Brains and Behavior

a modeling tutorial

Lucy Lai | Lab Meeting | April 24, 2018

acknowledgements

much of this was inspired by/pulled from

Weiji Ma's 2018 CCN Tutorial Talk

Weiji is awesome.

agenda

1. motivation: *why are we doing this?*
2. types of models
3. how to model stuff: *with examples*
    - model building/generative models: *where do we begin?*
    - model fitting: *what does it mean to fit models? what do you do?*
    - model comparison: *how do we know our model is good?*

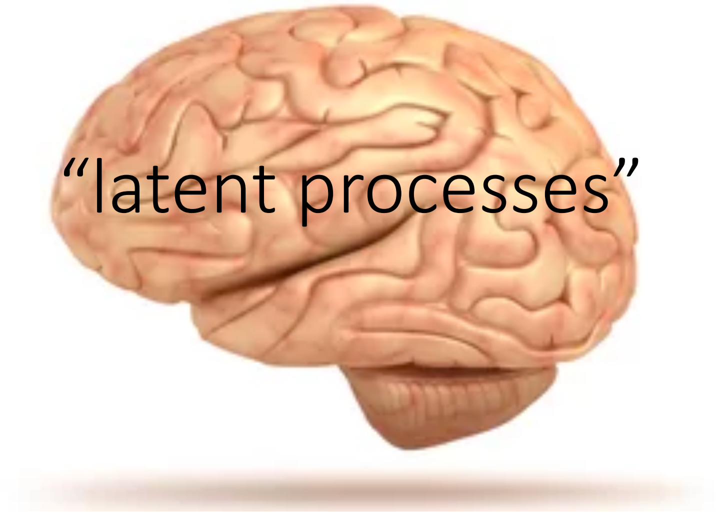
agenda

1. **motivation:** *why are we doing this?*
2. types of models
3. how to model stuff: *with examples*
    - model building/generative models: *where do we begin?*
    - model fitting: *what does it mean to fit models? what do you do?*
    - model comparison: *how do we know our model is good?*

why fit models?

- to infer *latent* causes of behavior that we cannot observe
- understanding the black box (brain)

input



output (behavior)

- to quantify/predict behavior with as few parameters as possible
- behavior is the MOTIVATION for studying the brain (imho)

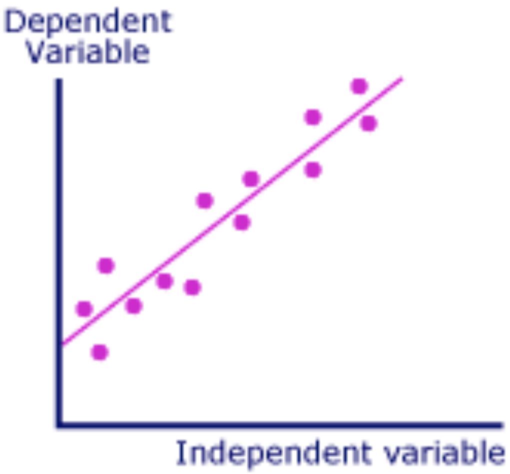
agenda

1. motivation: *why are we doing this?*
2. types of models
3. how to model stuff: *with examples*
   - model building/generative models: *where do we begin?*
   - model fitting: *what does it mean to fit models? what do you do?*
   - model comparison: *how do we know our model is good?*

types of models

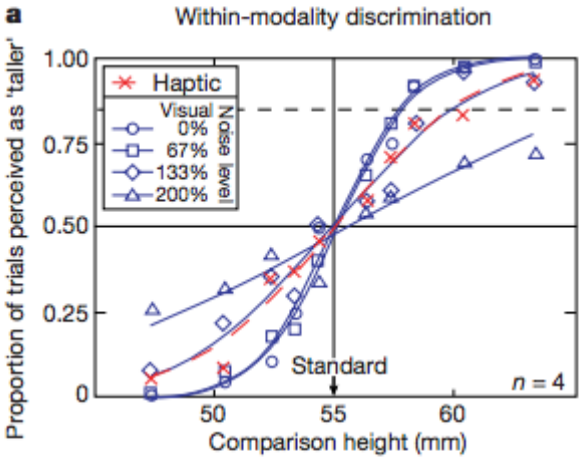
Descriptive model: summary of data, doesn't tell us much about what's going on; the parameters don't anything

- linear regression



Google

- psychometric curve

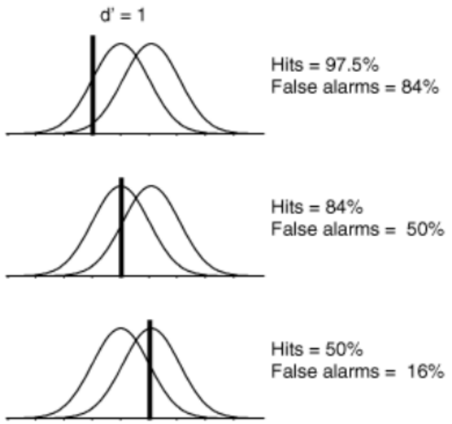


Ernst & Banks 2002

types of models

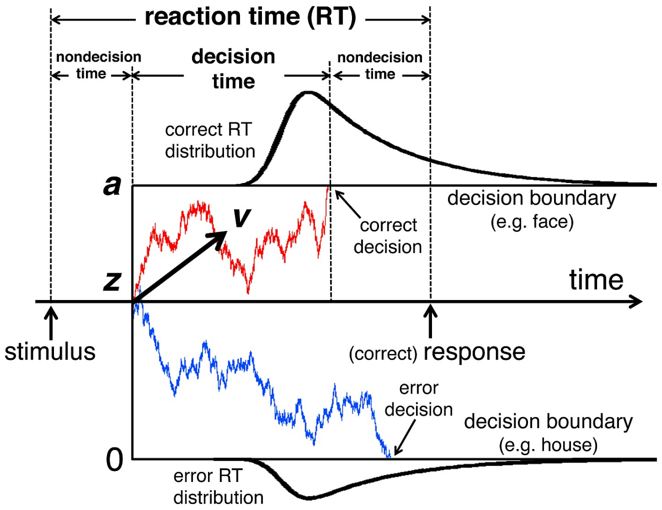
Process (or normative) model: model based on a description of the psychological hypothesis of how an observer perceives/decides

