Modeling Basics

The start-to-finish of Bayesian Cognitive Modeling

The Starters

who
All cognitive agents model (insects, animals, humans).

what
Learning from data to understand and make predictions about one's surroundings.

where
Everywhere, in all situations.

when
All day, every day.

why
The ability to understand and predi our environments ensures our survival.

how
Frequentists vs Bayesian

The Frequentist Approach

Based on the data observed, calculate the likelihood of the occurrence of a given event. No additional information.

The prior

What information do we have before we see any data?

The Bayesian approach

Bayes' Theorem: Posterior $\propto$ Likelihood * Prior

Hamish's Scottish Taverne Menu

Haggis

Haggis deluxe

Haggis supreme

Haggis a la mode

Haggis au gratin

Chicken Tikka Masala

Haggis flambée

Haggis surprise

Make-your-own Haggis

Something else involving Haggis

Frequentist approach

Looks at the likelihood of an event given the data alone (i.e. how *frequently* does an event occur)

Likelihood odds ratio, non-Haggis to Haggis:

1:9

Applying Bayes Theorem

But! You have more information.

Prior odds ratio	4: 1
Likelihood ratio	1: 9
Posterior odds ratio	4: 9

So there’s about a 40% chance that you’re not eating sheep stomach. Success.

Our goal

Inferring generative models from observed data.

Bayes Theorem - prior

Bayes Theorem - prior

Bayes Theorem

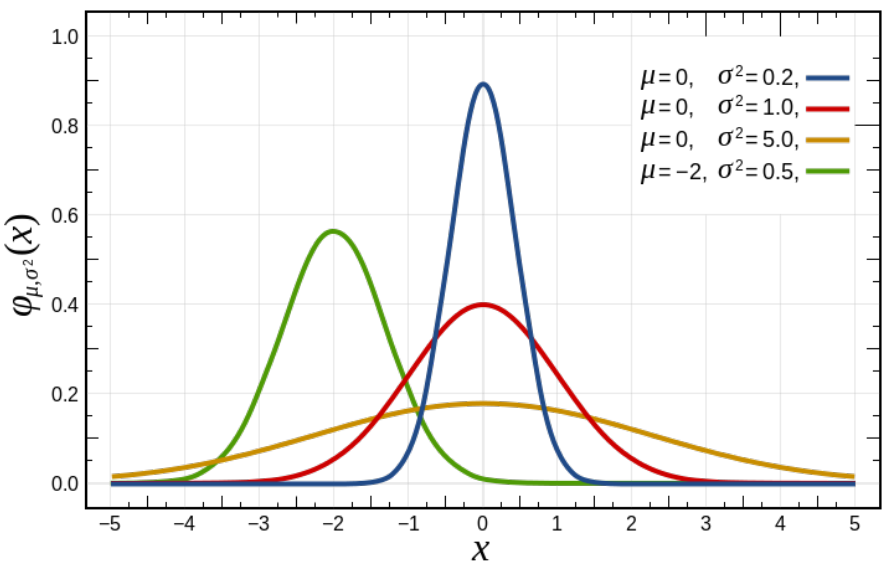
3 elements:

A Model
Parameters
Data

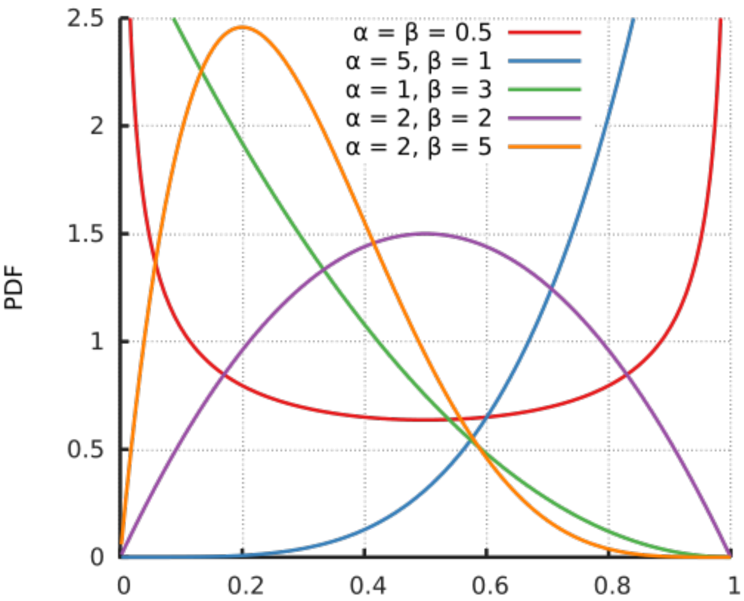
Bayes Theorem: model M

What distribution family do you choose for your prior and likelihood?

Gaussian?

