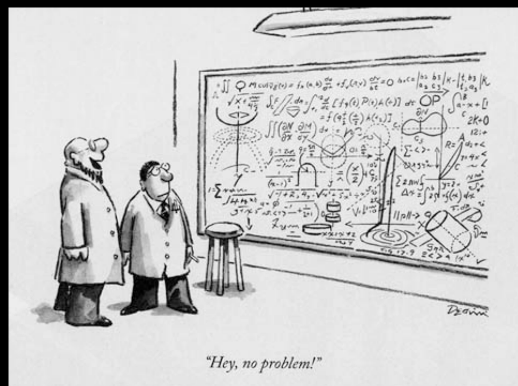
Model Fitting Workshop



1. Bayes rule
2. RL & trial-by-trial model fitting
3. model comparison

plan

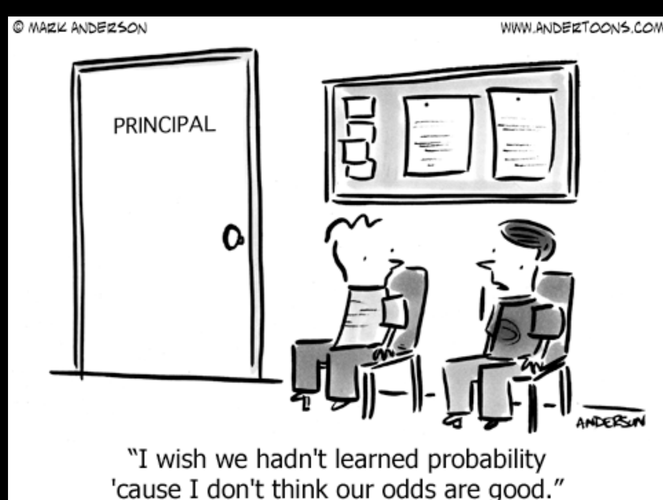
- Snap introduction to Bayes rule
- Basic RL modeling & fitting the parameters of a model
- April: Model comparison

- background papers: Niv & Schoenbaum (2008) - dialogues on prediction errors
  Niv (2009) - RL in the brain
- primary paper: Daw (2011) - trial by trial data analysis using computational models
- newer: Wilson & Collins (2018) - Ten simple rules for the computational modeling of behavioral data

The only stupid question is the one you did not ask

-Rich Sutton

Prologue: Bayes rule and Bayesian inference



first: the intuition

- https://youtu.be/BrK7X_XlGB8

I have two kids.
What is the probability that I have a daughter?

A. 2/3
B. 1/2
C. 3/4
D. none of the above

I have two kids. I have a son.
What is the probability that I have a daughter?

A. 2/3

B. 1/2

C. 3/4

D. it depends

more formally

- random variables (discrete, continuous)
- probability distributions
- two laws of probability:
  - total probability: $p(A) = \sum_B p(A|B)p(B)$
  - joint probability:
    $p(A,B) = p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$

probability distributions as beliefs

what is the probability that this coin is fair?

some beliefs are easier to calculate than others

- you toss a coin 2 times and get heads 2 times
- what is the probability that the coin is fair?
- ... let's ask the opposite: if the coin is fair, what is the probability that 2 of 2 tosses come up heads?

some beliefs are easier to calculate than others

- Brian says: "I'll pass" - what is Brian talking about?
- What is the probability that Brian is talking about an exam?
- If Brian is talking about an exam, how probable is this sentence?

Bayes rule

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

$$P(model \mid data) = \frac{P(data \mid model) \cdot P(model)}{P(data)}$$

$$P(hypothesis \mid evidence) = \frac{P(evidence \mid hypothesis) \cdot P(hypothesis)}{P(evidence)}$$

Bayes rule

posterior likelihood prior

$$P(model\,|\,data) = \frac{P(data\,|\,model) \cdot P(model)}{P(data)}$$

marginalization (ignore!)

Bayesian inference in real life

- interpretation of Bayes rule: we care about both prior and likelihood in inference
- eg. test for disease came out positive
  accuracy of test is 99%
  disease a-priori in 1/10000 people
  what are the odds that the patient has the disease?

Act I: fitting models to trial-by-trial behavior

Bandit tasks

a simple task often used in the laboratory:

- repeated choice between n options (n-armed bandit)
- ...whose properties (reward amounts, probabilities) are learned through trial-and-error

overall approach:

1. learn values for options
2. choose the best option

Bandit tasks

suppose we ran this experiment on a person

- what are the data?
- how do you suggest we model the data?
- what does the model predict?

Writing down a full model of learning

what can we measure (green)?
what do we not know (red)?

```
3 bandits/options
rewards: 1/0

Ntrials = 20
V = [0.5,0.5,0.5]; % initial values
% Input: choice - a vector of Ntrials choices 1/2/3 on each trial
%        R - a vector of Ntrials rewards 1/0 on each trial
alpha = 0.1
beta = 1

For t = 1:Ntrials
        % (log) probability of choice on trial t (we need this to calculate log likelihood, not to
        % model the learning process in the subject's head)
        log_choice[t] = log(exp^(beta*V[choice[t]])/sum(exp^(beta*V))); % normalize by
                                                            %dividing by sum over all possible actions
        % beta is softmax inverse temp (range 0-infinity; unknown – we will fit this)

        % learn from what happened
        PE = R[t] - V[choice[t]]                % prediction error
        V[choice[t]] = V[choice[t]] + alpha*PE % alpha = learning rate (between 0-1; unknown)
End

negLL = -sum(log_choice) % LL = log likelihood of the data given model and parameters
Return (negLL)
```

computational models are basically detailed hypotheses about behavior (and about the brain)

We can test these hypotheses!