- signal detection theory



D. Heeger

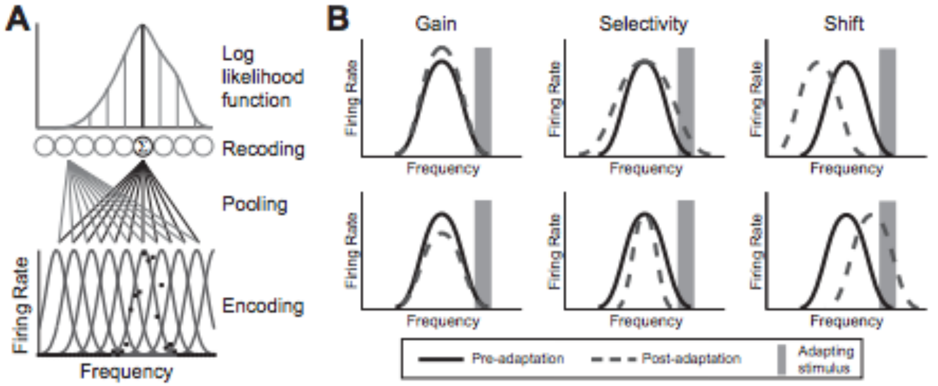
- drift-diffusion



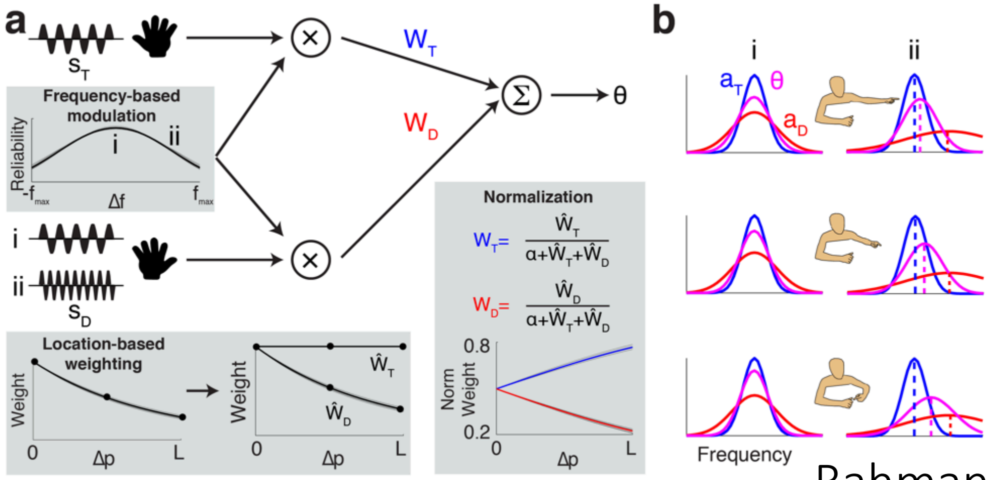
Ratcliff 1978

types of models

Process (or normative) model: model based on a description of the psychological hypothesis of how an observer perceives/decides

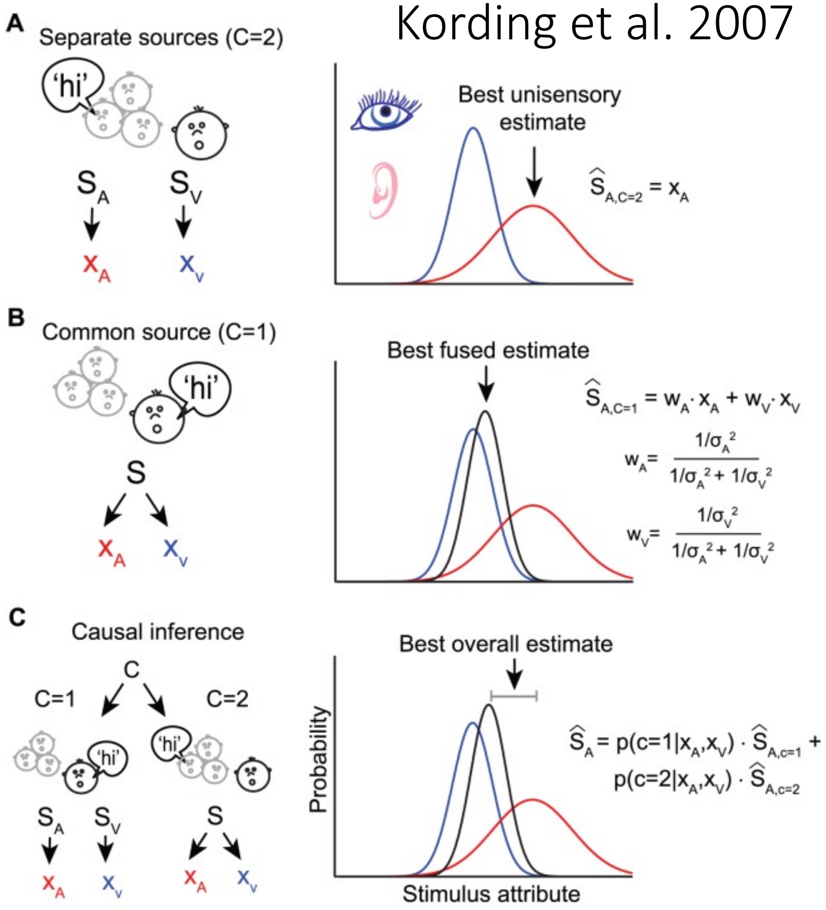


Crommett 2017



Rahman 2018

Bayesian causal inference; Kording et al. 2007



agenda

1. motivation: *why are we doing this?*
2. types of models
3. how to model stuff: *with examples*
    - model building/generative models: *where do we begin?*
    - model fitting: *what does it mean to fit models? what do you do?*
    - model comparison: *how do we know our model is good?*

