

Experiments, Conditional Probability, and Bayes' Theorem

Hadijat Oke 117818512 Thomas Urdinola 117971262

Contents

Simulating Experiments	1
Die Rolls	1
Coin Tosses for Binomial Distribution	4
Coin Tosses for Negative Binomial Distribution.	8
Simulating the Monty-Hall Problem	10
Monty-Hall Three doors	10
Monty-Hall with Ten doors.	12
Monty-Hall 10-doors (modified).	14
(Bonus) Monty Hall with n-doors.	16
Bayes's Theorem and Witnesses	17
Reliable Witnesses with fixed probability	17
Working with two types of witnesses	20
Working with four types of witnesses	22
BONUS: Working with witnesses coming from a Beta distribution	24

Simulating Experiments

In this problem we will simulate two important experiments: coin tosses and die rolls.

We will explore empirical probability of events using random samples. The two important functions to understand are “sample” and “replicate”.

We use the sample function to set up our experiment, and replicate function to repeat it as many times as we want.

Die Rolls

Setup the experiment where we roll two fair six 6-sided die independently, and calculate the sum of the numbers that appear for each of them.

```
roll_dice <- function(){
  die1 <- sample(1:6, 1, replace = TRUE)
  die2 <- sample(1:6, 1, replace = TRUE)
  return(die1 + die2)
}

roll_dice()
```

```
## [1] 9
```

Use the replicate function to repeat this experiment 10000 times, store the output of the experiment in a variable “sum_die”.

```
sum_die <- replicate(1000,roll_dice())
sum_die
```

```
##      [1] 10 7 8 6 8 8 10 10 7 5 4 9 10 9 4 6 7 8 10 8 8 10 11 6
##      [25] 6 8 11 11 5 9 9 9 2 7 11 10 8 7 3 5 8 6 10 9 8 4 6 5
##      [49] 12 10 3 8 8 9 6 6 6 6 5 9 11 3 3 6 9 7 11 10 10 11 8 11
##      [73] 10 3 4 9 8 11 10 3 10 2 10 7 9 11 6 5 6 7 7 2 5 10 9 10
##      [97] 7 8 9 5 9 8 7 7 7 10 6 7 6 8 6 5 11 7 7 5 9 11 7 6
##     [121] 9 4 6 10 5 8 8 3 4 3 6 12 9 7 7 6 7 4 5 9 4 2 10 7
##     [145] 6 7 6 6 10 6 8 7 5 10 7 5 8 4 2 10 12 10 10 10 10 7 6 6
##     [169] 7 7 3 7 6 4 3 4 3 5 7 6 3 7 6 7 11 5 10 10 4 9 3 11
##     [193] 6 3 6 4 9 6 8 8 10 7 7 7 4 7 9 8 7 9 5 6 6 10 4 6
##     [217] 8 6 5 9 10 10 9 6 9 11 8 8 10 9 8 9 10 3 7 7 11 5 6 10
##     [241] 8 7 10 10 10 6 3 7 9 6 8 9 3 9 2 5 7 6 6 9 6 6 6 7
##     [265] 3 6 3 5 6 6 4 12 7 10 8 5 11 6 5 2 11 9 9 7 9 11 6 9
##     [289] 4 11 12 2 8 4 5 7 8 6 7 6 11 8 7 8 10 9 10 3 6 5 4 6
##     [313] 8 9 7 4 7 7 7 3 6 5 4 3 11 4 10 7 8 4 7 7 8 3 8 9
##     [337] 9 2 10 12 6 7 7 6 7 8 6 3 10 5 7 8 7 8 8 6 11 7 4 6
##     [361] 7 6 7 3 7 7 7 6 9 3 6 7 7 10 6 5 10 7 8 7 2 4 5 10
##     [385] 7 6 9 6 7 8 10 4 10 5 7 9 10 11 7 4 7 5 5 4 8 5 9 3
##     [409] 2 10 8 8 6 4 6 8 5 5 5 3 10 6 8 10 7 5 7 8 7 9 7 9
##     [433] 7 6 4 4 5 8 3 6 6 7 10 6 3 7 4 7 11 7 6 6 9 4 4 7
##     [457] 9 11 9 8 10 6 12 7 7 6 6 9 6 9 10 6 8 5 5 8 6 7 3 4
##     [481] 8 8 8 6 9 9 6 6 7 10 7 10 7 6 5 9 7 7 8 9 10 9 7 4
##     [505] 4 11 10 4 5 6 10 8 9 6 7 7 8 10 9 6 10 3 6 7 5 8 6 7
##     [529] 8 9 7 2 8 3 10 8 5 7 4 9 7 8 5 8 8 5 6 9 10 6 9 5
##     [553] 11 5 9 5 4 6 11 8 9 9 6 7 5 5 10 10 8 7 11 10 5 7 7 9
##     [577] 11 7 5 7 6 5 6 6 10 7 11 6 10 6 4 12 9 9 7 3 7 11 7 9
##     [601] 8 7 5 7 4 7 8 7 5 5 6 4 11 5 8 6 8 8 4 6 9 8 3 6
##     [625] 2 6 7 6 8 5 7 5 8 7 8 6 7 9 8 10 9 7 4 7 9 8 9 6
##     [649] 6 9 6 9 6 9 7 9 8 7 8 2 5 11 6 8 2 6 10 10 7 9 10 3
##     [673] 4 7 3 10 2 8 5 8 6 2 8 6 7 7 10 3 9 12 5 5 7 9 5 9
##     [697] 11 8 3 8 3 6 6 9 8 11 9 8 9 5 7 8 8 7 4 8 3 5 5 5
##     [721] 9 8 7 8 7 11 6 7 6 4 6 6 7 7 8 6 7 6 3 4 6 3 11 11
##     [745] 8 7 6 8 2 10 6 9 8 8 3 12 5 7 12 5 8 11 12 7 6 9 7 11
##     [769] 9 5 11 7 6 6 12 10 8 11 8 7 8 5 2 7 7 9 2 5 8 10 7 12
##     [793] 7 6 7 3 5 10 9 7 2 5 6 11 6 3 4 4 9 6 3 5 2 3 8 8
##     [817] 5 5 6 7 11 9 10 10 12 10 5 8 6 4 5 11 7 5 6 8 5 11 7 8
##     [841] 5 6 11 8 5 7 8 5 10 7 3 5 9 10 8 6 9 7 12 6 5 6 6 4
```

```
## [865] 7 8 11 9 11 5 5 4 7 10 6 9 10 9 4 4 5 2 5 5 3 5 6 3
## [889] 9 7 9 11 7 9 7 9 5 6 9 7 3 7 11 3 9 6 5 7 11 7 6 7
## [913] 8 8 4 10 6 10 2 8 7 8 10 5 6 8 5 7 8 8 10 8 6 5 8 9
## [937] 11 9 7 10 9 3 9 8 7 9 6 6 7 10 7 7 4 7 5 6 6 8 4 9
## [961] 6 4 2 6 9 11 4 5 4 10 6 5 6 6 6 7 10 9 4 6 10 9 6 8
## [985] 11 5 12 4 7 8 11 12 6 6 2 6 6 8 4 10
```

Use “sum_die” to calculate the empirical probability of getting the sums 2, 3, 4, ..., 12. Compare your answers with the theoretical values.

