On Bayesian workflow

Aki Vehtari

Department of Computer Science
Aalto University

Talk partially based on

Bayesian workflow.

Outline

• Workflow frameworks improve
  - efficiency
  - reliability
  - reproducibility
  - etc.
Outline

• Workflow frameworks improve
  - efficiency
  - reliability
  - reproducibility
  - etc.

• Practical workflows
  - shortcuts and workflow as engineering
Outline

• Workflow frameworks improve
  - efficiency
  - reliability
  - reproducibility
  - etc.

• Practical workflows
  - shortcuts and workflow as engineering

• Iterative process
  - as in software development
  - as a learning process
  - models as experiments
Outline

• Workflow frameworks improve
  - efficiency
  - reliability
  - reproducibility
  - etc.

• Practical workflows
  - shortcuts and workflow as engineering

• Iterative process
  - as in software development
  - as a learning process
  - models as experiments

• The argument against the iterative model building
  - and the counter argument
Outline

• Workflow frameworks improve
  - efficiency
  - reliability
  - reproducibility
  - etc.

• Practical workflows
  - shortcuts and workflow as engineering

• Iterative process
  - as in software development
  - as a learning process
  - models as experiments

• The argument against the iterative model building
  - and the counter argument

• Software assisted workflows
Workflow term

- Saunders and Blundstone, 1921. Railway engineering.
- Started to get more popular in 1990s based on Google books ngram viewer
Bioinformatics and scientific workflow in 2000’s

Curcin & Ghanem (2008). Scientific workflow systems - can one size fit all?

Fig. 1. Workflow example
Figure 1: Science workflow for the comparison of a molecular dynamics simulation with a high-energy X-ray microscopy of the same material system includes three interrelated computational and experimental workflows.

Workflows in general

- Workflow frameworks improve
  - efficiency
  - reliability
  - reproducibility
  - etc.
Bayesian workflow

  - “The Bayesian workflow”
Box & Youle, 1955

The Exploration and Exploitation of Response Surfaces: An Example of the Link between the Fitted Surface and the Basic Mechanism of the System

![Diagram of experimental process iteration]

*Figure 7. Diagramatic representation of process of experimental iteration.*
A. The Advancement of Learning

A(1) An Iteration Between Theory and Practice

A(2) A Feedback Loop
Data analysis, a subiteration in the process of investigation, is illustrated here.
From the talk slides ‘Some aspects of statistical design in quality improvement’
Talts (2018)

StanCon 2018 intro

Bayesian Workflow

Scope out your problem
What inputs and outputs can help you learn? What relationships can you see by eye?

Specify likelihood & priors
Use knowledge of the problem to construct a generative model and shape the scope of the parameters

Check the model with fake data
Generate data, fit model, and evaluate fit as a sanity check

Fit the model to real data
To recover parameters

Check diagnostics
Algorithms should come with diagnostics that let you know when they’re not working

Graph fit estimates
Understand your inferences

Check predictive posterior
Perform PPCs to understand predictions

Compare models
Iterate on model design, choose a model

Aki.Vehtari@aalto.fi – @avehtari@bayes.club
Bayesian workflows

- Pick an initial model (2.1)
  - Prior predictive check (2.5)
    - Prior is provisionally accepted
      - Fit the model (3)
      - Validate computation (4)
        - Convergence Diagnostics
        - Fake data simulation
        - Simulation-based calibration
      - Addressing computational issues (5)
        - Simplify the model
        - Implement model components separately
        - Run for small number of iterations
        - Run on a subset of data
      - Compare models (8)
        - Comparing inferences
        - Multiverse analysis
        - Model averaging/stacking
  - Prior contradicts domain knowledge
    - Computation is provisionally accepted
      - Evaluate and use model (9)
        - Posterior predictive check
        - Cross validation
        - Influence of individual data points
        - Influence of prior
        - Prediction
        - Poststratification
      - Computation is not valid
        - Model is not trustworthy
        - Modify the model (7)
          - Pick a new starting model
          - Replace model component
          - Enrich/expand the model
          - Use an approximation
          - Add more data
          - Modify priors
      - Model is provisionally accepted

Model building as software development process

Boehm (1996, 1988)

Figure 2. Spiral model of the software process.
Example: Birthdays

https://avehtari.github.io/casestudies/Birthdays/birthdays.html
Prototypes, shortcuts, and workflow as engineering

- When building prototypes or making initial experiments, we can use shortcuts
  - the final models tested with sufficient rigor
Example: Birthdays

https://avehtari.github.io/casestudies/Birthdays/birthdays.html
Different requirements and workflow as cooking

- Make a sandwich out of available ingredients
  - a generic model that is useful for almost any data
  - normal linear model is not probably the best dish given the ingredients, but can be delicious anyway
Different requirements and workflow as cooking

