Quantitative methods for cognitive scientists

An opinionated intro by Charles Zheng

Some quantitative methods

- two-sample t-test
- ANOVA
- linear/nonlinear regression, general linearized models, mixed-effects models
- probabilistic models
- recurrent neural networks

Stages of development in using methods

- Stage 1: Ignorance. *Fear and distrust formalized methods and rely only on intuition.*
- Stage 2: Method idolatry. *Found a method that works but do not fully understand it. Discard intuition in favor of dogma and blind belief in method outputs.*
- Stage 3: Holistic. *Understands deep principles; uses methods to extend (rather than replace) intuition.*

Psychology would be OK without quantitative methods

- Progress would be much slower
- Publication standards would be in terms of "you need X number of samples" rather than "<=0.05 significance"
- Experiments would be much simpler

Dealing with confounds is the main reason for using quantitative methods

Example: does mood influence risk-taking?

You could study this easily with a simple experiment

- Randomize subjects to treatment and control, ask them to rate mood
- In the treatment group, intervene with their mood (by showing a funny video, giving a compliment, etc.)
- In the control group, do nothing
- Ask them to rate mood, then have them decide whether or not to take a gamble (e.g. flip a coin, heads +1 dollar, tails -1 dollar)

- The only problem with the simple experiment is that it is not very *powerful* at finding effects
- You will need a lot of sessions with a lot of subjects to get a robust effect
- Why does it lack power?
- 1. You cannot get alot of repeats in one session
- 2. You may not be able to change their mood by a large amount

Here's another experiment

- Every subject plays the same task
- Task is a series of game rounds, and interspersed mood prompts
- Game round: take a certain amount \$C or gamble to receive either \$A or \$B
- Game rounds are grouped into a number of blocks, "favorable" and "unfavorable"

Pros to this new experiment

- Get more data per subject
- You can better characterize behavior at an individual level
- Subjects find it more engaging than repeats of "simple experiment"
- You may get larger mood variation

Cons to this new experiment

- More complicated
- The decision to gamble is confounded with the past trial values and outcomes
- If you wanted to study the relationship between mood and gambling using ANOVA or correlations, you might be better off with "simple experiment"

Causal inference 101

- The trial values are a potential confound for the effect of mood on gambling probability
- Two main ways to correct for this confound
- 1. Matching
- 2. Modeling the joint effect of trial and mood on gambling

Matching

- Within a subject, pair up trials where they saw the same values but where they had different mood ratings.
- One trial in each pair goes to a "low mood" group, the other to a "high mood" group
- Not all trials may be paired
- Test to see if the groups differ by gambling probability

Open question

Can we apply matching to the MMI data to investigate the effect of mood on behavior?

Modeling

- Come up with at least one plausible model of how both the variable of interest (mood) and all measured confounders (trials) affect the outcome (gambling)
- Actually, come up with multiple models and use data to eliminate bad models
- If you have a good model, you can make inferences about the effect of interest that are less contaminated by confounding than a naïve analysis

Cons to using modeling

- No guarantee that it will work (protect you from false conclusions) if your model is wrong
- "All models are wrong..." - George E. Box
- Difficult work that requires technical knowledge
- Danger of over-interpreting models

