Advanced Data Analysis in R

Bayesian Modeling in R

Michael DeWitt

2018-03-17 (Updated 2019-04-14)

Bayesian Modeling in R

A Thought Exercise

You are already Bayesian!

You just didn't know it!

A Coin

What is the probability a given coin is fair?

Frequentist

If you didn't answer 100% or 0% you're Bayesian!

What is Bayes?

- Named after Rev. Thomas Bayes

```
knitr::include_graphics(here::here("bayesian_modeling", "figs",
```

Bayes Theorem

Just derrived from identities of probability

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

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

Prior, Likelihood, Posterior Distribution

Prior

Prior is *subjective* and represents a range of potential values

Specified by a distribution

Informative

Restricts range of likely values to small range

More informative adds more “weight” to the prior vs the data

Noninformative/ Uninformative

Wider range of possibilities

More closely approximates the Maximum likelihood estimates

Likelihood

Likelihood is the distribution for the data generating process

Examples

- Poisson process -> poisson likelihood function
- Binomial process -> binomial likelihood function
- Normal distibution -> normal likelihood function
- Ordered categorical -> ordered categorical likelihood function

Posterior

$$P(A|B) \sim Prior * Likelihood$$

```
curve(expr = dbeta(x, shape1 = .1, shape2 = 50), from = 0, to =
```

Bayesian Workflow

Doing Bayesian Inferences

Bayesian inference and modeling techniques can be applied across the board

In MLE approachs you often make assumptions without even realising it!

Bayes requires you to be more deliberate

Write Your Model

What is your data generating process?

We are estimating the support for a given referendum. Thus our population has a choice, either 1 (support) or 0 (do not support). We do not have any good estimates from previous literature for overall support.

Write The DGP in Math

Likelihood Function

Series of bernouilli trials -> binomial likelihood function

$$P(\theta|Data) \sim Binomial(Data, \theta)$$

Prior

Additionally say that we have some data on

$$\theta \sim Beta(1, 1)$$

Simulate Your Data Generating Process

```r
n <- 1000
gender <-rep(x = 0:1, length.out = n)
income <- rnorm(n, 0, 1)
mu <- gender * 1.5 + income * 2
y <- rbinom(n,1, prob = plogis(mu))

dat <- data.frame(gender =gender, income = income, y = y)
```

Domain Specific Languages

BUGS
JAGS
Stan
Hand coded samplers

Enter `brms`

`brms` makes Bayesian Modeling Easy

Utilises Hamiltonian Monte Carlo with a No U-Turn Sampler

Specifying a Model in `brms`

```
library(brms)
(model <- bf(y ~ gender + income))
## y ~ gender + income
```

Inspect Priors

```
get_prior(model, dat, bernoulli())
##                  prior     class   coef group resp dpar
## 1                              b
## 2                              b gender
## 3                              b income
## 4 student_t(3, 0, 10) Intercept
##   nlpar bound
## 1
## 2
## 3
## 4
```

Specify Priors

```
my_priors <- c(prior(normal(0, 0.5), class = "b", coef = "gender
                     prior(normal(0, 0.5), class = "b", coef =
```

Model Family

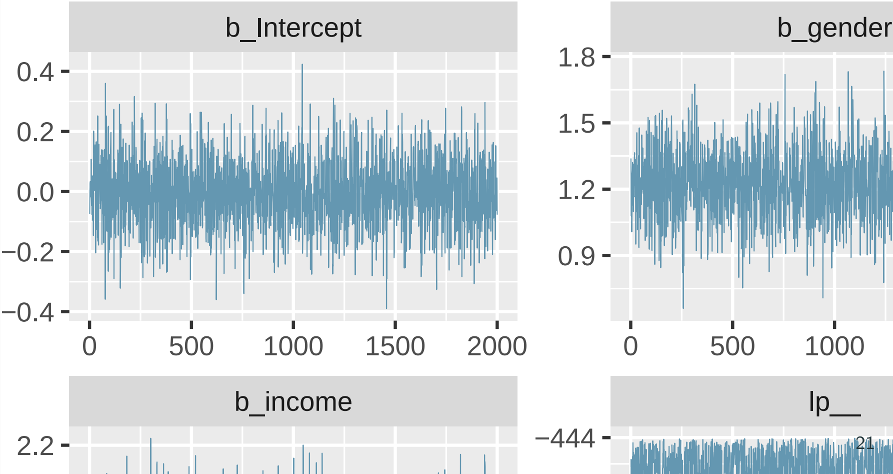
```
fit <- brm(model, my_priors,
           data =  dat,
           family = bernoulli(),
           inits = 1000, cores = 2, chains = 2, seed = 1234, ref
```

Posterior Checks

- Convergence
  - Trace Plots `tracplot`
  - Rhat metrics
  - Effective Sample Size
- Posterior Predictive Intervals
  - Was there a good fit between the model and the data

So Let's Check

```
library(tidybayes)
library(bayesplot)
mcmc_trace(as.matrix(fit))
```



Inferences

```
summary(fit)
##  Family: bernoulli
##   Links: mu = logit
## Formula: y ~ gender + income
##    Data: dat (Number of observations: 1000)
## Samples: 2 chains, each with iter = 2000; warmup = 1000; thin
##          total post-warmup samples = 2000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI
## Intercept    -0.01      0.11    -0.22     0.23
## gender        1.23      0.16     0.92     1.54
## income        1.85      0.12     1.63     2.08
##           Eff.Sample Rhat
## Intercept       2373 1.00
## gender          1787 1.00
## income          1595 1.00
##
```

Advantages of Bayesian Analysis

- Takes advantage of expert opinion
  - Especially helpful with small samples size studies
  - Reduces possibility of wildly odd results
- Easier communications (more intuitive to discuss probabilities)
- Studies can build on one another
  - Results from one study can be supplied directly as a prior into a replciation or another study

Drawbacks of Bayesian Inference

- Not as widely utilised in major publications
- Computationally intensive
- Picking a prior
- Heuristics exist

References