- Make a sandwich out of available ingredients
  - a generic model that is useful for almost any data
  - normal linear model is not probably the best dish given the ingredients, but can be delicious anyway
- Follow a recipe to make a dish with pre-listed ingredients
  - many different pre-described models that match your data and task
  - e.g. rstanarm, brms, shinybrms
Different requirements and workflow as cooking

- Make a sandwich out of available ingredients
  - a generic model that is useful for almost any data
  - normal linear model is not probably the best dish given the ingredients, but can be delicious anyway
- Follow a recipe to make a dish with pre-listed ingredients
  - many different pre-described models that match your data and task
  - e.g. rstanarm, brms, shinybrms
- Develop a new recipe to be used in a fine dining restaurant
  - may require several iterations (depending on how much the customers are willing to pay)
  - write the model in probabalistic programming language
Iterative workflow as a learning process

A. The Advancement of Learning

A(1) An Iteration Between Theory and Practice

A(2) A Feedback Loop

- Incremental learning reduces cognitive load
Example: Golf putting

Data on putts in pro golf

Distance from hole (feet)

Probability of success

Example by Andrew Gelman CC-BY-NC 4.0
https://mc-stan.org/users/documentation/case-studies/golf.html
Example: Golf putting

Fitted logistic regression

Logistic regression, \( a = 2.23, b = -0.26 \)
Example: Golf putting

Data on putts in pro golf
Example: Golf putting

Two models fit to the golf putting data

[Graph showing the probability of success versus distance from the hole with two models: Logistic regression and Geometry-based model.]
Example: Golf putting

Checking already-fit model to new data

- Probability of success vs. Distance from hole (feet)
- Old data (blue dots)
- New data (red dots)
Example: Golf putting

Checking already-fit model to new data

Probability of success vs. Distance from hole (feet)

Old data vs. New data

Aki.Vehtari@aalto.fi – @avehtari@bayes.club
Example: Golf putting

Checking model fit

Probability of success vs. Distance from hole (feet)
Models as experiments

A. The Advancement of Learning

A(1) An Iteration Between Theory and Practice

A(2) A Feedback Loop

- New facts are learned by using the models
- New learned facts can be used to choose the next model
Example: Birthdays

https://avehtari.github.io/casestudies/Birthdays/birthdays.html
Software development process vs. Models as experiments

• The end result may also be a series of models
  - it may be useful to report results of simpler experiments, too
  - presenting the results of many models is also part of the workflow
Cost and benefit of model building
Cost and benefit of model building
Cost and benefit of model building
Cost and benefit of model building

Stick figure and silhouette from Weech et al. (2014) doi:10.1167/14.12.10. Photo CC0.
Cost and benefit of model building

- Sufficient accuracy for model and computation
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Watch and listen more (4.5h) in

- Model assessment, selection and averaging
  https://www.youtube.com/watch?v=Re-2yVd0Mqk
- Use of reference models in variable selection
  https://www.youtube.com/watch?v=N0ce8J8slFY
- These are a few of my favorite inference diagnostics
  https://www.youtube.com/watch?v=vLx6lUlZ0fc
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Counter arguments
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Counter arguments

- Integration over the model space
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Counter arguments

- Integration over the model space
  + filter zero weight models
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Counter arguments

- Integration over the model space
  + filter zero weight models
- Missing perfect calibration is not a problem
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Counter arguments

- Integration over the model space
  + filter zero weight models
- Missing perfect calibration is not a problem
- Diagnostics

Use well specified models and priors (they are inseparable)

Watch and listen more (4.5h) in
- Model assesment, selection and averaging
  https://www.youtube.com/watch?v=Re-2yVd0Mqk
- Use of reference models in variable selection
  https://www.youtube.com/watch?v=N0ce8J8slFY
- These are a few of my favorite inference diagnostics
  https://www.youtube.com/watch?v=vLx6lUlZ0fc
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Counter arguments

- Integration over the model space
  + filter zero weight models
- Missing perfect calibration is not a problem
- Diagnostics
- Use well specified models and priors (they are inseparable)
Argument against iterative model building

- Double dipping, inference after model selection, etc.

