Modelling with RStan

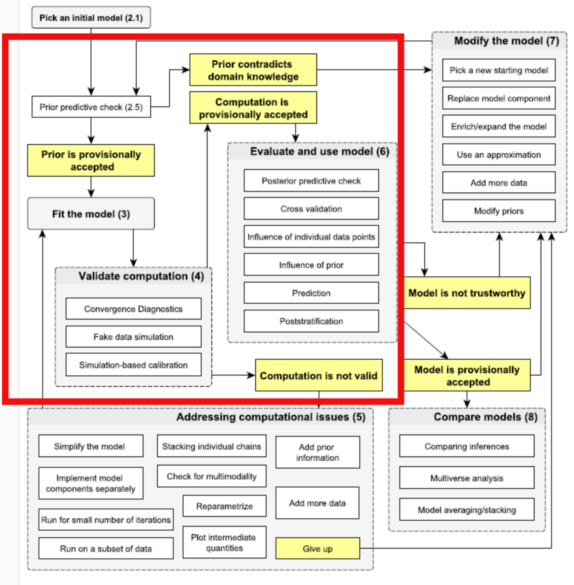
Sydney Dimmock

June 13, 2022

Modelling with Stan

1. Choose an initial model.
2. Construct Stan model.
3. Sample the model.
4. Check for convergence and good sampling quality.
    - Unsuccessful? Try a different parameterisation and review of the prior model!
5. Extract and interpret posterior parameters.

Bayesian Workflow



Adapted from Gelman et al. 2020

Structure of a Stan Model

Stan Example

```
data{
    ...
}
parameters{
    ...
}
transformed parameters{
    ...
}
model{
    ...
}
generated quantities{
    ...
}
```

Data and Model

We have a dataset containing the number of observed head counts for twenty biased coins:

ID	Flips	Heads	P
1	35	20	0.50
2	61	61	0.99
3	72	31	0.46
4	62	51	0.77
5	48	6	0.25
6	41	39	0.95
7	93	29	0.35

... and so on.

Data and Model

We can construct a mathematical model for this data:

$$y_i \sim \text{Binomial}(n_i, \theta_i), \quad i \ldots N$$
$$\theta_i \sim \text{Beta}(\alpha, \beta)$$

How does this translate into a Stan model?

Constructing a Stan Model

Stan Example

```
data{
    int N;
    int y[N];
    int n[N];
}
```

Constructing a Stan Model

Stan Example

```
data{
    int N;
    int y[N];
    int n[N];
}
parameters{
    real<lower=0, upper=1> theta[N];
}
```

Constructing a Stan Model

Stan Example

```
data{
    int N;
    int y[N];
    int n[N];
}
parameters{
    real<lower=0, upper=1> theta[N];
}
model{
    theta ~ beta(2,2);
}
```

Constructing a Stan Model

Stan Example

```
data{
    int N;
    int y[N];
    int n[N];
}
parameters{
    real<lower=0, upper=1> theta[N];
}
model{
    theta ~ beta(2,2);
    y ~ binomial(n, theta);
}
```

Constructing a Stan Model

Stan Example

```
data{
    int N;
    int y[N];
    int n[N];
}
parameters{
    real<lower=0, upper=1> theta[N];
}
model{
    theta ~ beta(2,2);
    y ~ binomial(n, theta);
}
generated quantities{
    int yrep[N];
    for(i in 1:N){
        yrep[i] = binomial_rng(n[i], theta[i]);
    }
}
```

Sampling Statements

The model section is made up of sampling statements:

Stan Example

```
y ~ normal(0,1)
```

Using the distributions are short hands for incrementing the log density:

