Teaching and Learning Bayesian Statistics with {bayesrules}

bit.ly/dogucu-talks
Southern California R Users Group and R Ladies Irvine

Mine Dogucu

2021-10-26

MineDogucu

mdogucu

minedogucu.com

Alicia Johnson

Macalester College

Miles Ott

Smith College

[Bayes Rules! An Introduction to Bayesian Modeling with R]

[{bayesrules}]

Who are you?

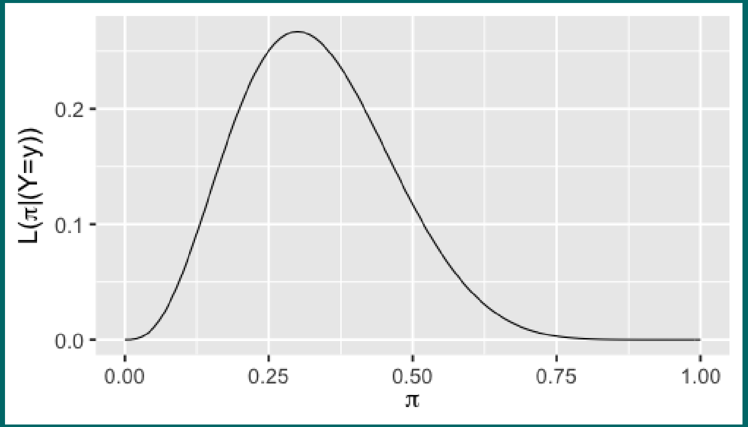
A quick example

Let π be the proportion of spam emails where $\pi \in [0, 1]$.

What do you think π is? How certain are you?

Binomial Likelihood

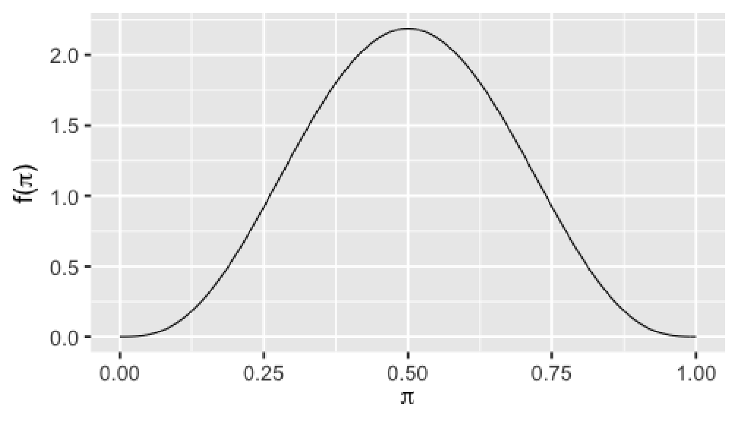
```
plot_binomial_likelihood(y = 3, n = 10)
```

