Fundamentals of Stan

StanCon 2023

mc-stan.org

Instructor:

Charles Margossian
Flatiron Institute

Teaching Assistant:

Manny Mokel
Flatiron Institute & Duke University

Outline

- Review of Bayesian Analysis
- Markov chain Monte Carlo
- The Stan language
  - Example: Bayesian linear regression
  - Example: Disease transmission model

Logistics

An `R` notebook to do the exercises can be found at:

https://github.com/charlesm93/stanTutorial

You can run the `R` code on your local machine or on the `Colab` cloud server.

I

Review of Bayesian Analysis

What is a Bayesian model?

What is a Bayesian model?

Defined as a joint distribution

$$p(\theta, y)$$

over observed variables y and unknowns θ.

What is a Bayesian model?

Defined as a joint distribution

$$p(\theta, y) = p(\theta)p(y \mid \theta)$$

over observed variables y and unknowns θ.

What is a Bayesian model?

Defined as a joint distribution

$$p(\theta, y) = p(\theta)p(y \mid \theta)$$

over observed variables y and unknowns θ.

$p(y \mid \theta)$ is the *likelihood*:

For a fixed θ, defines a data generating process.

What is a Bayesian model?

Defined as a joint distribution

$$p(\theta, y) = p(\theta)p(y \mid \theta)$$

over observed variables y and unknowns θ.

$p(y \mid \theta)$ is the *likelihood*:

For a fixed θ, defines a data generating process.

$p(\theta)$ is the *prior*:

quantitative assumptions and understanding about θ
information from previous analysis
regularization tool

RESEARCH ARTICLE

Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe

Anthony Hauser[1], Michel J. Counotte[1], Charles C. Margossian[2], Garyfallos Konstantinoudis[3], Nicola Low[1], Christian L. Althaus[1], Julien Riou[1,4]*

RESEARCH ARTICLE

Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe

Anthony Hauser[1], Michel J. Counotte[1], Charles C. Margossian[2], Garyfallos Konstantinoudis[3], Nicola Low[1], Christian L. Althaus[1], Julien Riou[1,4]*

Likelihood:

Epidemiological model of the disease dynamic

Measurement model: test results, hospital deaths.

RESEARCH ARTICLE

Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe

Anthony Hauser[1], Michel J. Counotte[1], Charles C. Margossian[2], Garyfallos Konstantinoudis[3], Nicola Low[1], Christian L. Althaus[1], Julien Riou[1,4]*

Likelihood:

Epidemiological model of the disease dynamic
Measurement model: test results, hospital deaths.

Prior:

Constraints on interpretable parameters
Meta-analysis for asymptomatic rate

Bayesian inference

Given observations y, want to learn about θ

Bayesian inference

Given observations y, want to learn about θ

Proposition: Learn a posterior distribution

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{p(y)}$$

Bayesian inference

Given observations y, want to learn about θ

Proposition: Learn a posterior distribution

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{p(y)}$$

For some transformation f of the parameters θ, can learn

$$p(f(\theta) \mid y)$$

Bayesian inference

Given observations y, want to learn about θ

Proposition: Learn a posterior distribution

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{p(y)}$$

For some transformation f of the parameters θ, can learn

$$p(f(\theta) \mid y)$$

f may be predictions about the future, a useful summary (e.g. R_0 infection rate), etc.

Bayesian inference

Given observations y, want to learn about θ

Proposition: Learn a posterior distribution

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{p(y)}$$

For some transformation f of the parameters θ, can learn

$$p(f(\theta) \mid y)$$

f may be predictions about the future, a useful summary (e.g. R_0 infection rate), etc.

The variance in $p(f(\theta) \mid y)$ accounts for both the modeled noise and the uncertainty in our estimation of θ.

Example: normal-normal model

$$\begin{aligned} p(\theta) &= \text{normal}(\mu, \tau) \\ p(y_n \mid \theta) &= \text{normal}(\theta, \sigma) \end{aligned}$$

Suppose we have N i.i.d observations $y_1, y_2, \cdots, y_N$.

Example: normal-normal model

$$\begin{aligned} p(\theta) &= \text{normal}(\mu, \tau) \\ p(y_n \mid \theta) &= \text{normal}(\theta, \sigma) \end{aligned}$$

Suppose we have N i.i.d observations $y_1, y_2, \cdots, y_N$.
Then

$$p(\theta \mid \mathbf{y}) = \text{normal}\left(\frac{\mu/\tau^2 + N\bar{y}/\sigma^2}{1/\tau^2 + N/\sigma^2}, \frac{1}{1/\tau^2 + N/\sigma^2}\right)$$

Example: normal-normal model

$$\begin{aligned} p(\theta) &= \text{normal}(\mu, \tau) \\ p(y_n \mid \theta) &= \text{normal}(\theta, \sigma) \end{aligned}$$

Suppose we have N i.i.d observations $y_1, y_2, \cdots, y_N$.

Then

$$p(\theta \mid \mathbf{y}) = \text{normal}\left(\frac{\mu/\tau^2 + N\bar{y}/\sigma^2}{1/\tau^2 + N/\sigma^2}, \frac{1}{1/\tau^2 + N/\sigma^2}\right)$$

In practice, the posterior is <u>not</u> tractable.