Stan Example

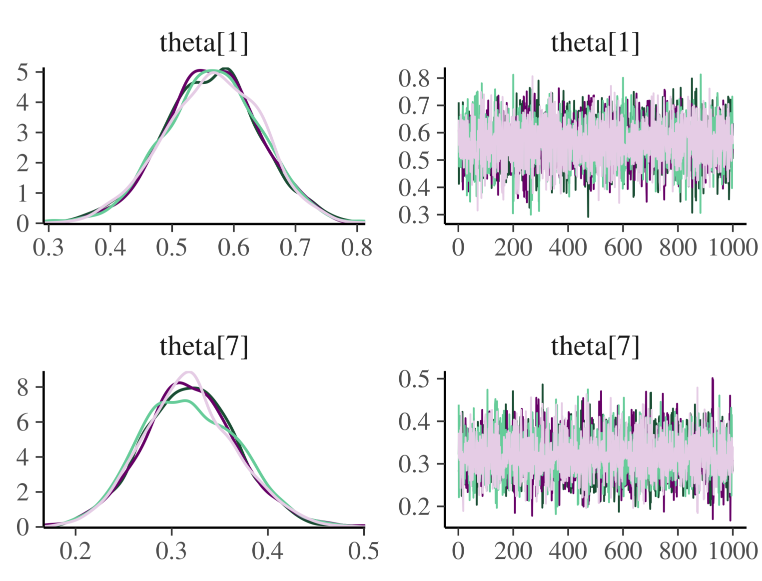
```
target += normal_lpdf( y | 0, 1);
```

This is how you can define your own densities within Stan.

Running the Model from R

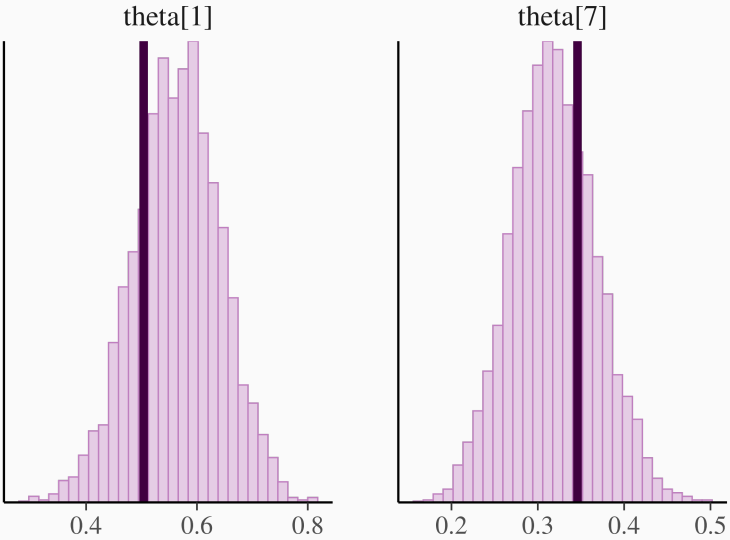
```r
library(rstan)
library(readr)
library(bayesplot)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

data <- read_csv("data/coin.csv") # Read in the data

# Prepare data for Stan
stan_data <- list(N = nrow(data),
                  y = data$Heads,
                  n = data$Flips)

# Sample the model
fit <- stan(file="models/binom_coin.stan", data=stan_data, chains=4, iter=1000)

# Extract posterior samples
theta <- extract(fit, "theta")$theta
```

Plotting



theta[1]
5
4
3
2
1
0
0.3
0.4
0.5
0.6
0.7
0.8
theta[1]
0.8
0.7
0.6
0.5
0.4
0.3
0
200
400
600
800
1000
theta[7]
8
6
4
2
0
0.2
0.3
0.4
0.5
theta[7]
0.5
0.4
0.3
0.2
0
200
400
600
800
1000

Plotting



Plotting

Visualisations can be created quickly using `bayesplot`.

The syntax is consistent across most of the library functions requiring:

1. Stan fit object
2. Parameters of interest, specified by string or regular expression.

Stan Example

```
mcmc_areas(fit, c('theta[1]','theta[2]'));
```