Stan Demo

BSDA 2024

Väinö Yrjänäinen

2023-09-30

Today's goals

- Get Stan installed
- Replicate simple examples with Stan
- Import own data and tinker with the models (if time allows)

Stan Installation

Installing stan

- The easiest choice for most of you is rstan
  - R and RStudio have nice stan features
- You can also use pystan or other versions if you want to

Before installing

- Check your R version
- What operating systems are you using?

Installing rstan

- Instructions: https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started
- Requires C++ compliler

```
install.packages("rstan")
```

Installing pystan

- Create and activate conda environment
- Install pystan and arviz for R hat etc.[1]

```
pip install pystan
pip install arviz
```

[1] https://discourse.mc-stan.org/t/pystan-parameter-summary-ess-and-rhat/22538

Verifying the installation

```
example(stan_model, package = "rstan", run.dontrun = TRUE)
```

- Run the eight schools model[2][3]

[2] https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started#example-1-eight-schools

[3] https://pystan.readthedocs.io/en/latest/#quick-start

Examples

Three things

- Normally distributed data, no priors
- Binomial example
- Delayed flights

Normally distributed data

$$y_i \sim \mathcal{N}(\mu, \sigma^2)$$
$$i \in \{1, ..., N\}$$

Add prior

- Flat prior?
- $\sigma^2 \sim \text{Exp}(\lambda)$?

Binomial model

$$n_{j,i} \sim B\left(N_j, \theta_j\right)$$
$$i \in \{1, ..., N\}$$
$$j \in \{0, 1\}$$

j are the treatment and control groups

Delayed flights

date	from	to	planned-dep	dep	planned_arr	arr
14/09/2023	HEL	ARN	09:15	09:28	09:15	09:17
…	…	…	…	…	…	…
04/09/2023	HEL	ARN	09:15	09:24	09:15	09:08

Delayed flights

- I want to know when I arrive when planned arrival is 9.15
  - Model the departure of the next flight (posterior predictive)
- Use the Gaussian model

Posterior predictive

For each posterior sample of (μ, σ^2), we draw

$$\tilde{y}_i \sim \mathcal{N}(\mu, \sigma^2)$$

We also calculate whether you make it on time

$$\text{ontime} = \mathbb{I}(\tilde{y}_i \leq 9.5)$$

Use your own data

- Find some univariate data
  - Steps you took last week, points scored by your football team in the last 5 games, heights of people in your family, absolutely whatever
- Use the normal.stan model

Change the model

- Try out some new priors
- Use eg. ‘gamma’ instead of ‘normal’