STATS 451 Homework 1

Yuzhou Peng

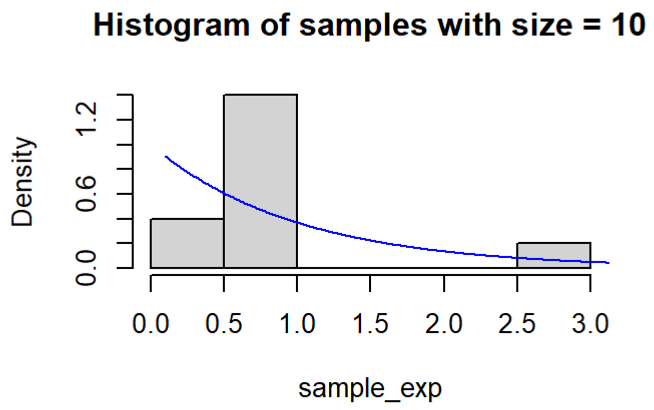
2024-01-27

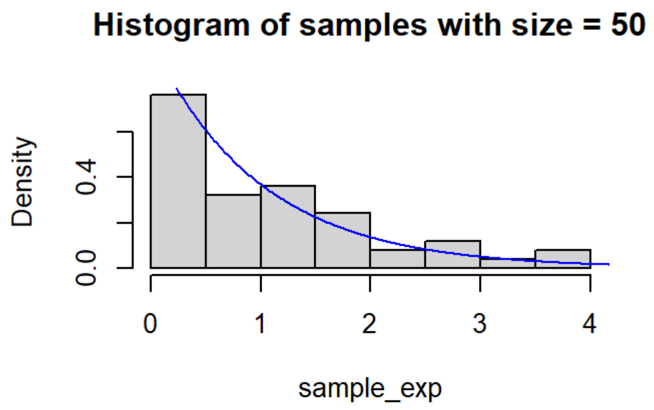
Problem 1

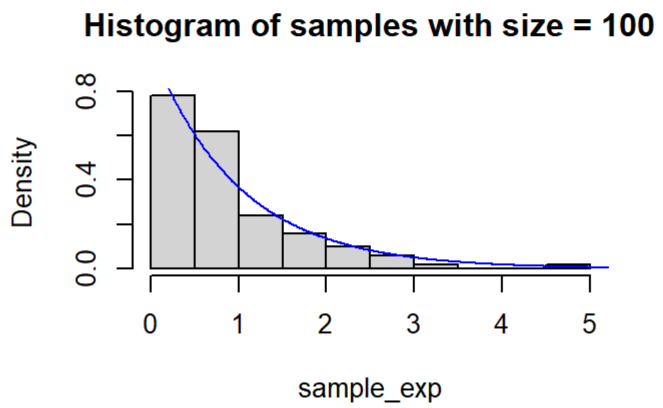
```
sample_size_1 = c(10, 50, 100)
```

(a)

```
set.seed(127)

par(mfrow = c(2, 2))
for(i in sample_size_1){
  sample_exp <- rexp(i, rate = 1)
  sample_exp_range <- range(sample_exp)
  sample_exp_step <- seq(0.1, 6, by = 0.01)
  hist(sample_exp,
       main = paste("Histogram of samples with size =", i),
       freq = FALSE)
  lines(sample_exp_step, dexp(sample_exp_step), col = "blue")


}
```