Need to estimate summary quantities: expectation values, variance, quantiles, $\cdots$

Bayesian learning

Suppose we obtain data over two observations, $\mathbf{y}_1$ and $\mathbf{y}_2$.

Bayesian learning

Suppose we obtain data over two observations, $\mathbf{y}_1$ and $\mathbf{y}_2$. Then the following two procedure are equivalent,

(1)

- Start with a prior $p(\theta)$.
- Compute the posterior $p(\theta \mid \mathbf{y}_1, \mathbf{y}_2)$.

or

(2)

- Start with a prior $p(\theta)$.
- Compute the posterior $p(\theta \mid \mathbf{y}_1)$
- Use $p(\theta \mid \mathbf{y}_1)$ as a new prior.
- Compute the posterior $\tilde{p}(\theta \mid \mathbf{y}_2)$.

Bayesian learning

Suppose we obtain data over two observations, $\mathbf{y}_1$ and $\mathbf{y}_2$. Then the following two procedure are equivalent,

(1)

- Start with a prior $p(\theta)$.
- Compute the posterior $p(\theta \mid \mathbf{y}_1, \mathbf{y}_2)$.

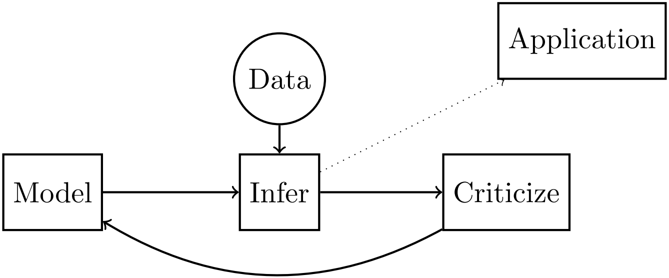
or

(2)

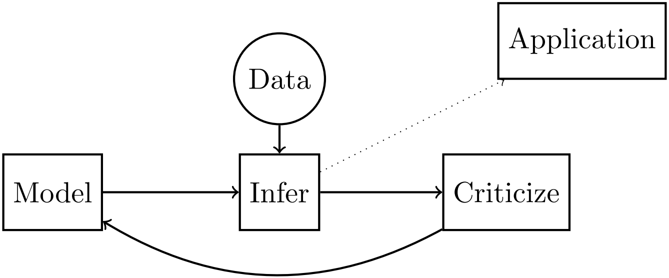
- Start with a prior $p(\theta)$.
- Compute the posterior $p(\theta \mid \mathbf{y}_1)$
- Use $p(\theta \mid \mathbf{y}_1)$ as a new prior.
- Compute the posterior $\tilde{p}(\theta \mid \mathbf{y}_2)$.

$$\tilde{p}(\theta \mid \mathbf{y}_2) = p(\theta \mid \mathbf{y}_1, \mathbf{y}_2)$$

Bayesian workflow



Bayesian workflow

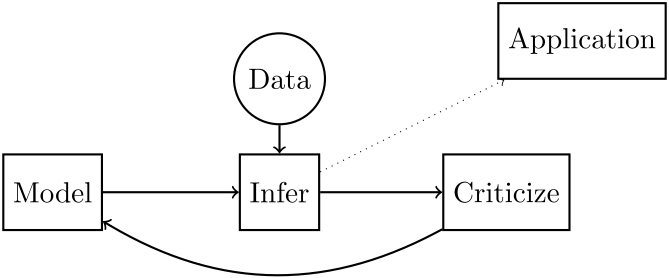


Model

Likelihood $p(y \mid \theta)$

Prior $p(\theta)$

Bayesian workflow



Model

Likelihood $p(y \mid \theta)$

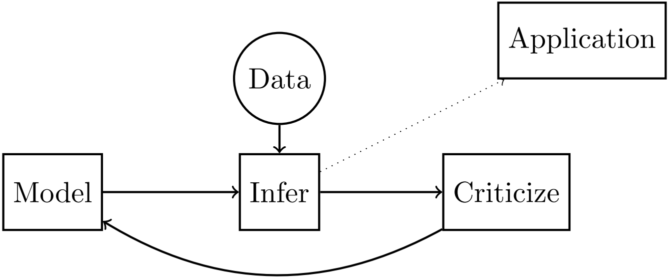
Prior $p(\theta)$

Inference

Approximate

$p(f(\theta) \mid y)$

Bayesian workflow



Model

Likelihood $p(y \mid \theta)$

Prior $p(\theta)$

Inference

Approximate

$p(f(\theta) \mid y)$

Criticize

Does the model capture the info we care about?

If not, how can we improve the model?

RESEARCH ARTICLE

Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe

Anthony Hauser[1], Michel J. Counotte[1], Charles C. Margossian[2], Garyfallos Konstantinoudis[3], Nicola Low[1], Christian L. Althaus[1], Julien Riou[1,4]*

The published model is the ~15th iteration.

RESEARCH ARTICLE

Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe

Anthony Hauser[1], Michel J. Counotte[1], Charles C. Margossian[2], Garyfallos Konstantinoudis[3], Nicola Low[1], Christian L. Althaus[1], Julien Riou[1,4]*

The published model is the ∼15th iteration.

Grinsztajn et al. Bayesian workflow for disease transmission model in Stan, *Statistics in Medicine* (2021)

Gelman et al. Bayesian workflow, *arXiv:2011.01808* (2020)