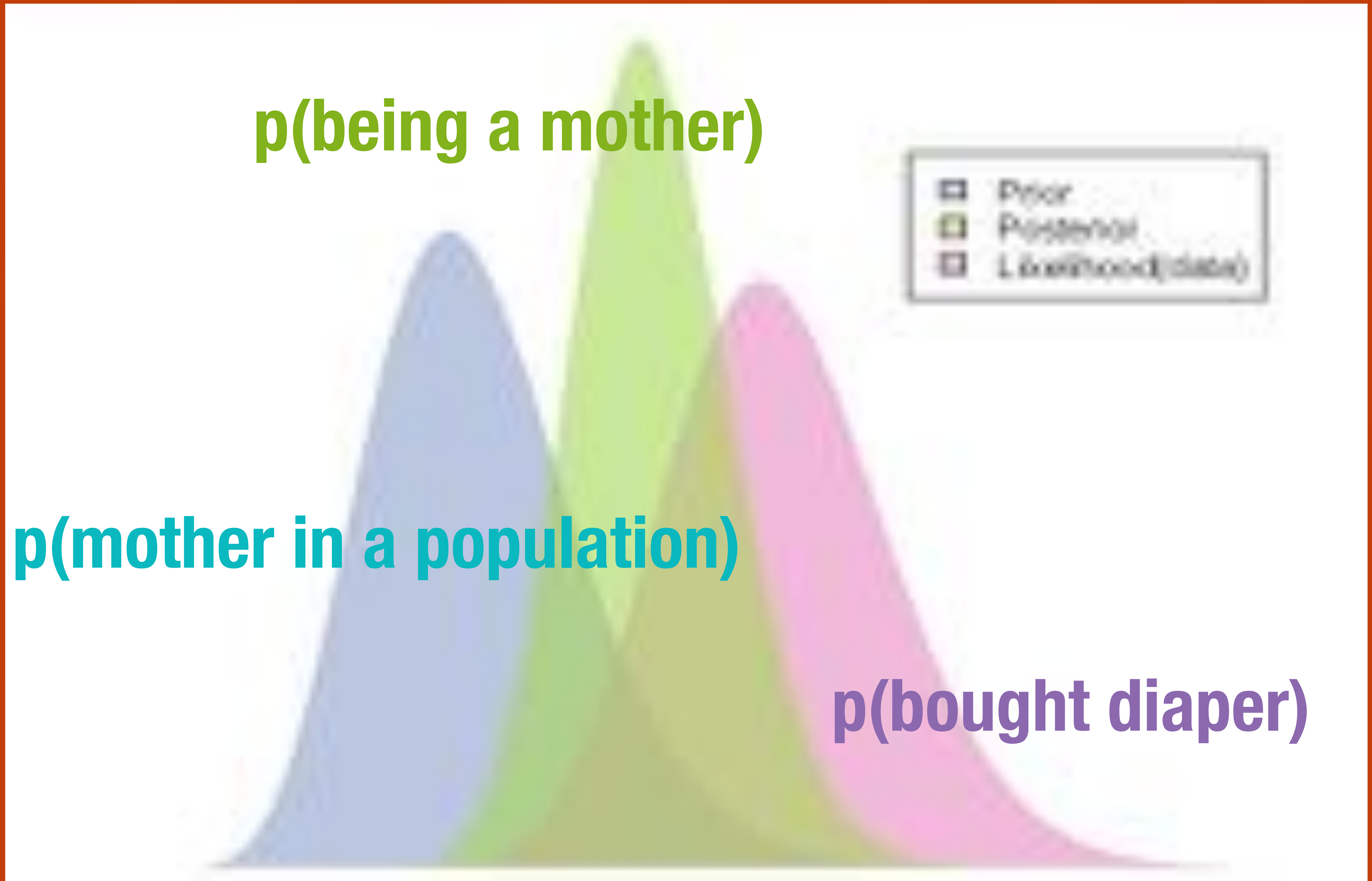
P(A) = Being a young mother: 1%

P(A | B): M bought diaper,
p of M being a mother: 10%

P(B | A): probability of a young mother buying baby formula

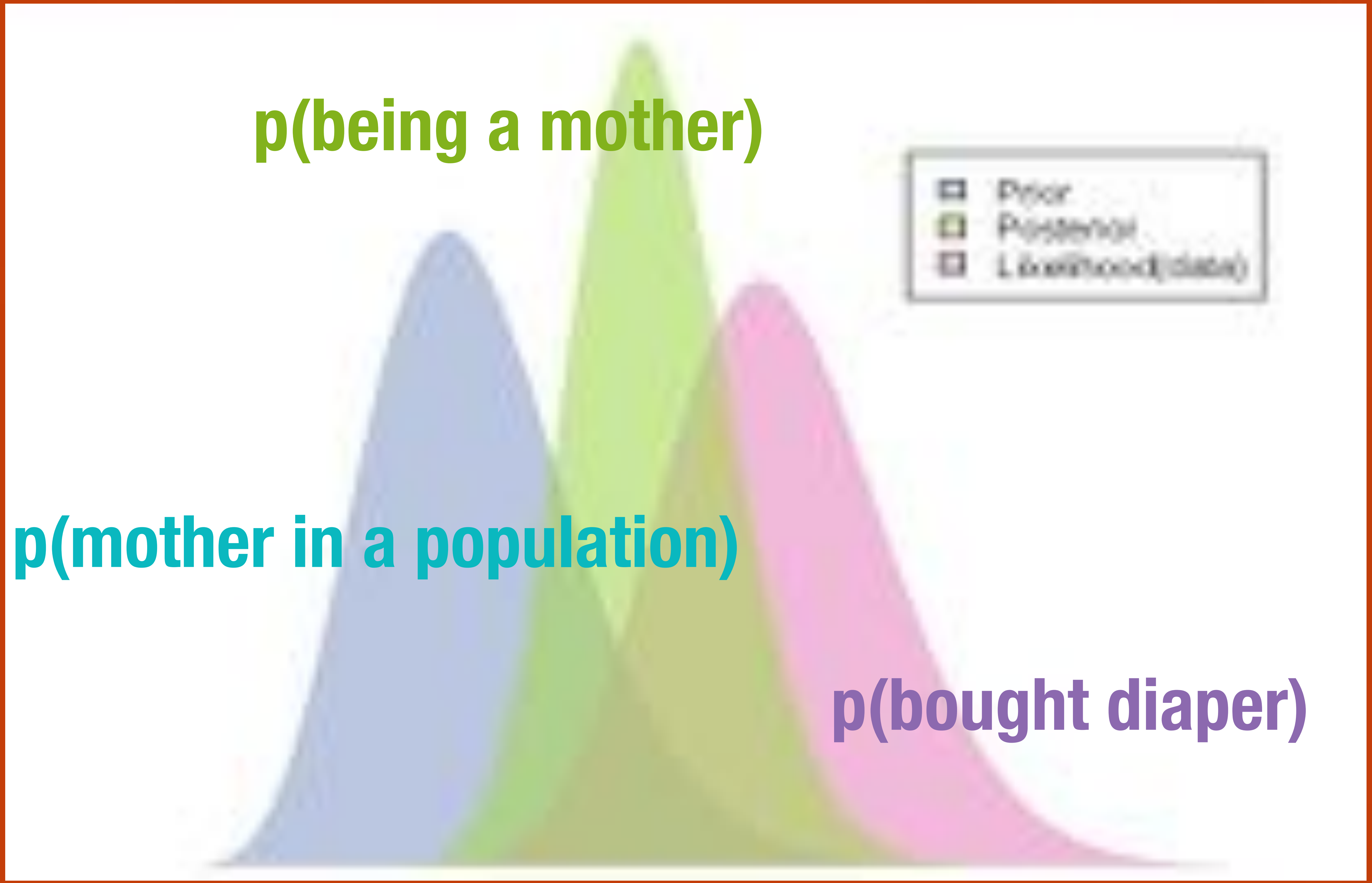
P(B) = Bought baby formula

Data collected



Prior + Data ——> Posterior

Initial guess + data -> more valid guesses



Prior + Data ——> Posterior

pre-existing data + present information -> inferences

$$\text{Prior} + \text{Data} \xrightarrow{?} \text{Posterior}$$

pre-existing data + present information -> inferences

“Generative model:

a story of how your data came to be.”

–Rasmus Baath

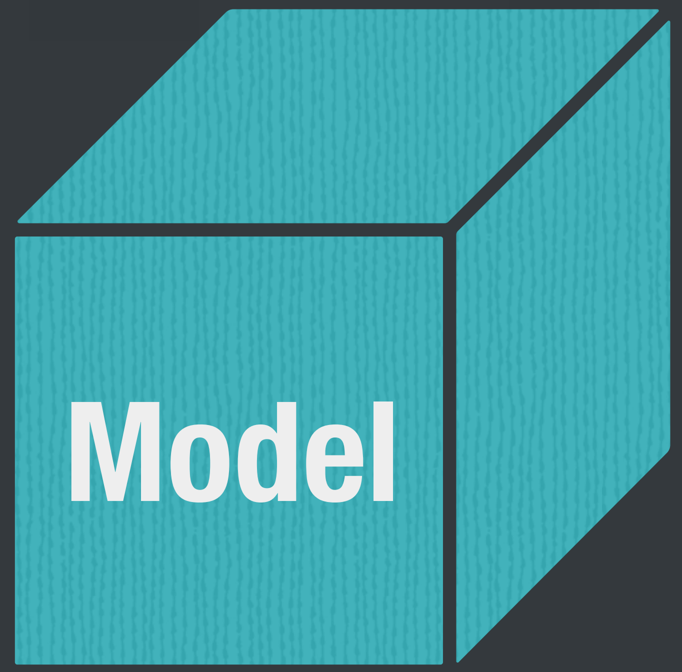
https://www.datacamp.com/courses/fundamentals-of-bayesian-data-analysis-in-r