Pros to using modeling

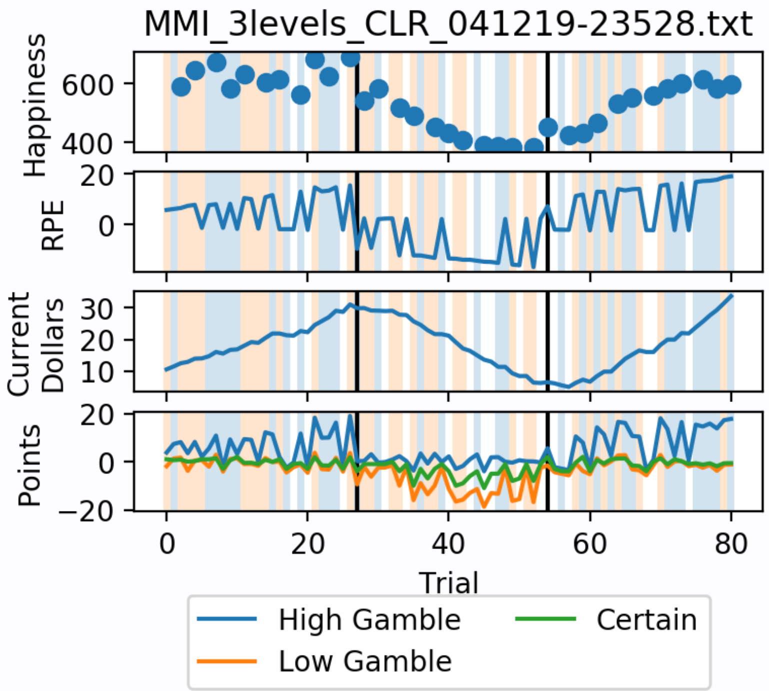
- "...but some [models] are useful" - G. E. Box
- If you find good models, and can control for confounds...
- ...this allows you to use more complicated experiments, giving you more power to detect effects.
- Having a quantitative model allow you to test a theory very precisely and thoroughly

Modeling approaches

- Approach I: curve-fitting
- Approach IIa: probabilistic models
- Approach IIb: probabilistic models *with uncertainty quantification*

Approach I: curve-fitting

Suppose you have some data that you want to explain



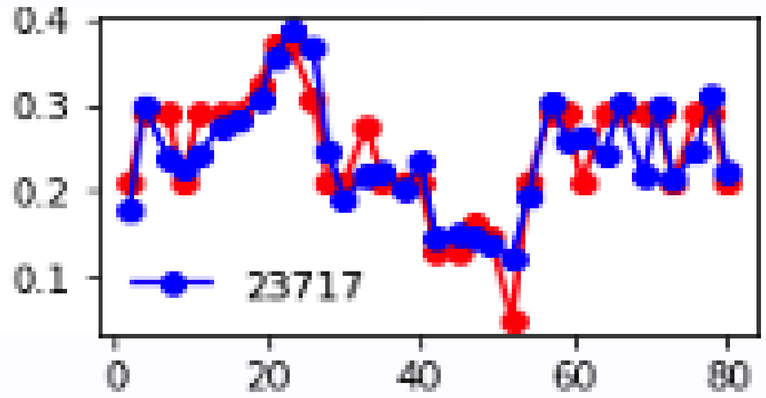
You think there is some simple function of the known variables that *approximates* your data, like