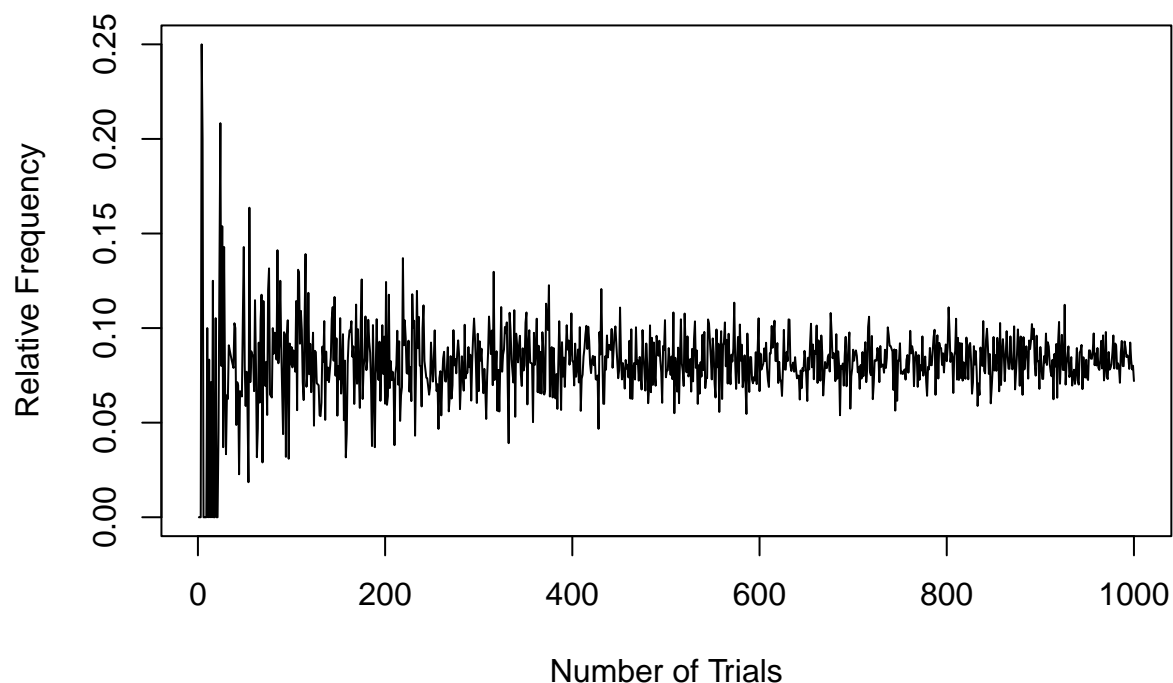
```
X <- 1000
sum_die <- replicate(X, roll_dice)
sum_sample <- sample(c(1,2,3,4,5,6), size = X, replace = TRUE)
sum(sum_sample[1]+sum_sample[2])
```

```
## [1] 4
```

Plot the following: relative frequency of getting sum equal to 4 as a function of number of trials, for this experiment.

```
num_trials <- 1000
relative_frequencies <- numeric(num_trials)
for (i in 1:num_trials) {
  outcomes <- replicate(i, roll_dice())
  relative_frequencies[i] <- sum(outcomes == 4) / i
}
plot(1:num_trials, relative_frequencies, type = "l", xlab = "Number of Trials", ylab = "Relative Frequency")
```

Relative Frequency of Sum 4 vs. Number of Trials



Coin Tosses for Binomial Distribution

Suppose we toss a coin with $P(H) = 0.8$, n times independently. Let X be the the random variable that counts the number of heads in n tosses.

n = 2

Setup a function that simulates the random variable X when we toss the coin two times.

```
toss <- function(){  
  coin_toss <- sample(c("H", "T"), size = 2, replace = TRUE, prob = c(1/2, 1/2))  
  return(coin_toss)  
}  
  
toss()
```

```
## [1] "T" "T"
```

Use the above function to calculate the emprical probability distribution table for X . Compare this with the theoretical probability distribution of X .

```
num_trials <- 10000
```

```

# Initialize a table to store the counts of each outcome
outcome_table <- table(replicate(num_trials, paste(toss(), collapse = "")))

# Calculate empirical probabilities
empirical_probs <- outcome_table / sum(outcome_table)

# Create a data frame for the probability distribution table
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)

probability_table

```

```

##      Outcome Probability.Var1 Probability.Freq
## 1      HH              HH          0.2440
## 2      HT              HT          0.2582
## 3      TH              TH          0.2495
## 4      TT              TT          0.2483

```

Calculate empirical estimates for the expected value and the variance of X . Compare these with the actual theoretical calculations.

```

num_trials <- 10000

sum_outcomes <- 0
sum_squares <- 0

# Simulate coin tosses and calculate the sum and sum of squares
for (i in 1:num_trials) {
  result <- sum(as.integer(toss() == "H")) # Count "H" as 1, "T" as 0
  sum_outcomes <- sum_outcomes + result
  sum_squares <- sum_squares + result^2
}

# Calculate empirical estimates for the expected value and variance
mean_estimate <- sum_outcomes / num_trials
variance_estimate <- (sum_squares / num_trials) - (mean_estimate^2)

mean_estimate

```

```
## [1] 0.9989
```

```
variance_estimate
```

```
## [1] 0.5024988
```

$n = 3$

Setup a function that simulates the random variable X when we toss the coin three times.

```

toss <- function(){
  coin_toss <- sample(c("H", "T"), size = 3, replace = TRUE, prob = c(1/2, 1/2))
  return(coin_toss)
}