From this morning

```
#now lets make some modeled response values
b0 <- 0
sd.b <- 1   #across sites variation: random effect residual
sd.e <- .15  #within sites variation, model residual

for(i in d$site){
      dd <- d[d$site == i,]
      nn <- nrow(dd)
      #introduce between site variation
      b1.site <- 0 + rnorm(n=1, mean=0, sd=sd.b) #site mean
      #introduce within site variation
      d$x[d$site == i] <- b0 + b1.site + rnorm(n=nn, sd=sd.e)
}
```

Generative model: 熟悉的陌生人

Parameters → 

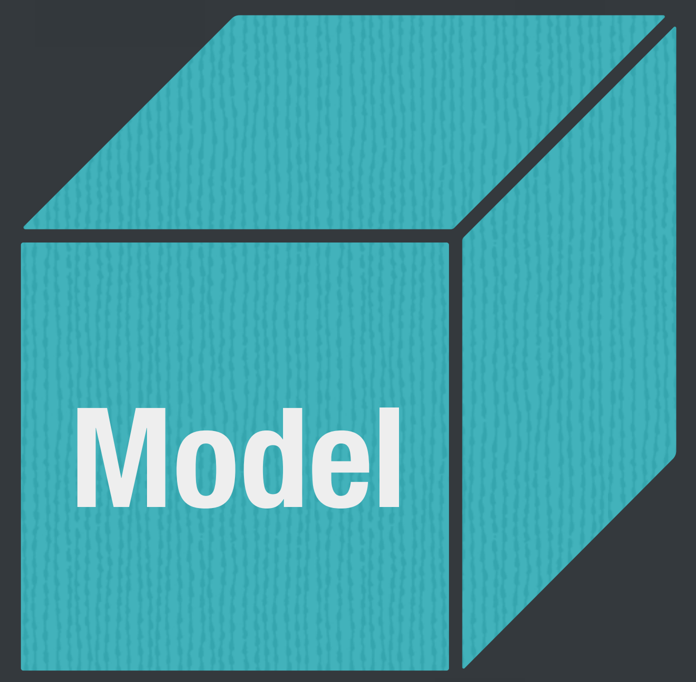
+ Some Randomness → Data

10, 0, 1 + rnorm() = 0.22 -0.19 1.31 -1.67 -0.43 0.70 0.19 1.45 -0.31 0.64

rbinom()
rpoisson()
rbeta()