HOW TO MODEL STUFF

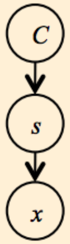
The four steps of Bayesian modeling

Example: categorization task

STEP 1: GENERATIVE MODEL

a) Draw a diagram with each node a variable and each arrow a statistical dependency. Observation is at the bottom.

b) For each variable, write down an equation for its probability distribution. For the observation, assume a noise model. For others, get the distribution from your experimental design. If there are incoming arrows, the distribution is a conditional one.



World state of interest
$p(C)=0.5$

Stimulus
$p(s|C)=\mathrm{N}(s;\mu_C,\sigma_C^2)$

Observation
$p(x|s)=\mathrm{N}(x;s,\sigma^2)$

STEP 2: BAYESIAN INFERENCE (DECISION RULE)

a) Compute the posterior over the world state of interest given an observation. The optimal observer does this using the distributions in the generative model. Alternatively, the observer might assume different distributions (natural statistics, wrong beliefs). Marginalize (integrate) over variables other than the observation and the world state of interest.

$$p(C|s)\propto p(C)\,p(x|C)=p(C)\int p(x|s)\,p(s|C)\,ds=\ldots=\mathrm{N}(x;\mu_C,\sigma^2+\sigma_C^2)$$

b) Specify the read-out of the posterior. Assume a utility function, then maximize expected utility under posterior. (Alternative: sample from the posterior.) Result: decision rule (mapping from observation to decision). When utility is accuracy, the read-out is to maximize the posterior (MAP decision rule).

$$\hat{C}=1 \text{ when } \mathrm{N}(x;\mu_1,\sigma^2+\sigma_1^2)>\mathrm{N}(x;\mu_2,\sigma^2+\sigma_2^2)$$

STEP 3: RESPONSE PROBABILITIES

For every unique trial in the experiment, compute the probability that the observer will choose each decision option given the stimuli on that trial using the distribution of the observation given those stimuli (from Step 1) and the decision rule (from Step 2).

$$p(\hat{C}=1|x)=\Pr_{x|s;\sigma}\left(\mathrm{N}(x;\mu_1,\sigma^2+\sigma_1^2)>\mathrm{N}(x;\mu_2,\sigma^2+\sigma_2^2)\right)$$

- Good method: sample observation according to Step 1; for each, apply decision rule; tabulate responses. Better: integrate numerically over observation. Best (when possible): integrate analytically.
- Optional: add response noise or lapses.

STEP 4: MODEL FITTING AND MODEL COMPARISON

a) Compute the parameter log likelihood, the log probability of the subject's actual responses across all trials for a hypothesized parameter combination.

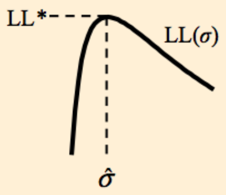
$$\mathrm{LL}(\sigma)=\sum_{i=1}^{\#\text{trials}}\log p(\hat{C}_i|s_i;\sigma)$$

b) Maximize the parameter log likelihood. Result: parameter estimates and maximum log likelihood. Test for parameter recovery and summary statistics recovery using synthetic data.

c) Obtain fits to summary statistics by rerunning the fitted model.

d) Formulate alternative models (e.g. vary Step 2). Compare maximum log likelihood across models. Correct for number of parameters (e.g. AIC). (Advanced: Bayesian model comparison, uses log marginal likelihood of model.) Test for model recovery using synthetic data.

e) Check model comparison results using summary statistics.



this specific to BAYESIAN models

Ma and Kording (to appear), 2018

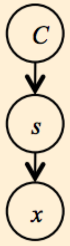
HOW TO MODEL STUFF

The four steps of Bayesian modeling

Example: categorization task

STEP 1: GENERATIVE MODEL

a) Draw a diagram with each node a variable and each arrow a statistical dependency. Observation is at the bottom.
b) For each variable, write down an equation for its probability distribution. For the observation, assume a noise model. For others, get the distribution from your experimental design. If there are incoming arrows, the distribution is a conditional one.



World state of interest
$p(C)=0.5$

Stimulus
$p(s\mid C)=\mathrm{N}(s;\mu_C,\sigma_C^2)$

Observation
$p(x\mid s)=\mathrm{N}(x;s,\sigma^2)$

STEP 2: BAYESIAN INFERENCE (DECISION RULE)

a) Compute the posterior over the world state of interest given an observation. The optimal observer does this using the distributions in the generative model. Alternatively, the observer might assume different distributions (natural statistics, wrong beliefs). Marginalize (integrate) over variables other than the observation and the world state of interest.

$$p(C\mid s)\propto p(C)\,p(x\mid C)=p(C)\int p(x\mid s)\,p(s\mid C)\,ds=\ldots=\mathrm{N}(x;\mu_C,\sigma^2+\sigma_C^2)$$

b) Specify the read-out of the posterior. Assume a utility function, then maximize expected utility under posterior. (Alternative: sample from the posterior.) Result: decision rule (mapping from observation to decision). When utility is accuracy, the read-out is to maximize the posterior (MAP decision rule).

