Interactive Visualizations Tools for Prior Setting In Bayesian Analysis: Challenges For Evaluation

Abhraneel Sarma
University of Michigan
Ann Arbor, MI 48104, USA
abhsarma@umich.edu

Matthew Kay
University of Michigan
Ann Arbor, MI 48104, USA
mjskay@umich.edu

Abstract
In this position paper, we describe our efforts in creating tools to support the creation of Explorable Multiverse Analysis Reports, and for improving Bayesian analysis using interactive visualizations. We describe the challenges that we faced in evaluating these tools and raise questions that we faced in our efforts to design a study to evaluate these tools. We believe that answering these questions will help improve the design of data analysis tools to better fit analysts' workflow and evaluate their effectiveness.

Author Keywords
Bayesian Data Analysis, Prior setting, Interactive visualization

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
In our recent work, we presented the idea of an *Exploratory Multiverse Analysis Report* (EMAR) to increase the transparency of research papers [3]. In scientific research, authors make several decisions regarding data substitution, processing, modelling and presentation. We argued that all such decisions should be made explicit and presented to the reader, and EMARs, which are interactive research

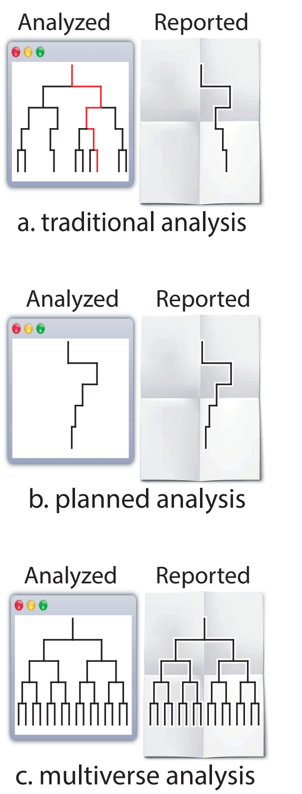

Figure 1: Three reporting strategies: (a) Not reporting the different options that have been tried, or the options that would have been chosen had the data been different are forms of opacity. (b) Pre-registration has been advocated as one possible solution. (c) Multiverse Analysis makes the analysis process completely transparent.

papers, can facilitate this. The reader can use interactions to explore the consequences of different decisions on the robustness of statistical findings.

We created five prototypes to demonstrate different approaches for creating EMARs. These were implemented using HTML and JavaScript, while the analysis, relevant images and CSV files were generated using R. However, computational notebooks, such as R Markdown and Jupyter, which have steadily been gaining popularity, allow creation of interactive statistical reports which can embed HTML and JavaScript, and can also be used to create EMARs.

We currently lack the tools which allow the creation of EMARs, and implementing them can require significant statistical and programming expertise. As we advocate for creating tools to facilitate EMARs, we believe their adoption is only be possible if these tools can be easily and seamlessly integrated into the current data analysis workflow. Since computational notebooks have become exceedingly common for analysis and communication [4], we posit that tools for creating EMARs should be incorporated into them.

Currently, we are creating one such tool — an R package which allows researchers to implement an interactive visualization tool in the R Markdown notebook to show the effect of different prior specifications in Bayesian models. In this paper, we discuss the barriers that exist in evaluating such tools, particularly in ensuring that they fit into analysts' existing workflows and assessing their effectiveness.

Background

Explorable Multiverse Analysis Reports

It has been claimed that interactivity can be used to help readers appreciate the influence of analysis choices on outcomes [1]. In our previous paper, we argued that such interactive learning can help improve the current communication of scientific results through research papers [3], by presenting the results of not just a single analysis, but a multiverse analysis [6, 7]; and by allowing the reader to explore the multiverse by interacting with the document through the *explorable explanations* [8] framework.

A *multiverse analysis* [6, 7] is a philosophy of statistical reporting which calls for making all the choices in data substitution and processing, modelling and presentation explicit; the authors then present the results of all permutations of choices involved in the analysis. This shows how robust or fragile one's statistical results are.

Explorable explanations [8] are narratives which allow the reader to become an "active reader" by interacting with the document and playing "with the author's assumptions and analyses, and see the consequences".

In our work, we explored the design space of EMARs using five examples, and showed how combining those two ideas can complement existing reporting approaches and constitute a step towards more transparent research papers. However, we did not address questions around building tools to support the creation of EMARs.

Tools for supporting creation of EMARs

The biggest barrier that exists towards the adoption of EMARs is perhaps the need for extensive programming and statistical knowledge. Hence, tools which reduce this barrier can greatly benefit the adoption of EMARs. Further, we believe that such tools should be programmable, easily accessible and integrated with the predominant workflow for data analysis and modelling.

We are currently creating an R package which allows users to not only implement one of the examples of EMAR from our paper [3], but we are also hoping this would help ana-

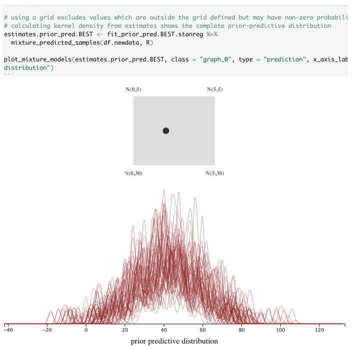

Figure 2: Interactive visualization to evaluate the effect of different priors on the posterior predictive distribution. This can be used within R Studio (shown in picture) or be compiled to a R Markdown document and viewed using a browser. The visualization remains interactive in both cases

lysts in fitting better Bayesian models. In the example, we demonstrated an approach of interactively evaluating the effect of prior distributions on the posterior distribution of the parameter to the reader, after model has been fit. We argued that this would help researchers justify their modelling decisions, and help the reader understand the model better by interacting with the prior.