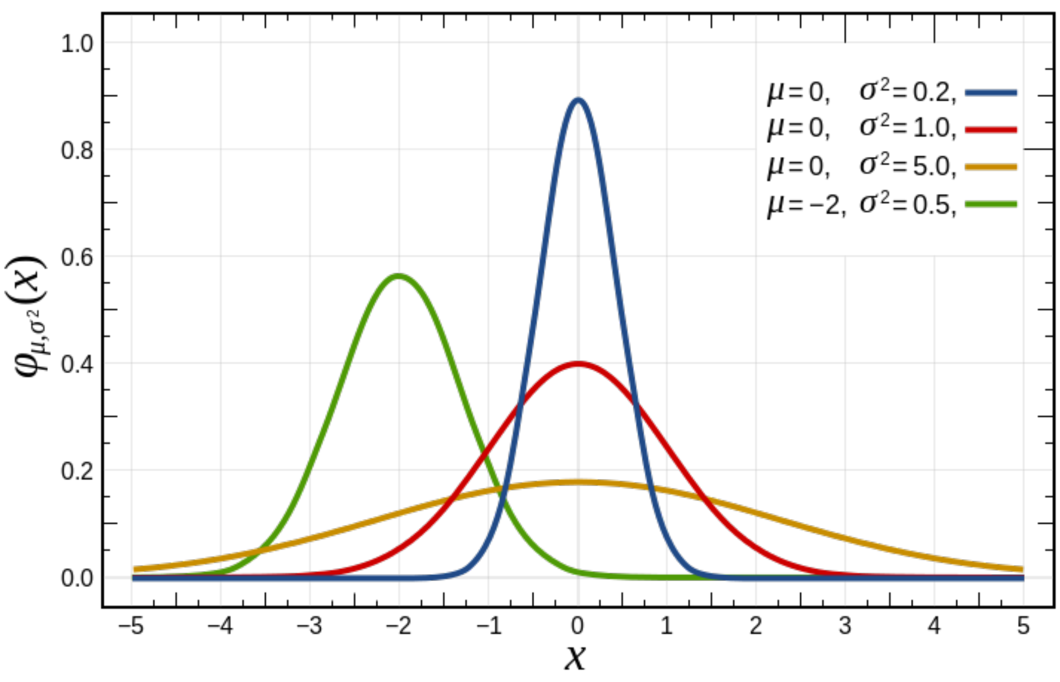


Beta?



Bayes Theorem: parameters θ

Parameters θ – this can be mean and variance, or density parameters, etc.



Bayes Theorem: data y

The information we get from our environment.

Bayes Theorem

Model M + Parameters θ + Data y

Prior: $p(\theta|M)$

Likelihood: $p(y|\theta, M)$

Posterior: $p(\theta|y, M)$

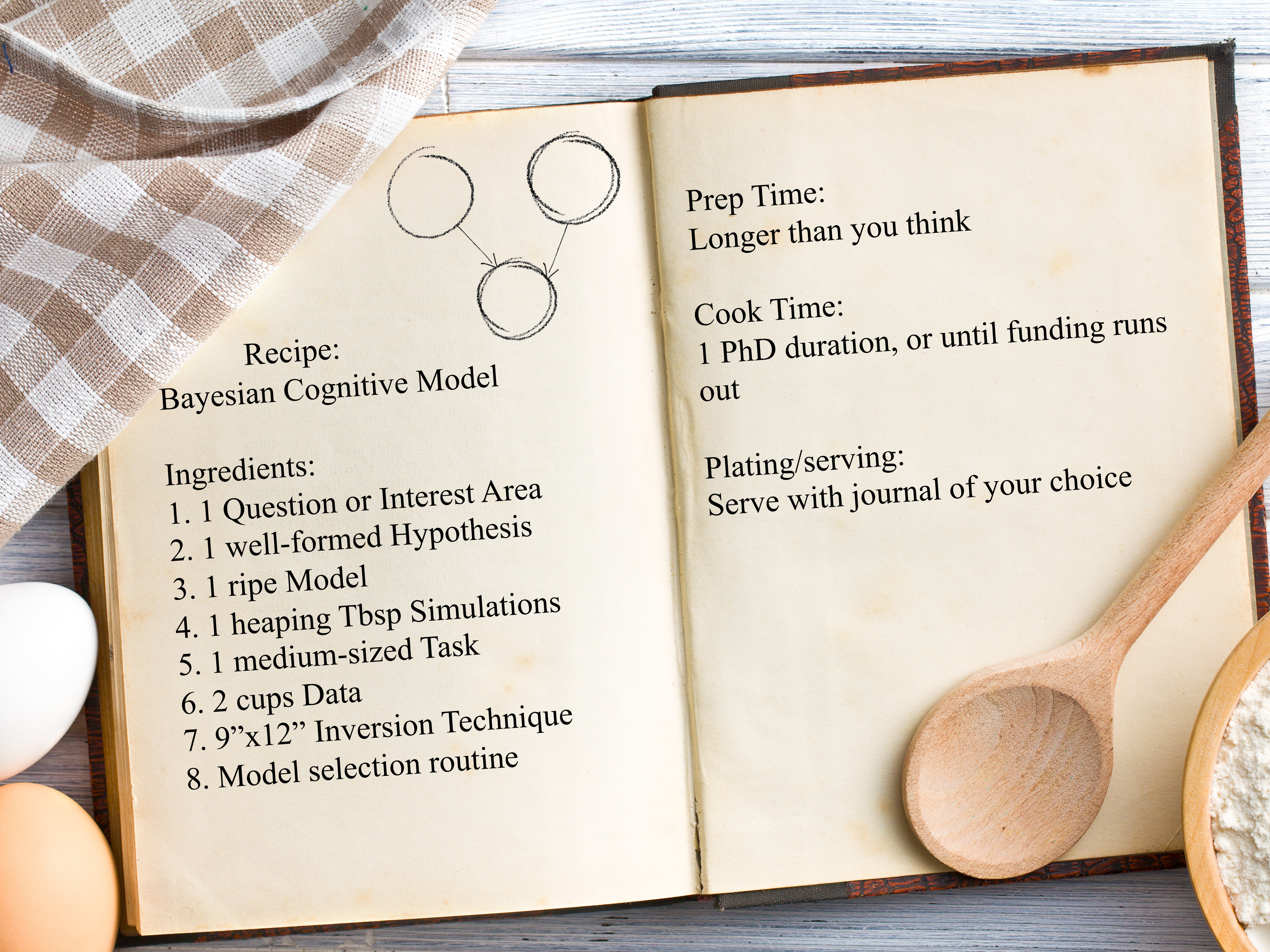
Model Evidence: $p(y|M) = \sum_{\theta'} p(y|\theta', M)p(\theta'|M)$