```

```
}
```

```
toss()
```

```
## [1] "T" "T" "H"
```

Use the above function to calculate the empirical probability distribution table for X . Compare this with the theoretical probability distribution of X .

```
num_trials <- 10000
```

```
# Initialize a table to store the counts of each outcome
```

```
outcome_table <- table(replicate(num_trials, paste(toss(), collapse = "")))
```

```
# Calculate empirical probabilities
```

```
empirical_probs <- outcome_table / sum(outcome_table)
```

```
# Create a data frame for the probability distribution table
```

```
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)
```

```
probability_table
```

```
##      Outcome Probability.Var1 Probability.Freq
## 1      HHH              HHH          0.1252
## 2      HHT              HHT          0.1240
## 3      HTH              HTH          0.1266
## 4      HTT              HTT          0.1288
## 5      THH              THH          0.1281
## 6      THT              THT          0.1224
## 7      TTH              TTH          0.1225
## 8      TTT              TTT          0.1224
```

Calculate empirical estimates for the expected value and the variance of X . Compare these with the actual theoretical calculations.

```
num_trials <- 10000
```

```
sum_outcomes <- 0
```

```
sum_squares <- 0
```

```
# Simulate coin tosses and calculate the sum and sum of squares
```

```
for (i in 1:num_trials) {
```

```
  result <- sum(as.integer(toss() == "H")) # Count "H" as 1, "T" as 0
```

```
  sum_outcomes <- sum_outcomes + result
```

```
  sum_squares <- sum_squares + result^2
```

```
}
```

```
# Calculate empirical estimates for the expected value and variance
```

```
mean_estimate <- sum_outcomes / num_trials
```

```
variance_estimate <- (sum_squares / num_trials) - (mean_estimate^2)
```

```
mean_estimate
```

```
## [1] 1.5118
```

```
variance_estimate
```

```
## [1] 0.7406608
```

```
n = 4
```

Setup a function that simulates the random variable X when we toss the coin four times.

```
toss <- function(){  
  coin_toss <- sample(c("H", "T"), size = 4, replace = TRUE, prob = c(1/2, 1/2))  
  return(coin_toss)  
}  
  
toss()
```

```
## [1] "T" "H" "T" "H"
```

Use the above function to calculate the empirical probability distribution table for X . Compare this with the theoretical probability distribution of X .

```
num_trials <- 10000  
  
# Initialize a table to store the counts of each outcome  
outcome_table <- table(replicate(num_trials, paste(toss(), collapse = "")))  
  
# Calculate empirical probabilities  
empirical_probs <- outcome_table / sum(outcome_table)  
  
# Create a data frame for the probability distribution table  
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)  
  
probability_table
```

```
##      Outcome Probability.Var1 Probability.Freq  
## 1      HHHH              HHHH             0.0649  
## 2      HHHT              HHHT             0.0627  
## 3      HHTH              HHTH             0.0637  
## 4      HHTT              HHTT             0.0616  
## 5      HTHH              HTHH             0.0631  
## 6      HTHT              HTHT             0.0605  
## 7      HTTH              HTTH             0.0601  
## 8      HTTT              HTTT             0.0613  
## 9      THHH              THHH             0.0585  
## 10     THHT              THHT             0.0638  
## 11     THTH              THTH             0.0615  
## 12     THTT              THTT             0.0629  
## 13     TTHH              TTHH             0.0635  
## 14     TTHT              TTHT             0.0633  
## 15     TTTH              TTTH             0.0631  
## 16     TTTT              TTTT             0.0655
```

Calculate empirical estimates for the expected value and the variance of X . Compare these with the actual theoretical calculations.

```
num_trials <- 10000

sum_outcomes <- 0
sum_squares <- 0

# Simulate coin tosses and calculate the sum and sum of squares
for (i in 1:num_trials) {
  result <- sum(as.integer(toss() == "H")) # Count "H" as 1, "T" as 0
  sum_outcomes <- sum_outcomes + result
  sum_squares <- sum_squares + result^2
}

# Calculate empirical estimates for the expected value and variance
mean_estimate <- sum_outcomes / num_trials
variance_estimate <- (sum_squares / num_trials) - (mean_estimate^2)

mean_estimate
```

```
## [1] 2.01
```

```
variance_estimate
```

```
## [1] 1.0147
```

Coin Tosses for Negative Binomial Distribution.

Suppose we toss a coin with $P(H) = 0.8$ until we get r heads.

$r=1$

Setup an experiment that counts the number of tails until the first head. Run 10000 simulations of this experiment, and use these to find estimate for the expected number of tails until the first head. Match this value with the theoretical/calculated value.

```
count_tails_until_head <- function() {
  tails_count <- 0
  while (TRUE) {
    # Simulate a coin flip (0 for tails, 1 for heads)
    coin_flip <- sample(0:1, 1)
    tails_count <- tails_count + 1
    if (coin_flip == 1) { # Head
      break
    }
  }
  return(tails_count)
}

num_simulations <- 10000
```



```

# Vector to store the results
results <- numeric(num_simulations)

# Run the simulations
for (i in 1:num_simulations) {
  results[i] <- count_tails_until_head()
}
expected_val <- mean(results)
expected_val

```

```
## [1] 1.9817
```

```
r = 2
```

Setup an experiment that counts the number of tails until the second head. Run 10000 simulations of this experiment, and use these to find an estimate for the expected number of tails until the second head. Match this value with the theoretical/calculated value.

```

count_tails_until_second_head <- function() {
  count_tails <- 0
  head_count <- 0

  while (head_count < 2) {
    # Simulate a coin flip (0 for tails, 1 for heads)
    flip <- sample(0:1, 1)

    if (flip == 1) {
      head_count <- head_count + 1
    }

    count_tails <- count_tails + 1
  }

  return(count_tails)
}
num_simulations <- 10000

# Vector to store the results
results <- numeric(num_simulations)

# Run the simulations
for (i in 1:num_simulations) {
  results[i] <- count_tails_until_second_head()
}
expected_val <- mean(results)
expected_val