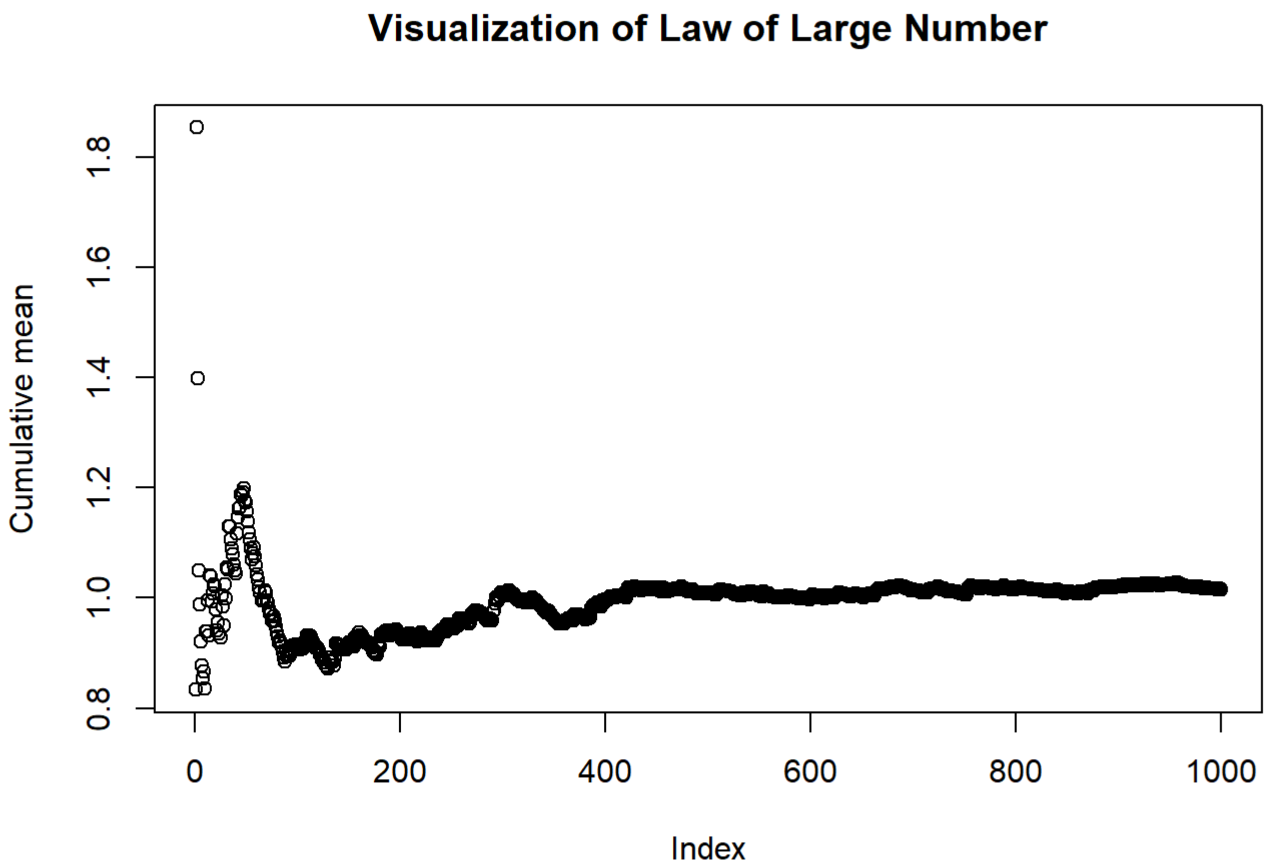




(b)