Linear model

X →



→ Predicted Data

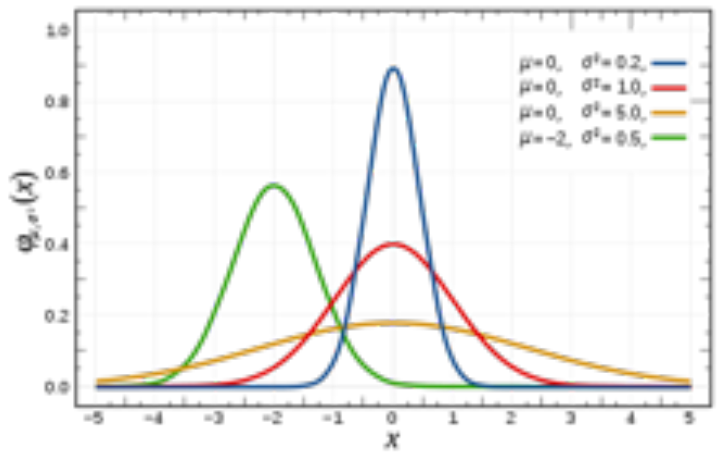
Location
log.diversity
+ LMM = Biomass

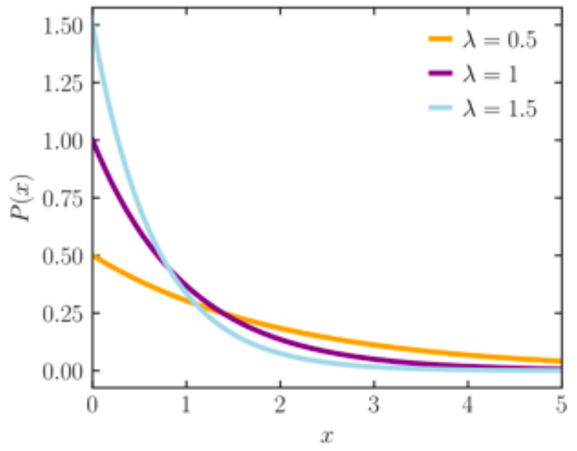
Distributions recap

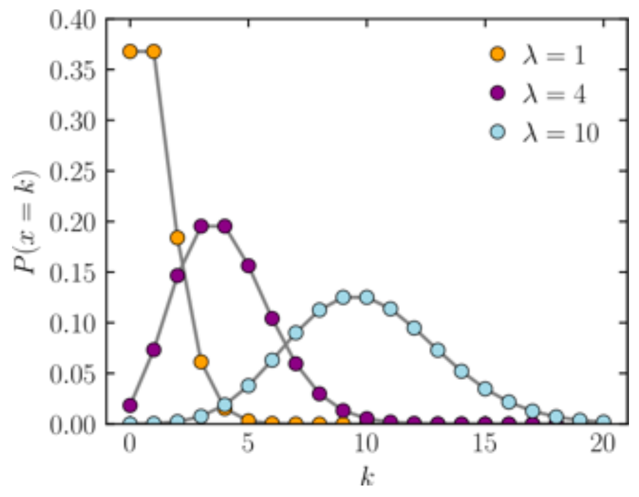
Distribution	Data Type	Range	Example	Parameters
Normal	Continuous	[-infinite, infinite]	“normal” data	mean, sd
Exponential	Continuous	(0, infinite]	growth	rate
Poisson	Discrete	[0, infinite]	Count data	lambda
Binomial	Binary	0/1	success/failure	n, rate
Uniform	Continuous	[-infinite, infinite]	unloaded dice	min, max
Beta	Continuous	[0, 1]	ratio	shape1, shape2