Estimating model parameters

why estimate parameters?

1. May measure quantities of interest (learning rates in different populations, how variance in the task affects learning rate etc.)
2. have to use these to generate hidden variables of interest (eg. prediction errors) in order to look for these in the brain

how to estimate parameters?

we want: $P(\alpha,\beta \mid D,M)$

wwBd?

$P(\alpha,\beta \mid D,M) \propto P(D \mid \alpha,\beta,M)$ This we know!

$P(D \mid \alpha,\beta,M) = \prod P(c_t \mid \alpha,\beta,M)$

Estimating model parameters

$P(D \mid \alpha,\beta,M) = \prod P(c_t \mid \alpha,\beta,M)$

...this is a probability distribution

we often consider a point estimate: the maximal likelihood point $\operatorname{argmax}_{\alpha,\beta} P(D \mid \alpha,\beta,M)$

(equivalently, can maximize the log likelihood)

Estimating model parameters

$P(D \mid \alpha,\beta,M) = \prod P(c_t \mid \alpha,\beta,M)$

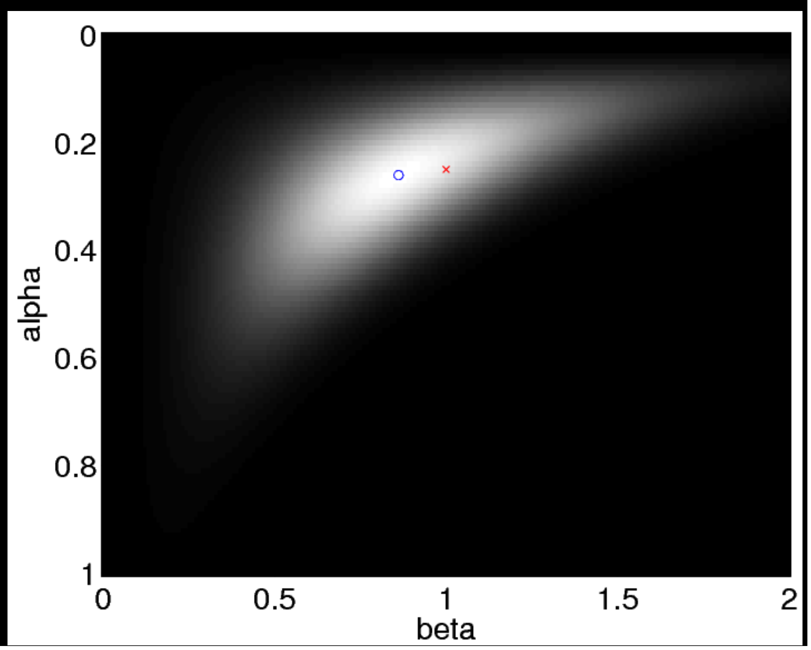
pragmatics:
fmincon, fminunc, fminsearch
minimize -LL
local minima: multiple starting points

simulated 1000 trials

$\alpha=0.25,\beta=1$

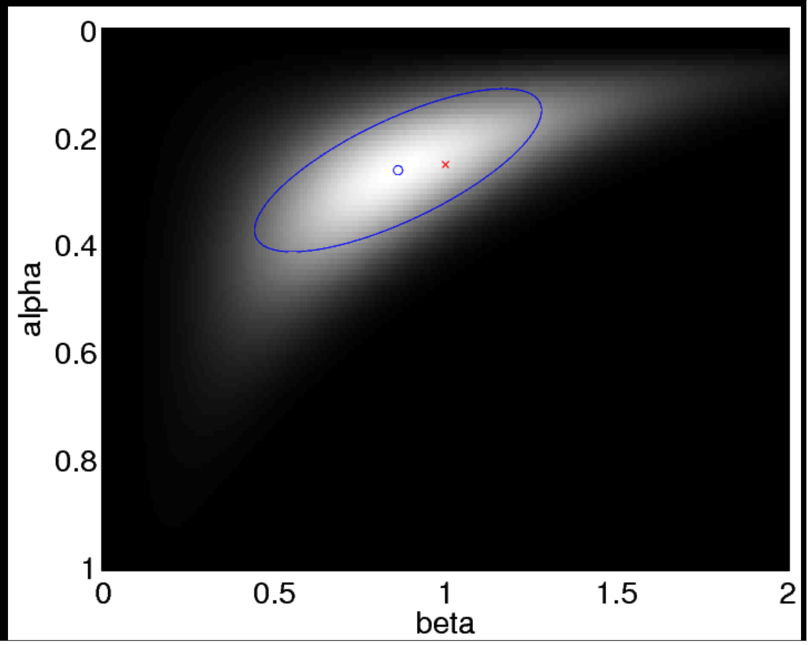
recovered:

$\alpha=0.26,\beta=0.86$



Error bars on estimates

<u>intuition</u>: how fast likelihood is changing as you change the parameters



pragmatics:
inverse Hessian (2nd derivative matrix) of –LL estimates parameter covariance matrix (your minimizing routine will give you this)
- error bars along the diagonal (sqrt)
- covariation off the diagonal
- can also look at variation in fits across subjects

Summary so far

- Learning models are detailed hypotheses about trial-by-trial overt and covert variables
- these can be tested against the brain to help understand what different brain regions/networks are doing/computing
- at CCNP, we use these methods heavily to measure individual differences and correlate to psychopathology

- Next time: how do we decide what is the best model to use?