```

```
## [1] 3.9757
```

r = 3

Setup an experiment that counts the number of tails until the third head. Run 10000 simulations of this experiment, and use these to find an estimate for the expected number of tails until the third head. Match this value with the theoretical/calculated value.

```
count_tails_until_second_head <- function() {  
  count_tails <- 0  
  head_count <- 0  
  
  while (head_count < 3) {  
    # Simulate a coin flip (0 for tails, 1 for heads)  
    flip <- sample(0:1, 1)  
  
    if (flip == 1) {  
      head_count <- head_count + 1  
    }  
  
    count_tails <- count_tails + 1  
  }  
  
  return(count_tails)  
}  
num_simulations <- 10000  
  
# Vector to store the results  
results <- numeric(num_simulations)  
  
# Run the simulations  
for (i in 1:num_simulations) {  
  results[i] <- count_tails_until_second_head()  
}  
expected_val <- mean(results)  
expected_val
```

```
## [1] 5.9926
```

Simulating the Monty-Hall Problem

An interesting application of conditional probability is the Monty-Hall problem, whose solution intrigued many, including career mathematicians (life is hard!). There is a lot of trivia around this problem in the internet, a good place to start would be the Wikipedia article here: https://en.wikipedia.org/wiki/Monty_Hall_problem

In this project we will understand the solution of this problem using simulations.

Monty-Hall Three doors

Run a simulation to check that the probability of winning increases to $2/3$ if we switch doors at step two in the regular 3-door Monty-Hall problem.

Set up the experiment two functions “monty_3doors_noswitch” and “monty_3doors_switch” (these functions will have no input values):

```

monty_3doors_noswitch <- function(){
  prize <- sample(1:3, 1)
  choice = 1
  if (prize == choice){
    1
  } else{
    0
  }
}

monty_3doors_switch <- function(){

  prize <- sample(1:3,1)
  choice = 1

  if(prize == choice){
    notprize = sample(2:3,1)
    if(notprize == 2){
      choice = 3
    } else{
      choice = 2
    }
  } else if (prize == 2){
    notprize = 3
    choice = 2
  } else{
    notprize = 2
    choice = 3
  }

  if(choice == prize){
    1
  } else {
    0
  }

}

monty_3doors_switch()

```

```
## [1] 1
```

Use your two functions and the replicate function to compute the empirical probability of winning for the two experiments. Compare your answers with the actual theoretical predictions.

```

num_trials <- 10000

# Initialize a table to store the counts of each outcome
outcome_table <- table(replicate(num_trials, paste(monty_3doors_switch(), collapse = "")))

# Calculate empirical probabilities
empirical_probs <- outcome_table / sum(outcome_table)

# Create a data frame for the probability distribution table
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)

probability_table

```

```

##      Outcome Probability.Var1 Probability.Freq
## 1         0                0         0.3395
## 2         1                1         0.6605

```

```

num_trials <- 10000

# Initialize a table to store the counts of each outcome
outcome_table <- table(replicate(num_trials, paste(monty_3doors_noswitch(), collapse = "")))

# Calculate empirical probabilities
empirical_probs <- outcome_table / sum(outcome_table)

# Create a data frame for the probability distribution table
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)

probability_table

```

```

##      Outcome Probability.Var1 Probability.Freq
## 1         0                0         0.6665
## 2         1                1         0.3335

```

```

print("The first table is the probabilitiy associated with you switching which results in a win frequency of 0.3395,
      the second table is the probability associared with you now switching which results in a win frequency of 0.3335.
      This reflects the actual theoretical probabilitites.")

```

```

## [1] "The first table is the probabilitiy associated with you switching which results in a win frequency of 0.3395,

```

Monty-Hall with Ten doors.

Repeat the Monty Hall experiment now with 10 doors. Recall the game is as follows: Step 1: you choose one door at random. Step 2: Monty opens 8 (out of 9 doors) that do not have the prize. Step 3: you either switch or don't switch.

Set up the experiment two functions “monty_10doors_noswitch” and “monty_10doors_switch” (these functions will have no input values):

```

monty_10doors_noswitch <- function(){
  prize <- sample(c(1:10), 1)
  choice = 1
  if (prize == choice){
    1
  } else{
    0
  }
}

monty_10doors_switch <- function(){
  prize <- sample(c(1:10), 1)
  choice = 1

  if(prize == choice){
    close<- sample(c(2:10),1)
    choice = close
  } else {
    choice = prize
  }

  if(choice == prize){
    1
  } else {
    0
  }
}

```

Use your two functions and the replicate function to compute the empirical probability of winning for the two experiments. Compare your answers with the actual theoretical predictions (you need to calculate these).

```

num_trials <- 10000

# Initialize a table to store the counts of each outcome
outcome_table <- table(replicate(num_trials, paste(monty_10doors_switch(), collapse = "")))

# Calculate empirical probabilities
empirical_probs <- outcome_table / sum(outcome_table)

# Create a data frame for the probability distribution table
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)

probability_table

```

```

## Outcome Probability.Var1 Probability.Freq
## 1      0              0          0.0975
## 2      1              1          0.9025

```

```

num_trials <- 10000

# Initialize a table to store the counts of each outcome
outcome_table <- table(replicate(num_trials, paste(monty_10doors_noswitch(), collapse = "")))

# Calculate empirical probabilities
empirical_probs <- outcome_table / sum(outcome_table)

# Create a data frame for the probability distribution table
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)

probability_table

```

```

##      Outcome Probability.Var1 Probability.Freq
## 1         0                0          0.8989
## 2         1                1          0.1011

```

```

print("The first table represents the probability of switching after 8 doors are opened. This results
      the second table represents the probability of not switching after 8 doors are opened. This resu
      The accurately aligns with the actual probabilities. For not switching you have to choose the p
      is a 1/10 chance. However, if you picked a door without a prize intially (9/10) then switched y

```

```

## [1] "The first table represents the probability of switching after 8 doors are opened. This results

```

Monty-Hall 10-doors (modified).

Consider the following modified Monty-Hall game with 10 doors. Step 1: you choose one door at random. Step 2: Monty opens 7 (out of 9 doors) that do not have the prize. Step 3: you either stick with your original choice, or choose between one of the two unopened doors.