Counter arguments

- Integration over the model space
  + filter zero weight models
- Missing perfect calibration is not a problem
- Diagnostics
- Use well specified models and priors (they are inseparable)

Watch and listen more (4.5h) in

- Model assessment, selection and averaging
  https://www.youtube.com/watch?v=Re-2yVd0Mqk
- Use of reference models in variable selection
  https://www.youtube.com/watch?v=N0ce8J8sIFY
- These are a few of my favorite inference diagnostics
  https://www.youtube.com/watch?v=vLx6IUIZ0fc
Filtering from Riha, Siccha & Vehtari (2023)
Computer assisted workflows

• PDFs and blog posts with checklists are not enough
Computer assisted workflows

- PDFs and blog posts with checklists are not enough
- Code generation and libraries (e.g. brms, bambi)
Computer assisted workflows

- PDFs and blog posts with checklists are not enough
- Code generation and libraries (e.g. brms, bambi)
- Computational graphs (e.g., drake, targets, scientific workflow software)

Diagnostics (e.g. posterior, bayesplot, priorsense, loo, projpred, ArviZ)

- More robust inference and diagnostics with diagnostics
- Automated sub-workflows
- Allow shortcuts with automated diagnostic, checking whether the shortcut was a safe choice

Interplay between software, diagnostics, and documentation
- Different levels of detail
- Just-in-time learning

See also "What do we need from a PPL to support Bayesian workflow?" by Bob Carpenter
statmodeling.stat.columbia.edu/2021/10/22/carpenter-slides-papers-prob-prog-2021/
Computer assisted workflows

- PDFs and blog posts with checklists are not enough
- Code generation and libraries (e.g. brms, bambi)
- Computational graphs (e.g., drake, targets, scientific workflow software)
- Diagnostics (e.g. posterior, bayesplot, priorsense, loo, projpred, ArviZ)
Computer assisted workflows

- PDFs and blog posts with checklists are not enough
- Code generation and libraries (e.g. brms, bambi)
- Computational graphs (e.g., drake, targets, scientific workflow software)
- Diagnostics (e.g. posterior, bayesplot, priorsense, loo, projpred, ArviZ)
- More robust inference and diagnostics with diagnostics
  - automated sub-workflows
  - allow shortcuts with automated diagnostic, checking whether the shortcut was a safe choice

See also "What do we need from a PPL to support Bayesian workflow?" by Bob Carpenter
statmodeling.stat.columbia.edu/2021/10/22/carpenter-slides-papers-prob-prog-2021/
Computer assisted workflows

- PDFs and blog posts with checklists are not enough
- Code generation and libraries (e.g. brms, bambi)
- Computational graphs (e.g., drake, targets, scientific workflow software)
- Diagnostics (e.g. posterior, bayesplot, priorsense, loo, projpred, ArviZ)
- More robust inference and diagnostics with diagnostics
  - automated sub-workflows
  - allow shortcuts with automated diagnostic, checking whether the shortcut was a safe choice
- Interplay between software, diagnostics, and documentation
  - different levels of detail

See also “What do we need from a PPL to support Bayesian workflow?” by Bob Carpenter
statmodeling.stat.columbia.edu/2021/10/22/carpenter-slides-papers-prob-prog-2021/
Computer assisted workflows

- PDFs and blog posts with checklists are not enough
- Code generation and libraries (e.g. brms, bambi)
- Computational graphs (e.g., drake, targets, scientific workflow software)
- Diagnostics (e.g. posterior, bayesplot, priorsense, loo, projpred, ArviZ)
- More robust inference and diagnostics with diagnostics
  - automated sub-workflows
  - allow shortcuts with automated diagnostic, checking whether the shortcut was a safe choice
- Interplay between software, diagnostics, and documentation
  - different levels of detail
  - just-in-time learning

See also "What do we need from a PPL to support Bayesian workflow?" by Bob Carpenter
statmodeling.stat.columbia.edu/2021/10/22/carpenter-slides-papers-prob-prog-2021/
Computer assisted workflows

- PDFs and blog posts with checklists are not enough
- Code generation and libraries (e.g. brms, bambi)
- Computational graphs (e.g., drake, targets, scientific workflow software)
- Diagnostics (e.g. posterior, bayesplot, priorsense, loo, projpred, ArviZ)
- More robust inference and diagnostics with diagnostics
  - automated sub-workflows
  - allow shortcuts with automated diagnostic, checking whether the shortcut was a safe choice
- Interplay between software, diagnostics, and documentation
  - different levels of detail
  - just-in-time learning
- See also “What do we need from a PPL to support Bayesian workflow?” by Bob Carpenter
  statmodeling.stat.columbia.edu/2021/10/22/carpenter-slides-papers-prob-prog-2021/
Computer assisted workflows

- There are software for sub-workflows
- Software for the whole is very challenging as there are so many possibilities
Computer assisted workflows

- There are software for sub-workflows
- Software for the whole is very challenging as there are so many possibilities
  - LLMs (e.g. ChatGPT) can only help when there has been enough material in the internet for them to learn, and even then they may provide misleading recommendations
On Bayesian workflow

- Workflow frameworks improve
  - efficiency
  - reliability
  - reproducibility
  - etc.
- Practical workflows
  - shortcuts and workflow as engineering
- Iterative process
  - as in software development
  - as a learning process
  - models as experiments
- The argument against the iterative model building
  - and the counter argument
- Computer assisted workflow + different levels of information

Aki.Vehtari@aalto.fi – @avehtari@bayes.club