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: 25.04.2022

Status: Revised version

Number of ECs for the thesis: 32EC

Ethics Review Board (ERB) code: -

2. Title and Summary of the Research Project

2.1 Title: Assessing the performance of Occam's window for Bayesian model averaging

2.2 Summary of proposal

When we select a statistical model and use it to make inferences about its parameters, we usually ignore the uncertainty derived from the model selection process, leading to overconfident inferences. There are techniques that address this, like Bayesian model averaging. However, when the space of possible models is vast, such as with graphical models that are popular in psychology, it is not evident how to efficiently find the most relevant ones. Occam's window is a model search algorithm that explores the space of possible models.

The goal of this project is to assess in broad terms if Occam's window is a suitable method to explore the model space, specifically in the context of graphical models. To this end we will develop an Occam's window implementation, and conduct a simulation study exploring how the algorithm performs under different conditions and how it compares to other alternative model search techniques.

Keywords: Bayesian inference, Bayesian model averaging, model selection, model search algorithms, Occam's window

Word count: 148/150

3. Project description

3.1 Prior research

When we perform statistical inferences, such as hypotheses tests about the inclusion of a parameter in a model or whether a parameter lays within an interval, we typically select a statistical model and then use that model to perform the inference. However, this single-model approach underestimates the total uncertainty in our inferences, since it ignores the uncertainty derived from the model selection process. And, ignoring this uncertainty, leads to overconfident conclusions (Leamer, 1978; Draper et al., 1987; Hoeting et al., 1999; for a recent review of the issue see Kaplan, 2021). The aim of this project in general terms is to explore whether an algorithm called Occam's window can be useful to deal the issue of single-model inference. Specifically, we are motivated by the issue of deciding whether to include or not particular edges in graphical models that are popular in psychology. The number of possible graphical models grows exponentially with the number of variables, and current approaches to multi-model inference struggle because of the size of the model space.

Different Bayesian solutions have been proposed that allow us to model the uncertainty of the model selection process. These approaches can be categorized into two groups. The first group is using mixture models that encompass all possible models. To estimate the joint posterior distribution of all possible models researchers usually employ simulation based methods like Markov chain Monte Carlo model composition (MC^3, Madigan & York, 1995) or reversible jump Markov chain Monte Carlo (Green, 1995). However, it is often impossible to implement simulation based methods that produce good results in realistic time frames, and they tend to have stability issues (Yao et al., 2018). The second group of approaches to multi-model inference is to only combine the information from a set of candidate models $\mathcal{A}$, instead of using the whole model space. To combine a set of candidate models we can use Bayesian model averaging (BMA, Hinne et al., 2020; Hoeting et al., 1999; Leamer, 1978). This approach allows to separate the use of multiple models into two steps: identifying a set of candidate models $\mathcal{A}$ and then combining the uncertainty from those models. With BMA, the posterior probability of our target inference (e.g. whether a parameter is included in the model or not) given the observed data, $p(\Delta|D)$, is the weighted average of that inference across all candidate models $p(\Delta|M_k, D),\ M_k \in \mathcal{A}$. BMA uses the posterior probability of candidate models $p(M_k|D)$ as model weights, and our target inference $p(\Delta|D)$ becomes

$$p(\Delta|D) = \sum_{\forall M_k \in \mathcal{A}} p(\Delta|\mathcal{M}_k, D)p(M_k|D).$$

From Bayes theorem we know that the posterior probability of a model is the product of the prior probability of that model $p(M_k)$ times the marginal likelihood of the data under that model $p(D|M_k)$, divided by the sum of that same product for all candidate models

$$p(M_k|D) = \frac{p(D|\mathcal{M}_k)p(M_k)}{\sum_{\forall M_l \in \mathcal{A}} p(D|M_l)p(M_l)}.$$

Lastly, to calculate the marginal likelihood we need to integrate the product of the likelihood function of each model $p(D|\theta_k, M_k)$ and the prior distribution of the model parameters

$p(\theta_k|M_k)$ over the whole parameter space

$$p(D|M_k) = \int p(D|\theta_k, M_k) p(\theta_k|M_k) d\theta_k.$$

This is often not possible to do analytically, and we expand more about different ways of approximating the marginal likelihood later in this section.

When we do not have strong theoretical arguments to pre-select a set of candidate models $\mathcal{A}$ to average with BMA, we can use model search algorithms. One possible algorithm is the topic of this thesis: Occam's window (Madigan & Raftery, 1994; Raftery et al., 1997), which is based on Occam's razor principle. Occam's razor (also known as the law of parsimony) states than when one is presented with competing hypotheses that explain equally well a particular phenomena, one should choose the simplest one. In general terms, Occam's window algorithm first selects a set of models that fit the data reasonably well, and then discards all models that have simpler counterparts that fit the data equally well. The final result is the set of simplest models that explain the data well.

