

# Simulation Exercise

*Will Swales*

*8 July 2018*

## Overview

The purpose of this report is to document a demonstration of the Central Limit Theorem. We accomplish this by investigating the sampling distribution of averages of 40 random variates drawn from an exponential distribution with parameter  $\lambda = 0.2$ . 1000 simulations are used to create the sampling distribution.

## Simulations

We begin by generating `mns`, a vector of 1000 sample means. Each sample mean is calculated as the arithmetic mean of 40 independent and identically distributed (iid) variates drawn randomly from an exponential distribution with parameter  $\lambda = 0.2$ .

We have used the seed 123 for random number simulation.

```
## initialise parameters
lambda = 0.2
n = 40
nosims = 1000
set.seed(123)

## generate mns, a vector of sample means of size n from an exponential distribution
## with parameter lambda
mns = NULL
for (i in 1 : nosims) mns = c(mns, mean(rexp(n, lambda)))
```

## Sample Mean versus Theoretical Mean

One of the results of the Central Limit Theorem is that if  $X_i$  are iid then the sample mean  $\bar{X}_n$  is approximately normally distributed with mean  $\mu$  and variance  $\sigma^2/n$ . Therefore the expectation of the sampling distribution,  $E[\bar{X}_n]$ , should be approximately equal to the theoretical mean,  $\mu$ .

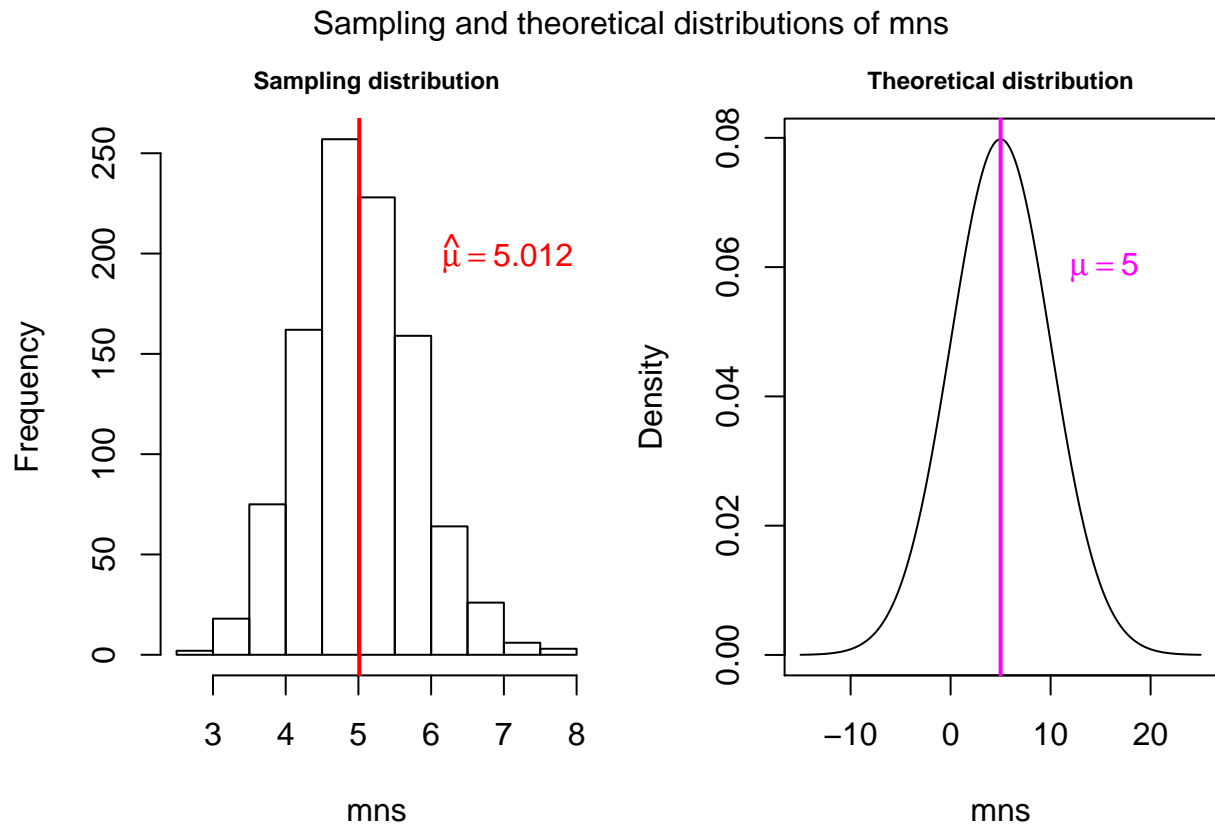
The theoretical mean,  $\mu$ , of the exponential distribution is equal to  $1/\lambda = 1/0.2 = 5$ . Our estimate of the mean,  $\hat{\mu}$  is calculated as the expectation of the sampling distribution,  $E[\bar{X}_n]$ . This can be evaluated in R:

```
mu_hat <- mean(mns)
mu_hat
```

```
## [1] 5.011911
```

