Q1. (30 pts) Let X be a random variable whose set of outcomes is a unit interval between 0 and 1 (i.e., a set of real numbers between 0 and 1). Suppose that for any $x \in [0,1]$, $\mathbb{P}(X \leq x) = x$. Using R, we can simulate this random variable by running the following code: runif(1). Using this code "runif(1)", how are you going to simulate outcome of a coin flip? (Hint: Use ifelse function in R. The ifelse function in R works identically to the if function in excel. For example, if you run "ifelse(1+1 == 3, 'stupid', 'genious')", R will print out 'stupid' on your screen. On the other hand, if you run "ifelse(1+1 <= 3, 'stupid', 'geneious')", it will print out 'geneious' on your screen.).

Now, let $X_1, X_2, \ldots$ be a collection of i.i.d. random variables such that

$$X_i = \begin{cases} 2 & \text{with probability } 0.3 \\ 1 & \text{with probability } 0.4 \\ 0 & \text{with probability } 0.3. \end{cases}$$

Simulate $\frac{\sum_{i=1}^{n} X_i}{n}$ and discuss how the simulated $\frac{\sum_{i=1}^{n} X_i}{n}$ behaves as $n \to \infty$. For this part of problem, submit your r code as well.

Q2. (20 pts) Suppose hourly wage in a population is exponentially distributed with rate θ. In other words, if we let X be a random variable for a person in the population's hourly wage, we have

$$\mathbb{P}(X \leq x) = 1 - \exp(-\theta x).$$

Let X_i be the hourly wage of the i^{th} person in the population. Due to the privacy, you cannot directly collect hourly wage from people but you can collect data that indicate whether a person's hourly wage exceeds \$5 USD or not. That is the data you collect from the i^{th} person in the population is

$$Y_i = \begin{cases} 1 & \text{if } X_i \geq 5 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose you collect the data from n people and you have $Y_1 = y_1$, $Y_2 = y_2$, ..., $Y_n = y_n$. Use this data, estimate θ. (Hint: use the maximum likelihood estimator approach.)

Q3. (30 pts) Without using equations, explain what unbiased estimator is.

Q4. (20 pts) Suppose you want to estimate β_0 and β_1 in

$$y = \beta_0 + \beta_1 x + \epsilon$$

where ϵ is a mean zero random noise. You collect x and y from one subject. Can you estimate β_0 and β_1 from this dataset with one row by OLS algorithm?