Bayes' Theorem: Posterior = Likelihood * Prior / Model Evidence

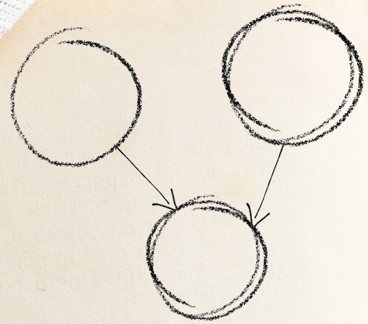
$$p(\theta|y, M) = \frac{p(y|\theta, M)p(\theta|M)}{p(y|M)}$$

Model evidence is difficult to compute, so we work with:

$$\boxed{p(\theta|y, M) \propto p(y|\theta, M)p(\theta|M)}$$


Recipe:
Bayesian Cognitive Model
Ingredients:
1. 1 Question or Interest Area
2. 1 well-formed Hypothesis
3. 1 ripe Model
4. 1 heaping Tbsp Simulations
5. 1 medium-sized Task
6. 2 cups Data
7. 9”x12” Inversion Technique
8. Model selection routine
Prep Time:
Longer than you think
Cook Time:
1 PhD duration, or until funding runs out
Plating/serving:
Serve with journal of your choice

Our problem:
The
Tricky Coin



Recipe:
Bayesian Cognitive Model

Ingredients:

1. 1 Question or Interest Area
2. 1 well-formed Hypothesis
3. 1 ripe Model
4. 1 heaping Tbsp Simulations
5. 1 medium-sized Task
6. 2 cups Data
7. 9”x12” Inversion Technique
8. Model selection routine

Prep Time:
Longer than you think

Cook Time:
1 PhD duration, or until funding runs out

Plating/serving:
Serve with journal of your choice

Step 1: Question

What is the probability of heads?

Step 2: Hypothesis

3 competing hypotheses encoded in priors

(i) We have no information about the coin = Uniform prior

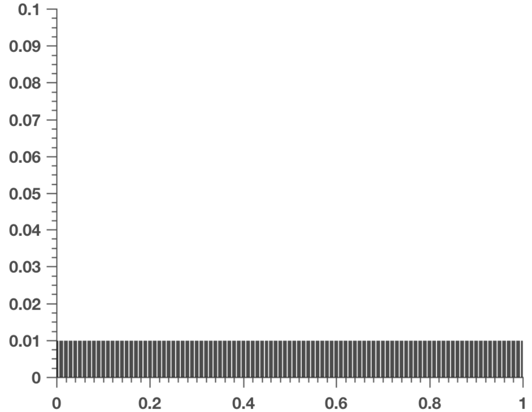
(ii) We believe the coin to be fair = prior around 0.5

(iii) We have a suspicion that the coin is unfair = prior around 0.2

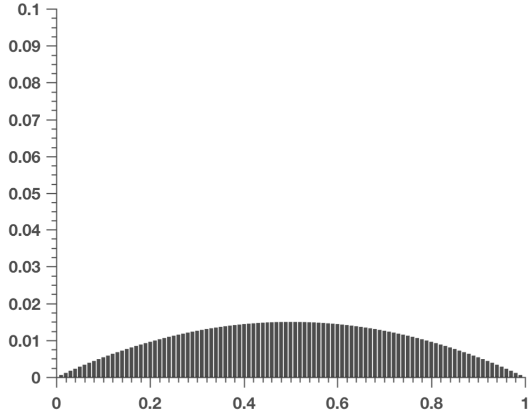
Model class: Beta distribution

$$p(\theta|M) \propto x^{a-1} \cdot (1-x)^{b-1}$$

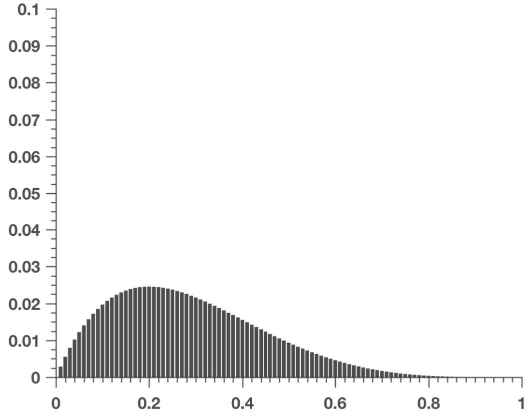
Prior 1

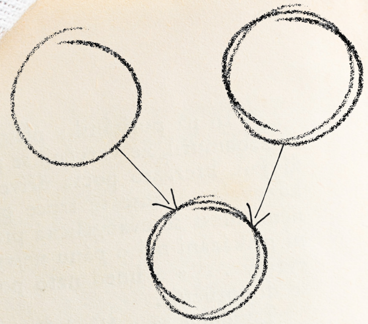


Prior 2



Prior 3





Recipe:
Bayesian Cognitive Model

Ingredients:

1. 1 Question or Interest Area
2. 1 well-formed Hypothesis
3. 1 ripe Model
4. 1 heaping Tbsp Simulations
5. 1 medium-sized Task
6. 2 cups Data
7. 9”x12” Inversion Technique
8. Model selection routine

Prep Time:
Longer than you think

Cook Time:
1 PhD duration, or until funding runs out

Plating/serving:
Serve with journal of your choice

Step 3: Model

Beta-binomial model.