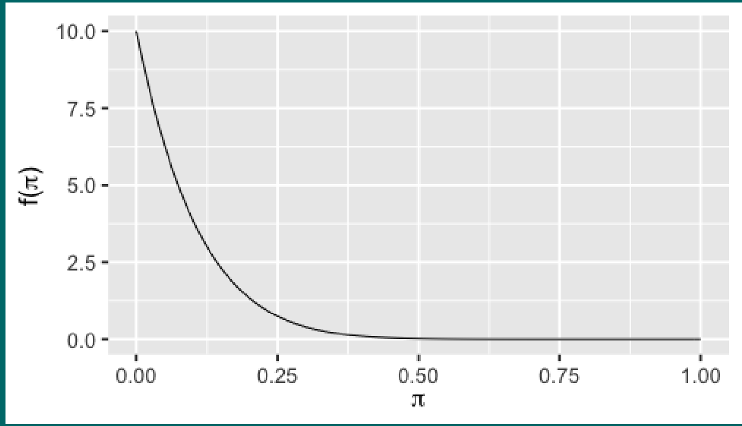


Prior Model

```
plot_beta(alpha = 4, beta = 4)
```

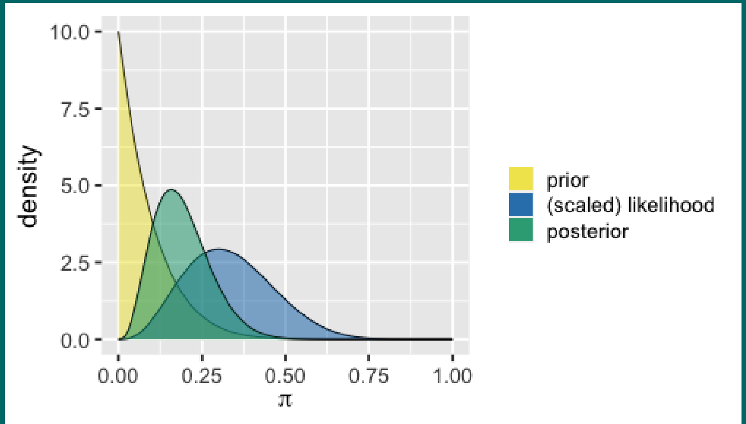
```
plot_beta(alpha = 1, beta = 10)
```





Posterior Model

```
plot_beta_binomial(alpha = 1, beta = 10, y = 3, n = 10)
```



`

Target Audience of the Book

- Advanced Undergraduate Students in Statistics / Data Science Programs
- Equally trained learners
- Prior course/training in statistics is required
- Familiarity with probability, calculus, and tidyverse is recommended.

Our Motivation

- Bayesian methods are becoming more popular due to computing advances and reevaluation of subjectivity.
- Lack of resources for the target audience.

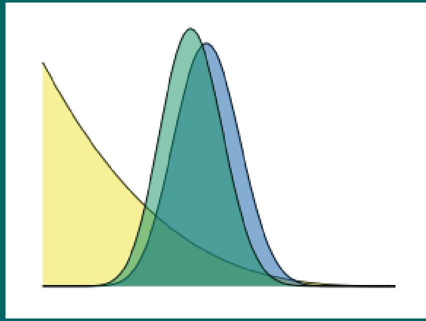
Unit 1

Unit 2

Unit 3

Unit 4

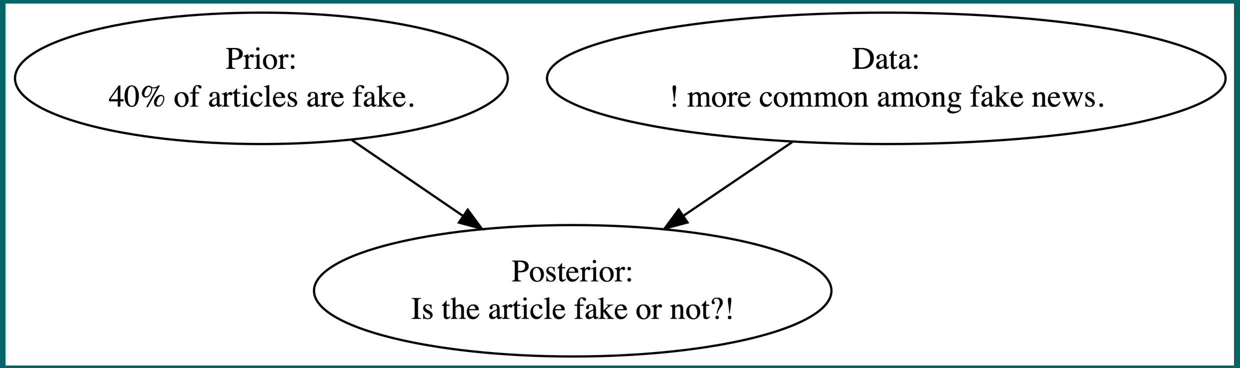
Bayesian Foundations



- Bayes' Rule
- The Beta-Binomial Bayesian Model
- Balance and Sequentiality in Bayesian Analysis
- Conjugate Families

Pedagogical Approach

Checking Intuition



Active Learning

Quizzes

Quiz Yourself

Hands-on Programming

Metropolis-Hastings Algorithm

Computing and Math Together

Compute for a Single Case

Then Use Built-In Functions

Accessibility and Inclusion

Accessibility and Inclusion Criteria	Questions
Accessibility	Is the cost affordable for learners from diverse socioeconomic backgrounds?
	Are plots distinguishable to color blind learners?
	Is alt text provided for images?