Formally, the first step equals constructing the set of models

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

with posterior probabilities $p(M_k|D)$ not significantly lower than the model with highest posterior probability of all models $M_l \in \mathcal{A}'$. The constant c specifies the range of posterior probabilities that are acceptable, the size of the window of models that fit well enough. For the second step the algorithm identifies the set of models

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

that have at least one submodel M_l in $\mathcal{A}'$ with greater posterior probability. The final set of candidate models is the set of models in the first set that are not present in the second $\mathcal{A} = \mathcal{A}' \setminus \mathcal{B}$. Computationally, the algorithm is a deterministic greedy search over the model space, but we are omitting the computational details from this document. To calculate posterior model probabilities $p(M_k|D)$ we need to compute the marginal likelihood $p(D|M_k)$ of each model, similarly to BMA. However, in most cases it is not possible to calculate marginal likelihoods analytically, and we require of approximate solutions.

Since Occam's window uses marginal likelihoods to compare models many times during the model search, we need efficient ways of approximating them. The first and crudest approximation is to use the Bayesian information criterion (BIC, Schwarz, 1978; Kass & Raftery, 1995). The BIC of a model M_k is defined as

$$\text{BIC}(M_k) = -2 \log p \left(D | \widehat{\theta}, M_k \right) + d_{M_k} \log n,$$

where $p \left(D | \widehat{\theta}, M_k \right)$ is the likelihood function evaluated at the maximum likelihood estimates of the model's parameters, d_{Mi} is the number of parameters in the model and n is the sample size. Kass and Raftery (1995) show that the logarithm of the marginal likelihood of a model can be approximated as

$$\log p \left(D | M_k \right) \approx \log p \left(D | \widehat{\theta}, M_k \right) - \frac{1}{2} d_{M_k} \log n,$$

which means that

$$\log p \left(D | M_k \right) \approx \frac{\text{BIC}(M_k)}{-2}$$

and that the ratio of marginal likelihoods between two models—the Bayes factor—is

$$2 \log B_{ij} = -\text{BIC}(M_i) + \text{BIC}(M_j).$$

Bridge sampling offers another approach to approximate the marginal likelihood (Bennett, 1976; Gronau et al., 2017). Bridge sampling generally provides accurate approximations of the marginal likelihoods, but is also very computationally demanding and not usable with a model search algorithm, because it is a simulation based method and has to draw samples. A method between BIC and bridge sampling in terms of accuracy and computational demands is the Laplace approximation (Kass & Raftery, 1995; LeCam, 1953). This method approximates the posterior distribution with a normal distribution centered around the posterior mode, which can be estimated using expectation-maximization algorithms. The standard Laplace approximation is accurate to the second moment of the posterior distribution, but it is possible to extend it to get more accurate approximations at the cost of more computational resources or further assumptions (Hubin & Storvik, 2016; Rue et al., 2009; Ruli et al., 2016; Tierney & Kadane, 1986; Tierney et al., 1989). Lastly, note that in the context of Occam's window and BMA, it is possible to use a faster but less accurate approximation during model search, and use a slower but more accurate approximation during the BMA step.

One of the drawbacks of Occam's window is that it overestimates the posterior probability of the selected "best" candidate models and it underestimates —essentially nullifies—the posterior probability of the rest of the models. This is by design and acknowledged by Madigan and Raftery (1994), and it is a trade-off we have to make to avoid having to combine information from the complete model space. Occam's window is implemented for linear regression models using priors that allow to analytically calculate the marginal likelihoods (Raftery et al., 1997) in the R package BMA (Raftery et al., 2015). There is also an extension of Occam's window to allows to model streams of data that become available sequentially (Onorante & Raftery, 2016).

The most common alternative model search algorithms to Occam's window, in a Bayesian framework, are Bayesian adaptive sampling (BAS) and birth-death Markov chain Monte Carlo (BDMCMC). BAS samples without replacement from the space of possible models, and uses the marginal likelihoods of the sampled models to iteratively estimate the marginal likelihoods of the models that remain unsampled (Clyde et al., 2011). BAS is available for (generalized) linear models as an R package (Clyde, 2021). BDMCMC (Mohammadi & Wit, 2015; Mohammadi et al., 2017) samples from the joint posterior space of all possible models, and uses a Poisson process to model the rate at which the Markov chains jump from one model to another. BDMCMC is available in the R package BDGraph (Mohammadi & Wit, 2019) for graphical models. However, BDMCMC shares the same limitations as other simulation based methods for graphical models: it can have stability issues and it is prohibitively slow to use in most cases.