$$M(T) = M_0 + \sum_{t=1}^{T} \lambda^{T-t}(\beta_{RPE} RPE(t) + \beta_E E(t))$$

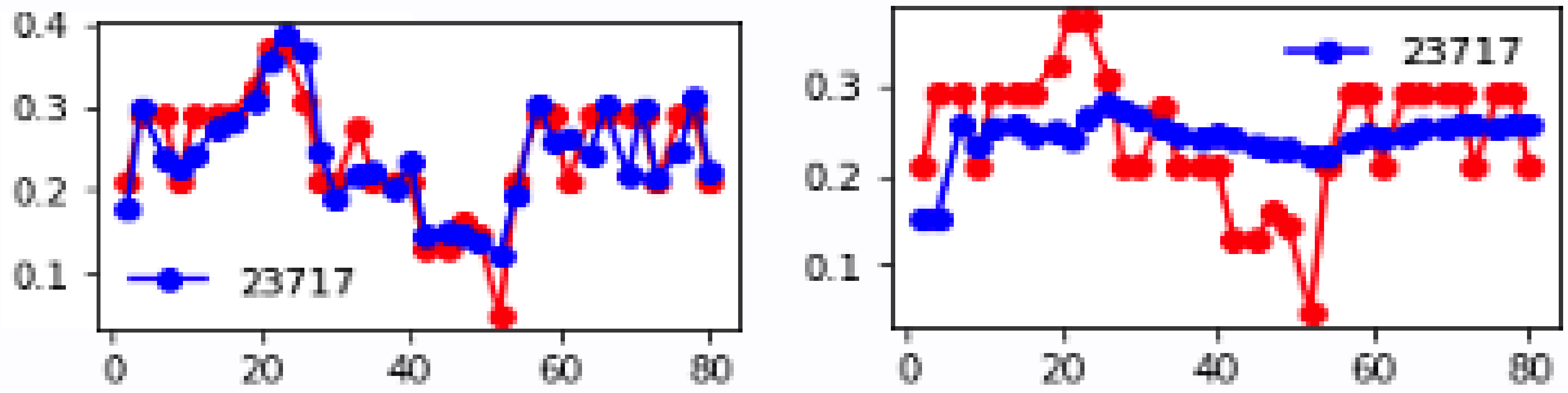
where

$$E(t) = \frac{1}{t-1} \sum_{u=1}^{t-1} A(u)$$
$$RPE(t) = A(t) - E(t)$$

You can find parameters of that function which best match the fit to your data (e.g. in squared-error loss or absolute-error loss)



Different theories may suggest different families of functions. You can see which family fits the data best.



- Pitfalls: A family may fit the data better simply because it is more complicated, and hence *overfits* more

Validated curve-fitting

Predicting on held-out data is a better assessment of which function family is a better fit to the data

Regularized curve-fitting

Without sufficient amount of data, parameter estimates become very unstable. Regularization constrains the parameters (e.g. to be small) which stabilizes estimates.

https://quantumcomputingtech.blogspot.com/2019/06/machine-learning-regularization-term.html

Methods in python

- Nonlinear least-squares: in `scipy`
- Regularized methods: elastic net, multinomial logistic regression in `sklearn`
- Custom function families: use `pytorch` (or `tensorflow`)

A problem with joint models

- Suppose you want to predict both mood and choice...
- Mood measured in MSE, choice measured in accuracy...
- How can you combine both into a single metric?

A problem with assessing model reliability

- A reviewer asks you to assess how reliable the parameter fits you obtain via curve-fitting
- But you don't know the ground truth!
- If only you could *simulate* the data with a known ground truth... (Wilson and Collins 2019)

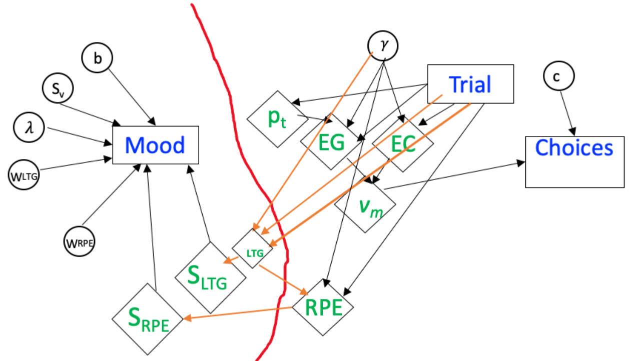
Approach II: probabilistic models

- Specify a probability model for generating your data from parameters
- Advantage 1: we have an omnibus measure of fit for *all* predictions--the data likelihood $p(D|\theta)$.
- Advantage 2: you have a recipe for generating data from your model

Example: A likelihood model for mood and choice

LTA_nonlinear_choice
{LTA Gamma (W/O RT p0 and $\boldsymbol{\beta}$) Model}

○ Parameter node
□ Data node
◇ Intermediate "Latent" node



***Mood at time t*:**
M(t)~ Normal(mean = $b + W_{LTG} * S_{LTG}(t) + W_{RPE} * S_{RPE}(t)$, variance = S_v)

- $S_{LTG}(t) = \sum_{u=0}^{t} \lambda^u * LTG(t-u)$
- $S_{RPE}(t) = \sum_{u=0}^{t} \lambda^u * RPE(t-u)$
- Weights for regressors: W_{LTG} , W_{RPE}
- *Baseline Mood Bias: b*
- *Discount Rate: λ*
- $LTG(t) = \frac{1}{t}\sum_{i=1}^{t}(V_i^{\gamma})$
- V_i^{γ} is actual outcome