$$\hat{C}=1 \text{ when } \mathrm{N}(x;\mu_1,\sigma^2+\sigma_1^2)>\mathrm{N}(x;\mu_2,\sigma^2+\sigma_2^2)$$

STEP 3: RESPONSE PROBABILITIES

For every unique trial in the experiment, compute the probability that the observer will choose each decision option given the stimuli on that trial using the distribution of the observation given those stimuli (from Step 1) and the decision rule (from Step 2).

$$p(\hat{C}=1\mid x)=\Pr_{x\mid s;\sigma}\left(\mathrm{N}(x;\mu_1,\sigma^2+\sigma_1^2)>\mathrm{N}(x;\mu_2,\sigma^2+\sigma_2^2)\right)$$

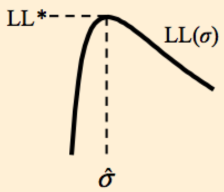
- Good method: sample observation according to Step 1; for each, apply decision rule; tabulate responses. Better: integrate numerically over observation. Best (when possible): integrate analytically.
- Optional: add response noise or lapses.

STEP 4: MODEL FITTING AND MODEL COMPARISON

a) Compute the parameter log likelihood, the log probability of the subject's actual responses across all trials for a hypothesized parameter combination.

$$\mathrm{LL}(\sigma)=\sum_{i=1}^{\#\text{trials}}\log p(\hat{C}_i\mid s_i;\sigma)$$

b) Maximize the parameter log likelihood. Result: parameter estimates and maximum log likelihood. Test for parameter recovery and summary statistics recovery using synthetic data.
c) Obtain fits to summary statistics by rerunning the fitted model.
d) Formulate alternative models (e.g. vary Step 2). Compare maximum log likelihood across models. Correct for number of parameters (e.g. AIC). (Advanced: Bayesian model comparison, uses log marginal likelihood of model.) Test for model recovery using synthetic data.
e) Check model comparison results using summary statistics.



not everything in the world is Bayesian,
so I will focus on the two more general modelling concepts

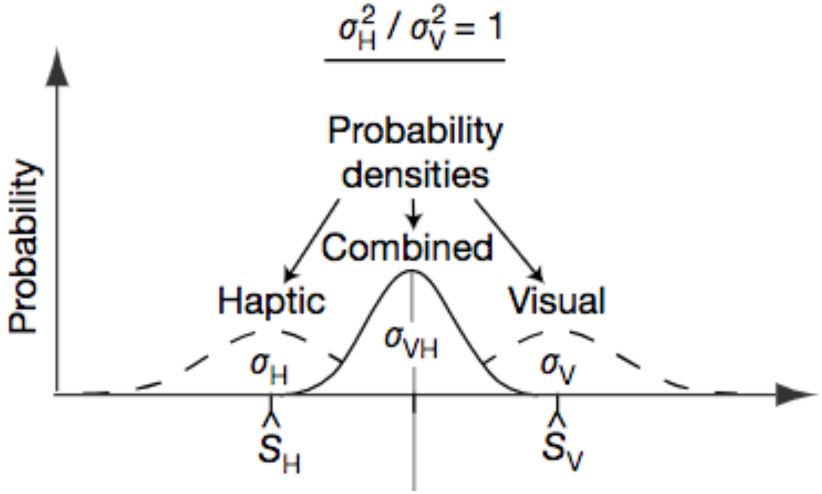
HOW TO MODEL STUFF

<u>Important to understand:</u> This kind of tutorial is NOT for experiments that have known (fully specified) parameters. Ernst and Banks 2002 is one of these experiments.

Fully specified means: parameters that are determined by the experimenter or can be directly measured.

Ernst and Banks 2002: The visual/haptic sensitivities are determined by unisensory experiments (JND). The weights and VH sensitivities are also determined by the model they constructed. YOU DO NOT FIT THIS MODEL. All parameters are known.

Ernst and Banks 2002



$$\sigma_{VH}^2 = \frac{\sigma_V^2 \sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

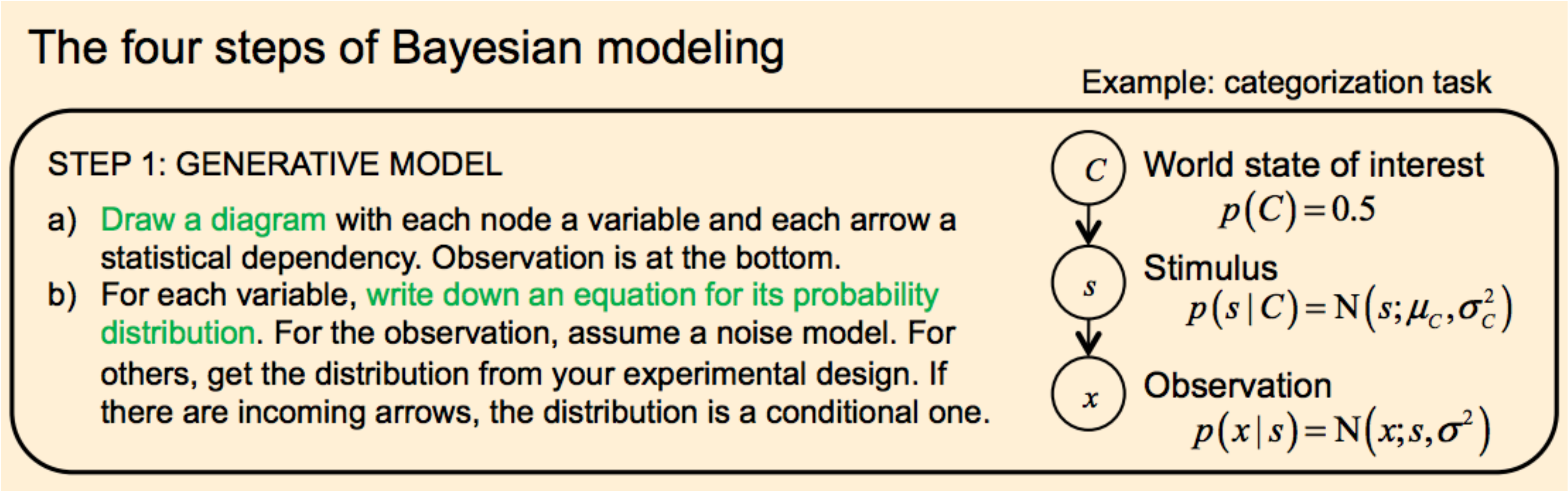
$$w_V = (\mathrm{PSE} - S_H)/(S_V - S_H)$$

