PSY-3102-2 Modelling Challenge

(Submission due April 30th)

For this assignment, you are provided with a rich behavioral dataset that captures many aspects of cognition ranging from reinforcement learning, change point detection, perception under uncertainty, decision making under uncertainty, recognition memory, etc. You will work in teams of 2-4 students to come up with your own models of one or more of these cognitive processes. You should follow the process we learnt about in the initial half of the course (described in detail in Wilson & Collins 2019) to build, evaluate and fit your models to data. You are free to do as much or as little as you want – but at the very minimum, you should formulate and implement one model and use it to say something meaningful about the cognitive process you are studying.

Submission:

Your submission package should include three things:

1. A likelihood function for each model that you are submitting. As a reminder, the likelihood function should take data from one subject as input, and provide the negative log likelihood for the model. Each model will have its own likelihood function.
2. A document briefly describing your models and your methods. This will typically contain any equations, and a sketch of the simulations, model fits, model comparisons or model validations you carried out.
3. A recorded 10 to 15 min video presentation of your models, methods, results and conclusions. I'm okay with a longer presentation if you are submitting multiple models.

Please submit your package online via the classroom and let me know by email as soon as you submit.

Grading: Your grade will be based on the understanding that you demonstrate of the concepts and skills taught in this course, and, in particular, the quality of the process that you employ in building, evaluating and using your models. To get the highest grade, you would need to: conduct a model-free analysis of the data, come up with at least two models (okay to use pre-existing models from the literature), simulate them to understand the influence of parameters on behavior, carry out parameter recovery simulations, carry out model recovery simulations, fit the data to your model, carry out model comparison, validate the winning model with a posterior predictive check, then interpret the parameter estimates, and make a competent presentation of your methods and results. To get the minimum grade, you need to build and implement at least one model (even if its one of the models from Assignment 1), fit it to the real data, interpret the parameter estimates, and present your methods and results competently. This can be a lot of work, so typically I would expect you to spend some time working on this for some time every day until the deadline.

Meetings: Each group should plan to meet with me once every week until the deadline. Please find a slot and make an appointment. Please also use Ritu's office hours to get help with programming etc.

Advice: Nobody should feel like they don't know what to do. At the very minimum, you can take some of the models you worked on for Assignment 1 and adapt them to model this dataset. That is fine. It likely won't get you the highest grade, but if you did everything required in Assignment 1 and 2 for this dataset, you would get a very healthy grade. Tailor your ambitions and your grade expectations to the resources and attention you are willing to give the assignment.

Dataset: (courtesy of Serre lab at Brown University)

You will find the data in a single zipped file called ChallengeData.zip. When you unzip it, you should have a list of 50 subs for both data phases, indexed by phase ('learn' vs 'memory') and by subject number.

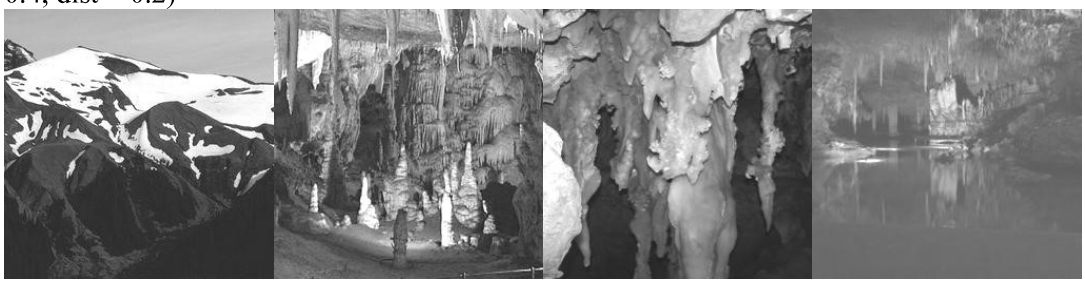
The experiment is comprised of two tasks: a betting task (the learning phase - represented in the data as learnSub[n].mat) and a recognition memory task (the memory ophase - represented in the data as memorySub[n].mat).

Betting task:

In the learning phase, the subjects have 160 trials during which they see the "stakes" (i.e. reward magnitude: how much reward could they win if they take the bet), then they see a stimulus that belongs to either category 1 (manmade) or category 2 (natural).

Example stimuli: four images in category 1, increasingly closer to decision boundary (i.e. dist = 0.8 dist = 0.6, dist = 0.4, dist = 0.2)

Example stimuli: four images in category 2, increasingly closer to boundary (dist = 0.8 dist = 0.6, dist = 0.4, dist = 0.2)



The stimuli are coded in the data as a 4-digit number (e.g 1002, 2043 etc.). The category and distance from decision boundary can be extracted from this number as follows:

- The first digit is always the category. All the 1000s are man-made, all the 2000s are natural.
- The last two digits code boundary distance:
    - 00-20: distance = 0.8 (easiest to classify)
    - 21-40: distance = 0.6
    - 41-60: distance = 0.4
    - 61-80: distance = 0.2 (hardest to classify)

Once they see the stimulus, subjects can decide whether or not to take the bet. The time they take to decide is the reaction time. (If they take longer than 3.5s the trial cuts off and the RT gets a value of -1) The RT is coded in the data in milliseconds (so an RT of 2046 means they took 2.046s).

After taking the bet, the subjects see whether they won or not. Then they see a screen with their total score so far. Then the next trial begins, showing them the new "stakes" etc.

The data structure in the betting task (learnSub.mat) is as follows:

- Column 1: trial index
- Column 2: reaction time (RT – milliseconds)
- Column 3: stimulus they saw (4-digit number that codes category and boundary distance)
- Column 4: take bet/ don't take bet (1/0)
- Column 5: win or lose (1/0)
- Column 6: stakes, or reward magnitude (1,5,20,100)
- Column 7: reward probability associated with category of seen stimulus (0.8 or 0.2, with change points)
- Column 8: current running score

Recognition Memory task:

In the memory phase, subjects see all 160 stimuli from phase 1, plus 160 foils (i.e. novel images they had not seen before) in the same category/distance boundary. They are asked to say whether or not this is a known stimulus (0/1 response), and how confident they are (on a 4-point scale, 1-4).

The data structure from the recognition memory task (memorySub.mat) is as follows:

- Column 1: trial index
- Column 2: reaction time (RT- milliseconds)
- Column 3: distance from boundary of stimulus they're being tested on
- Column 4: two-digit code for stimulus category, and whether stim is familiar or a foil
    - 11: man-made familiar
    - 12: man-made foil
    - 21: natural familiar
    - 22: natural foil
- Column 5: subject's answer
    - On odd-number trials, this is a 0/1 coding the yes/no for recognition
    - On even-number trials, this is a 1-4 coding confidence for the previous response
    - If you are having trouble parsing this into recognition/confidence, please talk to me or Ritu.