Statistical Inference Part 1

Gill Collier

06/11/2020

Gill Collier - 01 November 2020

Coursera: Statistical Inference - Course Project

Part 1: Simulation Exercise Instructions

In this project you will investigate the exponential distribution in R and

compare it with the Central Limit Theorem. The exponential distribution can

be simulated in R with rexp(n, lambda) where lambda is the rate parameter.

The mean of exponential distribution is 1/lambda and the standard deviation

is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will

investigate the distribution of averages of 40 exponentials. Note that you

will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of

the distribution of the mean of 40 exponentials. You should:

1. Show the sample mean and compare it to the theoretical mean of the

distribution.

2. Show how variable the sample is (via variance) and compare it to the

theoretical variance of the distribution.

3. Show that the distribution is approximately normal.

```r
library(ggplot2)
```

```
## Warning: replacing previous import 'vctrs::data_frame' by 'tibble::data_frame'
## when loading 'dplyr'
```

Set variables

```r
set.seed(123)
lambda <- 0.2
n <- 40
```

Generate the sample means

```r
mns = NULL
for (i in 1 : 1500) mns = c(mns, mean(rexp(n, lambda)))
summary(mns)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.943   4.465   4.964   4.996   5.517   7.717
```

Calculate the mean of these means

```r
m_mns <- mean(mns)
```

Calculate the theoretical mean

```r
t_mns <- lambda^-1
t_mns
```

```
## [1] 5
```

```r
abs(m_mns - t_mns)
```

```
## [1] 0.003612815
```

The central limit theorem states that the sampling distribution of a sample

mean is approximately normal if the sample size is large enough, even if

the population distribution is not normal.

The CLT in this simulation appears to be valid as this shows that increasing

the number of samples narrows the gap between the simulation mean and the

theoretical mean

Calculate the sample variance

```
s_var <- var(mns)
s_var
```

```
## [1] 0.5952215
```

Calculate the theoretical variance

```
t_var <- (lambda * sqrt(n)) ^ -2
t_var
```

```
## [1] 0.625
```

Compare the sample variance to the theoretical variance

```
s_var - t_var
```

```
## [1] -0.02977846
```

This comparison shows there is only a small difference between the simulation variance and the theorectical variance

The distribution can be shown in a histogram. This shows the sample means from the simulation with an overlay of a normal distribution.

```
hist(mns, prob = TRUE, col = "light blue", main =
             "Histogram of Means from Simulation", xlab = "simulation mean",
     breaks = 20)
```