This shows that our estimate of the mean,  $\hat{\mu} = 5.012$ , is close to the true value,  $\mu = 5$

The plots below show the sampling and theoretical distributions of `mns`, highlighting the sample estimate of the mean,  $\hat{\mu}$ , and theoretical value of the mean  $\mu$ . We can also see that the shape of the sampling distribution approximates that of the theoretical distribution.



## Sample Variance versus Theoretical Variance

The Central Limit Theorem states that the variance of the sample mean,  $Var[\bar{X}_n]$ , should be approximately equal to the theoretical variance  $\sigma^2$ , divided by  $n$ .

The theoretical variance,  $\sigma^2$ , of the exponential distribution is equal to  $1/\lambda^2 = 1/0.2^2 = 25$ . As  $n = 40$ ,  $\sigma^2/n = 25/40 = 0.625$

Our estimate of the sample variance,  $\hat{s}^2$ , is calculated as the variance of the sampling distribution,  $Var[\bar{X}_n]$ . This can be evaluated in R:

```
var_hat <- var(mns)
var_hat
```

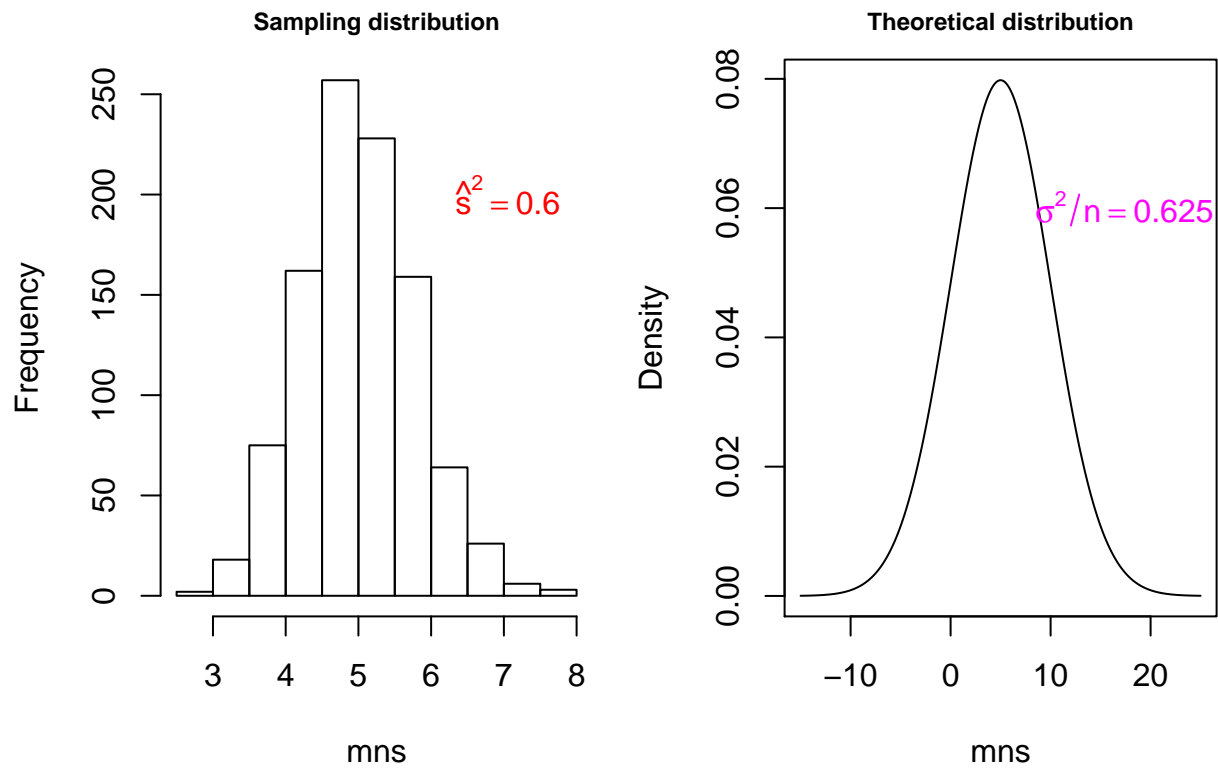
```
## [1] 0.6004928
```

Note that the var function in R evaluates the sample variance (i.e. with  $n-1$  degrees of freedom) as opposed to the population variance.