Accessibility and Inclusion Criteria	Questions
Inclusivity of scholars	Do the cited scholars represent diversity across identities, experiences, and expertise?
	Are scholars cited using the correct names and pronouns?

Accessibility and Inclusion Criteria	Questions
Inclusivity of students	Do examples avoid the necessity of specialized knowledge?
	Do names and pronouns reflect diverse cultural and personal identities?
	Are there examples that could potentially speak to younger as well as older students?
	Does the delivery embrace mistakes and critical thinking?
	Are efforts made to accommodate different academic experiences and create a shared foundation?

More on accessibility and inclusion is available as a preprint

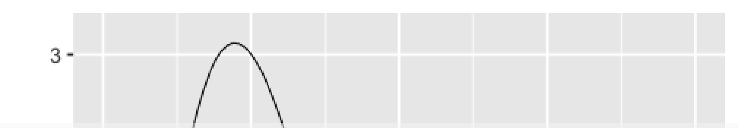
R packages

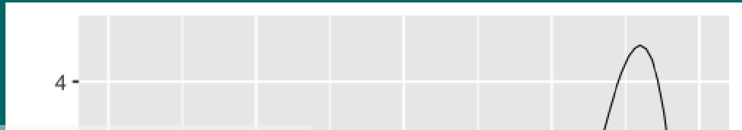
[library(bayesrules)]

```
devtools::install_github("bayes-rules/bayesrules")
```

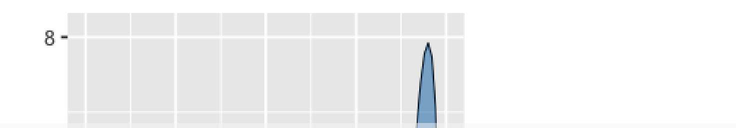
```
plot_beta(alpha = 3, beta = 8)
```

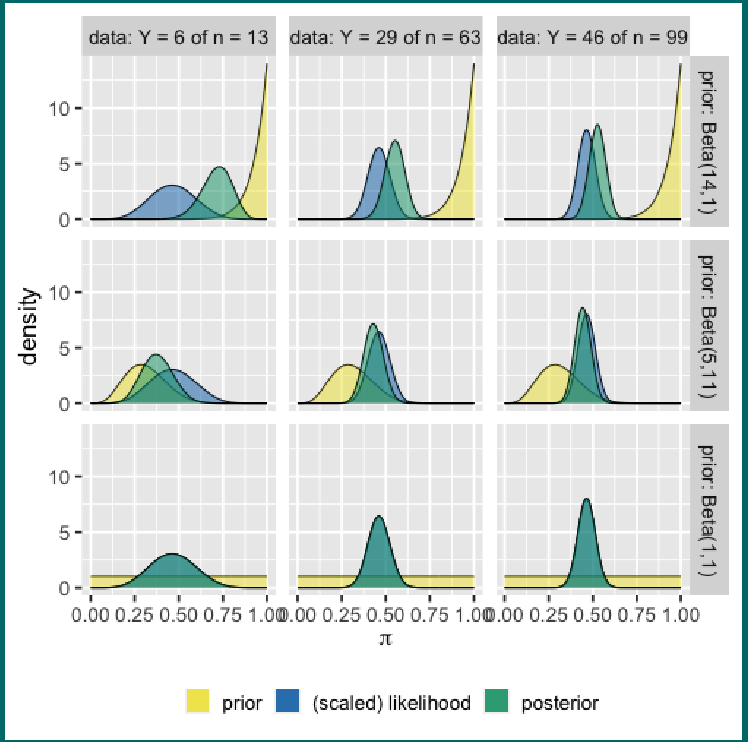
```
plot_beta(alpha = 10, beta = 2)
```





```
plot_beta_binomial(alpha = 3, beta = 8, y = 19, n = 20)
```




data: Y = 6 of n = 13
data: Y = 29 of n = 63
data: Y = 46 of n = 99
prior: Beta(14,1)
prior: Beta(5,11)
prior: Beta(1,1)
density
π
prior
(scaled) likelihood
posterior