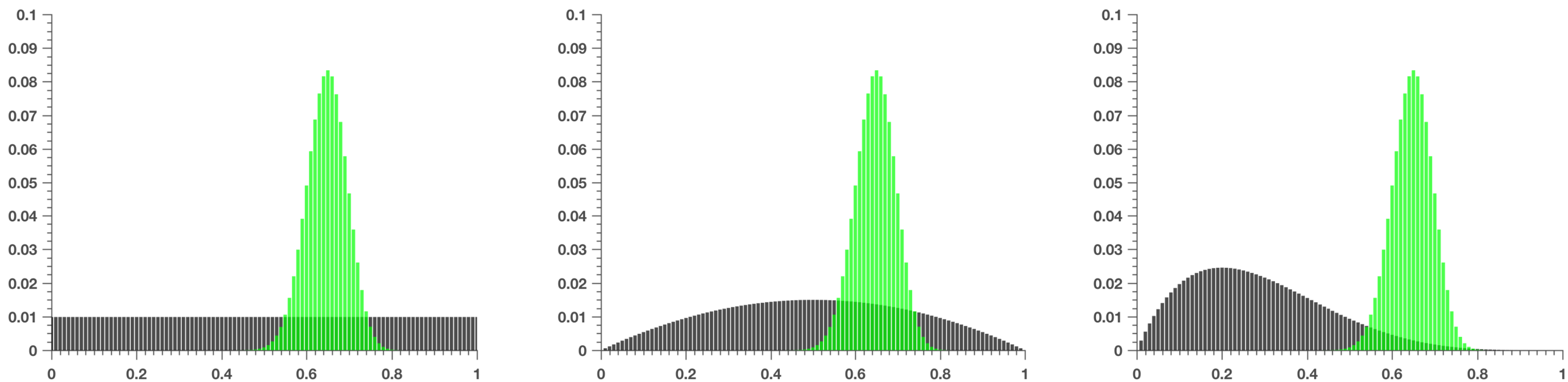
Prior: Beta distribution

$$p(\theta|M) \propto x^{a-1} \cdot (1-x)^{b-1}$$

Likelihood: Binomial distribution

$$p(y|\theta, M) \propto x^{k} \cdot (1-x)^{n-k}$$

For 65 successes, our likelihood looks as follows:



