Dear Julia (and Jonas),

Thank you very much for your feedback, it was really useful.

I have edited my proposal based on your comments. In this letter I will address all points raised in both reviews, which also cover the two main points from the revision letter.

1-3.1 Prior research

Julia

- My only quarrel is that a lot of space went to concepts that where not necessarily needed to understand the problem, and therefore other aspects were not sufficiently explained. For example, I was waiting for model stacking to be relevant, but I think this entire paragraph was essentially irrelevant to understand the project.

The paragraph on model stacking was a remnant from a previous version of the proposal were model stacking, and how it compares to BMA, occupied half of the total word count. I felt that, even after cutting stacking from the project, it was disingenuous to omit the whole debate and criticism of BMA in a project that heavily relies on BMA. But, now I agree that it feels out of place in the current version of the proposal and that it impedes the reading flow. I have deleted it.

I do not think there are any other concepts that are irrelevant for this project, but I have tried to explain more clearly the relevant concepts in the new version of the proposal.

- On the other hand, BDMCMC and BAS are only mentioned in the methods section, even though it makes sense to explain them in more detail if they are used as benchmark.

I agree that it makes more sense to explain the alternative methods in the *Prior Research* section. I have also expanded on the limitations of BDMCMC and why it is not currently a satisfactory approach to model selection in the context of graphical models. However, I do not think that it is worth explaining them in more detail in the proposal. They are fairly complex methods and very different from Occam's window, so it would take a significant amount of space to be more detailed. Moreover, we have selected them because they are the most used Bayesian model search algorithms and I am going to treat their R and C++ implementation as *black-boxes* to benchmark against. I am implementing Occam's window in Julia, so our comparisons are not going to be very fair and I am only using them to have an external general point of reference to compare against. I think that the way they are currently presented is enough to get a general idea of how they work and how they differ from Occam's window.

Jonas

- In the summary you say you study Occam's window in general and specifically in the context of graphical models which sounds like a contradiction.

I have clarified what I meant. We want to asses it's performance in general/broad terms (i.e. if it is a somewhat usable method or if it just does not work), but with the specific motivation of seeing if it works for graphical models.

- If possible, already give some intuition for how Occam's window works.

The summary is already at the maximum word count, and I think that the information it presents now is more important than an intuition about how the algorithm works.

- In 3.1 you say that estimating a single model "essentially ignores the uncertainty …" why essentially? I think this statement could be cleared up.

I agree that "essentially" was an unnecessary word and that the sentence is clearer without it.

- In 3.1 "can help with the issue of single-model inference" be more specific; you already explained what the problem is.

I think the sentence is a good compromise to introduce what the main goal of the project is, without having explained yet what BMA or model-search algorithms do.

- Section 3.1 starts with the problem, but I think a paragraph before on the background is needed (what are graphical models, what are the general issues one has to tackle to estimate them)

I have made a bit more explicit what the main limitation with graphical models we are addressing is (i.e. calculating how likely it is that a node is present or *real*). However, I do not think that it is necessary to explain what graphical models are in detail, in the same manner that I have not explained the fundamentals of Bayesian inference or model search algorithms.

- Why is it unclear how to implement the first group of methods?

Because it is really hard, have stability issues and they tend to not work in realistic time frames. I have clarified this in the proposal.

- In 3.1: "weighted average" what does this refer to? That the models not considered are weighted to zero?

It means that the contribution of each model to the final prediction is not equal, and that they are weighted based on their posterior probability. I hope that after not mentioning stacking at

all and talking directly about BMA instead of "model combination" procedures it is clearer to understand.

- I was a bit surprised by the paragraph "one method to combine …"; is this still about the second type of method? Or is this a general method? If the latter is the case, maybe this material should go first

Similarly to the last point, I hope that in the new version the transitions between topics are easier to follow.

- Misunderstandings about how Occam's window work

  – "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"; shouldn't we be selecting those models that fit equally well but are simpler based on the principle of parsimony?

    Yes, and that is what we are doing: {Models that fit well} \ {Models that have simpler counterparts that fit well} = {Those simpler models that fit well}.

  – I don't understand how the set A' is constructed; it's definition seems to depend on Ml, which is itself a member of A'.

    Yes, it does. The definitions in the proposal are formal definitions of what the algorithm does, and should not be interpreted as computational steps. I have made that clearer in the proposal. $\mathcal{A}'$ contains all models that fit the data somewhat as good as the best model of them all.

  – Can you provide a bit more intuition for how big A=A' \ B is? I would imagine one model fits best, which means B contains all models, which means A contains only one model. Clearly, I'm misunderstanding the method, but from the text it is not entirely clear to me.

    $\mathcal{A}'$ can contain multiple models, and $\mathcal{B}$ can only contain models that are in $\mathcal{A}'$. The dimensions of $\mathcal{A}$ depend on multiple factors and I am not sure what sizes are

normal. In any case, it should be significantly less models than what a simulation-based method would explore.

The explanations in the proposal are correct, and the premises of these questions are wrong. However, that is clearly the proposal's (my) fault, and not Jonas'. I have expanded this section with more detailed explanations and hopefully it is easier to understand now.