Set up the experiment two functions “monty_10doors_mod_noswitch” and “monty_10doors_mod_switch” (these functions will have no input values):

```

monty_10doors_mod_noswitch <- function(){
  prize <- sample(c(1:10), 1)
  choice = 1

  if(prize == choice){
    1
  } else {
    0
  }
}

#my_door_and_prize_door <- c(my_door, prize_door)
#not_my_door_nor_prize_door <- doors[!doors %in% my_door_and_prize_door]

monty_10doors_mod_switch <- function(){
  prize <- sample(c(1:10),1)
  choice = 1

```

```

if(prize == choice){
  #you can't win if you choose the prize initially
  0
} else {

  if(prize == 2){
    door_choose <- sample(c(3:10),1)
  } else if(prize == 10){
    door_choose <-sample(c(2:9),1)
  } else {
    end = prize -1
    start = prize +1
    door_choose <-sample(c(2:end, start:10 ), 1)
  }

  choose <- sample(c(door_choose, prize), 1)

  if(prize == choose){
    1
  } else {
    0
  }

}

}

```

Use your two functions and the replicate function to compute the empirical probability of winning for the two experiments. Calculate the theoretical probability of winning a prize if you switch, and if you don't switch. Compare your values to the empirical probabilities.

```

num_trials <- 10000

# Initialize a table to store the counts of each outcome
outcome_table <- table(replicate(num_trials, paste(monty_10doors_mod_switch(), collapse = "")))

# Calculate empirical probabilities
empirical_probs <- outcome_table / sum(outcome_table)

# Create a data frame for the probability distribution table
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)

probability_table

```

```

## Outcome Probability.Var1 Probability.Freq
## 1 0 0 0.5469
## 2 1 1 0.4531

```

```

num_trials <- 10000

# Initialize a table to store the counts of each outcome
outcome_table <- table(replicate(num_trials, paste(monty_10doors_mod_noswitch(), collapse = "")))

# Calculate empirical probabilities
empirical_probs <- outcome_table / sum(outcome_table)

# Create a data frame for the probability distribution table
probability_table <- data.frame(Outcome = names(empirical_probs), Probability = empirical_probs)

probability_table

```

```

##      Outcome Probability.Var1 Probability.Freq
## 1         0                0          0.8948
## 2         1                1          0.1052

```

```

print("The first table represents the probability of switching after 7 doors are opened. This results in
      the second table represents the probability of not switching after 7 doors are opened. This results in
      The accurately aligns with the actual probabilities. For not switching you have to choose the prize
      is a 1/10 chance. However, if you picked a door without a prize initially (8/10) then switched you
      to the prize. So (8/10) * (1/2) = 0.4 ")

```

```
## [1] "The first table represents the probability of switching after 7 doors are opened. This results in

```

Not for submission: Play with this modified setup, for example Monty opens k doors at step 2 etc with k any number between 1 and 8.

(Bonus) Monty Hall with n-doors.

This is for your curiosity, you don't have to turn this in. However, if you work this general case out, you can use it to solve the earlier versions of the problem.

Repeat the Monty Hall experiment now with n doors. Recall the game is as follows: Step 1: you choose one door at random. Step 2: Monty opens $n-2$ (out of $n-1$ doors) that do not have the prize. Step 3: you either switch or don't switch.

Set up the experiment two functions "monty_ndoors_noswitch" and "monty_ndoors_switch" (these functions will have input value as n):

```

monty_ndoors_noswitch <- function(n){
}

monty_ndoors_switch <- function(n){
}

```

Use your two functions and the replicate function to compute the empirical probability of winning for the two experiments. Compare your answers with the actual theoretical predictions.

Bayes's Theorem and Witnesses

Recall the hit-and-run example from the asynchronous lecture and HW4. We noted that the Bayes' probabilities are calculated as the following formula

$$P(B|W) = \frac{pq}{pq + (1-p)(1-q)},$$

where $P(B) = q$ and $P(W|B) = p$ (check notes). We also noted in HW4 that if $P(W|B) = p > \frac{1}{2}$ we will have

$$P(B|W) > P(B).$$

In this problem, we will explore how changing the level of Witness-Reliability, that is p , will affect the chance of getting a Blue Cab driver arrested.

Recall that $P(B)$ is the probability that the Blue cab is at fault with no added information. We will now run multiple iteration of Bayes' rule, and at every iteration, we will update the value of $P(B)$ with the value of $P(B|W)$ from earlier iteration. From our experience from the homework, if $p > \frac{1}{2}$, we must have $P(B) \rightarrow 1$ as the number of iterations increase (intuitively, the probability that the Blue cab is at fault will go to 1 as the number of "reliable witness" who say that "the Blue Cab is at fault" increases).

Write a function called "bayes" that takes input q, p and outputs $P(B|W)$ (use the formula above)

```
bayes <- function(q, p){  
  return(p*q/(p*q + (1-p)*(1-q)))  
}
```

```
bayes(0.01, 0.8)
```

```
## [1] 0.03883495
```

Reliable Witnesses with fixed probability

In this part, we will assume that we are using testimony of reliable witnesses with fixed reliability p .