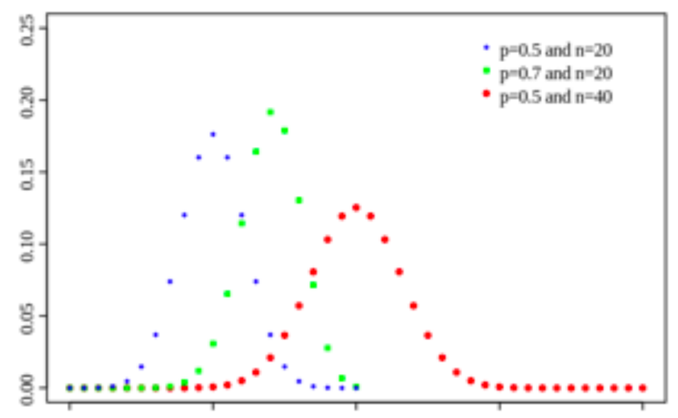
Distributions recap

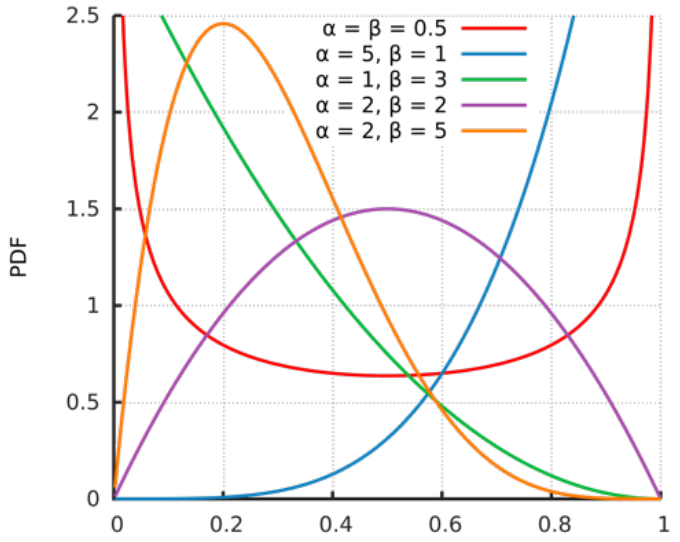
Probability density/mass function

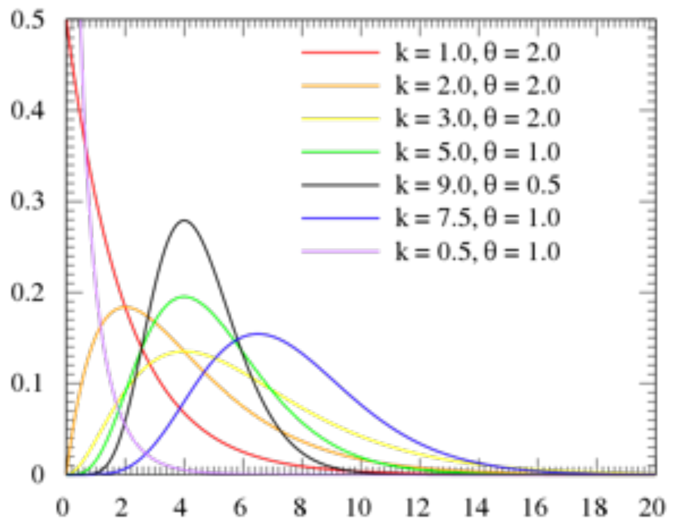




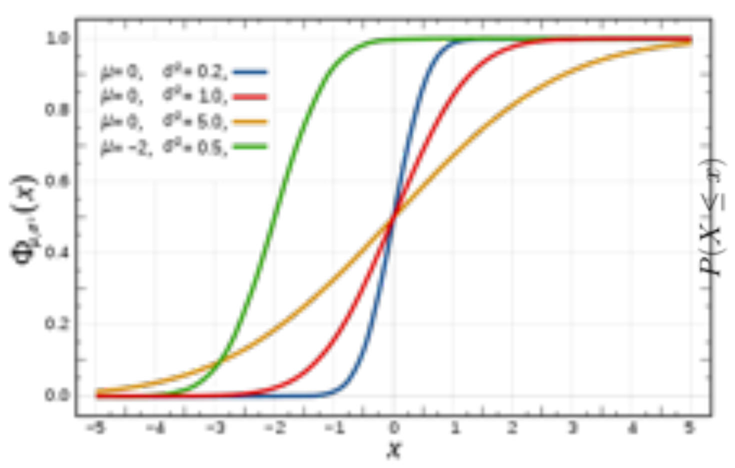


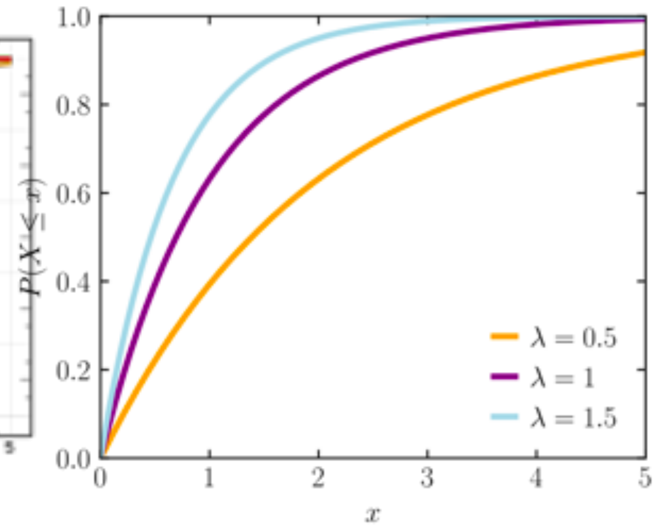


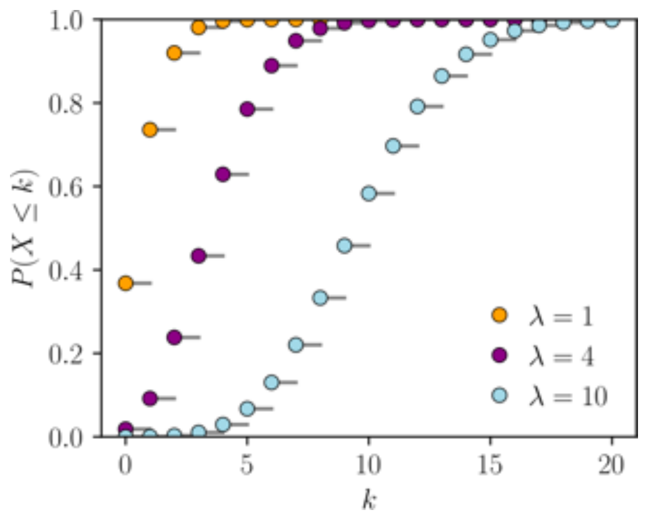


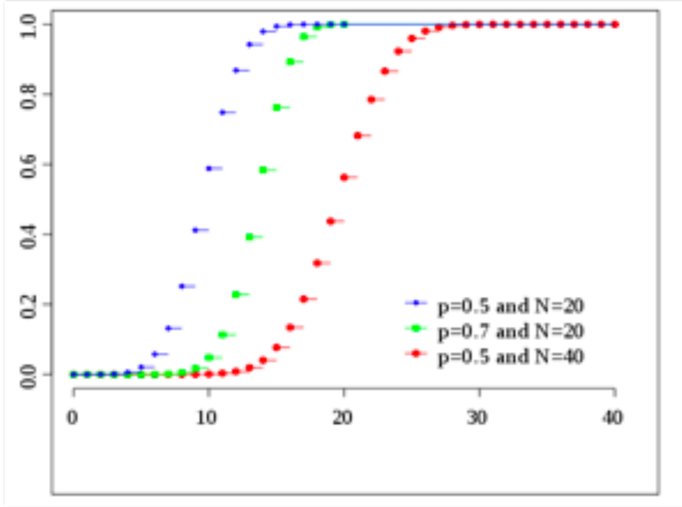


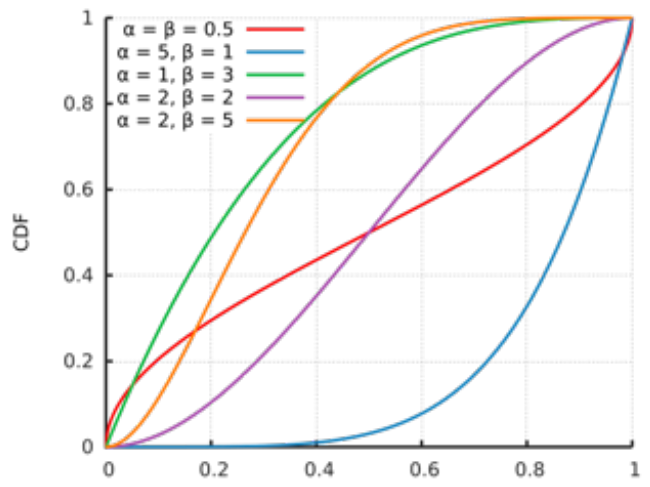
Cumulative distribution function

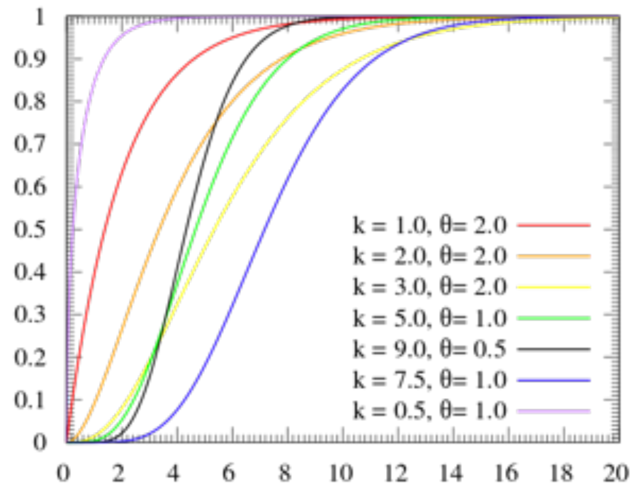












R function on distribution recap

notation	function	Input	Output	Example
d	density	x	height	d(0) = 0.4
p	probability	x	area under curve till x	p(0)=0.5
q	quantile	area under curve till x	x	q(0.5)=0
r	generator	n	n datapoints	r(10) =

Methods so far for model fitting

- Least squares (LM models): normal data
- Maximum Likelihood (GLM): not for random effect
- REML ()
- Bootstrapping (restricted to observed data)
- GLS: normal data