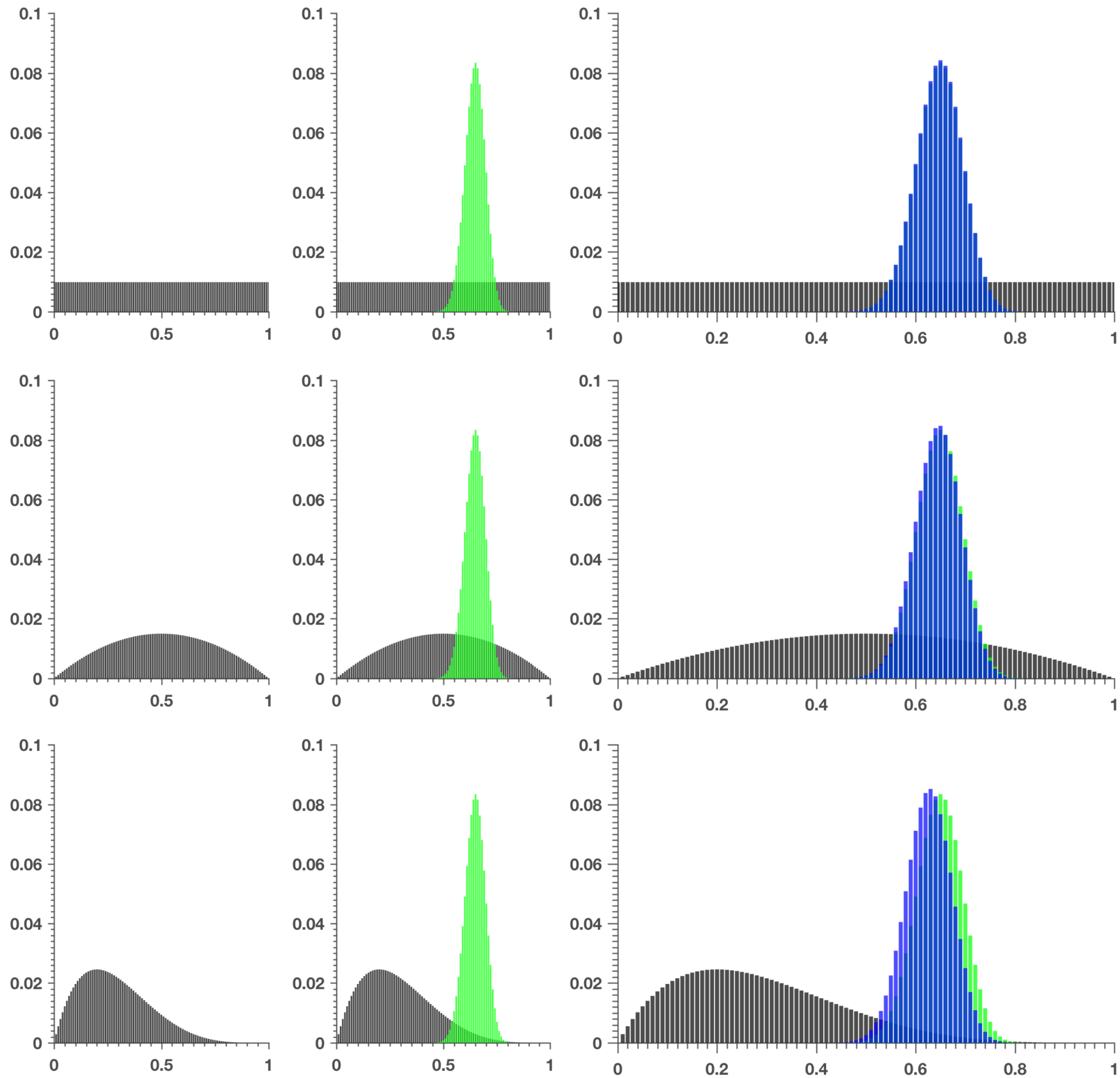
Step 4: Simulations

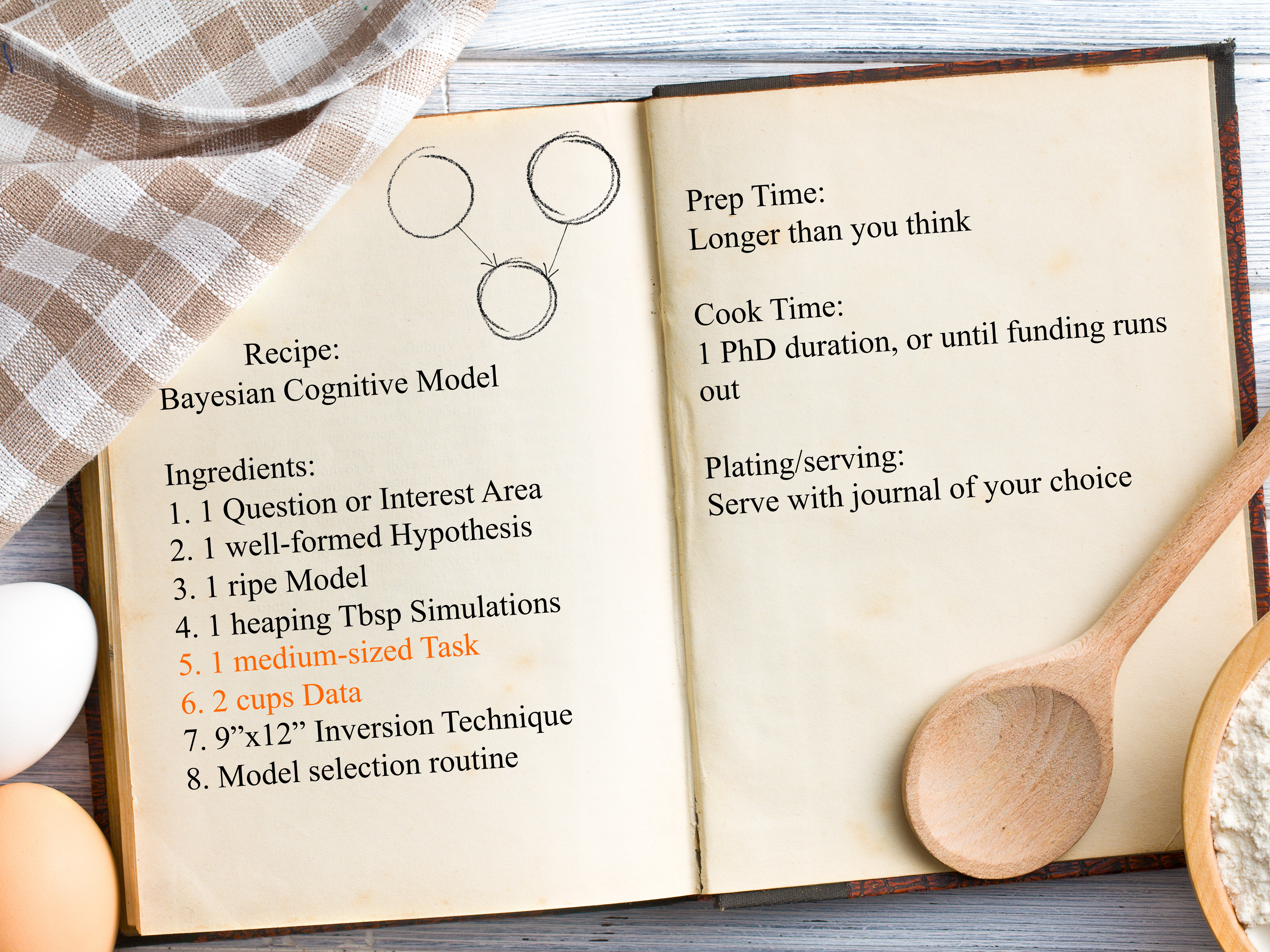
Simulate 65 successes

Simulations Hypothesis 1:

Simulations Hypothesis 2:

Simulations Hypothesis 3:



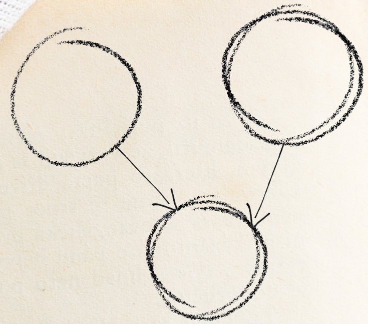

Recipe:
Bayesian Cognitive Model
Ingredients:
1. 1 Question or Interest Area
2. 1 well-formed Hypothesis
3. 1 ripe Model
4. 1 heaping Tbsp Simulations
5. 1 medium-sized Task
6. 2 cups Data
7. 9”x12” Inversion Technique
8. Model selection routine
Prep Time:
Longer than you think
Cook Time:
1 PhD duration, or until funding runs out
Plating/serving:
Serve with journal of your choice

Step 5: Task - Coin toss

Our task is to observe 100 coin flips and estimate the probability of heads.

Step 6: Data

Say we now play the task. We observe a set number of heads, say 45 heads out of 100 tosses.



Recipe:
Bayesian Cognitive Model

Ingredients:
1. 1 Question or Interest Area
2. 1 well-formed Hypothesis
3. 1 ripe Model
4. 1 heaping Tbsp Simulations
5. 1 medium-sized Task
6. 2 cups Data
7. 9”x12” Inversion Technique
8. Model selection routine

Prep Time:
Longer than you think

Cook Time:
1 PhD duration, or until funding runs out

Plating/serving:
Serve with journal of your choice

Step 6: Model Inversion

We now want to assess which of our competing hypotheses (i.e. our 3 priors) is most plausible. In order to do so, we invert our Beta-Binomial model on the observed data. This involves approximating the posterior distribution.