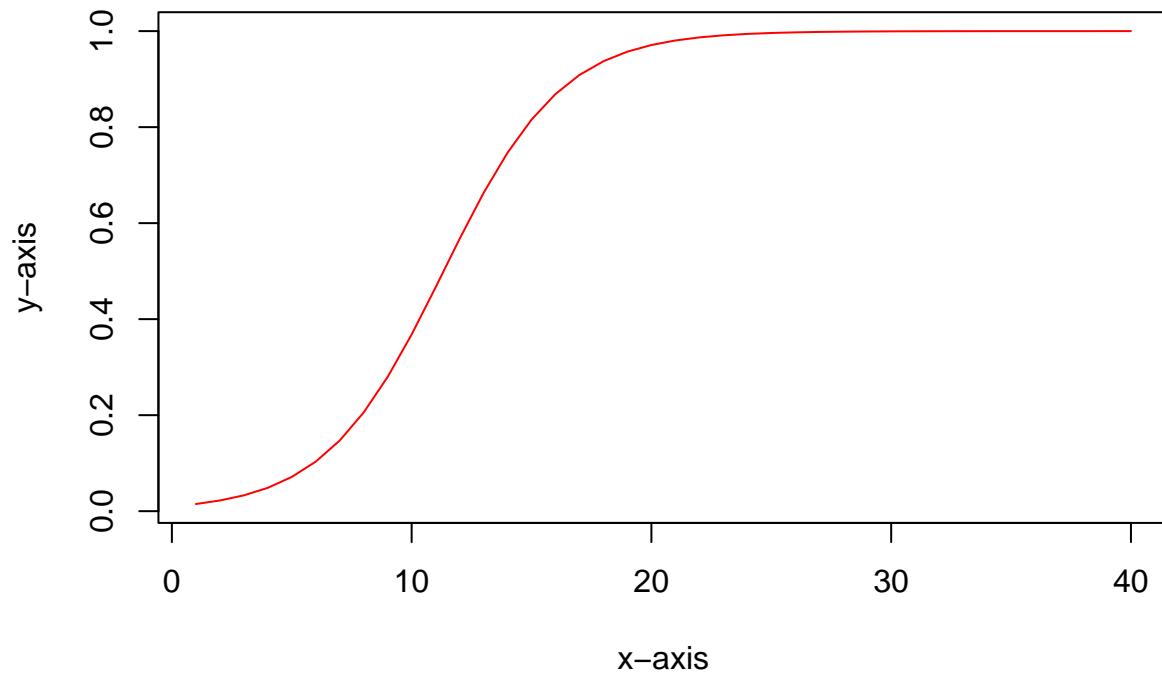
Initial Plots

Initialize $p = 0.6$, $q = 0.01$, and $q_vals = c()$, the empty vector. Run a for loop (1 in 1:40) inside it, update q at every iteration to $q = \text{bayes}(p, q)$, and concatenate the vector q_vals with this new value of q .

```
p = 0.6  
q = 0.01  
q_vals = c()  
  
for (x in 1:40){  
  q = bayes(p,q)  
  q_vals <-c(q_vals, q )  
}
```

Construct a plot with x-axis 1:40 and y-axis q_vals .

```
plot(1:40,q_vals,type="l",col="red",xlab="x-axis", ylab="y-axis")
```

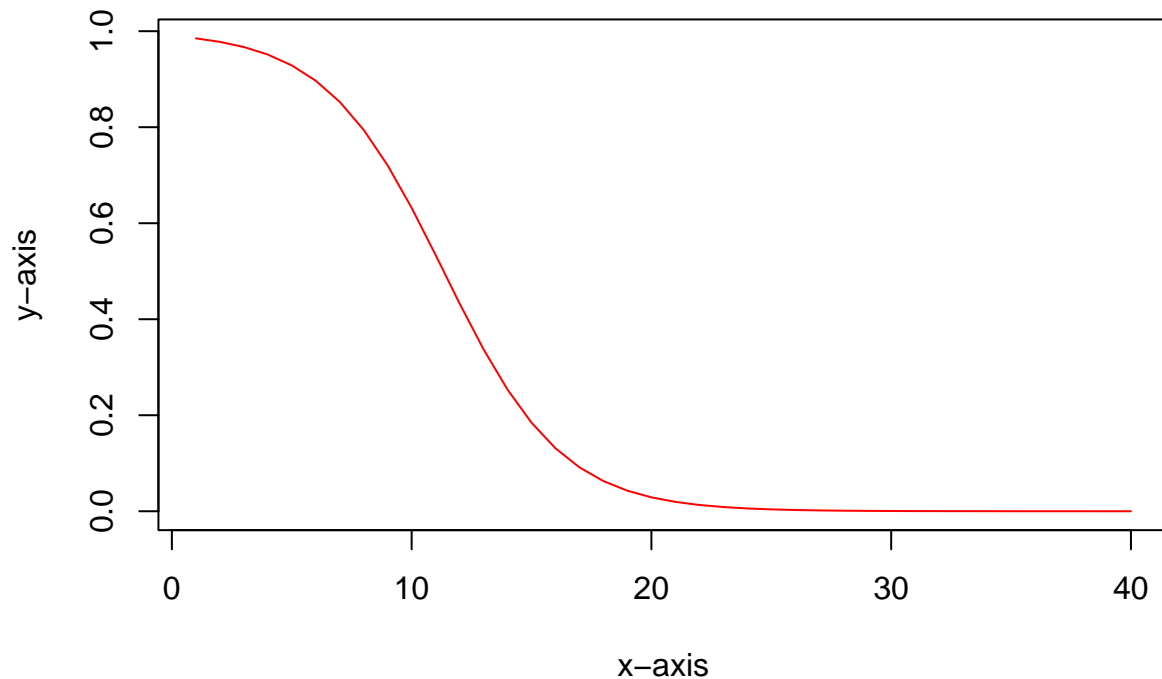


Repeat the above experiment with $p = 0.4$, $q = 0.99$.

```
p = 0.4
q = 0.99
q_vals = c()

for (x in 1:40){
  q = bayes(p,q)
  q_vals <-c(q_vals, q )
}

plot(1:40,q_vals,type="l",col="red",xlab="x-axis", ylab="y-axis")
```



How many reliable Witnesses needed?

Suppose the police decide to arrest the Blue cab driver if $P(B|W) \geq 0.9$.

Suppose we initialize $p = 0.6$ (witness reliability), $q = 0.01$ (proportion of Blue cabs in the city), how many witnesses with reliability level p would we need for the police to consider arresting the Blue cab?

You will need a counter variable initialized to zero and a while loop inside which you update q using the bayes function, and update the counter by 1 at every iteration of the while loop.

```
count = 0
p = 0.6

q = bayes(0.6,0.01)

while(q < 0.9 ){
    q = bayes(p,q)
    count = count+1
}
print("You need this amount of witnesses")
```

```
## [1] "You need this amount of witnesses"
```

```
print(count)
```

```
## [1] 16
```

Working with two types of witnesses

In this problem we will work with two types of witnesses (1) reliable witnesses with fixed reliability level $p_1 > \frac{1}{2}$ and (2) unreliable witness with fixed reliability level $p_2 < \frac{1}{2}$.