- Internal representation factor of Reward : γ

- Expectation of Certain Value: $EC = V^{\gamma}$
- Expectation during Gambling: $EG = P * V_+^{\gamma} + (1 - P) * V_-^{\gamma}$
- Update rule of winning probability:

  p_t = Nw / Nt

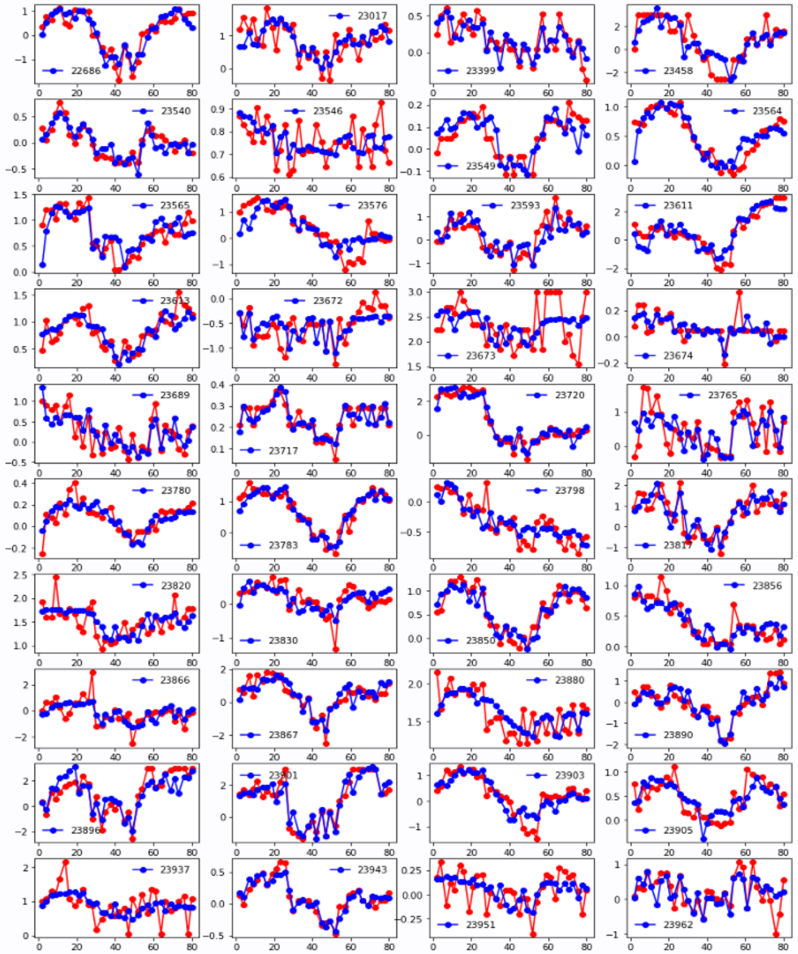
  Nw = sum of wins until trial t

  Nt = sum of trials until trial t
- Reward Prediction Error: $RPE(t) = V_a^{\gamma}(t) - LTG(t)$
- Mean Drift Rate: $v_m = c(EC - EG)$
- Likelihood Function : Binomial(Choice| $invlogit(v_m)$)