3.2 Key questions

The goals of this project are to develop an efficient Occam's window implementation for graphical models that are popular in psychological research, like the Gaussian graphical model (GGM) and the Ising model, and benchmark its performance. We want to know whether it can produce results that are good enough to be used, while also being able to run in an adequate time frame.

To this end we will first implement Occam's window algorithm for simpler models, such as linear regression and logistic regression, and then for graphical models. This will allow us to test the model search algorithm without having to deal with the extra complexity of graphical models. Later, we will explore with a simulation study the possible trade-offs between accuracy and computational speed of Occam's window versus alternative model search algorithms, and also how different marginal likelihood approximations impact the trade-offs.

Word count: 1424/1200

4. Procedure

4.1 Operationalization

To address our research questions we will first implement Occam's window model search algorithm in steps, and then conduct a simulation study. We plan on implementing our algorithm and running our simulations in the Julia programming language (Bezanson et al., 2017). There are more simulation conditions that are potentially interesting than how many we can realistically tackle during this project, and the number of conditions that we can test will depend on how smoothly the project progresses.

Regarding which models to use during our simulations, linear regression is the obvious simplest choice to start developing the algorithm. Logistic regression is a next step that increases the complexity of the procedure, and the GGM and the Ising model are the ones that motivate this project. First, we will implement Occam's window algorithm using the BIC approximation for the marginal likelihood, since it is the simplest method and it will allow us to test our implementation while developing it. Next, for linear regression models and the

GGM there are convenient prior distributions for the model parameters that allow to calculate the marginal likelihoods analytically. Finally, for the logistic and Ising models we will have to implement Laplace approximations of the marginal likelihoods. We plan to test all the marginal approximations we are implementing in the simulation study.

We will rely on the R implementations of BAS for linear models and BDgraph for graphical models as benchmarks.

4.2 Sample characteristics

We plan on generating data from a set of models and evaluating how well each simulation condition recovers the characteristics of the true data-generating models. In general terms, we will consider conditions with different sample sizes and sparsity levels in the covariance matrices of the data-generating models. However, we do not think it makes sense to commit to specific data-generating processes at this stage of the project.

4.4 Data analysis

This project is inherently exploratory and, similarly to the last section, we do not think it makes sense to commit at this stage to a specific analysis plan. In general terms, to assess how well each model search algorithm performs we will use BMA to calculate the total posterior probabilities of including specific edges that are (or not) present in the data-generating model. The total posterior probability of including (or not) a specific edge in a model is the sum of the posterior probabilities of all candidate models that contain (or not) that edge. Then, we can use a threshold (most likely it will be just $p(\Delta|D) = 0.5$) to consider if the procedure considers that an edge is present or not, and analyze in terms of sensitivity and specificity the results. This can potentially be extended and consider the area under the curve (AUC) in a plot of sensitivity against specificity for different threshold values. To assess computational costs we will use real runtime in order to not penalize algorithms that benefit from parallel computations. If instead we used CPU time, we would be penalizing all parallelizable algorithms by a factor of the number of parallel processes or threads.

4.4 Modifiability of procedure

The scope of this project is highly flexible, and we can adapt which conditions to include or exclude in our simulation study depending on how fast we progress. In section 6.1 ”Time schedule” we detail the milestones we aim to complete before certain deadlines.

Word count: 543/1000

5. Intended results

The main goal of this project is to assess in general terms how Occam’s window performs. The main limitation of current methods in the context of graphical models, like BDMCMC from BDGraph, is that they are prohibitively slow. We anticipate that Occam’s window will produce results faster, and we think that it can be a useful tool that is currently underused. If our analysis concludes that the results Occam’s window are good enough in terms of sensitivity and specificity, while also being significantly faster than the alternatives, we will show that the algorithm can be a useful tool to supplement the use of BMA to avoid the problem of single-model inference. In case that our results show that the performance of Occam’s window does not compensate for its shortcomings, we would have provided an updated assessment of its performance that is currently lacking in the literature. To our knowledge there are no simulation studies evaluating how Occam’s window performs under different conditions, or how it compares to other model search algorithms.

Moreover, we expect to contribute software that implements BMA and Occam's window, and that integrates with the rest of the Julia ecosystem.

Word count: 197/250

6. Work plan

6.1 Time schedule

This thesis project consists of 28 EC, excluding the thesis proposal. This is equivalent to approximately 18 weeks working full time. We aim to complete and present the project by the 15th of July 2022. In broad terms we plan to achieve the following milestones each month:

April
- Week 1-3: Address feedback on the proposal and implement Occam's window algorithm for linear regression models using BIC as an approximation to the marginal likelihood.
- Week 4: Implement analytical evaluations of the marginal likelihood for linear regression models.

