

Bayesian workflow for non-trivial projects

Systems Biology for Scientific Computing: week five

Introduction

Recap

Last time we found out what to do after running MCMC:

- Check **diagnostics** such as \hat{R} and divergent transitions.
- **Find things out** by checking expectations and their MCSEs.
- **Evaluate** the model using decision theory.

Plan for today

How to structure a Bayesian statistics project that can scale to many models, datasets and transformations?

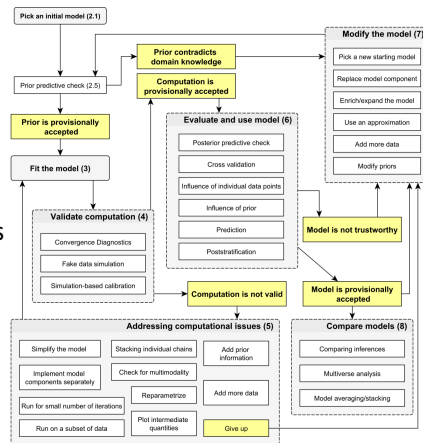
Theory

Bayesian Workflow

Reading: (Gelman et al. 2020)

Main takeaways:

- Work iteratively, starting with a small model
- Do checks at every stage
- Use fake data to find and understand problems
- Draw a lot of graphs
- Modeling is software development



Tasks in a statistical analysis

- Get raw data(s)
- Get prepared data(s)
- Implement statistical model(s)
- Interface between prepared data(s) and model(s)
- Specify and perform inference(s)
- Investigate data(s) and inference(s)
- Document what was done

Practice

General considerations for a scalable statistics workflow

Use **files** (don't leave things in your or the computer's memory)

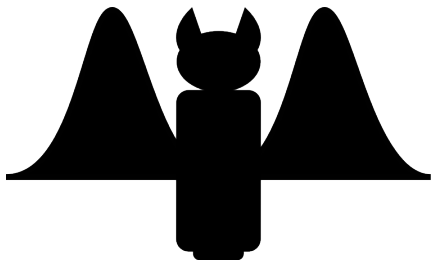
Separate tasks into self-contained **modules** where possible.

Use **automation** to save work and ensure reproducibility.

<https://github.com/teddygroves/bibat>

A Python package that I wrote for setting up
scaleable Bayesian analysis projects.

“bibat” stands for “Batteries-Included Bayesian
Analysis Template”



bibat

Installation

```
$ pip install bibat
```

Usage

```
$ bibat
```

Documentation

<https://bibat.readthedocs.io/en/latest/>



Challenge

Implement the [monod analysis](#) using bibat!

Find some raw data [here](#).

References

References

Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. “Bayesian Workflow.” *arXiv:2011.01808 [Stat]*, November.
<http://arxiv.org/abs/2011.01808>.