Markov Chain Monte Carlo (MCMC):

Mechanism: sample from a proposal distribution and generate a set of plausible samples from the posterior

Advantages: more precise

Drawbacks: computationally intensive (often takes a long time)

Variational Bayes:

Mechanism: Analytic inversion of a model under the mean field approximation

Advantages: extremely fast

Drawbacks: often inexact

More info:

Next talk & Tuesday afternoon

Inversion continued

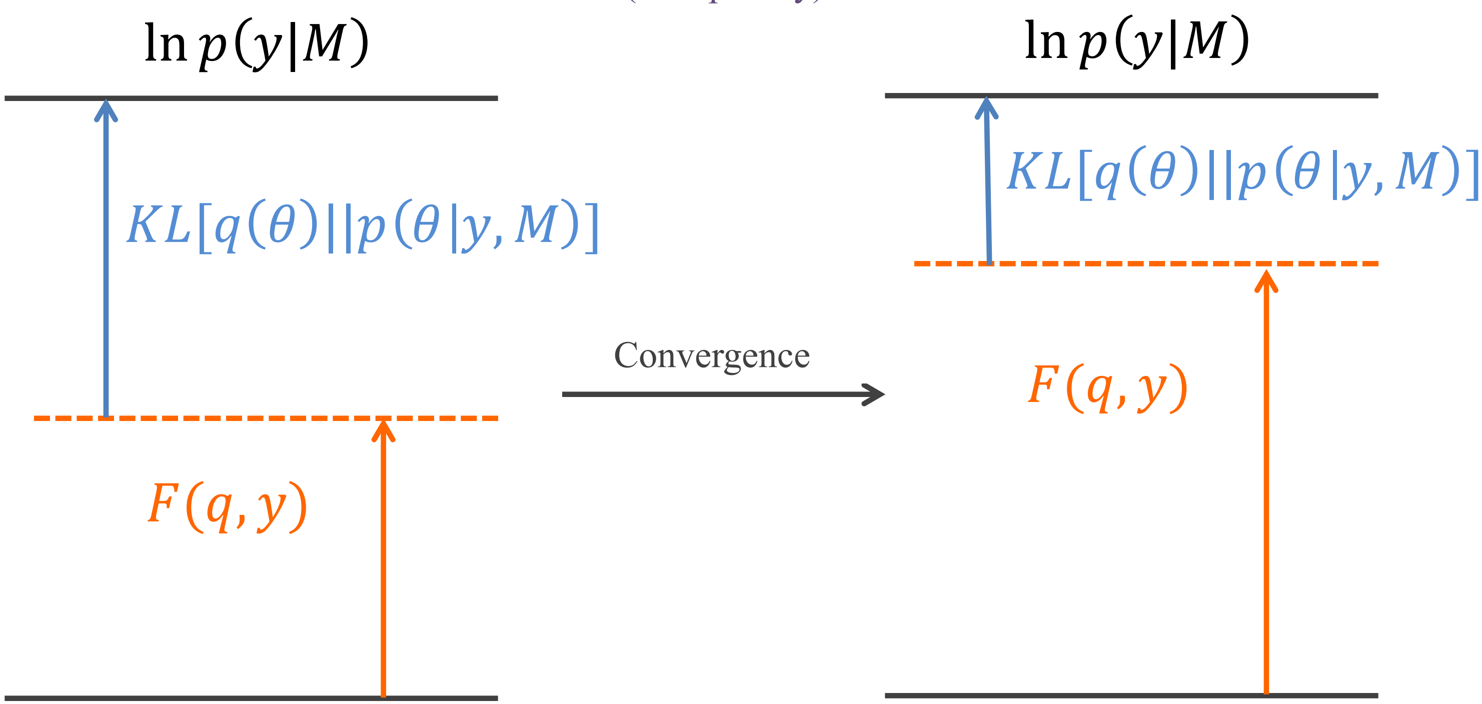
Variational Bayes uses free energy as its variational principle to calculate an approximate posterior. Free energy is the lower bound on model evidence.

$$\ln p(y|M) = KL[q(\theta)||p(\theta|y, M)] + F(q, y)$$

$$F[q(\theta)] = \int q(\theta) \ln p(y, \theta) d\theta - \int q(\theta) \ln q(\theta) d\theta$$

Expected Log-joint (Accuracy)

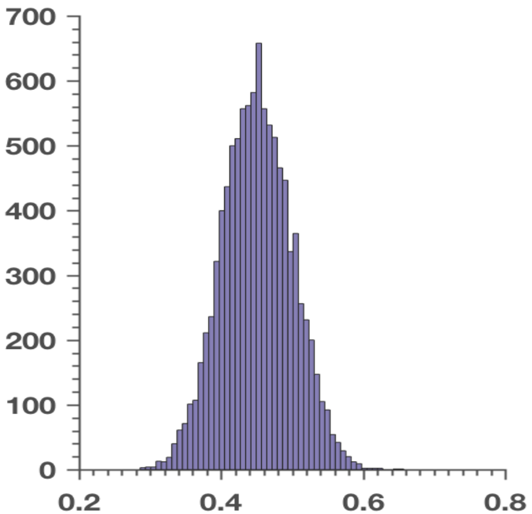
Entropy (Complexity)

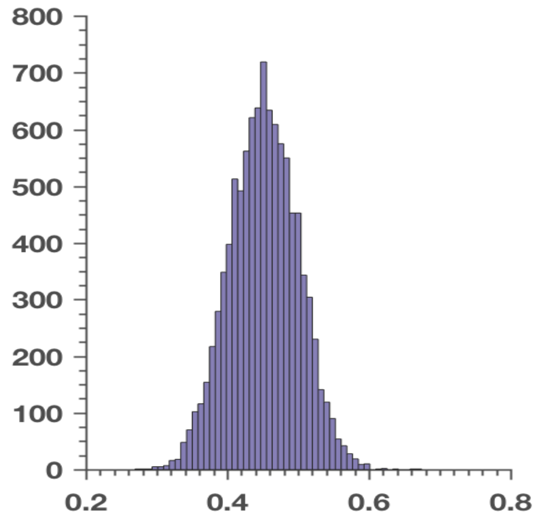


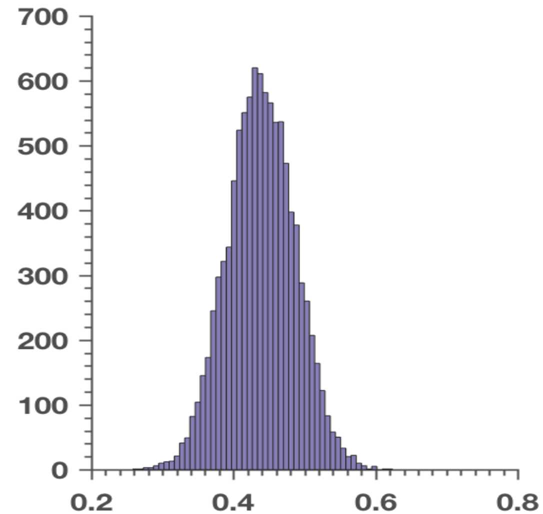
Inversion results

Inversion results:

