University of Amsterdam
Psychology

Graduate School of Psychology

Form 2A - Research Master's Psychology: Thesis Research Proposal

1. General Information

1.1 Student information

Student name: David Coba

Student Id card number: 12439665

Address: -

Postal code and residence: -

Telephone number: -

Email address: coba@cobac.eu

Major: Psychological methods

1.2 Supervisor information

Supervisor name: Maarten Marsman

Second assesor name: Jonas Haslbeck

Specialization: Psychological Methods

1.3 Other information

Date: 1.04.2022

Status: First draft

Number of ECs for the thesis: 32EC

Ethics Review Board (ERB) code: -

2. Title and Summary of the Research Project

2.1 Title: Occam's window something something

2.2 Summary of proposal

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Word count: /150

3. Project description

- Introductory paragraph with goals
  - Check if Occam's window works / is useful in general terms.
  - Benchmark how it performs under different conditions against different alternatives.
  - Simple models + graphical models
  - Contribute software

3.1 Prior research

Single model inference vs Multiple model

- Single model bad, ignores uncertainty of the model selection process

Big encompassing model vs combination of multiple models

- Computational feasibility of sampling across sub-models
  - e.g. RJMCMC, MC^3

  - Stability issues, with either similar models and/or not massive samples

- Combination of multiple models allows to separate the two steps, model search + combination

$$p(\Delta|D) = \sum_k p(\Delta|\mathcal{M}_k, D) w_k, \ \exists k : M_k \in \mathcal{A}$$

- $\mathcal{A}$ is the set of considered models

Occam's window as a model search algorithm before combination of multiple models

- Depends on marginal likelihoods
- Explain algorithm
  - Only models that predict reasonably (relatively) well + Occam's razor
  - Set of considered models
    * They use $p(M|D)$ instead of just the marginal likelihood $p(D|M)$ to incorporate prior information about how likely each model is

$$\mathcal{A}' = \left\{ M_k : \frac{\max\{p(M_l|D)\}}{p(M_k|D)} \leq c \right\}$$

- They also discard models that have submodels which have higher posterior probability

$$\mathcal{B} = \left\{ M_k : \exists M_l \in \mathcal{A}', M_l \subset M_k, \frac{p(M_l|D)}{p(M_k|D)} > 1 \right\}$$

- The set of considered models is $\mathcal{A} = \mathcal{A}' \setminus \mathcal{B}$

- Greedy search with posterior model probabilities
- $M_0 \subset M_1$, they differ by only 1 edge
- If M_0 is rejected, all submodels of M_0 are also rejected
  - A model is submodel of another if all the edges in the first one are included in the second one
- Full algorithm as an appendix

BMA vs stacking

- Two methods of Bayesian model combination
- BMA uses the posterior probability of models as weights, dependent on the marginal likelihoods / BFs

$$p(\Delta|D) = \sum_{k=1}^{K} p(\Delta|\mathcal{M}_k, D) p(\mathcal{M}_k|D)$$

$$p(\mathcal{M}_k|D) = \frac{p(D|\mathcal{M}_k)p(\mathcal{M}_k)}{\sum_{l=1}^{K} p(D|\mathcal{M}_l)p(\mathcal{M}_l)}$$

$$p(D|\mathcal{M}_k) = \int p(D|\theta_k, \mathcal{M}_k)p(\theta_k|\mathcal{M}_k)d\theta_k$$

- Stacking minimizes an utility function (LOOCV) to assign weights to models
  - Common for point predictions, yao2018bayesianstacking extends it to whole posterior distributions

  - Fancy way that reuses sampling draws and includes uncertainty about LOOCV estimates
  - The main practical difference is that if the data-generating model is not in $\mathcal{A}$, BMA will select the single model that minimizes the KL divergence while stacking will select the combination of models that minimize the (log posterior predictive) loss.
    - Different asymptotic behavior
    - Occam's window was conceived with BMA as the combination step
      - You need BFs for both
      - Stacking estimation of weights re-uses model estimation samples
  - Same discussion as BF/marginal approach vs posterior predictive based approaches
    - BFs untrained models vs ppd-based trained models

- Mention the pseudo-BMA approximations that use estimations of the posterior model probabilities
- Full comparison between models is out of the scope of the proposal, rooted on differences in philosophical positions and scientific goals.
- In this case our ultimate scientific goals are about the conditional dependencies structures in the data, inclusion/exclusion which edges
- BMA more sensible to the models that are considered than stacking
- No-one believs that a GGM or an ISING model are the data generating process
- We are going to make trade-offs during the model search phase between computational feasibility and exactness
- Stacking more robust option for model combination (?)
  - Although posterior distribution of parameters might be wonky, we were planing on using the sum of weights (posterior model probabilities in BMA) of the models that include a particular parameter

Other alternatives in the literature

- BAS
  - Sample without replacement from the space of models
  - Choose an initial approximation for the marginals
  - Update those approximations with the actual marginal likelihoods of the already-sampled models
  - Still require analytical calculations of marginals
- BDgraph
  - For graphical models,
    - Pseudolikelihood
  - Birth-death MJMJ as an alternative to RJMCMC, the sampler explores the space of models
    - Higher acceptance rates, poisson/exponential modeling
  - Fast analytical approximations to avoid having to sample from the distributions of parameters
    - Trade-offs

Approximations to the marginal likelihood during model search

- BIC
  - Can be approximated with the BIC if we assume an unit information prior
    - It could be more conservative and favor simpler models
    - The log-marginal likelihood can be approximated as $\log f(x|M_i) \approx \mathcal{L}(\widehat{\theta}) - \frac{1}{2}\text{dim}(\theta) \log n$
    - $2 \log B_{ij} = -\text{BIC}(M_i) + \text{BIC}(M_j)$

3.2 Key questions

- Same as in the introduction.

Word count: /1200

4. Procedure

4.1 Operationalization

4.2 Sample characteristics

4.4 Data analysis

4.4 Modifiability of procedure

Word count: /1000

5. Intended results

Word count: /250

6. Work plan

6.1 Time schedule

6.2 Infrastructure

6.3 Data storage

6.3 Budget

Word count: /500

8. Further steps

Make sure your supervisor submits an Ethics Checklist for your intended research to the Ethics Review Board of the Department of Psychology at https://www.lab.uva.nl/lab/ethics/

7. Signatures

- □ I hereby declare that both this proposal, and its resulting thesis, will only contain original material and is free of plagiarism (cf. Teaching and Examination Regulation in the research master's course catalogue).
- □ I hereby declare that the result section of the thesis will consist of two subsections, one entitled "confirmatory analyses" and one entitled "exploratory

analyses” (one of the two subsections may be empty):

1. The confirmatory analysis section reports exactly the analyses proposed in Section 4 of this proposal.
2. The exploratory analysis section contains not previously specified, and thus exploratory, proposal analyses.

Location:	Student’s signature:	Supervisor’s signature:
Amsterdam		