

Assignment-2 Q&A Bilge Salman (03796071), Salim Kaplan (0378856) GitHub link: <https://github.com/salimkaplan/BayesIntro24-Assignments/tree/main>

```
# load packages here
library(tidyverse)

-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(tibble)
library(ggplot2)
theme_set(theme_light())
```

Task Set 1

For tasks 1.1-1.2, suppose there are 3 companies, Company A to C. Company A has a customer satisfaction rate of .70, Company B of .50, and Company C of .80. Further suppose that you receive 10 customer reviews (6 positive, 4 negative) for the same company, but you don't

know for which company. Assume that Company B is twice as likely to obtain reviews than Company A and C.

Task 1.1

Show that the posterior probability that Company A was rated is ≈ 0.29 .

```
# Assuming that Company B is twice as likely to obtain reviews than Company A and C.
# Prior probabilities
prior_A <- 0.2
prior_B <- 0.4
prior_C <- 0.2

# Likelihoods using dbinom (6 positive, 4 negative)
likelihood_A <- dbinom(6, 10, 0.7)
likelihood_B <- dbinom(6, 10, 0.5)
likelihood_C <- dbinom(6, 10, 0.8)

# Updating & Standardization
sum_posterior <- sum(likelihood_A * prior_A, likelihood_B * prior_B, likelihood_C * prior_C)

posterior_A <- (likelihood_A * prior_A) / sum_posterior
posterior_B <- (likelihood_B * prior_B) / sum_posterior
posterior_C <- (likelihood_C * prior_C) / sum_posterior

paste("Posterior probability that Company A is approximately", round(posterior_A, 2))
```

```
[1] "Posterior probability that Company A is approximately 0.29"
```

Task 1.2

Suppose you receive 10 more reviews (9 positive and 1 negative). Show that the posterior probability that Company C received the reviews increases by ≈ 33 percentage points, when considering all 20 rather than only the first 10 reviews. To obtain the updated posterior, compute the likelihood of the 10 most recent reviews only.

```
# New likelihoods using dbinom (9 positive and 1 negative)
likelihood_A_new <- dbinom(9, 10, 0.7)
likelihood_B_new <- dbinom(9, 10, 0.5)
likelihood_C_new <- dbinom(9, 10, 0.8)
```

```

# Updating
prior_A <- posterior_A
prior_B <- posterior_B
prior_C <- posterior_C

# Updating & Standardization
sum_posterior_new <- sum(likelihood_A_new * prior_A, likelihood_B_new * prior_B,
                        likelihood_C_new * prior_C )

posterior_A_new <- (likelihood_A_new * prior_A) / sum_posterior_new
posterior_B_new <- (likelihood_B_new * prior_B) / sum_posterior_new
posterior_C_new <- (likelihood_C_new * prior_C) / sum_posterior_new

paste("Company C received the reviews increases by", round((posterior_C_new - posterior_C)

```

```

[1] "Company C received the reviews increases by 0.33"

```

Task Set 2

For tasks 2.1 and 2.2, suppose there are Factory A and Factory B, producing the same product. The company C receives equally many shipments from both factories. Even though the machines, processes, and standards are virtually identical, the factories differ in their defect rates. Shipments from Factory A entail defective products 10% of the time, shipments from Factory B entail defective products 20% of the time.

Task 2.1

You receive a shipment from one of the factories, and upon inspection, you find that the shipment contains defective products. Compute the probability that the next shipment from this company will also contain defective products.

```
# Prior probabilities
prior_A <- 0.5
prior_B <- 0.5

# Likelihoods using defect rates
likelihood_A <- 0.1
likelihood_B <- 0.2

# Updating & Standardization
sum_posterior <- sum(likelihood_A * prior_A, likelihood_B * prior_B)

posterior_A <- (likelihood_A * prior_A) / sum_posterior
posterior_B <- (likelihood_B * prior_B) / sum_posterior

# Calculate the probability that the next shipment will also contain defective products
probability <- sum(posterior_A * likelihood_A, posterior_B * likelihood_B)

# Print the probability
paste("Probability that the next shipment contain defective products is", probability)
```

```
[1] "Probability that the next shipment contain defective products is 0.166666666666667"
```

Task 2.2

Suppose the R&D department came up with a Machine Learning algorithm that (imperfectly) identifies the factory based on the shipped products. But the classification algorithm is imperfect. This is the information you have about the algorithm:

- The probability it correctly identifies a Factory A product is 93%.
- The probability it correctly identifies a Factory B product is 87%.

