Analysis of Exponential Distribution

Thammanoon Kawinfruangfukul

Overview

This report summarizes an analysis of sample and theoretical means and variances of the exponential distribution. Moreover, this report analyzes a distribution of samples based on the Central Limit Theorem.

Simulation

The report investigates the distribution of 1000 averages of 40 exponentials.

```
# System configuration
set.seed(952)
lambda = 0.2
simulations = 1000

# Calculate means of samples
means = NULL
for (i in 1 : simulations) {
    means = c(means, mean(rexp(40, lambda)))
}
```

Sample Mean versus Theoretical Mean

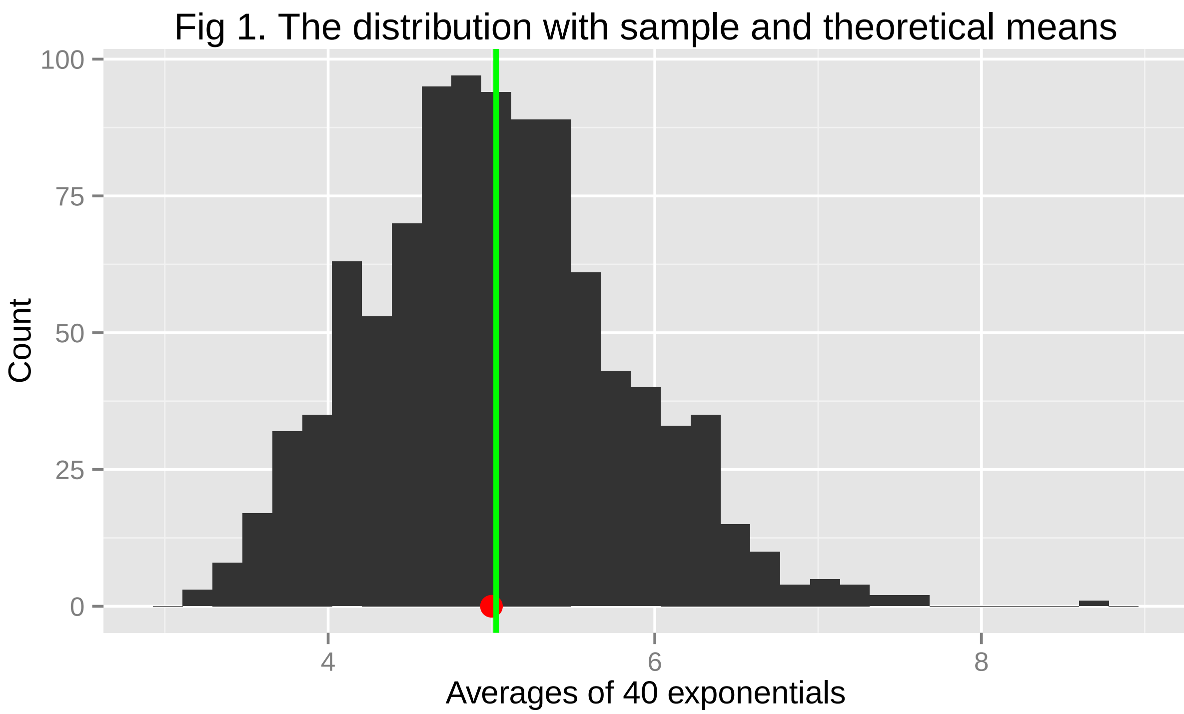
```
# Calculate theoretical mean
dist_mean = 1/lambda

# Calculate sample mean
sample_mean = mean(means)
```

The theoretical mean is 5 and sample mean is 5.02891. Based on the means, the sample mean is close to the theoretical mean. It is said that the sample mean is an unbiased estimator of the population mean. In other words, the sample mean is an estimate of the population mean.

