Stan for the people

Two day introductory workshop
on Bayesian modeling

McGill University
January 25th 2019

mc-stan.org

Instructor:

- Charles Margossian
  Columbia University, Department of Statistics

Teacher assistant:

- Aaron Erlich
  McGill University, Department of Political Science

Outline

Day 1

- Morning
  - Introduction to Bayesian analysis
  - Algorithms and computational considerations
- Afternoon
  - Introduction to the Stan programing language
  - Expressing a model in Stan
  - Examples: Bayesian linear and logistical regression

Outline

Day 2

- Morning
  - Conversational Stan
  - Hierarchical models
- Afternoon
  - Model parametrization
  - Discussion and concluding remarks

Logistics

The worksho package includes:

- R scripts and Stan files to do the exercises
- These slides
- Outline of the course
- Additional documentation

We will be using:

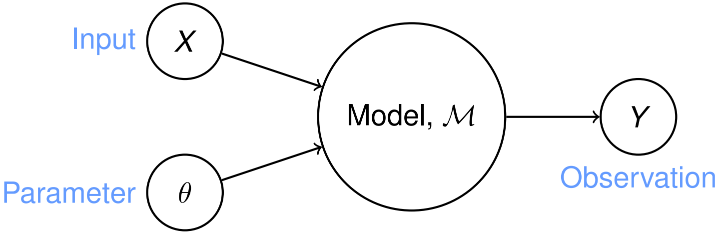
- RStan 2.18.2.
- ggplot, plyr, tidyr, dplyr

I

Introduction to Bayesian Analysis

What is a model?

The model is a *story of how the data is generated*:



$$\mathcal{M} : (X, \theta) \to Y$$

$$\mathcal{M} : (X, \theta) \to Y$$

- Tells us how to *simulate* data.
- Characterized by the distribution $P_\theta(Y|X)$.

Inference problem

- We have some data $Z = (X, Y)$ and a model $\mathcal{M}$.
- We want to learn about θ.
- How can we *reverse engineer* the data generating process?

Bayesian inference

The two core notions of Bayesian statistics:

1. Unknown quantities are viewed as random variables, i.e. they are described in terms of probability distributions.
2. Bayes rule provides a formal mechanism for combining prior knowledge with new data.

Bayesian inference

Proposition – make the requisite inference using the *posterior distribution*:

$$P(\theta|Z)$$

Bayes' rule

Consider two random variables A and B.

Recall

$$P(A, B) = P(A|B)P(B)$$

Then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Bayes' rule

$$P(\theta|Z) = \frac{P(Z|\theta)P(\theta)}{P(Z)}$$

Bayes' rule

$$P(\theta|Z) = \frac{P(Z|\theta)P(\theta)}{P(Z)}$$

“Likelihood” distribution

$$P(Z|\theta) = P(Y|\theta, X)$$

- Characterizes $\mathcal{M}$ and tells us how to simulate Y, given X and θ.
- Also tells us how “likely” it would be to simulate Y, given X and θ.

Bayes' rule

$$P(\theta|Z) = \frac{P(Z|\theta)P(\theta)}{P(Z)}$$

Prior distribution

$$P(\theta)$$

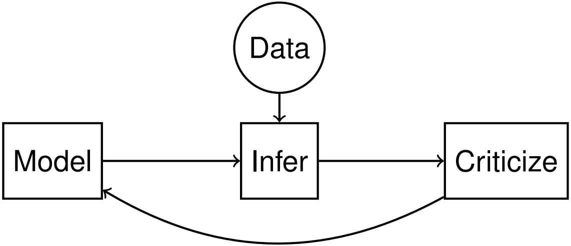
- One of the challenges of Bayesian analysis is the need for a prior distribution.
- But this is also an immense advantage:
  - It quantitatively incorporates in $\mathcal{M}$ our prior understanding / assumptions about the parameters
  - It acts as a regularization tool.

Bayes' rule

$$P(\theta|Z) = \frac{P(Z|\theta)P(\theta)}{P(Z)}$$

Beyond the posterior

Box's loop:



Beyond the posterior

Model:

- Specify a data generating process.
- Specify a prior distribution.

Inference:

- Compute or approximate $p(\theta|z)$.
- Do any number of operations on $p(\theta|z)$.

Criticize:

- Does our model capture the information we care about?
- If not, how can we improve the model?

Beyond the posterior

The following articles discuss Bayesian modeling frameworks:

- Philosophy and the practice of Bayesian statistics [Gelman and Shalizi, 2013]
- Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models [Blei, 2014]
- Visualization in Bayesian workflow [Gabry et al., 2018]
- Towards a principled Bayesian workflow [Betancourt, 2018]

References I

[Betancourt, 2018] Betancourt, M. (2018).
Towards a principled bayesian workflow.

[Blei, 2014] Blei, D. (2014).
Build, compute, critique, repeat: Data analysis with latent variable models.
Annual Review of Statistics and Its Application, 1.

[Gabry et al., 2018] Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2018).
Visualization in bayesian workflow.
Royal Journal of Statistics, section A, 182:1 –14.

[Gelman and Shalizi, 2013] Gelman, A. and Shalizi, C. R. (2013).
Philosophy and the practice of bayesian analysis.
British Journal of Mathematical and Statistical Psychology, 66.