Posterior 1 Posterior 2 Posterior 3



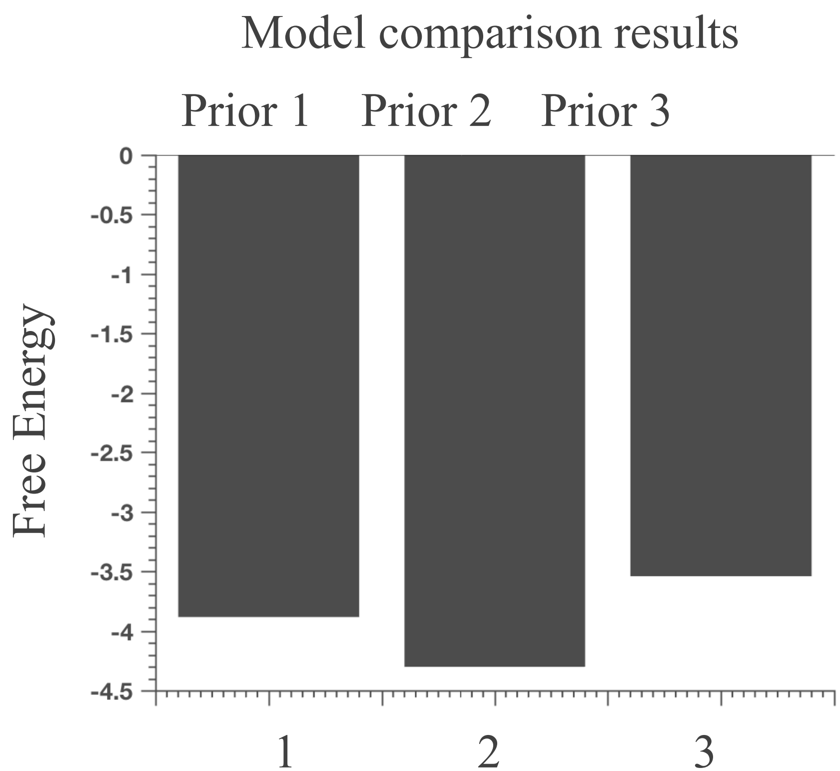




Step 8: Model selection

Free energy comparison allows us to select the best model.

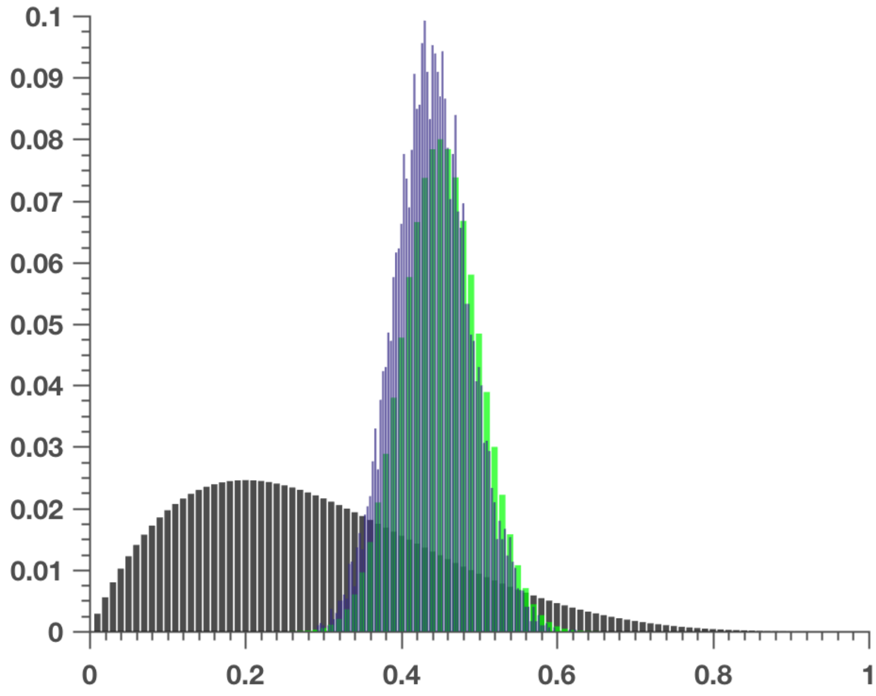
$$F[q(\theta)] = \underbrace{\int q(\theta)\ln p(y,\theta)d\theta}_{\substack{\text{Expected Log-joint} \\ \text{(Accuracy)}}} - \underbrace{\int q(\theta)\ln q(\theta)d\theta}_{\substack{\text{Entropy} \\ \text{(Complexity)}}}$$



More info:
Tuesday Morning

Winning model

Prior ~ Be(2,5)
Likelihood ~ Bin(45,100)