Translating prior beliefs about the data into probability distributions is an important aspect of performing Bayesian analyses, yet identifying priors can be difficult [5]. We realized that our approach can be useful for tasks such as understanding the implications of the analysts prior beliefs' on the model, or performing sensitivity analysis, and thus help them fit better models. Instead of adopting iterative model fitting and prior evaluation approaches that are currently encouraged [2], our R package provides functions which allow the users to define reasonable limits on priors (only parameterisation using location and scale is currently supported). The functions then defines a small set of reasonable priors along the two parameters, and fits a corresponding number of models. We then allow the user to create an interactive visualization which simulates the prior predictive, posterior, and posterior predictive distributions and evaluate the effect of different priors along the continuous values of the location and scale parameters.

Such tools can be and should be created for other approaches to create EMARs as well. As we advocate for tools which support EMARs and are integrated with computational notebooks, it is important to ensure that they can easily fit into the analysts' existing workflow without making changes.

Discussion

After creating our interactive visualization tool, we needed to conduct an evaluation to understand: (1) how the tool will situate within existing workflows; and (2) how to measure the effectiveness of the tool. Although, we faced these problems in creating a tool for one of the examples, we can expect them to arise in developing any tool for creating EMARs or even, tools for computational notebooks in general. We raise questions to discuss how to overcome these challenges, which we hope will help us adopt a more human-centered approach towards developing and evaluating such tools.

Determining the workflow

We based the design of our package to situate within the workflow proposed by Michael Betancourt [2]. Although the proposed workflow can be considered the "best practice" in performing a Bayesian analysis, it is still prescriptive and perhaps not how most researchers conduct their analysis. A human-centered approach to this problem would have been to study how users specify priors in their analysis.

Moreover, we can expect the workflows of scientists and analysts to vary based on experience. While beginners or intermediates might use guidelines prescribed by popular textbooks for their modelling decisions, experts might make these decisions based on heuristics identified from their experience. Hence, we would like to discuss the best practices for studying users' existing workflows and how any differences in workflows between users can be resolved.

Evaluating the impact on the workflow

In our study design to evaluate the tool, we are adopting the following two approaches: (1) through case studies, we attempt to understand how the tool fits in to the workflow of experts in Bayesian statistics; and (2) we present participant with a published study, and ask them to use the tool to specify prior distributions, to evaluate the effectiveness of the tool in helping people set stronger and more informative priors. There are considerable challenges to conducting an evaluation using these approaches.

To evaluate the effectiveness of such a tool, we need to conduct a quantitative comparison between the priors that users would set using our tool, compared to a control which prescribes participants the best practices to choosing and adjusting priors. The target users for our tool, analysts who use Bayesian statistical methods or are considering adopting Bayesian statistical methods, constitute a specialised group which makes them difficult to recruit. Sampling a large number of such experts can be challenging due to financial constraints as well as availability. Thus, how can we recruit a large number of participants for such a study?

One possible way to overcome this would be to relax the criteria for expert users, and recruit participants who have a certain minimum experience with statistics. Such a sample, however, could exhibit large variations in prior knowledge which might affect the results from such a study. Hence, we ask: How do we sample participants with similar prior knowledge? Alternatively, how do we provide them with the requisite knowledge such that variations resulting from prior knowledge can be reduced.

It is important to understand how the tool would fit into the workflow and to determine how it impacts participants' existing workflows. Hence, we would need to set up an experiment that elicits realistic behavior. How do we conduct such a study? Or, is that even possible? If we cannot observe the realistic use of a new intervention, how would we set up a case study or qualitative study that gives us reliable information about what works and what doesn't?

REFERENCES

1. Christie Aschwanden. 2015. Science is not broken. `https://fivethirtyeight.com/features/science-isnt-broken/`. Accessed: 2018-09-15. (2015).
2. Michael Betancourt. 2018. Towards A Principled Bayesian Workflow. `https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html`. (October 2018). Retrieved Febuary 18, 2019.
3. Pierre Dragicevic, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier. 2019. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.
4. Jeffrey M. Perkel. 2018. Why Jupyter is data scientists' computational notebook of choice. `https://www.nature.com/articles/d41586-018-07196-1`. (30 October 2018). Retrieved Febuary 11, 2019.
5. Chanda Phelan, Jessica Hullman, Matthew Kay, and Paul Resnick. 2019. Some Prior(s) Experience Necessary: Templates for Getting Started with Bayesian Analysis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.
6. Uri Simonsohn, Joseph P Simmons, and Leif D Nelson. 2015. Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications. Available at SSRN: https://ssrn.com/abstract=2694998 or http://dx.doi.org/10.2139/ssrn.2694998. (November 2015).
7. Sara Steegen, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science* 11, 5 (2016), 702–712.
8. Bret Victor. 2011. Explorable Explanations. Online. `http://worrydream.com/ExplorableExplanations/`. (2011).