Initial Plots

Initialize $p_1 = 0.6$, $p_2 = 0.45$, $q = 0.01$, and $q_vals = c()$, the empty vector. Run a for loop (1 in 1:90) inside it, update q at every even iteration to $q = \text{bayes}(p_2, q)$, and at every odd iteration to $q = \text{bayes}(p_1, q)$ (You will need to use a “if” command), and finally concatenating the vector q_vals with the new value of q at every iteration.

```
p_1 = 0.6
p_2 = 0.45
q = 0.01
q_vals = c()

for(x in 1:90){

  if(x %% 2 == 0){
    q = bayes(p_2,q)
  } else {
    q = bayes(p_1,q)
  }
  q_vals = c(q_vals, q)

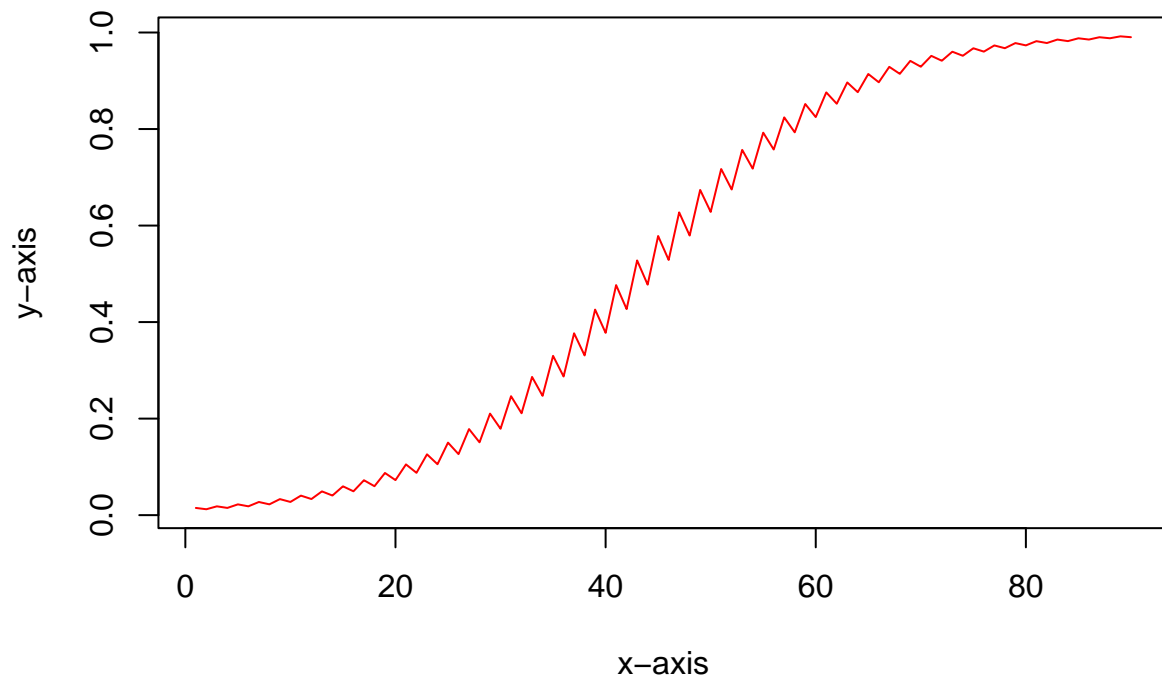
}

print(q_vals)
```

```
## [1] 0.01492537 0.01224490 0.01825558 0.01498612 0.02231200 0.01832963
## [7] 0.02724475 0.02240213 0.03323098 0.02735426 0.04047776 0.03336371
## [13] 0.04922442 0.04063823 0.05974341 0.04941777 0.07233924 0.05997548
## [19] 0.08734397 0.07261645 0.10510838 0.08767325 0.12598704 0.10549689
## [25] 0.15031638 0.12644182 0.17838506 0.15084383 0.21039718 0.17899025
## [31] 0.24643100 0.21108307 0.28639775 0.24719759 0.33000781 0.28724128
## [37] 0.37675248 0.33092021 0.42590932 0.37772126 0.47657554 0.42691792
## [43] 0.52772850 0.47760430 0.57830579 0.52875614 0.62729197 0.57931110
## [49] 0.67379748 0.62825558 0.71711699 0.67470320 0.75676045 0.71795278
## [55] 0.79245613 0.75751871 0.82413081 0.79313354 0.85187499 0.82472770
## [61] 0.87590146 0.85239458 0.89650421 0.87634900 0.91402226 0.89688620
## [67] 0.92881060 0.91434577 0.94121890 0.92908278 0.95157718 0.94144663
## [73] 0.96018737 0.95176684 0.96731913 0.96034471 0.97320900 0.96744924
## [79] 0.97806145 0.97331631 0.98205123 0.97814976 0.98532630 0.98212377
## [85] 0.98801107 0.98538580 0.99020951 0.98805982 0.99200807 0.99024941
```

Construct a plot with x-axis 1:90 and y-axis q_vals .

```
plot(1:90,q_vals,type="l",col="red",xlab="x-axis", ylab="y-axis")
```



What do you notice? (Compared to case with single witness)

It takes more witnesses to reach a higher level of confidence

How many witnesses needed?

Suppose the police decide to arrest the Blue cab driver if $P(B|W) \geq 0.9$.

Initialize $p_1 = 0.6$, $p_2 = 0.45$, $q = 0.01$. How many witnesses will we have to have if we are going to sequentially work with one reliable witness followed by an unreliable witness?

```
p_1 = 0.6
p_2 = 0.45
q = 0.01
count = 1

while(q<0.9){

    if(count %% 2 == 0){
        q = bayes(p_2,q)
    } else {
        q = bayes(p_1,q)
    }
    count = count + 1
}
```

```
print("We need this many witnesses")
```

```
## [1] "We need this many witnesses"
```

```
print(count-1)
```

```
## [1] 65
```

This problem can be generalized to choosing p from any given probability distribution on $(0,1)$, we will be working with the Beta distribution later in the semester, which can be an interesting candidate to sample our values for p (that is witness reliability).

Working with four types of witnesses

In this problem we will work with two types of witnesses (1) two reliable witnesses with fixed reliability level $p_1, p_2 > \frac{1}{2}$ and (2) two unreliable witnesses with fixed reliability level $p_3, p_4 < \frac{1}{2}$.

Initial Plots

Initialize $q = 0.01$ and we will sample p from the vector $c(0.6, 0.8, 0.45, 0.3)$ and suppose choosing a witness with either of these witness-reliabilities is equally likely.