agenda

1. motivation: *why are we doing this?*
2. types of models
3. how to model stuff: *with examples*
    - model building/generative models: *where do we begin?*
    - model fitting: *what does it mean to fit models? what do you do?*
    - model comparison: *how do we know our model is good?*
4. sentimental words of reflection and advice

step 1: generative model

Generative model: specifies what is going on in the "black box" from input to output. you can *simulate* the task using MATLAB.



We will now walk through an example Bayesian model with ***discrete*** response choices. Recall, Bayesian models are just one way to model behavior. You don't have to do it this way.

1: generative model

Generative model: specifies what is going on in the "black box" from input to output. you can *simulate* the task using MATLAB.

HOW TO MAKE ANY MODEL

1. **ask yourself,** "what is the task?"
2. **think,** "how might the brain solve this task?"
3. **math out the details** using probability theory, or a dictionary of other kinds of theories:
   - bayesian decision theory
   - signal detection theory
   - drift diffusion models
   - markov chains
   - neural encoding/decoding
   - spiking networks
   - model based or model free reinforcement learning

1: generative model

Question: what level do you choose to model your data at?

- "cognitive," high-level descriptions vs…
- "neural," low-level descriptions

Think: what insight do I want my model to be able to tell me?

Do I want to just describe/predict behavior at a high level? (cognitive)

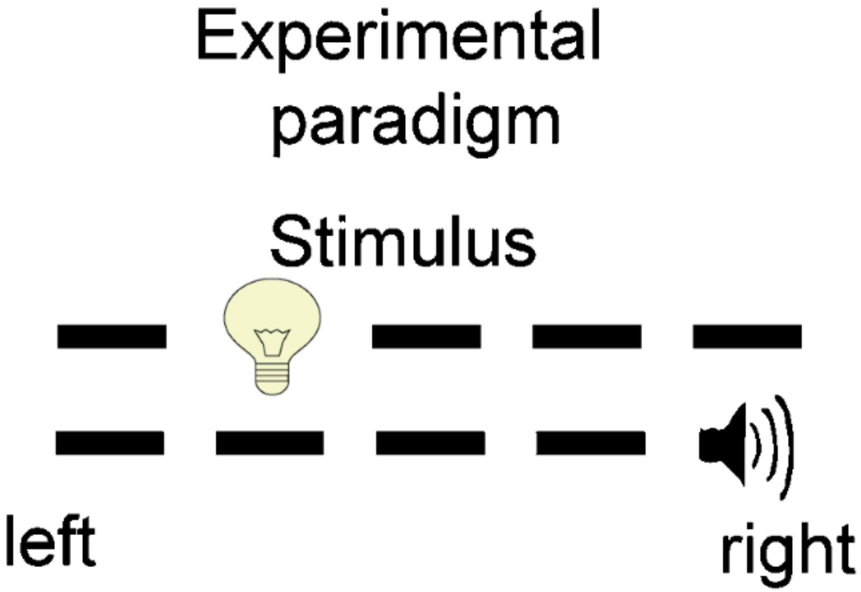
OR

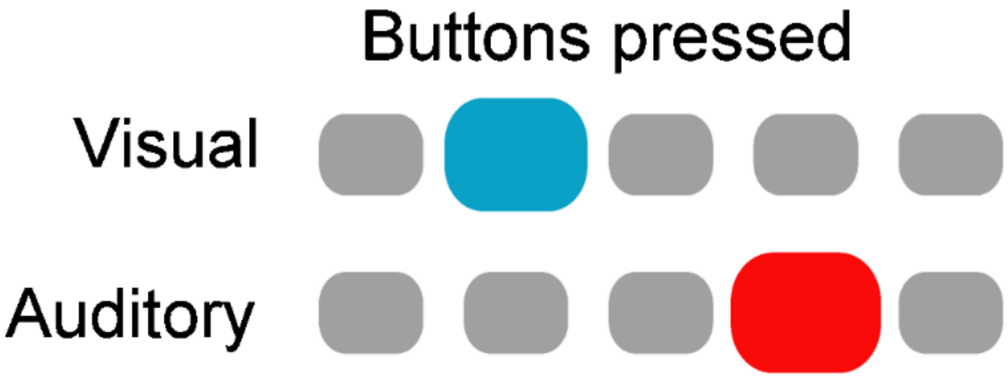
Do I want to understand the dynamics and/or encoding/decoding scheme of the neural populations that might implement this behavior? (neural)

1: generative model

Generative model: specifies what is going on in the “black box” from input to output. you can <u>*simulate*</u> the task using MATLAB.

Ask yourself: how do people localize sounds and flashes?