This shows that our estimate of the variance,  $\hat{s}^2 = 0.600$ , is quite close to the true value,  $\sigma^2/n = 0.625$ . Our estimate of the variance of the sampling distribution appears to have a larger error than our estimate of the mean. This is because the variance of the sample mean estimator,  $Var[E[\bar{X}_n]] = \sigma^2/n$ , is smaller than the variance of the sample variance estimator  $Var[s^2]$ .

The plots below show the sampling and theoretical distributions of `mns` with the sample estimate and theoretical value of the variance highlighted.

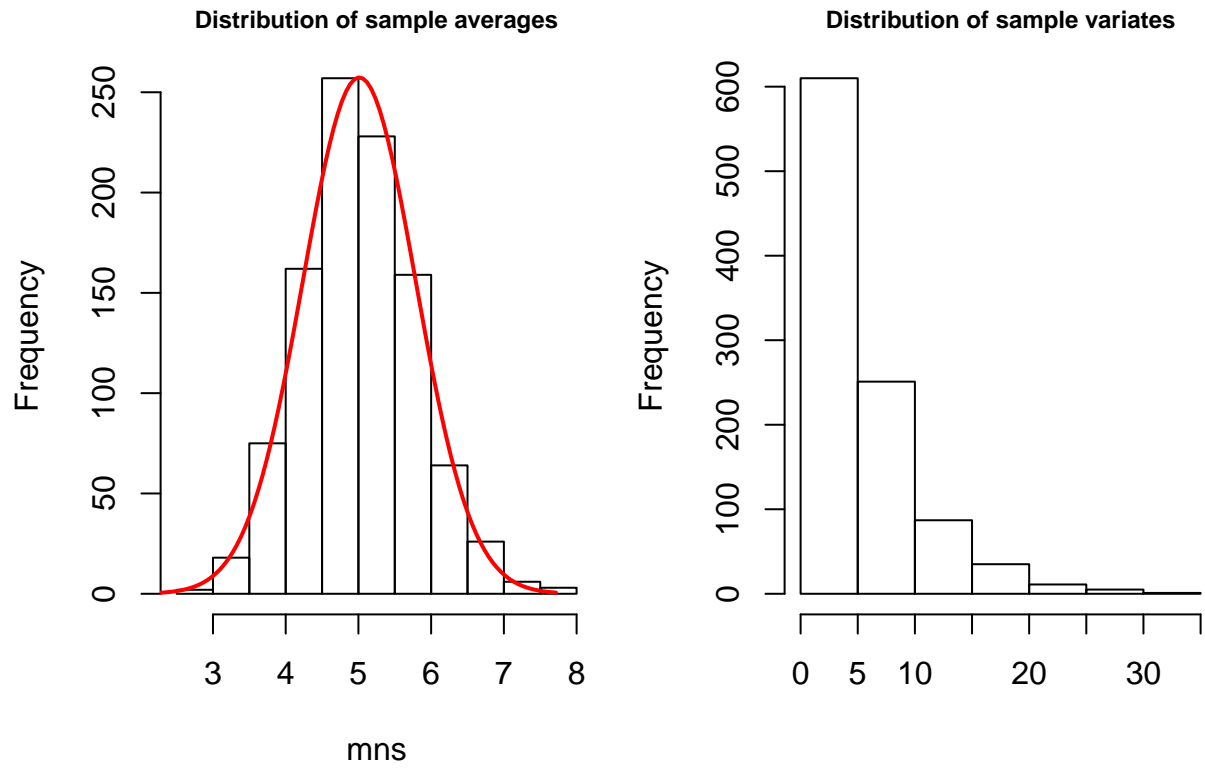
## Sampling and theoretical distributions of mns



### Distribution

We have already seen how the distribution of the sample mean is approximately normal (as predicted by the Central Limit Theorem). The plots below emphasise this point by showing the sampling distribution of the sample mean (overlaid with an appropriately scaled normal density function) alongside the sampling distribution of random variates from the exponential distribution.

## Comparison of sampling distributions



The plots show that the sampling distribution of the sample mean is clearly Gaussian, while also showing that the underlying distribution of the random variates themselves is exponential. This demonstrates the crux of the Central Limit Theory which is that the distribution of sample means of iid variates from **any** distribution will asymptotically approach a normal distribution with mean  $\mu$  and variance  $\sigma^2/n$ , as  $n$  increases.