Now let $q_vals = c()$, the empty vector and run a for loop (1 in 1:100) inside it, update q by sampling p from the vector $c(0.6, 0.8, 0.45, 0.3)$, and concatenating the vector q_vals with the new value of q at every iteration.

```
q = 0.01
q_vals = c()

for(x in 1:100){
  q = bayes(sample(c(0.6, 0.8, 0.45, 0.3), 1), q)
  q_vals = c(q_vals, q)
}

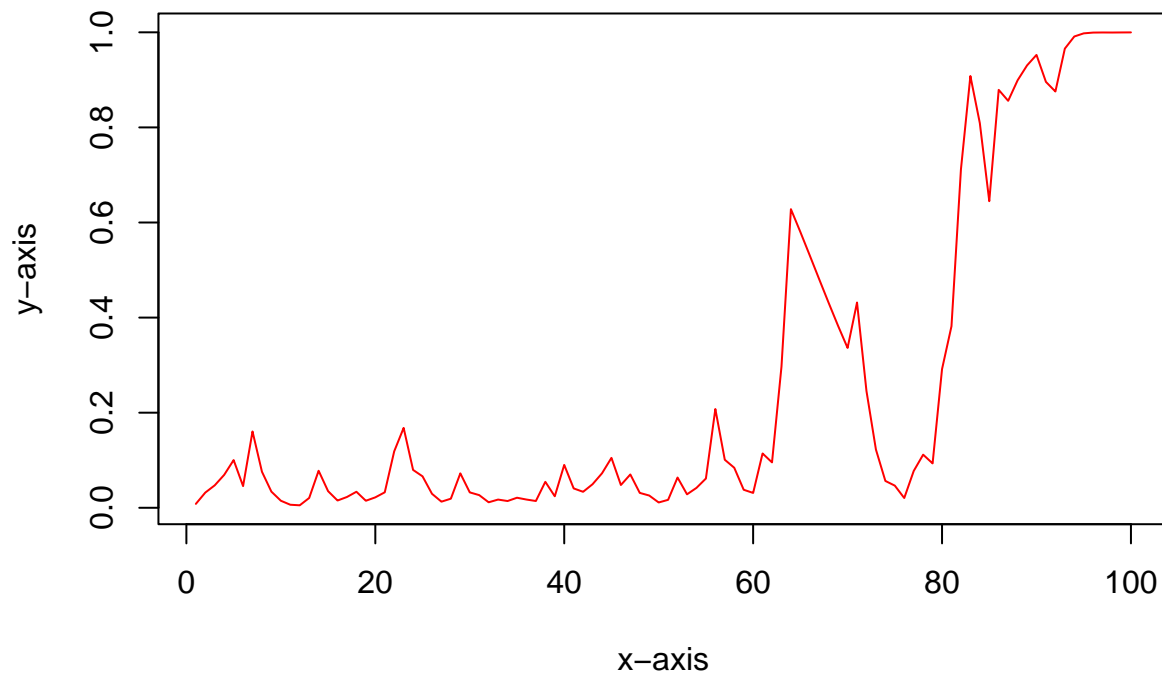
print(q_vals)
```

```
## [1] 0.008196721 0.032000000 0.047244094 0.069230769 0.100371747 0.045633803
## [7] 0.160555005 0.075759938 0.033937758 0.014832398 0.006411080 0.005251551
## [13] 0.020680390 0.077889222 0.034936033 0.015277579 0.022742642 0.033730403
## [19] 0.014739993 0.021948232 0.032564976 0.118666762 0.168030335 0.079661926
## [25] 0.066135851 0.029457179 0.012840648 0.019138099 0.072395839 0.032365728
## [31] 0.026637805 0.011592661 0.017288780 0.014189970 0.021135003 0.017358981
## [37] 0.014247771 0.054654951 0.024178685 0.090173881 0.040745479 0.033586025
## [43] 0.049546994 0.072523822 0.104978994 0.047862152 0.070115294 0.031303619
## [49] 0.025758659 0.011204345 0.016712888 0.063659739 0.028312675 0.041876199
## [55] 0.061526060 0.207756858 0.101033124 0.084210381 0.037914623 0.031236386
## [61] 0.114240191 0.095451872 0.296813326 0.628030151 0.580080700 0.530570318
## [67] 0.480450793 0.430721661 0.382351756 0.336205776 0.431733086 0.245625382
## [73] 0.122455560 0.056429593 0.046648275 0.020539625 0.077389827 0.111760189
## [79] 0.093336761 0.291675035 0.381827742 0.711872715 0.908111488 0.808995023
```

```
## [85] 0.644785397 0.878946324 0.855921341 0.899101801 0.930393477 0.952493395
## [91] 0.895754394 0.875473447 0.965661302 0.991188391 0.997782443 0.999444687
## [97] 0.999629723 0.999547476 0.999698272 0.999798828
```

Construct a plot with x-axis 1:100 and y-axis q_vals.

```
plot(1:100,q_vals,type="l",col="red",xlab="x-axis", ylab="y-axis")
```



What do you notice? How does this compare with the earlier cases?

The amount of witnesses increases faster than the earlier cases around the point 22.

How many witnesses needed?

Suppose the police decide to arrest the Blue cab driver if $P(B|W) \geq 0.9$.

Initialize $q = 0.01$ and we will sample p from the vector $c(0.6, 0.8, 0.45, 0.3)$ and suppose choosing a witness with either of these witness-reliabilities is equally likely. How many witnesses will we need to have if we are going to uniformly sample from each of these four types of witnesses at every step?

```
count = 0
p = 0.6

q = bayes(sample(c(0.6, 0.8, 0.45, 0.3), 1),0.01)

while(q < 0.9 ){
```

```

    q = bayes(sample(c(0.6, 0.8, 0.45, 0.3), 1),q)
    count = count+1
}

print("You need this many witnesses")

```

```
## [1] "You need this many witnesses"
```

```
print(count)
```

```
## [1] 60
```

BONUS: Working with witnesses coming from a Beta distribution

In this problem we will assume that the witness reliability of the people in the city follows the Beta distribution with parameters α, β . Choose α, β in such a way that the expected value of this Beta distribution is greater than .65.

Initial Plots

Initialize $q = 0.01$ and we will choose p by sampling the Beta distribution with parameters α, β chosen above.

Now let $q_vals = c()$, the empty vector and run a for loop (1 in 1:100) inside it, update q by sampling $p \sim \text{Beta}(\alpha, \beta)$, and concatenating the vector q_vals with the new value of q at every iteration.

Construct a plot with x-axis 1:100 and y-axis q_vals .

What do you notice? How does this compare with the earlier cases?

How many witnesses needed?

Suppose the police decide to arrest the Blue cab driver if $P(B|W) \geq 0.9$.

Initialize $q = 0.01$ and we will choose p by sampling the Beta distribution with parameters α, β chosen above. How many witnesses will we need to have in this setting?