Model fitting: the Bayesian way

Baby responds to his name

Generative model

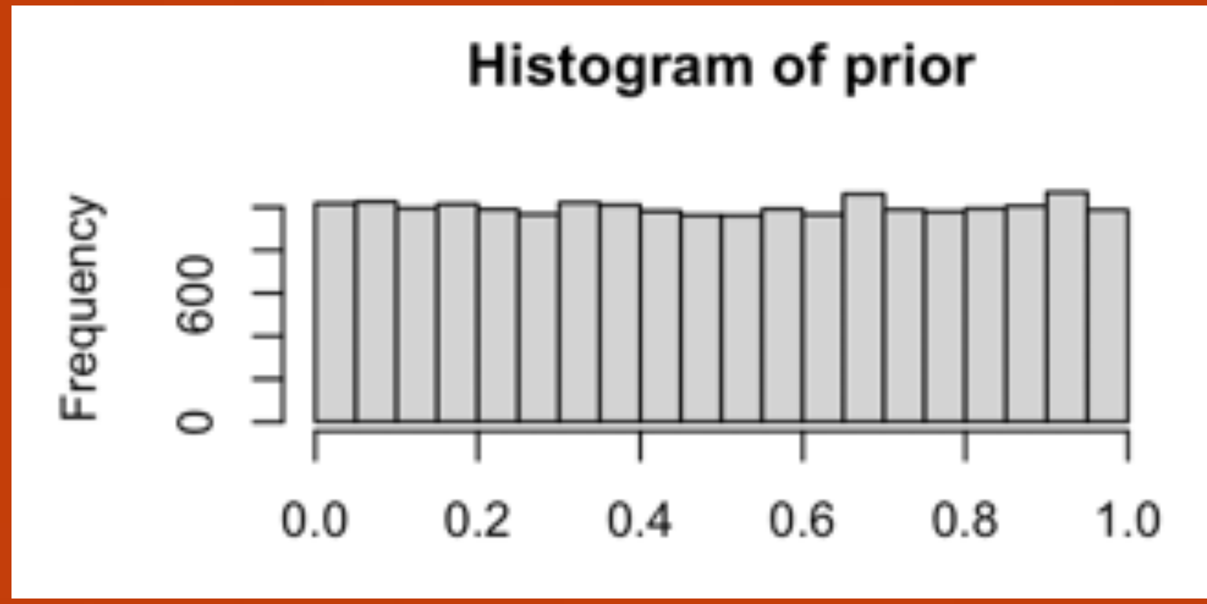
Prior + **Data** ——> **Posterior**

Name called: 10 times
Respond: 6 times

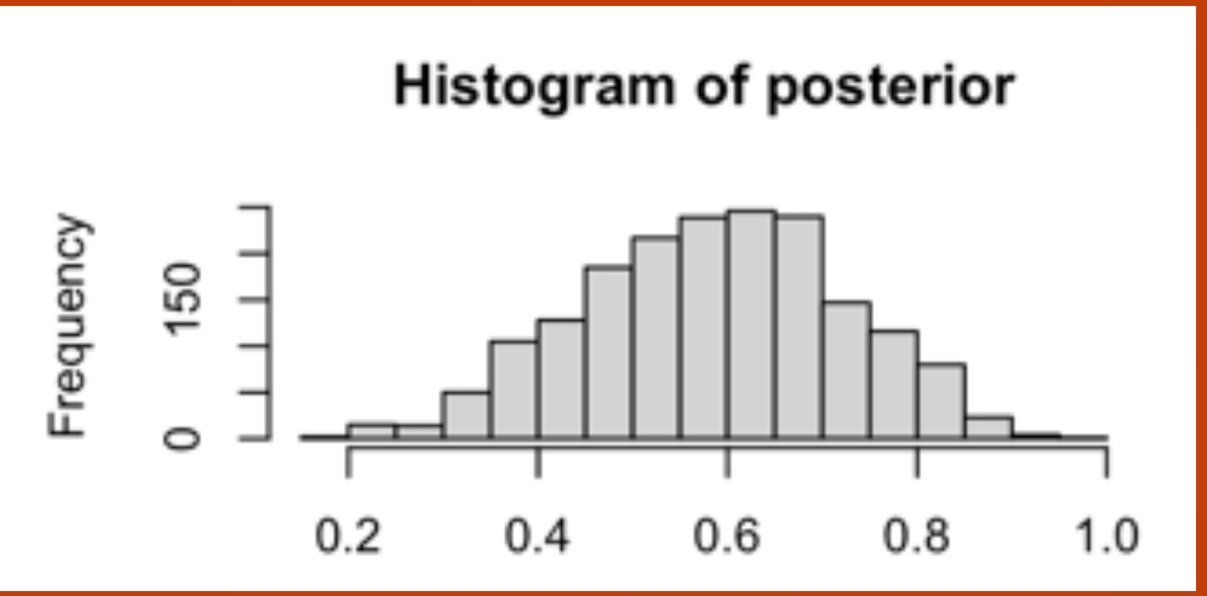
Exercise: 11.1 Baby Response to name

Go to R code

Exercise: 11.1 Baby Response to name



10 trials
6 responses



Prior

Data

Posterior

"Bayesianists assume that there is pre-existing data that can be combined with the present information to draw inferences (of the parameters)."

– Kyle Tomlinson

Exercise: 11.2 Expert advices on Child Development

Go to R code

"My grandson is really smart and he definitely responds to his name more than half of the time when I call it. You go to work and spend little time with him so he does not respond to you very well!."

–Grandma

“70-80% babies respond to their names starting month 6.”

–American Pediatric Association

Priors

- What information the model has before seeing the data
- We need to use probability to represent uncertainty in all parts of the model, meaning possibly different priors for different parameters.

Customized Priors

- Use existing defined distributions
- When we have enough data, the prior won't matter too much. While if we have limited data, the posterior will look like prior more. If we don't have data at all, the posterior will look exactly like prior (data-driven).

"If you're not using an informative prior, you're leaving money on the table."

