Stat 142 MP Light 6

First Semester, A.Y. 2025-26

First Name: Anne Christine

Last Name: Amores

Student Number: 2021-04650

Section: WFU

Monte Carlo Methods

I. Using Monte Carlo Integration, approximate the value of

$$I = \int_0^1 e^{x^2} dx.$$

Use $M = 1,000,000$ and `seed = 142`.

```
set.seed(142)
X <- runif(1000000, 0, 1)
mean(exp(X^2))
```

```
## [1] 1.462358
```

Thus, the approximate value of the definite integral above is 1.462358.

II. Monte Carlo Experiment

The Lognormal distribution is often used to model economic or financial variables such as income, wealth, and asset prices, which are typically positive and right-skewed. Unlike symmetric distributions such as the Normal, the Lognormal has a long right tail, meaning that extreme values occur more frequently. As a result, the sample mean can be strongly influenced by these large values, potentially leading to unstable estimates of the population mean.

Suppose our sample came from a Lognormal distribution with parameters meanlog = 0 and sdlog = 1 for the mean and standard deviation respectively. In this exercise, we want to examine how well the sample trimmed mean estimates the population mean (μ) of this distribution, depending on the percentage of lowest and highest sample values that are discarded when computing the trimmed mean.

Due to limited computational resources, in our simulation study we will only look at populations in which N = 1,000,000. From each population, one sample of size n = 500 will be obtained using simple random sampling without replacement (SRSWOR). Bias and variance will be estimated using the Monte Carlo method with M = 500 iterations.

Create a Monte Carlo experiment function with the name `mc_lognormal`, and make sure to add documentation inside this function. Use the `replicate()` function when writing this Monte Carlo experiment function, instead of loops.

```r
mc_lognormal <- function(trim, n = 500, N = 1000000, M = 500) {
  # Description
  # Performs a Monte Carlo simulation to estimate the sampling distribution of the
  # trimmed mean for samples drawn from a Lognormal population with parameters
  # meanlog = 0 and sdlog = 1

  # Parameters
  # trim -- a numeric value between 0 and 0.5 specifying the fraction of observations to
  # be trimmed from each end of the sorted sample before the mean is computed
  # (e.g. trim = 0.02 for a 2% trim)
  # n -- number of elements in the sample drawn from the population in each replication
  # N -- number of elements in the Lognormal population from which the samples are drawn
  # M -- number of replications

  # Value
  # Returns a numeric vector of length M containing the trimmed sample means computed
  # from samples drawn from a Lognormal(0, 1) population across M Monte Carlo
  # replications.
  pop <- rlnorm(N, meanlog = 0, sdlog = 1)
  samp_trimmedmean <- replicate(M, {
    samp <- sample(pop, n, replace = FALSE)
    mean(samp, trim = trim)
  })
  return(samp_trimmedmean)
}
```

After conducting your Monte Carlo experiments using `mc_lognormal`, answer the following questions below.

1. What are the estimated bias and variance of the sample 2% trimmed mean?

```r
pop_mean <- exp(0 + 1/2)
set.seed(1420)

trimmed2 <- mc_lognormal(trim = 0.02)

bias_trimmed2 <- mean(trimmed2) - pop_mean
var_trimmed2 <- mean((trimmed2-mean(trimmed2))^2)

bias_trimmed2
```

```
## [1] -0.1857261
```

```r
var_trimmed2
```

```
## [1] 0.0056088
```

2. What are the estimated bias and variance of the sample 25% trimmed mean?

```r
set.seed(14200)

trimmed25 <- mc_lognormal(trim = 0.25)
```

```
bias_trimmed25 <- mean(trimmed25) - pop_mean
var_trimmed25 <- mean((trimmed25-mean(trimmed25))^2)

bias_trimmed25
```

```
## [1] -0.5744179
```

```
var_trimmed25
```

```
## [1] 0.002897021
```

3. Compute for the estimated MSE for:

a. the sample 2% trimmed mean

```
mse_trimmed2 <- var_trimmed2 + bias_trimmed2^2
mse_trimmed2
```

```
## [1] 0.04010299
```

b. the sample 25% trimmed mean

```
mse_trimmed25 <- var_trimmed25 + bias_trimmed25^2
mse_trimmed25
```

```
## [1] 0.332853
```

4. Based on the estimated MSE in the scenarios considered, does increasing the % trimmed from the ends of a sample improve our estimate of the location parameter using the trimmed mean?

Since the MSE of the sample 25% trimmed mean is 0.332853, which is roughly 8 times greater than the MSE of the sample 2% trimmed mean (0.04010299), we can say that **increasing the % trimmed from the ends of a sample DOES NOT improve the estimate of the true population mean using the trimmed mean**. This is because, by using the 25% trim, we are cutting off too much of the Log-Normal's distribution's heavy right tail.