Plotting Functions

`plot_beta()`

`plot_binomial_likelihood()`

`plot_beta_binomial`

`plot_gamma()`

`plot_poisson_likelihood()`

`plot_gamma_poisson()`

`plot_normal()`

`plot_normal_likelihood()`

`plot_normal_normal()`

Summary Functions

`summarize_beta()`

`summarize_beta_binomial()`

`summarize_gamma()`

`summarize_gamma_poisson()`

`summarize_normal_normal()`

Model Evaluation Functions

Functions	Response	Model Type
`prediction_summary()` `prediction_summary_cv()`	Quantitative	rstanreg
`classification_summary()` `classification_summary_cv()`	Binary	rstanreg
`naive_classification_summary()` `naive_classification_summary_cv()`	Categorical	naiveBayes

Prediction Summary

```
prediction_summary(model, data,
                   prob_inner = 0.6,
                   prob_outer = 0.80)
```

```
       mae mae_scaled within_60 within_80
1 3.499055  0.5628169      0.75      0.85
```

```
prediction_summary_cv(model = model,
                      data = data,
                      k = 2,
                      prob_inner = 0.6,
                      prob_outer = 0.80)
```

```
$folds
  fold      mae mae_scaled within_60 within_80
1    1 3.628639  0.5984213       0.8       0.8
2    2 3.138409  0.3751545       0.8       0.9

$cv
       mae mae_scaled within_60 within_80

1 3.383524  0.4867879       0.8      0.85
```

```r
library(rstan)
```

```r
# STEP 1: DEFINE the model
stan_bike_model <- "
  data {
    int<lower=0> n;
    vector[n] Y;
    vector[n] X;
  }
  parameters {
    real beta0;
    real beta1;
    real<lower=0> sigma;
  }
  model {
    Y ~ normal(beta0 + beta1 * X, sigma);
  }
"
```

```r
# STEP 2: SIMULATE the posterior
stan_bike_sim <-
  stan(model_code = stan_bike_model,
  data = list(n = nrow(bikes),
              Y = bikes$rides, X = bikes$temp_feel),
  chains = 4, iter = 5000*2, seed = 84735)
```

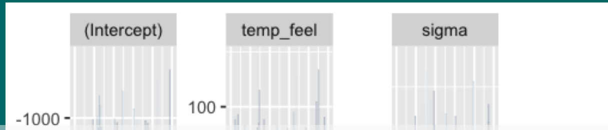
```
library(rstanarm)
```

```
normal_model_sim <- stan_glm(rides ~ temp_feel,
                             data = bikes,
                             family = gaussian,
                             chains = 4, iter = 5000*2,
                             seed = 84735)
```

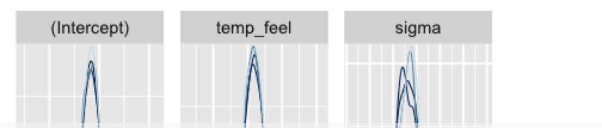
`

```
library(bayesplot)
```

```
mcmc_trace(normal_model_sim, size = 0.1)
```




```
mcmc_dens_overlay(normal_model_sim)
```



Resources

- Undergraduate Bayesian Education Resources
- Undergraduate Bayesian Education Network
- STATS 115 at UC Irvine

SoCal Data Science project

- Collaboration between UC Irvine, Cal State Fullerton, and Cypress College.
- Cal State Fullerton will develop and implement a course similar to UC Irvine's Stats 115 Introduction to Bayesian Data Analysis.
- Through this project our students get to work with real Bayesian and non-Bayesian projects of our academic and industry partners.

HDR DSC awards: #2123366 #2123380 #2123384

Questions?

bit.ly/dogucu-talks

MineDogucu

Images from freepnglogos.com