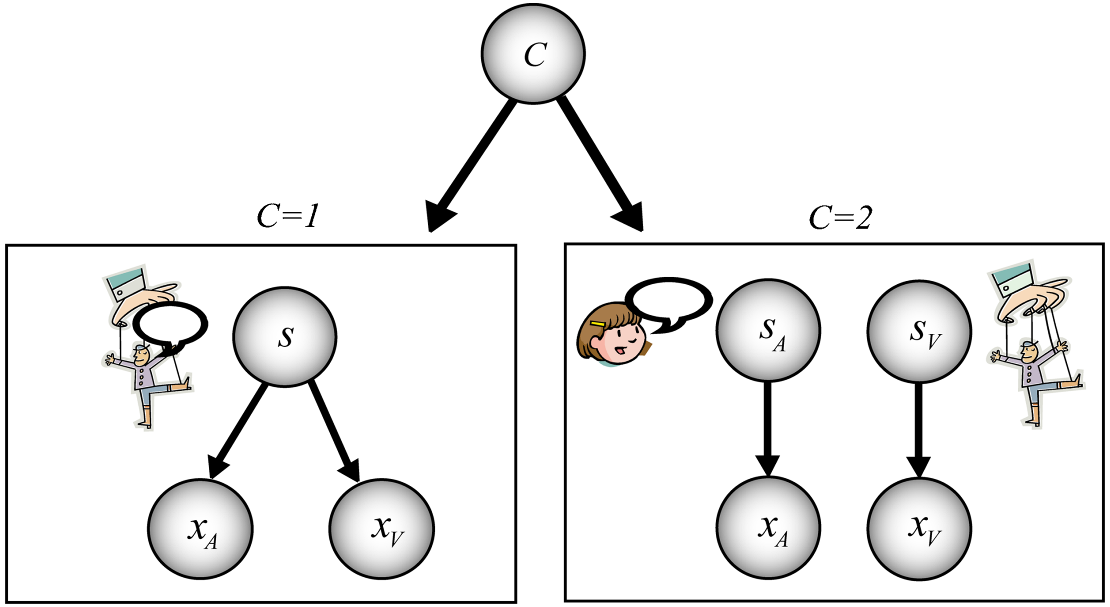
Kording et al 2007

1: generative model

Generative model: specifies what is going on in the "black box" from input to output. you can *simulate* the task using MATLAB.

Ask yourself: how do people localize sounds and flashes?

Think: the brain might *infer* the cause of the auditory and visual cues. then, the brain will use this information to *compute* the final location estimate (given how much they believe that the AV cues came from 1 cause or 2 causes)



Kording et al 2007

1: generative model

Math out the details: this will be very specific to the model you construct, and you will need to read papers/textbooks/internet to learn math and figure stuff out.

Kording et al 2007

1: generative model

Math out the details

1. compute probability that AV cues are from one cause

$$p(C=1|x_V,x_A)=$$

$$\frac{p(x_V,x_A|C=1)p_{\text{common}}}{p(x_V,x_A|C=1)p_{\text{common}}+p(x_V,x_A|C=2)(1-p_{\text{common}})}$$

2. derive the optimal estimates under C=1 or C-2

$$\hat{s}_{V,C=1}=\hat{s}_{A,C=1}=\frac{\frac{x_V}{\sigma_V^2}+\frac{x_A}{\sigma_A^2}+\frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2}+\frac{1}{\sigma_A^2}+\frac{1}{\sigma_P^2}}.$$

$$\hat{s}_{V,C=2}=\frac{\frac{x_V}{\sigma_V^2}+\frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2}+\frac{1}{\sigma_P^2}},\quad \hat{s}_{A,C=2}=\frac{\frac{x_A}{\sigma_A^2}+\frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2}+\frac{1}{\sigma_P^2}},$$

2. weight the final location estimate by the P(C=1|X) and P(C=2|X)

$$\hat{s}_V=p(C=1|x_V,x_A)\hat{s}_{V,C=1}+(1-p(C=1|x_V,x_A))\hat{s}_{V,C=2},$$

$$\hat{s}_A=p(C=1|x_V,x_A)\hat{s}_{A,C=1}+(1-p(C=1|x_V,x_A))\hat{s}_{A,C=2},$$

four "free" parameters:

σ_P , σ_V , σ_A , and p_{common}

Kording et al 2007

agenda

1. motivation: *why are we doing this?*
2. types of models
3. how to model stuff: *with examples*
    - model building/generative models: *where do we begin?*
    - model fitting: *what does it mean to fit models? what do you do?*
    - model comparison: *how do we know our model is good?*
4. sentimental words of reflection and advice

2: model fitting

What does it mean to fit models?

four ”free” parameters:

σ_P , σ_V , σ_A , and p_{common}

- You have some parameters that you don’t know.
- You want to know them.

- You also want your model to output something similar to your actual data.
- That way you perhaps have a reason to claim that the “internal variables” (aka parameters) that you made up, actually exist.

2: model fitting

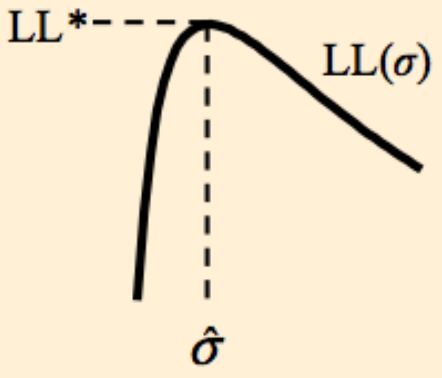
What does it mean to fit models?

STEP 4: MODEL FITTING AND MODEL COMPARISON

a) Compute the parameter log likelihood, the log probability of the subject's actual responses across all trials for a hypothesized parameter combination.

$$\mathrm{LL}(\sigma) = \sum_{i=1}^{\#\text{trials}} \log p\left(\hat{C}_i \mid s_i; \sigma\right)$$

b) Maximize the parameter log likelihood. Result: parameter estimates and maximum log likelihood. Test for parameter recovery and summary statistics recovery using synthetic data.



c) Obtain fits to summary statistics by rerunning the fitted model.

