

gw02-sols

```
library(bayesrules) # R package for our textbook
library(tidyverse) # Collection of packages for tidying and plotting data
library(janitor) # Helper functions like tidy and tabyl
```

1 BR Exercise 2.13

See handwritten

```
pi <- c(.4, .5, .6, .7)
prior <- c(.1, .2, .44, .26)
lik <- dbinom(47, size = 80, prob = pi)
post_unnorm <- prior*lik
post_norm <- prior*lik/sum(post_unnorm)

tibble(pi, prior, lik, post_unnorm, post_norm)
```

```
# A tibble: 4 x 5
      pi prior    lik post_unnorm post_norm
  <dbl> <dbl>   <dbl>         <dbl>    <dbl>
1  0.4  0.1 0.000301 0.0000301 0.000649
2  0.5  0.2 0.0264    0.00527    0.114
3  0.6  0.44 0.0880    0.0387    0.834
4  0.7  0.26 0.00929   0.00242    0.0520
```

2 BR Exercise 2.18

The smallest Y where $P(\pi = .6|Y) > .4$ is 6

```
tibble(
  y = 0:10, # possible values of Y
  p_y = dbinom(y, size = 10, prob = .6)) |> # f(Y|pi = .6)
  rowwise() |>
  mutate(
    norm_constant = (.25*dbinom(y, 10, .2) + .5*dbinom(y, 10, .4) + .25*dbinom(y, 10, .6)), # N
    posterior = (.25*p_y)/norm_constant # posterior: f(pi = .6 | Y)
  )
```

A tibble: 11 x 4

Rowwise:

| | y | p_y | norm_constant | posterior |
|----|-------|----------|---------------|-----------|
| | <int> | <dbl> | <dbl> | <dbl> |
| 1 | 0 | 0.000105 | 0.0299 | 0.000877 |
| 2 | 1 | 0.00157 | 0.0877 | 0.00449 |
| 3 | 2 | 0.0106 | 0.139 | 0.0191 |
| 4 | 3 | 0.0425 | 0.168 | 0.0630 |
| 5 | 4 | 0.111 | 0.175 | 0.159 |
| 6 | 5 | 0.201 | 0.157 | 0.319 |
| 7 | 6 | 0.251 | 0.120 | 0.523 |
| 8 | 7 | 0.215 | 0.0752 | 0.715 |
| 9 | 8 | 0.121 | 0.0356 | 0.850 |
| 10 | 9 | 0.0403 | 0.0109 | 0.928 |
| 11 | 10 | 0.00605 | 0.00156 | 0.966 |

3 BR Exercise 2.17

```
# Define possible trees
trees <- data.frame(type = c("mold", "not mold"))

# Define the prior model
prior <- c(0.18, .82)

# Simulate 10000 trees
set.seed(84735)
trees_sim <- sample_n(trees, size = 10000,
                      weight = prior, replace = TRUE)

trees_sim <- trees_sim %>%
  mutate(data_model = case_when(type == "mold" ~ 0.8,
                                type == "not mold" ~ 0.1))
```

```

# Define tree species
data <- c("maple", "not maple")

# Simulate species

trees_sim <- trees_sim %>%
  group_by(1:n()) %>%
  mutate(species = sample(data, size = 1,
                          prob = c(data_model, 1-data_model)))

trees_sim %>%
  filter(species == "maple") %>%
  tabyl(type) %>%
  adorn_totals("row")

```

| | type | n | percent |
|----------|------|-----------|---------|
| mold | 1482 | 0.6327925 | |
| not mold | 860 | 0.3672075 | |
| Total | 2342 | 1.0000000 | |

4 MLE

If $Y \sim \text{Binom}(n, \pi)$, show that $\hat{\pi}_{MLE} = \frac{Y}{n}$

(see handwritten)

5 BR Exercise 3.1

For each part, use `plot_beta` and/or `summarize_beta` to justify your answer

6 BR Exercise 3.12

7 Choice of prior

I am interviewing Carleton students about whether or not they have used (knowingly) used ChatGPT on coursework in a non-approved way. I think the proportion has a 90% chance of being less than .25.

(a) Choose an informative prior that you think is reasonable for this belief

I then ask 20 students this question and 15 respond “yes”. Find the posterior using the prior from above, then using the 3 non/weakly informative priors below:

(a) $\text{Unif}(0,1)$ prior

(b) $\text{Beta}(2,2)$ prior

(c) “Reference” prior $\text{Beta}(.5, .5)$

Compare the posteriors for each of the priors above. Do results change if we instead observe 150/200 students responding “yes”?

8 BR 4.13

(see handwritten)