Following figure shows the distribution of the average 40 exponentials with sample and theoretical means. The green vertical bar is the sample mean and the red dot is the theoretical mean.

```
# Plot the distribution with sample and theoretical means
library(ggplot2)
base <- qplot(means, geom="histogram")
base <- base + labs(title = "Fig 1. The distribution with sample and theoretical means")
base <- base + labs(y = "Count", x = "Averages of 40 exponentials")
base <- base + geom_point(aes(x = dist_mean, y = 0), colour="red", size = 4)
base <- base + geom_vline(xintercept = sample_mean, size = 1, colour="green")
print(base)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

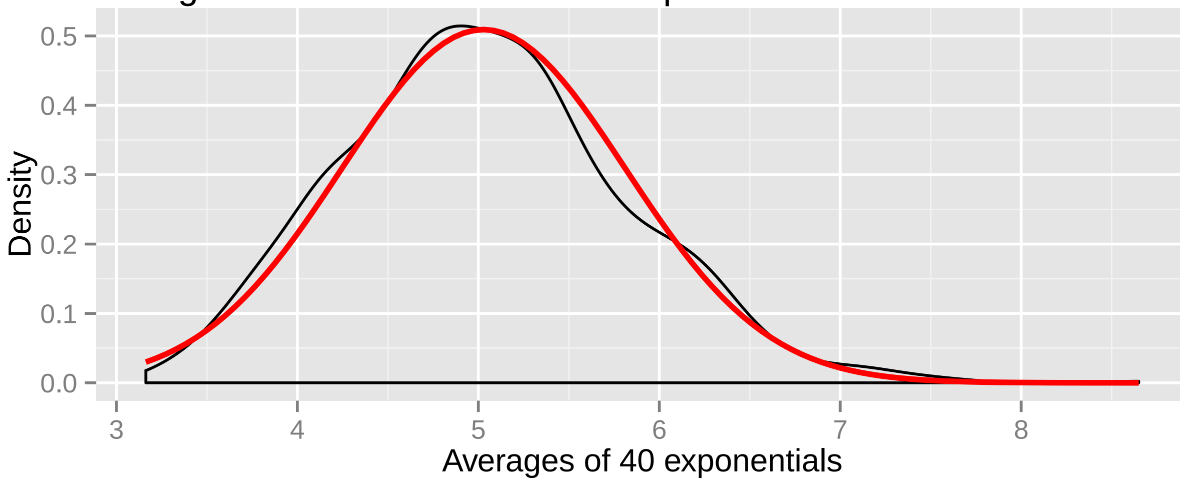


Sample Variance versus Theoretical Variance

Following figure shows the distribution of the mean of 40 exponentials with sample and theoretical variances. The black line is the sample variance and the red line is the theoretical variance.

```
# Calculate theoretical variance
dist_sd = 1/lambda
th_var = (dist_sd^2)/40

# Calculate sample variance
sample_var = var(means)

# Plot the distribution with sample and theoretical variances
base <- qplot(means, geom="density")
base <- base + labs(title = "Fig 2. The distribution with sample and theoretical variances")
base <- base + labs(y = "Density", x = "Averages of 40 exponentials")
base <- base + stat_function(fun=dnorm, size=1, color="red", arg=list(mean=sample_mean, sd=sd(means)))
print(base)
```

Fig 2. The distribution with sample and theoretical variances



The theoretical variance is 0.625 and sample variance is 0.6139094. The sample variance is close to the theoretical variance. This supports that the sample variance is an estimate of the population variance. Moreover, the figure 2 indicates that the sample variance centers around the population variance.

Distribution

Figure 2 shows that the distribution is similar to the normal distribution. According to Central Limit Theorem, when the sample size increases, the distribution will be closer to the normal distribution. Having larger sample size will make the sample mean be closer to the population mean, so it will help the sample variance be closer to the population variance and the normal distribution.

Fig 3. The distribution of 40,000 exponentials

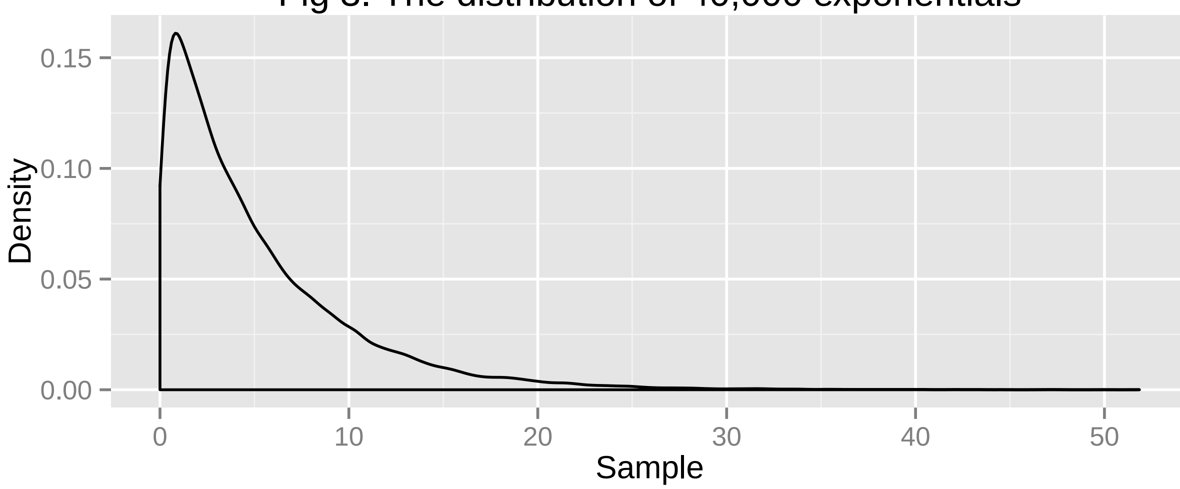


Figure 3 is the distribution of 40,000 exponentials. This distribution does not represent the normal distribution like the distribution of averages of 40 exponentials. If we can gather infinite samples, this distribution should be able to represent the exponential distribution. Figure 2 clearly shows Central Limit Theorem transforms the exponential distribution to the normal distribution.

Appendix

Code for generating figure 3

```r
# Generate samples
samples = rexp(40000, lambda)

# Plot the distribution of 40,000 exponentials
base <- qplot(samples, geom="density")
base <- base + labs(title = "Fig 3. The distribution of 40,000 exponentials")
base <- base + labs(y = "Density", x = "Sample")
print(base)
```