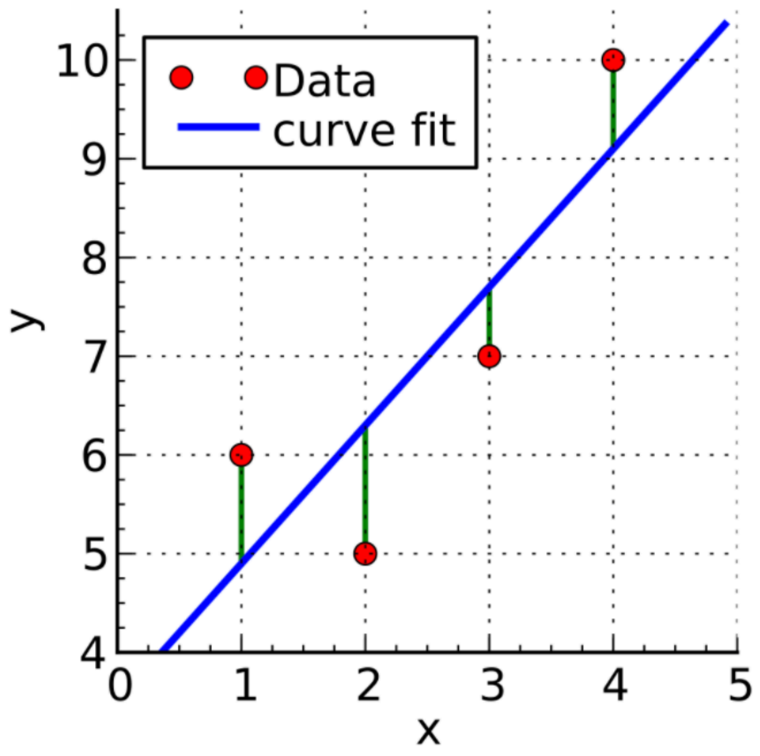
d) Formulate alternative models (e.g. vary Step 2). Compare maximum log likelihood across models. Correct for number of parameters (e.g. AIC). (Advanced: Bayesian model comparison, uses log marginal likelihood of model.) Test for model recovery using synthetic data.

e) Check model comparison results using summary statistics.

2: model fitting

HOW *NOT* TO FIT ANY MODEL

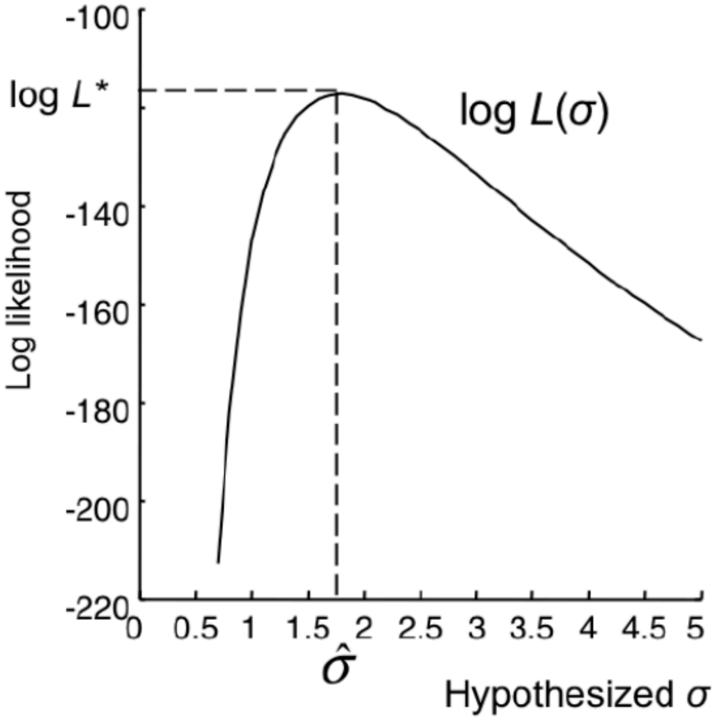
- minimize the squared error between your model output and your data output. it is arbitrary and insensitive to subject's choices



2: model fitting

HOW TO FIT ANY MODEL

1. **maximize** the *likelihood* of the model given the data
   - *likelihood* is the plausibility of a model parameter value given observed data
   - in other words: you have the data. you want to get the parameter values that are MOST LIKELY to have generated the data, given the data that you have.



2: model fitting

HOW TO FIT ANY MODEL

1. **maximize** the *likelihood* of the model given the data
   - *likelihood* is the plausibility of a model parameter value given observed data
   - in other words: you have the data. you want to get the parameter values that are MOST LIKELY to have generated the data, given the data that you have.

2. **do** *parameter recovery* to make sure that your model is *constrained*
3. **compare your models to other models** to show yours is superior (and to publish Nature papers muahahaha)

2: model fitting

Maximize the *likelihood* of the model given the data

- Most of the data from expts in our lab are *discrete responses*
  - *2 choices:* same v. different, which is longer interval?
  - *3 choices:* left v. right v. abort
  - *4 choices:* left v. right v. both v. none (Yash)
  - and so on…
- This allows us to use the *multinomial distribution*
  - not as scary as it sounds
  - for 2 choices, it reduces to the binomial distribution
  - we will look at binomial first, then multinomial