- When first mentioning the "passes" it's not clear what they refer to, and why we need another pass backwards.

Two passes are necessary because typically the algorithm uses a leaps-and-bounds algorithm to select the initial set of models that it uses for the model search. I have deleted the two sentences mentioning the two passes, since I agree that it is confusing to include only some descriptions of the computational process. I have chosen to omit the computational details from the proposal since they are extensive and unnecessary to understand the goals of the project and our plans. However, for the final version of the thesis document I plan on including a technical appendix where everything is fully specified.

- General comment: A figure / example with 3-4 nodes would really help; I think this would also be nice for the final thesis.

I have removed the mention to computational details and explained in more detail the idea behind Occam's window. I believe that including figures will not be worth for the proposal, but I agree that for the technical appendix in the final document they will be useful.

1-3.2 Research questions and hypotheses

Jonas

- These questions are very vague without the specifications in Section 4; maybe you can address this by giving a bit more information and referring to Section 4; for example, it should be clear in the research question whether the focus is regression or graphical models.

I have clarified this section, but the content remains the same.

2. Procedure

Julia

- I think there are some aspects of the project that could be described better at this point. For example, if I understand correctly, Occam's window is used to generate a set of candidate models. However, for the data analysis you want to compare posterior probabilities of including specific edges. To me, there is a step missing there, namely using BMA for the candidate models to estimate those posterior probabilities.

You are right. Occam's window was conceived as a model search algorithm specifically designed to be used with BMA, but it was not clear from the text that this is our plan. I have clarified it.

- How do you plan to perform the BMA?

During the model search we have calculated the marginal likelihoods of all considered models, and we can use the marginal likelihoods to calculate posterior weights. The simplest option is to assume a uniform prior distribution over the model space. Then, the posterior weights are just proportional to the marginal likelihoods of each model, constrained that all have to add up to 1. Our implementation will allow to specify non-uniform priors, but I have chosen to not include that condition in the simulation study because of feasibility constrains and we are prioritizing testing the other conditions.

- Likewise, once you have obtained the posterior probabilities, how exactly will you assess the success of the different algorithms?

I have added more information about this in the proposal. In summary:

1. Calculate the posterior probability of including each edge.

2. Use a 0.5 decision threshold to decide whether the algorithm says that an edge should (or not) be present.

3. Assess the sensitivity and specificity of the algorithm to find edges from the data-generating model.

- You write that BAS and BDMCMC are benchmarks, so how much better or worse should Occam's window perform to be successful?

BAS and BDMCMC are simulation-based methods and their results will most likely be closer to the real posterior distributions. However, in the context of graphical models BDMCMC is prohibitively slow. Occam's window will be successful if it is significantly faster, and it's results are somewhat usable. I do not think I can come up with a specific value of sensitivity or specificity or AUC that would mean it is *close enough*. Hopefully all this is clearer in the new version of the proposal.

- And, how will you aggregate the algorithms' performances across edges?

I am not certain that I understand this question. If it refereed to how to calculate the probability of including (or not) a specific edge with BMA, it is the sum of posterior probabilities of all models that contain (or not) that edge. I have made this more explicit in the proposal.

Jonas

- Which different marginal likelihood approximations you will consider?

We plan on trying all the approximations that we are implementing, but I agree that it was not obvious from the text that this is the case. I have clarified it.

- How did you select the alternative model selection procedures? Are they the best-performing ones? Or the ones available?

BAS and BDMCMC are the most common Bayesian model search algorithms for linear and graphical models respectively. I have clarified this in the text. I am not aware of any study comparing their performance.

- How will you assess performance?

See the responses above to Julia's comments.

- What are the simulation conditions you will consider?

I mention that we will use different sample sizes and sparsity patterns for the data generating models. I also mention which models we will consider, which model search algorithms we will use for each model and which marginal approximations we will use with Occam's window. I hope than in the revised version it is clear what our general plan is, including the variables that we are going to target when choosing simulation conditions. However, I think that it is not realistic to be more specific than in it's current state, and that if we were more specific we would undoubtedly deviate.

3. Indented results

Jonas

- Are there any theoretical expectations with respect to which method will do better than another? Maybe condition on the characteristics of the data generating model?

Not that we are aware of. The literature lacks a comparison between their performances, and we do not see a priori any theoretical reason to predict under which conditions each model would perform best.

- Do you expect trade-offs in estimation performance vs. computation time in the considered methods? If yes, which?

Yes. BAS and BDMCMC use fast but crude approximations to calculate the marginal likelihoods, while Occam's window can work with multiple approximations that have trade-offs between exactness and efficiency. On the other hand, Occam's window is more greedy when exploring the model space, and BAS and BDMCMC are more thorough. We

expect Occam's window to be significantly faster, and we want to test whether it's results are *good enough* to be used. Hopefully this is clearer in the new version of the proposal.

I would also like to add that so far the project is advancing according to plan, and that therefore I have not modified the project schedule.

Thanks again and have a nice Kingsday/week/weekend!

David Coba,
st. no. 12439665.