Approach IIa: Point estimates

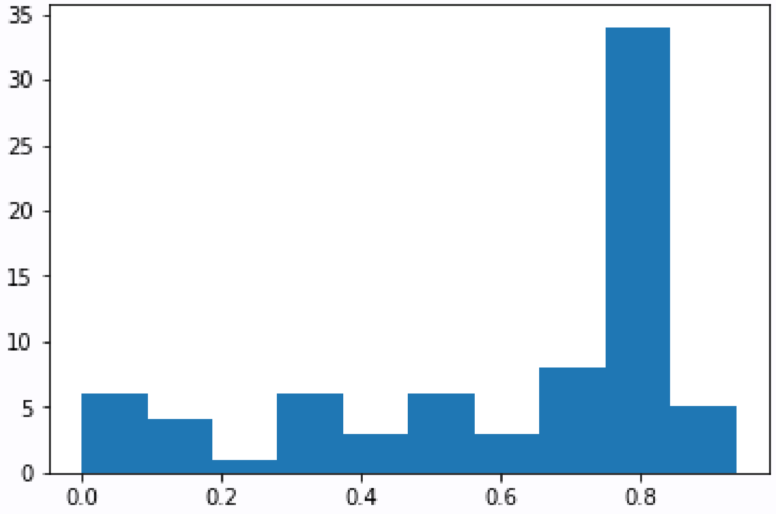
- **Maximum likelihood** (MLE): Find parameters θ maximizes the data likelihood $p(D|\theta)$
- **Maximum a posterior** (MAP): Maximize the prior-weighted likelihood $p(D|\theta)p(\theta)$
- MAP behaves like a regularized form of MLE
- Methods: `scipy.optimize`, `pytorch`, `tensorflow`

Example: MAP fits for LTA nonlinear choice model. (Fits to choice not shown.)



How uncertain are your parameter estimates?

Say you fit MAP/MLE and you find a range of parameters (e.g. for the discount factor λ) across subjects.



- Can we conclude from this that subjects actually differ from each other in the parameter λ?
- The problem is that there is noise in the estimate of λ. So even if all subjects had the same λ, you could get different estimates per subject.
- How tightly does your data and model allow you to constrain your parameter estimates?

Approach IIb: uncertainty quantification

- **Nonparametric Bootstrap**. Resample your data and see how much your parameter estimates vary within subject. *Requires IID assumption on data!*
- **Bayesian inference**. Assume a *prior distribution* over parameters, and condition on your data to get a *posterior distribution*.
- **Variational Bayes**. Also assumes a prior distribution, but uses optimization to obtain an *approximation* to the Bayesian posterior.

Nonparametric Bootstrap

This is a great method, but I will skip it because we cannot usually apply it in our non-i.i.d. settings. (But see the **Resources** at the end)

Link to an image: https://towardsdatascience.com/an-introduction-to-the-bootstrap-method-58bcb51b4d60

Bayesian inference

Demonstration of prior vs. posterior distributions in RunDEMC for MMI data.

Variational Bayes

Demonstration in Pyro for MMI data.

Resources

Curve-fitting

- Textbook: *Introduction to Statistical Learning* by James et al. An accessible introduction to the principles behind machine learning. http://faculty.marshall.usc.edu/gareth-james/ISL/
- Textbook: *The Elements of Statistical Learning* by Hastie et al. A graduate-level text on machine learning. Particularly good for understanding regularization. Free PDF online. https://web.stanford.edu/~hastie/ElemStatLearn/

Maximum likelihood

- "Probability concepts explained: Maximum likelihood estimation" by Jonny Brooks-Bartlett. A very gentle introduction to maximum likelihood. https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1

Bayesian methods

- "Probability concepts explained: Bayesian inference for parameter estimation." by Jonny Brooks-Bartlett. Explains Bayesian inference starting with coin-flipping examples and Bayes rule. https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-estimation-90e8930e5348
- "MCMC for dummies" by Thomas Wiecki. Explains how to sample posterior distributions using Markov Chain Monte Carlo. https://twiecki.io/blog/2015/11/10/mcmc-sampling/

Nonparametric bootstrap

- "Introduction to Bootstrapping in Statistics with an Example" by Jim Frost. This is a very applied intro to bootstrapping that runs the risk of oversimplifying, but it is good for a first intro. https://statisticsbyjim.com/hypothesis-testing/bootstrapping/
- "Bootstrap Confidence Intervals" by DiCiccio and Efron. If you want to implement bootstrap confidence intervals, this paper is a must-read. https://projecteuclid.org/euclid.ss/1032280214

Variational Bayes

- "Variational Inference: A Review for Statisticians" by Blei et al. https://arxiv.org/pdf/1601.00670.pdf
- "A Beginner's Guide to Variational Methods: Mean-Field Approximation" by Erik Jang. Incomplete but has some good insights. https://blog.evjang.com/2016/08/variational-bayes.html