2: model fitting

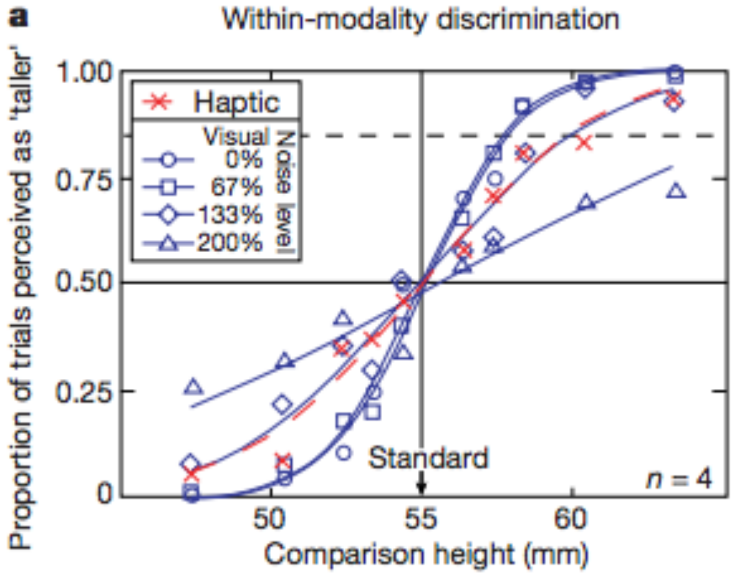
Maximize the *likelihood* of the model given the data

The Binomial Distribution

- tells you the probability of picking option 1 over option 2 in a trial type that is repeated multiple times
- simple logic of why we use it: there are two possible outcomes on each trial
  - ex: first interval or second interval (with the higher frequency)
  - the model outputs the probability that you'll choose a choice over another. we want to compare the actual data to the model's prediction to ***maximize*** the *likelihood* of the model given the data

2: model fitting

Maximize the *likelihood* of the model given the data



The Binomial Distribution

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)}$$

n = # of trial type repeats

x = # times the subject chose option 1

p = the model's probability of choosing option 1 (0 < p < 1)

2: model fitting

Maximize the *<u>log likelihood</u>* of the model given the data

<u>The Binomial Distribution</u>

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)}$$

$$\propto x \log(p) + (n-x)\log(1-p)$$

times subject chose option 1

the model's probability of choosing option 1

# of trial type repeats

2: model fitting

Maximize the *log likelihood* of the model given the data

The Multinomial Distribution

$$P(x) = \frac{n!}{\prod_{i=1}^{4} x_i} \prod_{i=1}^{4} p_i^{x_i}$$

$$\propto \sum_{i=1}^{4} x_i \log(p_i)$$

times subject chose option i

model's probability of choosing option i

n = # of trial type repeats

$\mathbf{x_i}$ = # times the subject chose option i. all the $\mathbf{x_i}$'s add to n

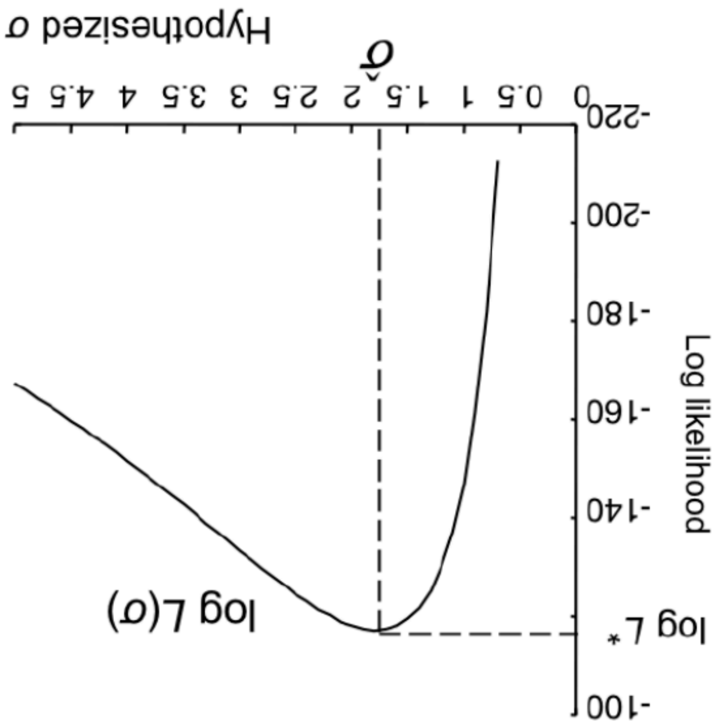
$\mathbf{p_i}$ = the model's probability of choosing option i ($0 < p < 1$). all the $\mathbf{p_i}$'s add to 1

2: model fitting

Here is a mind-blowing concept:

Maximize the *log likelihood* of the model given the data =

Minimize the *NEGATIVE log likelihood* of the model given the data



2: model fitting

MATLAB DEMO

2: model fitting

After you have your best fitting parameters…

Do ***parameter recovery*** to make sure that your model is *constrained*

- this simply means: feed your BEST FIT parameters back into the model → get model's simulated data
- then, take your model's simulated data and use it to FIT the model again
- you should get the SAME parameters back
- this checks to make sure that your model actually has a unique (constrained) parameter solution to your data, and is good at fitting your data. YAY!

agenda

1. motivation: *why are we doing this?*
2. types of models
3. how to model stuff: *with examples*
    - model building/generative models: *where do we begin?*
    - model fitting: *what does it mean to fit models? what do you do?*
    - model comparison: *how do we know our model is good?*
4. sentimental words of reflection and advice

2: model comparison

Compare your models to other models to show yours is superior (and to publish Nature papers muahahaha)

1. first, you need an alternative model. this can be just a simple change to your model (tweak the # of parameters or how the parameters function)
2. fit that model to the data in the same way
3. compare the models using a metric called BIC or AIC (there are plug-and-play functions in MATLAB for this).
    - **NOTE: if models have same # of parameters, you can directly compare their negative log likelihoods. the one with the smaller negLogLik wins!

2: model comparison

A small note on model validation:

- A model is REALLY GOOD if you can get parameters for a subject from that model, and predict that subject's performance on a completely new task
- That's when you know you've really discovered some general principle about how the brain works

questions?