Statistical Inference Course Project - Part 1

Roberto Baladrón Pardo

In this project, each simulation used a rate parameter $\lambda = 0.2$ used to produce n = 40 random variables. The mean and the standard deviation of the 40 observations is calculated 10,000 times:

```
set.seed(1230); lambda <- .2; n <- 40; no_sim <- 10000
mean_values <- NULL; mean_sds <- NULL
for (i in 1:no_sim){
  values <- rexp(n, lambda); means <- mean(values); sds <- sd(values)
  mean_values  <- c(mean_values, means); mean_sds <- c(mean_sds, sds)
}
avg <- mean(mean_values); s <- sd(mean_values)
```

For an exponential distribution, the expected value is $\mathbb{E}[X] = 1/lambda$ which is equal to 5 in this case. The mean for the experiment was calculated to be 5.00. As expected, due to the CLT, the expected value of the sample mean is equal to the mean it's trying to estimate.

The distribution of the sample mean is gausian, centered at 5 and concentrated at the center. Is plotted below:

```
## [1] 5.002512
```


Probability density function for 10000 simulations
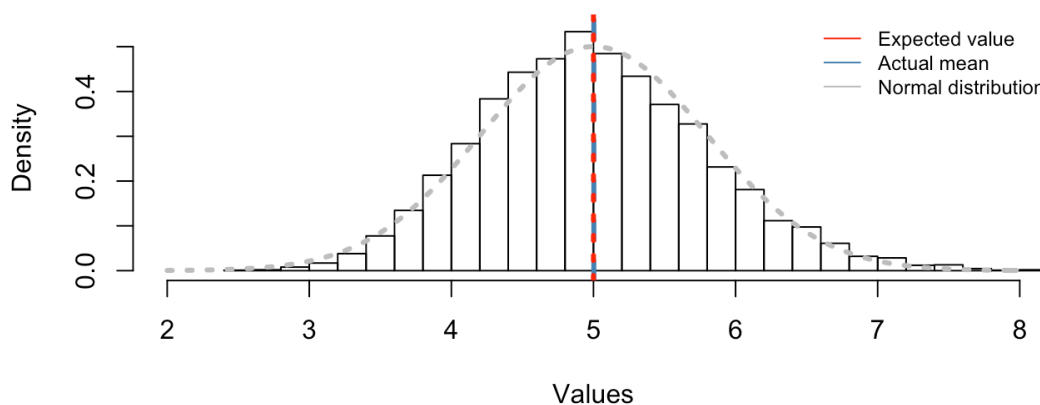
The variance of the sample mean was worked out to be 0.80. This corresponds with the standard error of the mean(i.e. $SE = sigma/\sqrt{n}$) which is equal to 0.79 for the 40 observations.

```
## [1] 0.7968704
```

There is little deviation between the actual quantile value and theoretical value; this indicates aggregate sample distribution is indeed normal. This is the plot of the mean of values:

Normal Q-Q Plot
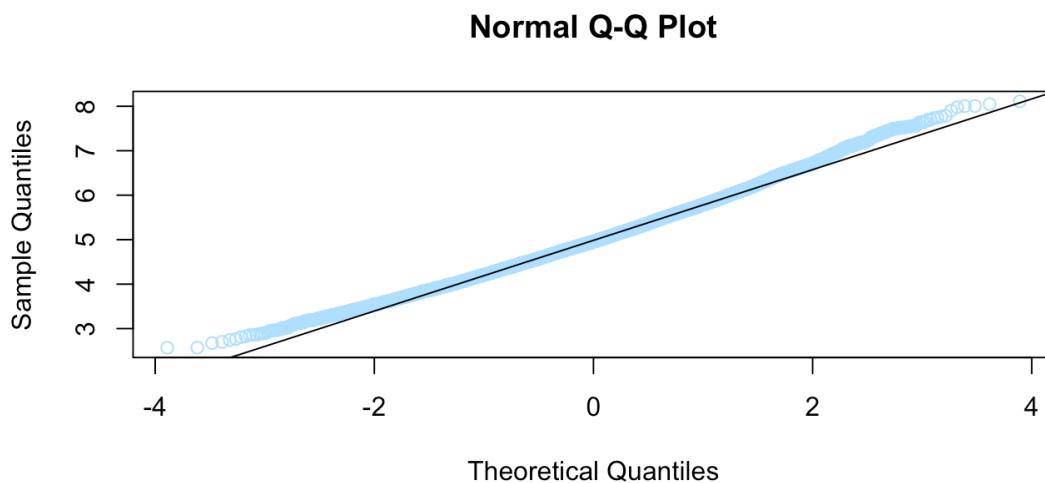
The 95% confidence interval of each simulation was worked out using each simulation's own standard deviaton and mean according to the equation $\bar{X} \pm 1.96 sigma/\sqrt{n}$. The coverage was computed as the percent of times the true mean fell within each interval's confidence interval.

```
upper <- mean_values +  1.96 * (mean_sds/sqrt(n))
lower <- mean_values -  1.96 * (mean_sds/sqrt(n))
sum(lower < 5 & 5 < upper)/no_sim * 100
```

```
## [1] 92.23
```

An extra experiment for 100 simulations was run, for visualization purposes. The confidence interval for each simulation was plotted below:

Plot of confidence interval coverage for 100 simulations