Law of Large Number

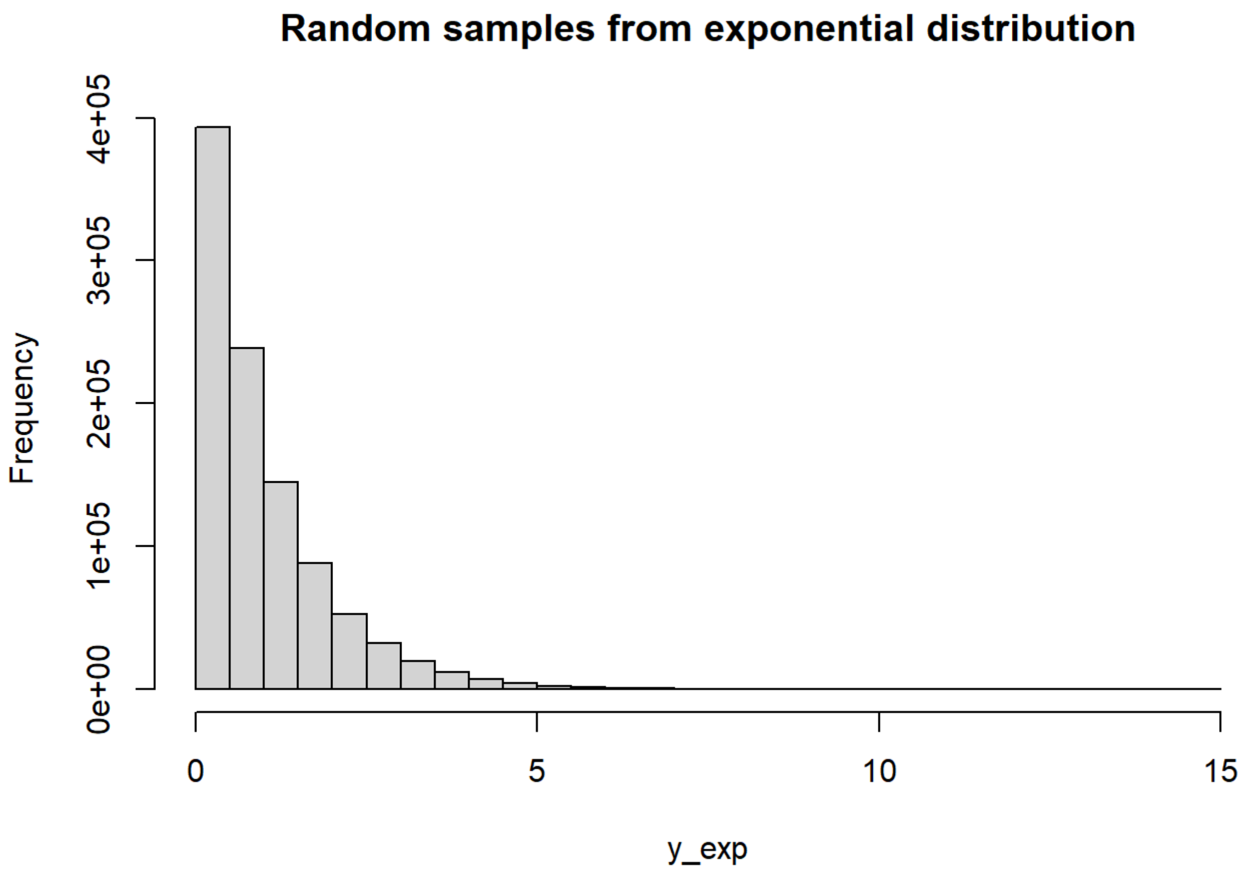
```
set.seed(127)

N <- 1000
x_exp <- rexp(N, rate = 1)
means_exp <- cumsum(x_exp) / (1:N)
plot(1:N, means_exp,
     ylab = "Cumulative mean",
     xlab = "Index",
     main = "Visualization of Law of Large Number")
```



