Investigation of the Central Limit Theorem Using the Exponential Distribution

Alex Van Russelt

17 December 2017

Overview

This document investigates the central limit theorem (CLT) through use of an exponential distribution. Random variables are generated using such a distribution, and then averaged. This is repeated a number of times, and it is apparent that the resulting distribution can be well approximated by a normal distribution, as predicted by the CLT.

Simulations

```r
set.seed(123)
simulation_count <- 1000 # The number of times to repeat the experiment
exponential_count <- 40 # The number of exponentials to average
lambda <- 0.2

means <- numeric(length=simulation_count)

for (i in 1:simulation_count){
  data = rexp(exponential_count, lambda)
  means[i] <- mean(data)
}

df_means <- data.frame(means)
```

The `rexp` function is used to generate 40 random variables using the exponential distribution with rate paramater `lambda` set equal to `0.2`. The mean is recorded. This is repeated 1000 times.

Comparison of sample mean to the theoretical mean

```r
sample_mean <- mean(means)
message(paste("Theoretical mean:", format(1/lambda, nsmall=2, digits=2)))
```

```
## Theoretical mean: 5.00
```

```r
message(paste("Sample mean:", format(sample_mean, nsmall=2, digits=2)))
```

```
## Sample mean: 5.01
```

The theoretical mean of an exponential distribution with rate parameter λ is given by $\frac{1}{\lambda}$. In this case, $\lambda = 0.2$, so the expected mean is 5.00. The sample mean is 5.01, which is very close to the theoretical value.

Variance of the sample compared to theoretical variance

```
sample_variance <- var(means)
standard_error <- (1 / lambda) ^ 2 / exponential_count
message(paste("Theoretical variance:", format(standard_error, nsmall=2, digits=2)))
```

```
## Theoretical variance: 0.62
```

```
message(paste("Sample variance:", format(sample_variance, nsmall=2, digits=2)))
```

```
## Sample variance: 0.60
```

The theoretical variance is given by the squared standard error of the means:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

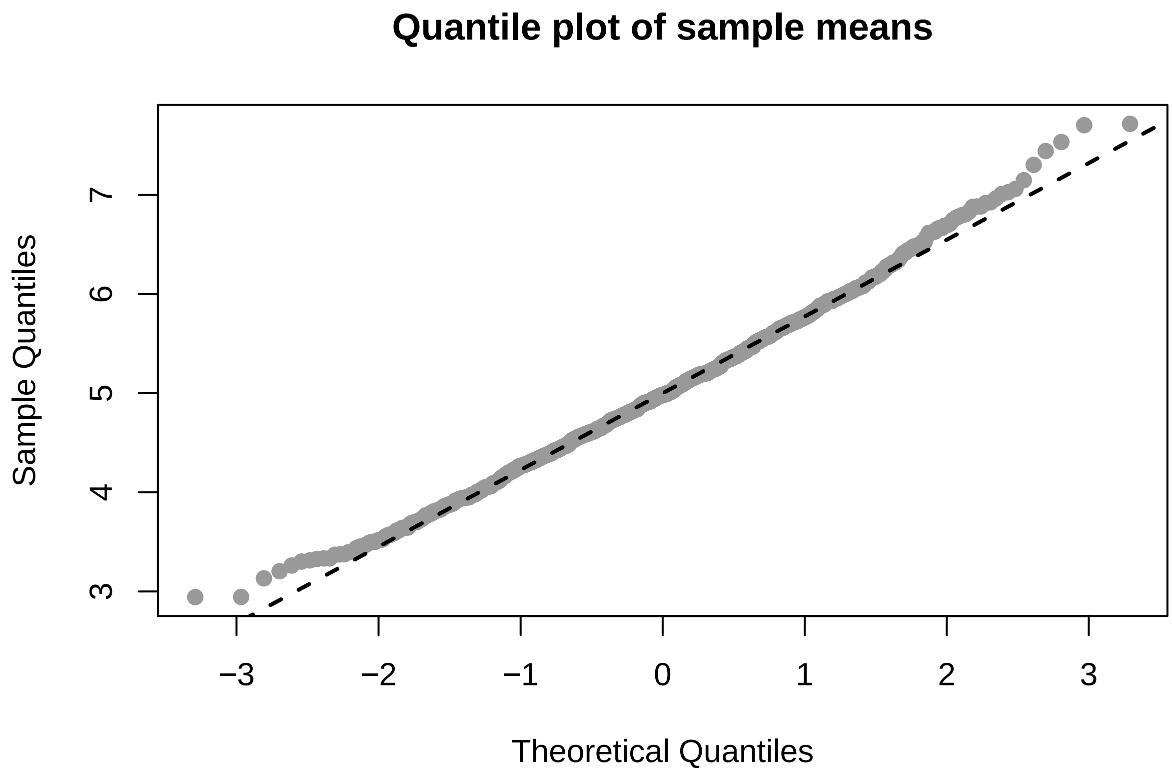
where:

1. $\sigma_{\bar{x}}$ is the standard error of the mean.
2. σ is the standard deviation of the population (which for an exponential distribution is equal to $\frac{1}{\lambda}$).
3. n is the number of observations (in this case, 40).

This produces a theoretical variance of 0.62. The sample mean is 0.60, which is very close to the theoretical value.

Comparing distribution of means to a normal distribution

```
qqnorm(means, col="#999999", pch=19,
       main='Quantile plot of sample means')
qqline(means, lwd=2, lty='dashed')
```



Here, a quantile plot is used to compare the quantiles of the means to that of a normal distribution. All data points lie close to the reference line, therefore, the distribution of the means is approximately normal.