Advanced Data Analysis in R

Bayesian Modeling in R

Michael DeWitt

2018-03-17 (Updated 2019-04-16)

Bayesian Modeling in R

A Thought Exercise

You are already Bayesian!

You just didn't know it!

What is the probability a given coin is fair?

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

Bayes Theorem

Just derived 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- What do we know (or not)?

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

Uninformative/ Weak
Wider range of possibilities

More closely approximates the Maximum likelihood estimates

## Likelihood - The Data Generating Process

Likelihood is the distribution for the data generating process

Examples

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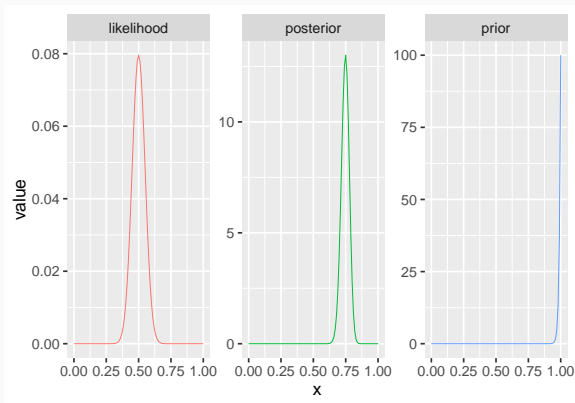- Ordered categorical -> ordered categorical likelihood function

Posterior - What we make inferences on!

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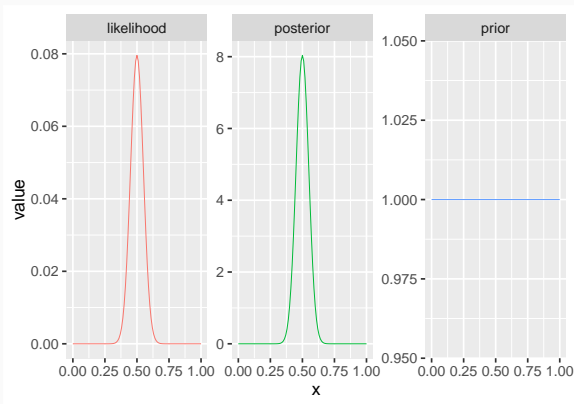
Strong Priors

Say, I had a stronger prior...

Bayesian Workflow

Bayesian inference and modeling techniques can be applied across the board

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

Bayes requires you to be more deliberate

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

$$P(y_i = 1 | income, gender_i) \sim logit^{-1}(\mu_i)$$

Where,

$$\mu = normal(\beta_1 * income + \beta_2 * gender + intercept, \sigma)$$

Prior
Additionally say that we believe that the impact of these two metrics aren't too strong

$$\beta_1 \sim N(0, 0.5)$$
$$\beta_1 \sim N(0, 0.5)$$

Simulate Your Data Generating Process

```
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

Implementation of Bayesian Data Analysis can be done manually...

But there exist domain specific languages to handle most cases:

- BUGS
- JAGS
- Stan
- Hand coded samplers

brms makes Bayesian modeling easy

Compiles traditional R and lme4 syntax to Stan

Utilises Hamiltonian Monte Carlo with a No U-Turn Sampler

Models can specified in-line or separately using the `bf` function

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

Inspect Priors

The get_prior function allows the user to see what priors can be
specified

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

Specify Priors

Priors can then be specified using the existing disribution families

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

Model Family

```
fit <- brm(model, my_priors,
           data =  dat,
           family = bernoulli(),
           inits = 1000, cores = 2,
           chains = 2, seed = 1234, refresh = 0)
```

Posterior Checks

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

So Let's Check

With trace plots we are looking to ensure that there are not divergent chains

```
library(tidybayes)
library(bayesplot)
mcmc_trace(as.matrix(fit))
```
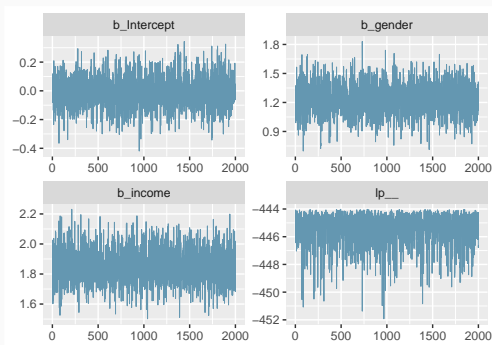
Rhat and Effective Sample Size

Rhat is a measure of chain missing (target < 1.01)

```
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 Eff.Sample Rha
## Intercept     -0.01      0.11    -0.22     0.21       1926 1.0
## gender         1.24      0.16     0.92     1.56       1488 1.0
## income         1.86      0.12     1.63     2.09       1388 1.0
##
## Samples were drawn using sampling(NUTS). For each parameter,
## is a crude measure of effective sample size, and Rhat is the
```
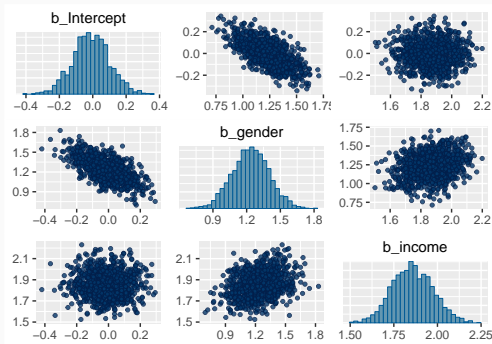
Convergence of Parameters

Here looking for the centeredness of our parameters

```
pairs(fit)
```

Posterior Predictive Checks

Not covered here, but more common when predicting continuous variables

Assess how well your model can predict the data that generated it.

See this tutorial for details

Inferences

Now that we have validated our fit, we can make 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 Eff.Sample Rha
## Intercept    -0.01      0.11    -0.22     0.21       1926 1.0
## gender        1.24      0.16     0.92     1.56       1488 1.0
## income        1.86      0.12     1.63     2.09       1388 1.0
##
## Samples were drawn using sampling(NUTS). For each parameter,
## is a crude measure of effective sample size, and Rhat is the
```
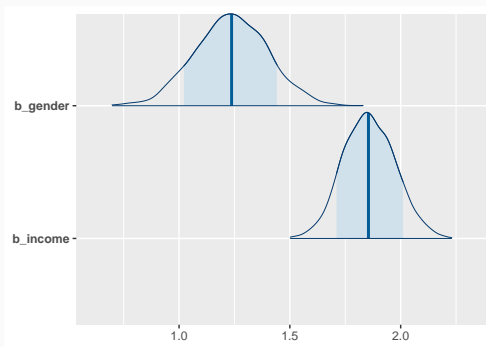
Inferences on Parameters

Posterior distribution of parameters

```
mcmc_areas(as.matrix(fit),
           pars = c("b_gender", "b_income"),
           prob = .8)
```

Convergence

If your model takes a long time to fit OR doesn't have good convergence. . .

- Increase number of iterations
- Change your `max_treedepth` and `adapt_delta`
- Add stronger priors
- Increase the "thinning"
- (re)-Scale parameters
- You may have a poorly fitting model

Drawbacks of Bayesian Inference

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

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 replication or another study

Check out https://michaeldewittjr.com/resources/ under the *Stan* tab for more worked examples