–Robert Weiss, UCLA

Estimation of multiple parameters

Excercise 11.2 Baby vs Border Collie

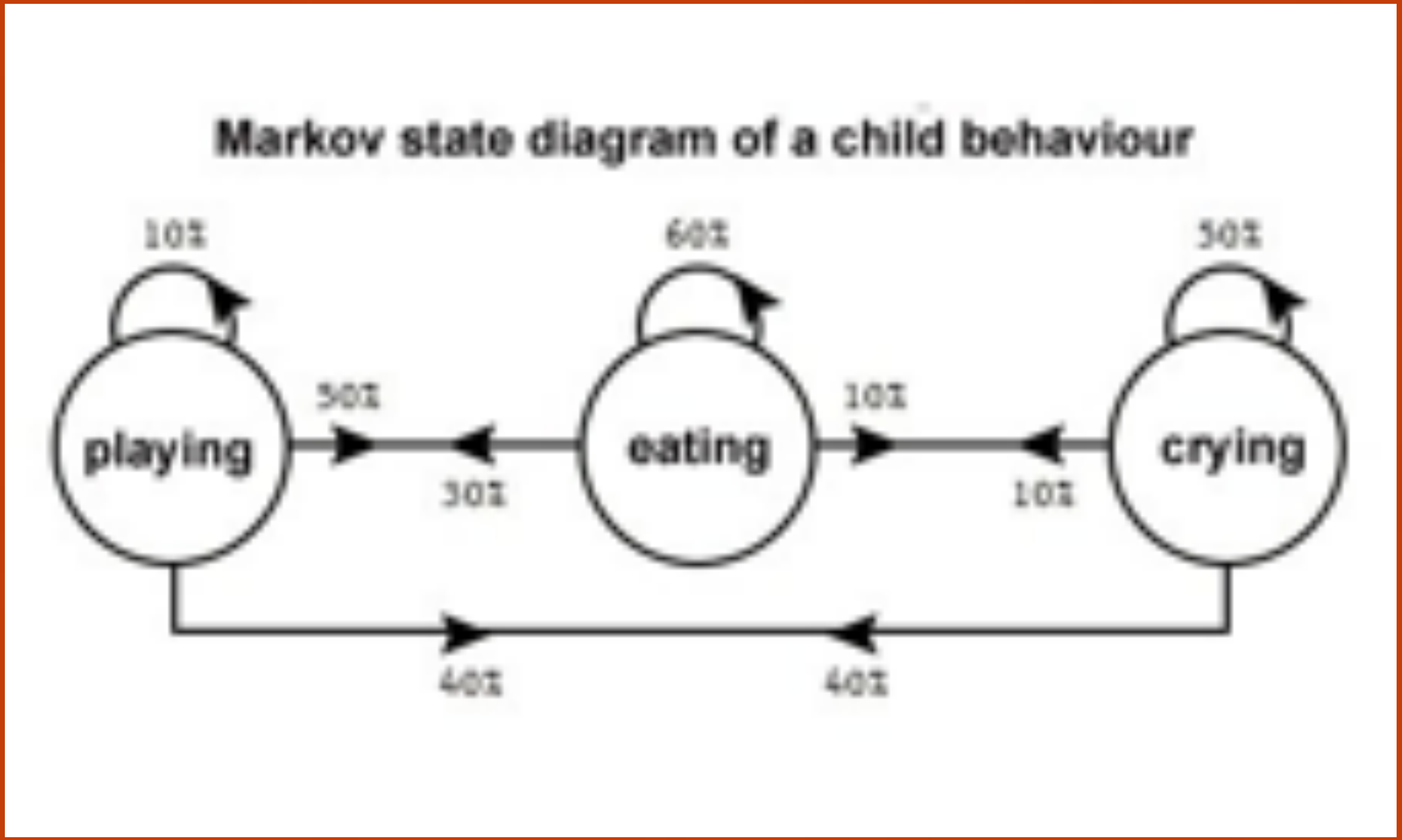
go to R code

How to fit a Bayesian model

- Approximate bayesian computation (only keep the priors that generate the data) Slow!
- Faster methods:
  - instead of throwing away a lot of simulations, it calculates the probability (likelihood) of generating the data with that prior.
  - Sampling the parameter space smarter (instead of just exhaustive search through all prior, MCMC is smarter, Ridges!)
  - The faster methods just try to do a close-enough job as approximate bayesian computation, but faster.
- Key: The rule for deciding which parameters in the prior distribution to keep in the posterior

What is MCMC any way?

Markov chain + Monte Carlo



Markov Chain

A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

Monte Carlo

% of land on earth

estimating pi

Please choose:

1. a letter between [A,X]
2. a number between [1,12]
3. Send it to the WeChat group at the form of letter_number

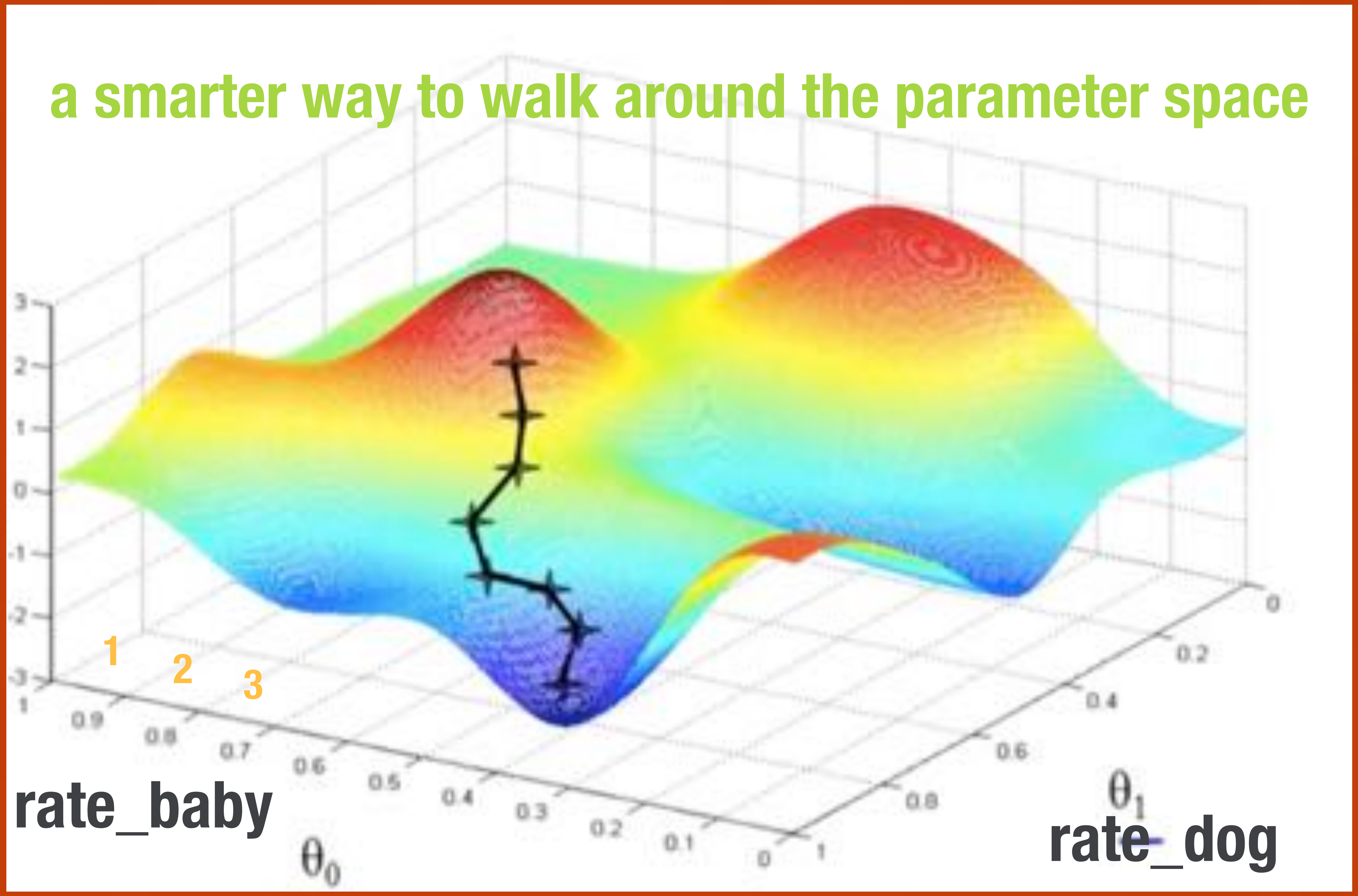
I choose L_7.