When applying the the algorithm to the shipped products, the test is positive for Factory A. Including the defect data from 2.1, compute the posterior probability that your shipment is from Company A.

```
# New likelihoods
likelihood_A_new <- 0.93

# likelihood of the shipped products being from Factory B
# 7% chance of incorrectly classifying products from Factory B as Factory A
likelihood_B_new <- 1 - 0.87

# Updating
prior_A <- posterior_A
prior_B <- posterior_B

# Updating & Standardization
sum_posterior_new <- sum(likelihood_A_new * prior_A, likelihood_B_new * prior_B)

posterior_A_new <- (likelihood_A_new * prior_A) / sum_posterior_new

paste("Posterior probability that the shipment is from Company A", posterior_A_new)
```

```
[1] "Posterior probability that the shipment is from Company A 0.781512605042017"
```

Task Set 3

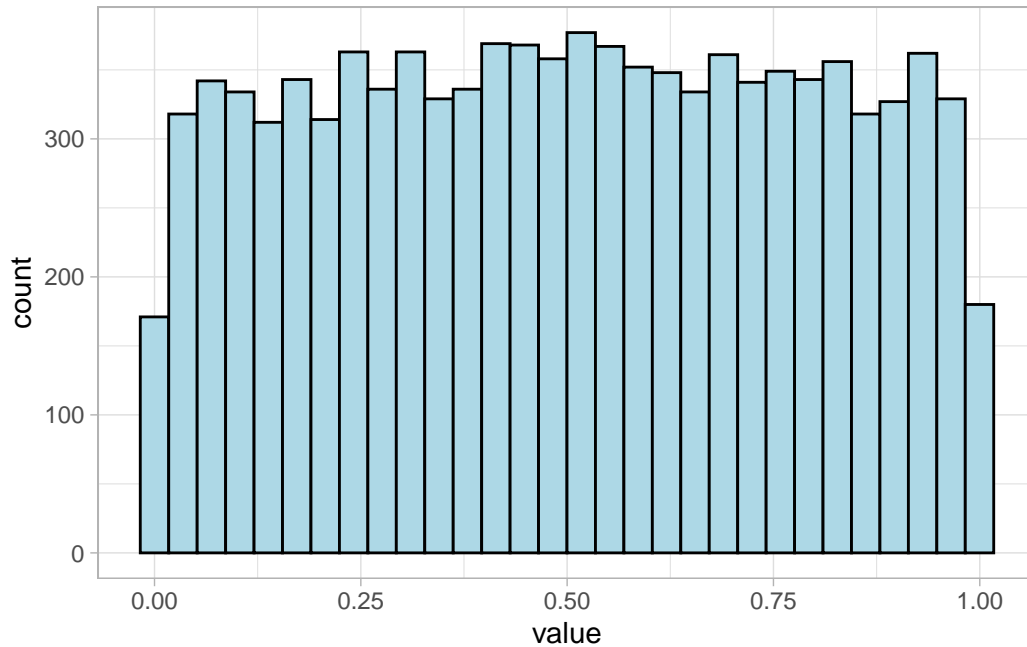
For Task 3.1 and 3.2, suppose, one last time, you want to estimate the proportions of land on the earth's.

Task 3.1

Specify a prior distribution and store 10,000 random samples from it in a vector `sample`. Plot the prior distribution and briefly explain your choice of the prior.

```
# Assuming we have no prior information about the phenomenon,  
# we will use beta distribution with parameters 1,1  
# which is equivalent to uniform distribution.  
  
# 10k random samples from the prior distribution  
sample <- rbeta(10000, 1, 1) %>%  
  as_tibble()  
  
sample %>%  
  ggplot() +  
  geom_histogram(aes(value), fill = "lightblue", color = "black")
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.



Task 3.2

Run the following code chunk that uses your object sample to obtain prior probabilities for the possible proportions of land 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1 that approximate your prior distribution.

```
prop <- seq(0, 1, length.out = 12)
priors <- vector("numeric", length(prop))
for (i in seq_along(prop)){
  priors[i] <- round( sum(sample >= prop[i] & sample < prop[i+1]) / 1e4 , 2)
}
poss <- tibble(prop_L = seq(0, 1, .1), prior = priors[1:11])
```

Use these priors to compute the posterior probability after observing 26 times land in 100 globe tosses. Take 1,000 samples from the posterior distribution and with each sample, predict the outcome of 100 globe tosses. Plot the posterior predictions in a histogram.

```
likelihood <- dbinom(26, size = 100, prob = poss$prop_L)

# Calculate the evidence
evidence <- sum(likelihood * poss$prior)
```

```

# Calculate the posterior probabilities
poss$posterior <- likelihood * poss$prior / evidence

# samples from the posterior distribution
samples <- sample(poss$prop_L, size = 1000, replace = TRUE, prob = poss$posterior)

# using samples to predict the outcome of 100 globe tosses
predictions <- rbinom(1000, size = 100, prob = samples) %>%
  as_tibble()

# final plot for histogram
ggplot() +
  geom_histogram(data = predictions, aes(value), fill = "lightblue", color = "black") +
  labs(x = "Number of Times Land",
       y = "Frequency")

```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