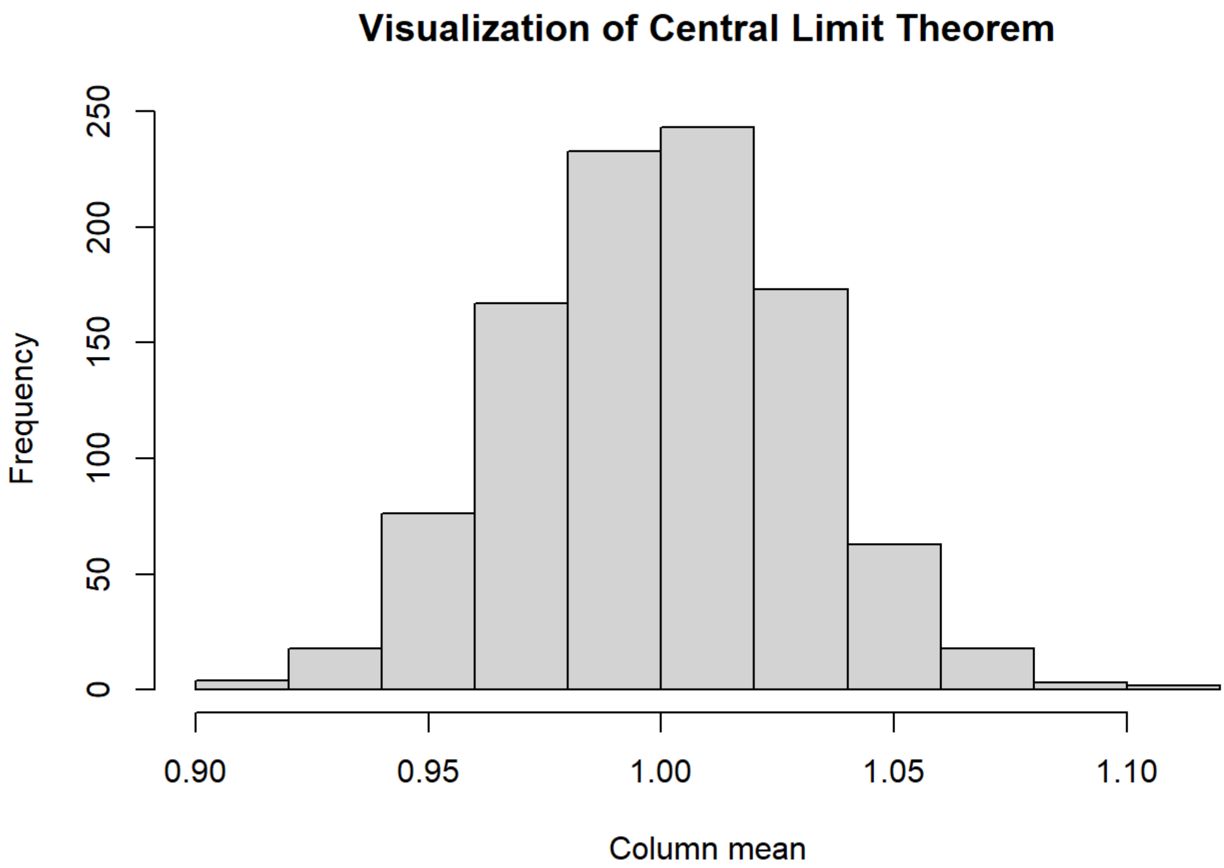
Central Limit Theorem

```
set.seed(127)
y_exp <- array(rexp(1000 * 1000, 1), c(1000, 1000))
```

```
hist(y_exp,
     main = "Random samples from exponential distribution")
```




```
hist(apply(y_exp, 1, mean),
     xlab = "Column mean",
     main = "Visualization of Central Limit Theorem")
```



Visualization of Central Limit Theorem
Frequency
0
50
100
150
200
250
0.90
0.95
1.00
1.05
1.10
Column mean

Problem 2

(a)

Let p_i be the probability of having a quiz on i-th class, and q_i be the binary random variable indicating whether the professor will have quiz on i-th class:

$$q_i = \begin{cases} 1, & \text{meaning having quiz on i-th class} \\ 0, & \text{meaning no quiz on i-th class} \end{cases}, \text{ and we have:}$$

$$p_i \overset{\text{iid}}{\sim} Uniform(0,1)$$
$$q_i|p_i \sim Bernouli(p_i), \text{ suggesting } q_i \text{ is independent.}$$

$$\begin{aligned}
\text{Thus, we have: } \mathbb{E}[\sum_{i=1}^{40} q_i] &= 40\mathbb{E}[q_i] \\
&= 40\int_0^1 \mathbb{E}(q_i, p_i = p)dp \\
&= 40\int_0^1 \mathbb{E}(q_i|p_i = p) * 1dp \\
&= 40\int_0^1 pdp \\
&= 20. \\
Var[\sum_{i=1}^{40} q_i] &= 40Var[q_i] \\
&= 40(\mathbb{E}[q_i^2] - (\mathbb{E}[q_i])^2) \\
&= 40(\int_0^1 \mathbb{E}(q_i^2|p_i = p)dp - \frac{1}{4}) \\
&= 40(\int_0^1 pdp - \frac{1}{4}) \\
&= 10. \\
s.d. = \sqrt{Var[\sum_{i=1}^{40} q_i]} &= \sqrt{10}.
\end{aligned}$$

(b) Monte-Carlo simulation

```
set.seed(127)
quiz_time_sample <- c()
for (i in 1:1000){
  p_quiz <- runif(40)
  q <- c()
  for (j in p_quiz){
    qj <- rbernoulli(n = 1, p = j)
    q <- append(q, qj)
  }
  quiz_time_sample <- append(quiz_time_sample, sum(q))
}
```

```
## Warning: `rbernoulli()` was deprecated in purrr 1.0.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
mean(quiz_time_sample)
```

```
## [1] 19.971
```

```
sd(quiz_time_sample)
```

```
## [1] 3.113495
```