computational algorithms that rely on repeated random sampling to obtain numerical results

Markov chain + Monte Carlo

computational algorithms that does repeated sampling
based on the previous sampling with some randomness

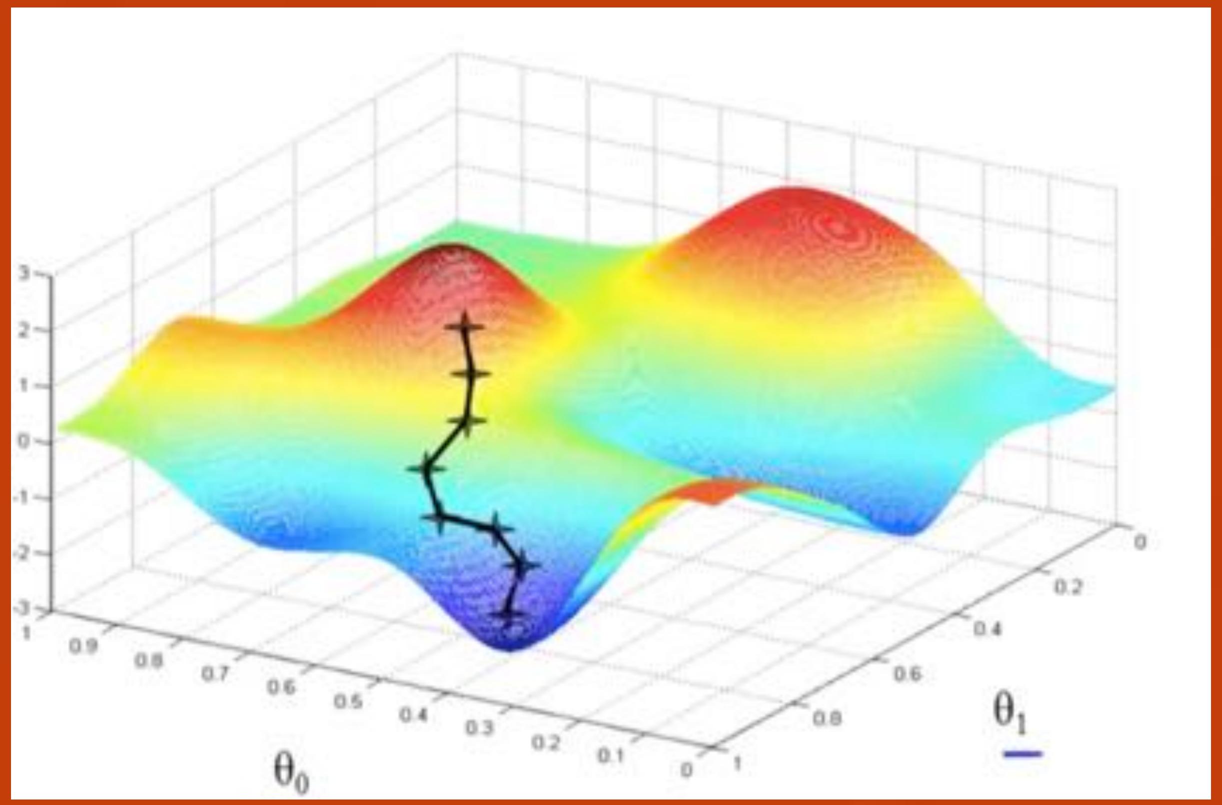
Markov Chain Monte Carlo:



Search through the parameter space

Approximate Bayesian Computation
Exhaustive search

1. Initial draw from prior (v1)
2. Take a step (s ~ N(0, sd) and propose another position in parameter space to add to the posterior (v2 = v1+ s)
3. Is my data more likely with v1 or v2?
   A. More likely with v2 (L (data | v2) > L (data | v1)):
      - add it to posterior, v2 become the new v1
   B. Less likely with v2 (L (data | v2) < L (data | v1)):
      - draw a random probability p from uniform(0,1)
      - If L(data | v2)/L (data | v1) > p, same as A.
      - If L(data | v2)/L (data | v1) < p, make a copy of v1 in the posterior.



Markov Chain Monte Carlo

Metropolis algorithm

Code this in R
with in class test example
Ravenzwaaij 2018

MCMC

- The reason that Bayesian analysis became popular again.
- Every step counts!
- Different tastes
  - Metropolis-hastings
  - Gibbs sampling
  - Hit-n-Run, the T-walk, particle monte carlo, etc.
  - One that Stan uses: Hamiltonian Monte Carlo

How to summarize the posterior

- Posterior mean, median, mode
- 95% credible interval
- Probability of events

“Bayesian data analysis is about more than just computing a posterior distribution.”

–Visualization in Bayesian workflow

Stan the language

- ☐ It has interfaces with your favorite language, be it R, python, matlab or julia.
- ☐ You define your model and Stan fit it using Hamiltonian Monte Carlo.
- ☐ Tailor made for defining generative models.
- ☐ brms is a R package that wraps stan with the syntax of glmer()

Let's learn some stan!
Go to R code

Bayesian Regression Model using Stan

Paul Buerkner

Summary

Why Bayesian?

- Potentially the most information-efficient method to fit a statistical model (thus computationally-intensive)
- Can deal with infinite number of parameters, thus good complex problems.
- Great flexibility when building models
- Include information of experts in prior
- Retains the uncertainty of the estimated parameters. While the maximum likelihood would only result in point estimation

Warning

- Computationally demanding
- Just a choice. You don't have to take it.

“大胆假设，小心求证”

Informative but conservative

–Take home message