MTH208a: Problem Set

Problem Set Questions

Last Updated: October 13, 2022

Practice R Questions

1. Define an `R` vector of the first 1000 odd integers.

2. Define an `R` vector of the first 500 Fibonacci numbers.

3. Write an `R` function that rolls a fair die and returns `1` if the die turns out to be even.

4. Write an `R` function that tosses a fair coin 15 times. The function should return `"lose"` if the number heads is less than 8 and `"win"` if the number of heads is more than or equal to 8.

5. Define a 5×5 matrix of all elements being 1.

6. Define a 5×5 diagonal matrix with the diagonals being `1:5`.

7. Define a 10×10 matrix where each entry is the result of a random roll of a fair die.

8. Write an `R` function that takes inputs `n` and `rho`. The function should return an $n \times n$ matrix with 1s on the diagonals and `rho` on the off-diagonals.

9. Write an `R` function that takes inputs `n` and `rho`. The function should return an $n \times n$ matrix with the i, j the element being $\rho^{|i-j|}$.

10. Write an `R` function that takes a matrix input and returns a smaller matrix with only the odd columns of the original matrix.

11. Define an 4 dimensional arrays with dimensions $10 \times 4 \times 6 \times 5$ where the entries are all 1s.

Worksheet Based Questions

1. Write an `R` function to calculate the area of a circle of radius r for a user-given value of r.

2. Write an `R` function that returns the larger of two inputs x and y.

3. Write an `R` program that saves the output of 1000 rolls of a fair die and returns the number of time the output was an even number.

4. Draw 1000 random number between $[0, 1]$ and calculate the proportion of numbers between 0.1 and 0.2.

5. A popular chips company came out with a new strategy. They were going to put different Harry Potter toys in their chips packets. You and your friends are very excited; you want to be the first to collect all figurines! The chips company has declared that they will put the toys in the packets based on the popularity of the character. So more popular characters are more likely to be in the packet. Below in the table is the probability of a packet of chips containing the type of toys.

Toys	Harry	Dumbledore	Hermione	Ron	Neville	Mcgonagall	Dobby
Prob	.25	.20	.20	.15	.10	.05	.05

Write an `R` function to calculate the number of chips packets that need to be bought to collect all toys. Repeat the simulation 1000 times and calculate the average number of chips packets that need to be bought to collect all toys.

6. You take half of a vitamin every morning. The vitamins are sold in a bottle of 100 (whole) tablets, so at first you have to cut the tablets in half. Every day you randomly pull one thing from the bottle. If it's a whole tablet, you cut it in half and put the leftover half back in the bottle. If it's a half-tablet, you take the vitamin. You just bought a fresh bottle. How many days, on average, will it be before you pull a half-tablet out of the bottle?

7. You are on a game show, being asked to choose between three doors. One door has a car, and the other two have goats. After you choose a door, the host, Monty Hall, opens one of the other doors, which he knows has a goat behind it. Monty then asks whether you would like to switch your choice of door to the other remaining door. Do you choose to switch or not to switch?

Whether you switch or not depends on which action has the largest probability of winning the car.

1. Write an `R` function, `MontyHall()` to simulate this game show for when the contestant decides to switch doors. The `R` function should return `1` if the contestant wins and `0` if they lose.

2. Repeat the simulation 1000 times to estimate the probability of winning if the contestant switches.

8. The website `https://editorial.rottentomatoes.com/guide/best-netflix-movies-to-watch-right-now/` has the "100 best movies on Netflix" ranked by "Tomatometer". The list goes from "#100" to "#1". Scrape this website and collect the following information:

1. Ranking

2. Name of Movie

3. Tomato % score

4. Year of movie

9. Scrape the complete table of summer olympic medals of the United States available at `https://en.wikipedia.org/wiki/United_States_at_the_Olympics`. It may be challenging to get everything, but try as much as possible.

10. From the website: `https://stats.stackexchange.com/questions?tab=Votes`, scrape the following information:

    1. The title of the questions

    2. The number of views

    3. The number of answers

    4. The notes of votes

11. Load the dataset `iris` by using command `data(iris)`. Draw side-by-side boxplots of all continuous variables based on the column `Species`. Do this in base R and in `ggplot2`.

12. Make a scatterplot of `Sepal.Length` and `Petal.Length` coloring each plot by `Species`. What can you conclude? Do this in both base R and in `ggplot2`.

13. Load the dataset `txhousing` in package `ggplot2`. Explore the dataset with plots similar to the way we did Worksheet 12. There are some missing values in the dataset. You may use function `complete.cases()` if required.

14. Construct a matrix `A` which is 1000x1000. Find the norm of every column of the matrix using a loop and then also using function `sapply`. Which is faster?

15. Improve the efficiency of the following code that calculates AB^Tx for matrices and A, B and vector x.

```
n <- 1000
m <- 1000
A <- matrix(runif(n*m), nrow = n, ncol = m)
B <- matrix(rnorm(n*m), nrow = n, ncol = m)
x <- runif(m)

ABtx <- (A %*% t(B)) %*% x
```

16. Improve the efficiency of the following code from Worksheet 3

```
n <- 50
m <- 1e3
A <- matrix(runif(n*m), nrow = n, ncol = m)
```

```r
# p_vec will store p_i eventually
p_vec <- numeric(length = m)

# running a loop for each column
# to find the norm (the numerator)
for(k in 1:m)
{
  p_vec[k] <- norm(A[ ,k], type = "2")
}
# divide by the sum
p_vec <- p_vec/sum(p_vec)

# choosing column
chosen <- sample(1:m, size = 1, prob = p_vec)
```