May
- Week 1: Buffer time and hopefully enjoy the UvA teaching-free days.
- Week 2: Implement analytical evaluations of the marginal likelihood for Gaussian graphical models.
- Week 3: Buffer time and start running simulations, including with BAS and BDGraph.
- Week 4: Continue running simulations and implement the Laplace approximation for logistic regression models.

June

- Week 1: Continue running simulations and implement the Laplace approximation for Ising models.
- Week 2: Continue running simulations and start analyzing results. Start writing the thesis.
- Week 3/4: Analyze results and thesis writing. Complete a first draft of the full thesis.

July

- Weeks 1/2: Complete writing the thesis and prepare the presentation.

The scope of this project is highly flexible, and we can adapt which conditions to include or exclude in our simulation study depending on how fast we progress.

6.2 Infrastructure

No special infrastructure is required to complete this project.

6.3 Data storage

We will keep the results of all our simulations under version control and with remote backups. We do not plan on collecting any data, and in the case we end up deciding to use empirical data we would use publicly available datasets.

6.3 Budget

In principle we will not require extra funds to complete this project. In the case that the computational resources that we have access to prove insufficient to conduct the simulations, we might consider using cloud computing services. In any case, such costs would not exceed the maximum budget.

Word count: 324/500

7. References

Bennett, C. H. (1976). Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, *22*(2), 245–268. https://doi.org/10.1016/0021-9991(76)90078-4

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Clyde, M. A. (2021). *BAS: Bayesian Variable Selection and Model Averaging using Bayesian Adaptive Sampling*. https://cran.r-project.org/package=BAS

Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, *20*(1), 80–101. https://doi.org/10.1198/jcgs.2010.09049

Draper, D., Hodges, J. S., Leamer, E. E., Morris, C. N., & Rubin, D. B. (1987). A research agenda for assessment and propagation of model uncertainty. *Rand Corporation, Report N-2683-RC*.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711–732. https://doi.org/10.1093/biomet/82.4.711

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*(nil), 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200–215. https://doi.org/10.1177/2515245919898657

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial [with comments by M. Clyde, David Draper and EI George, and a rejoinder by the authors]. *Statistical Science*, *14*(4), 382–417. https://doi.org/10.1214/ss/1009212519

Hubin, A., & Storvik, G. (2016). Estimating the marginal likelihood with integrated nested Laplace approximation (INLA). http://arxiv.org/abs/1611.01450v1

Kaplan, D. (2021). On the quantification of model uncertainty: A Bayesian perspective. *Psychometrika*, *86*(1), 215–238. https://doi.org/10.1007/s11336-021-09754-5

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*. Wiley.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publication in Statististics*, *1*(11), 277–330.

Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, *89*(428), 1535–1546. https://doi.org/10.1080/01621459.1994.10476894

Madigan, D., & York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, *63*(2), 215–232. https://doi.org/10.2307/1403615

Mohammadi, A. R., & Wit, E. C. (2015). Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Analysis*, *10*(1), 109–138. https://doi.org/10.1214/14-BA889

Mohammadi, A. R., Massam, H., & Letac, G. (2017). Accelerating Bayesian structure learning in sparse Gaussian graphical models. http://arxiv.org/abs/1706.04416v3

Mohammadi, A. R., & Wit, E. C. (2019). BDgraph: An R package for Bayesian structure learning in graphical models. *Journal of Statistical Software*, *89*(3), 1–30. https://doi.org/10.18637/jss.v089.i03

Onorante, L., & Raftery, A. E. (2016). Dynamic model averaging in large model spaces using dynamic Occam s window. *European Economic Review*, *81*, 2–14. https://doi.org/https://doi.org/10.1016/j.euroecorev.2015.07.013

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191. https://doi.org/10.1080/01621459.1997.10473615

Raftery, A. E., Hoeting, J., Volinsky, C., Painter, I., & Yeung, K. Y. Y. (2015). *BMA: Bayesian model averaging*. https://cran.r-project.org/package=BMA

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Ruli, E., Sartori, N., & Ventura, L. (2016). Improved Laplace approximation for marginal likelihoods. *Electronic Journal of Statistics*, *10*(2), 3986–4009. https://doi.org/10.1214/16-EJS1218

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464. http://www.jstor.org/stable/2958889

Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, *81*(393), 82–86.

Tierney, L., Kass, R. E., & Kadane, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, *84*(407), 710–716. https://doi.org/10.1080/01621459.1986.10478240

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions [with Discussion]. *Bayesian Analysis*, *13*(3), 917–1007. https://doi.org/10.1214/17-BA1091

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		