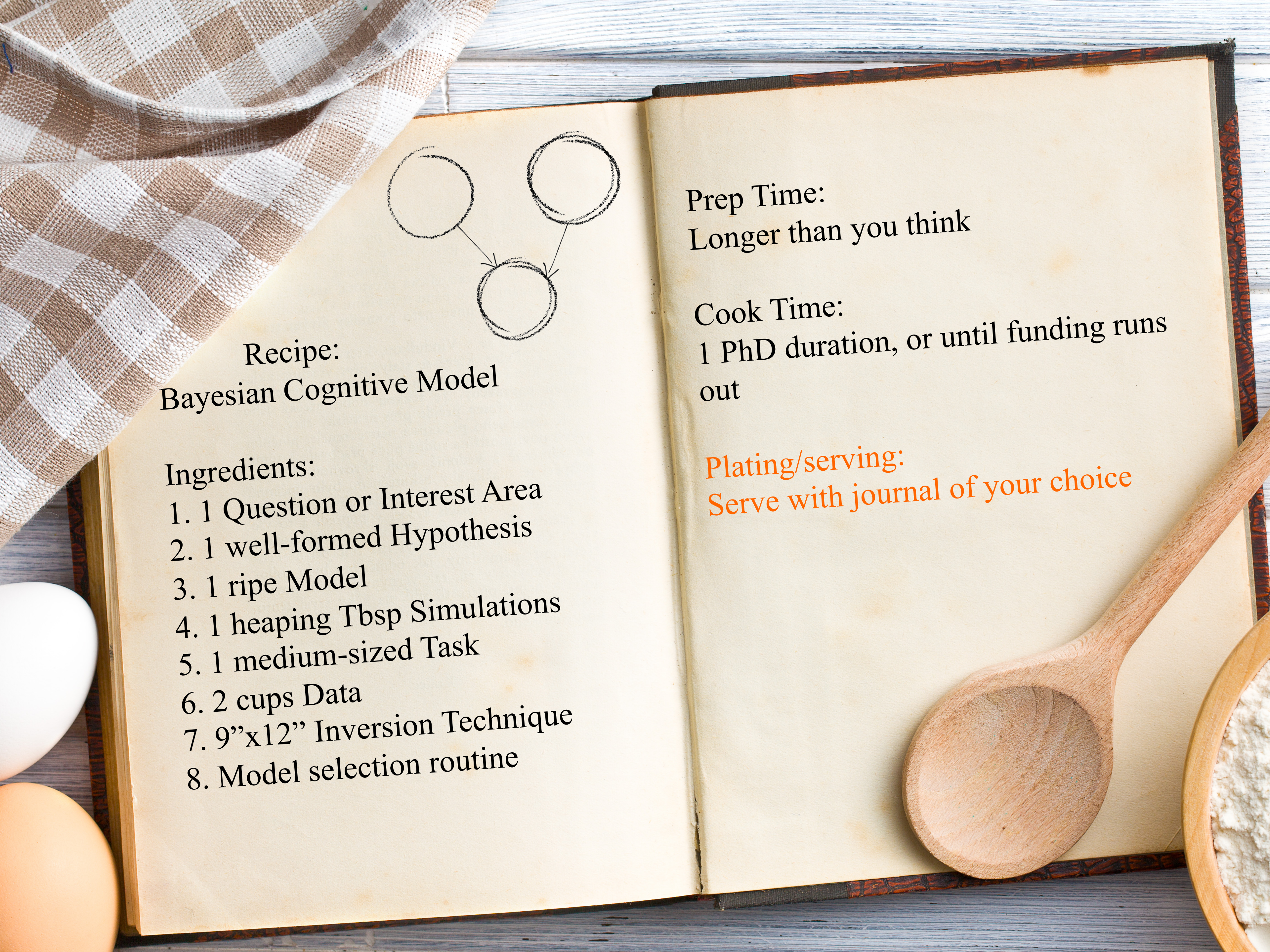

Recipe:
Bayesian Cognitive Model
Ingredients:
1. 1 Question or Interest Area
2. 1 well-formed Hypothesis
3. 1 ripe Model
4. 1 heaping Tbsp Simulations
5. 1 medium-sized Task
6. 2 cups Data
7. 9”x12” Inversion Technique
8. Model selection routine
Prep Time:
Longer than you think
Cook Time:
1 PhD duration, or until funding runs out
Plating/serving:
Serve with journal of your choice

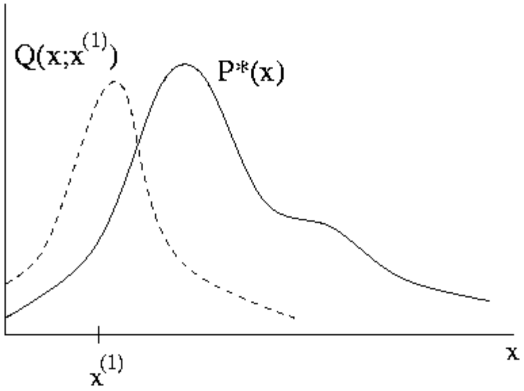
Plate and serve!

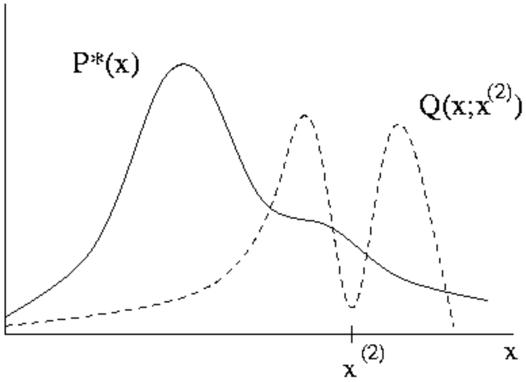
Thank you!

Questions?

Metropolis-Hastings

Goal:
Generate samples from the posterior distribution





Algorithm:

1. Generate a candidate parameter x* from the proposal distribution q(x)
2. Calculate the acceptance ratio $\alpha = f(x^*)/f(x)$ where f is the approximate posterior
3. If $\alpha \geq 1$, then automatically accept

Harmonic mean

Simple approximation of the model evidence, often likened to the free energy (but not equivalent)

Functional form:

$$P(y) \approx \frac{1}{\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}\frac{1}{P(y|t_i)}\right]}$$