Appendix - Study of Exponential Distribution and Comparison with Central Limit Theorem

Anna Teo

Code:

Simulations of a Exponential Distribution:

The exponential distribution is simulated in R using rexp(n, lambda) where lambda is the rate parameter which we will set it at 0.2 for all of the simulations here. In this study, we will investigate the distribution of mean and variance of 40 exponential numbers over a thousand times.

```
##  Simulate exponential numbers using rexp(n, lambda) and compute the mean and variance.
## This is then repeated for 1000 times.

lambda <- 0.2
n <- 40
mean_40 <- NULL
var_40 <- NULL
for (i in 1:1000) {
        value_40 <- rexp(n, lambda)
        mean_40 <- c(mean_40, mean(value_40))
        var_40 <- c(var_40, var(value_40))
        }
df <- data.frame(mean_40, var_40)
```

Plot the Sample Mean:

```
## Plot a histogram to show the distribution of the sample mean and to draw a red dashed
## line to show the mean of the distribution.

library(ggplot2)
g <- ggplot(df, aes(x=mean_40))
        + geom_histogram(alpha = .20, binwidth=0.2, colour = "black", aes(y = ..density..))
        + geom_vline(aes(xintercept=mean(mean_40, na.rm=T)),color="red",
                     linetype="dashed", size=1)
        + labs(x = "X")
        + labs(title = "Distribution of Mean of n = 40 exponential numbers")
        + theme(text = element_text(size=9))
g
```

Plot the Sample Variance:

```
## Plot a histogram to show the distribution of the sample variance and to draw a red
## dashed line to show the mean of the distribution.
```

```r
g <- ggplot(df, aes(x=var_40))
        + geom_histogram(alpha = .20, binwidth=2, colour = "black", aes(y = ..density..))
        + geom_vline(aes(xintercept=mean(var_40, na.rm=T)),color="red",
                     linetype="dashed", size=1)
        + labs(x = "X")
        + labs(title = "Distribution of Variance of n = 40 exponential numbers")
        + theme(text = element_text(size=9))
g
```

Distribution:

- We will simulate a standard normal random exponential variable n.
- $\mu = E[X] = \frac{1}{\lambda} = 5$
- $Var(X) = \frac{1}{\lambda^2} = 25$
- Standard Error $= \sqrt{\frac{25}{n}} = \frac{5}{\sqrt{n}}$
- Let's simulate n exponential numbers, take their mean, subtract off 5, and divide by $\frac{5}{\sqrt{n}}$ and repeat this over for 1000 simulations.
- We will perform the simulations for $n = 1$ and 40

```r
## Plot a histogram to show the distribution of the sample mean (properly normalised)
## for n = 1 and 40 and to show the normal distribution bell curve.

nosim <- 1000
lambda <- 0.2

## cfunc normalised the mean of the distrbution using sqrt(n) * (mean - E[X]/sqrt(Var(X)))
cfunc <- function(x, n) sqrt(n) * (mean(x) - 5) / 5

## create a data.frame of 2000 rows to store the properl normalised sample mean
## for n = 1 and 40
dat <- data.frame(
  x = c(apply(matrix(rexp(nosim * 1, lambda), nosim), 1, cfunc, 1),
        apply(matrix(rexp(nosim * 40, lambda), nosim), 1, cfunc, 40)
        ),
  size = factor(rep(c(1, 40), rep(nosim, 2))))

## plot histogram to show the distribution for n = 1 and 40 with normal distribution curve.
g <- ggplot(dat, aes(x = x, fill = size))
        + geom_histogram(alpha = .20, binwidth=.3, colour = "black", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, colour = "red") + theme(text = element_text(size=9))
g + facet_grid(. ~ size)
```