

A Demonstration of the Central Limit Theorem

A.K.N

Overview

In this report, we will simulate sampling a mean from an exponential distribution. We will investigate the sample mean and variance and compare it to the theoretical mean and variance. Finally, we will demonstrate the Central Limit Theorem by showing that the distribution of the mean is normally distributed.

Simulations

We wish to simulate sampling the mean of an exponentially distributed random variable. The rate parameter, λ , is set to 0.2. We run 1000 simulations of taking the mean of a sample of $n = 40$ exponentials. The sample means are calculated and saved to a data frame along with a set of 1000 random exponentials.

```
n = 40 #sample size
reps = 1000 #number of simulations
lambda = 0.2
set.seed(936907)
#Simulations
samples <- replicate(reps, rexp(n,lambda))
a <- rexp(reps, lambda); b <- apply(samples, 2, mean)
data <- data.frame(exp = a, means = b)
#Calculations
mu = 1/lambda; s = (1/lambda)/sqrt(n); sigma = s^2 #theoretical mu, sigma and s
obmean <- mean(data$means) #observed mu
sd_sample_mean <- sd(data$means); var_sample_mean <- var(data$means) #obs. s and sigma
```

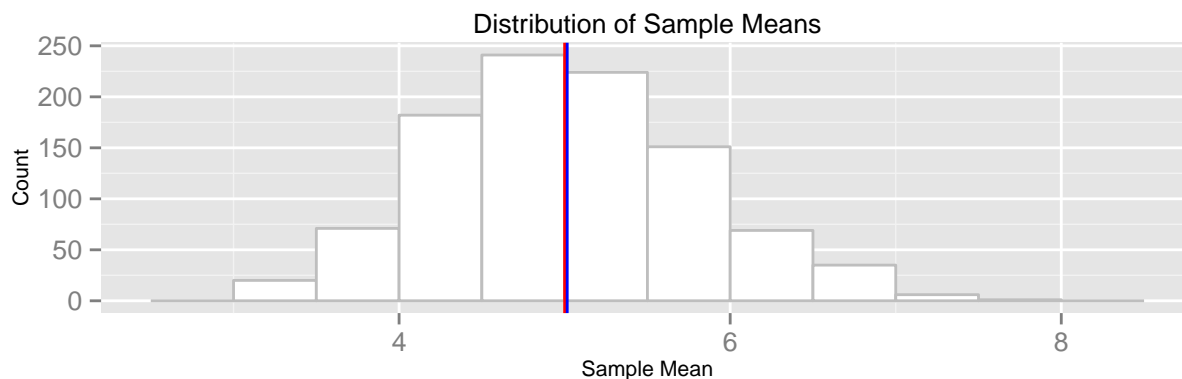
Sample Mean and Variance vs. Theoretical Mean and Variance

The exponential distribution has a theoretical mean equal to the inverse of rate parameter, $\frac{1}{\lambda} = 5$. The sample mean \bar{x} , is an estimate of the population mean. The standard deviation of an exponential distribution is also $s = \frac{1}{\lambda} = 5$. Since the variance is $\sigma = s^2$, the variance of an exponential distribution with $\lambda = 0.2$ would be $5^2 = 25$. However, in this case we wish to calculate the variance of the distribution of the sample means. Therefore:

$$s = \frac{1}{\lambda * \sqrt{n}} = \frac{5}{\sqrt{40}} = 0.79 \text{ and } \sigma = s^2 = 0.625$$

These theoretical values are compared to the values from the simulated sample means in *Figure 1*. The left panel shows a histogram of the distribution of sample means. Two lines mark the location of the center of the distribution (blue), and the theoretical center (red).

Variable	Theoretical	Observed
Mean	5	5.015
SD	0.791	0.792
Variance	0.625	0.627



Distributions

We will use our simulated data to test the central limit theorem. By the CLT, the simulated sample mean will have a normal distribution, despite the shape of the underlying distribution. If one takes enough samples, the mean of those samples will be *normally* distributed. First, with our simulated data, we can construct a 95% CI for sample mean and compare it to the 95% CI for a normal distribution of mean $\frac{1}{\lambda}$.

```
#Construct a CI for the sample mean
sample95CI <- obmean + c(-1,1)*qt(0.975, n-1) * (sd_sample_mean/sqrt(n))
normal95CI <- mu + c(-1,1)*1.96*(s/sqrt(n))
rbind(sample95CI, normal95CI)
```

```
##           [,1]      [,2]
## sample95CI 4.762268 5.268607
## normal95CI 4.755000 5.245000
```

Thus far, the parameters of the simulated distribution are very similar to the theoretical values. Next, we can visually compare the distribution of a large number of means of exponentials to the distribution of a large number of exponentials. In *Figure 2*, we use density plots to compare the shapes of these two distributions. In the left panel of *Figure 2*, we can see that the shape of the distribution of means is normal even though the exponential distribution is not. For reference, we overlay a plot of a normal distribution of $n = 1000$, centered around μ . The right panel provides another way to confirm the normality of the distribution of means: the quantile-quantile plot, or Q-Q plot. This plot shows that the sample quantiles are very close to the theoretical quantiles, another indication that the distribution is in fact normal.

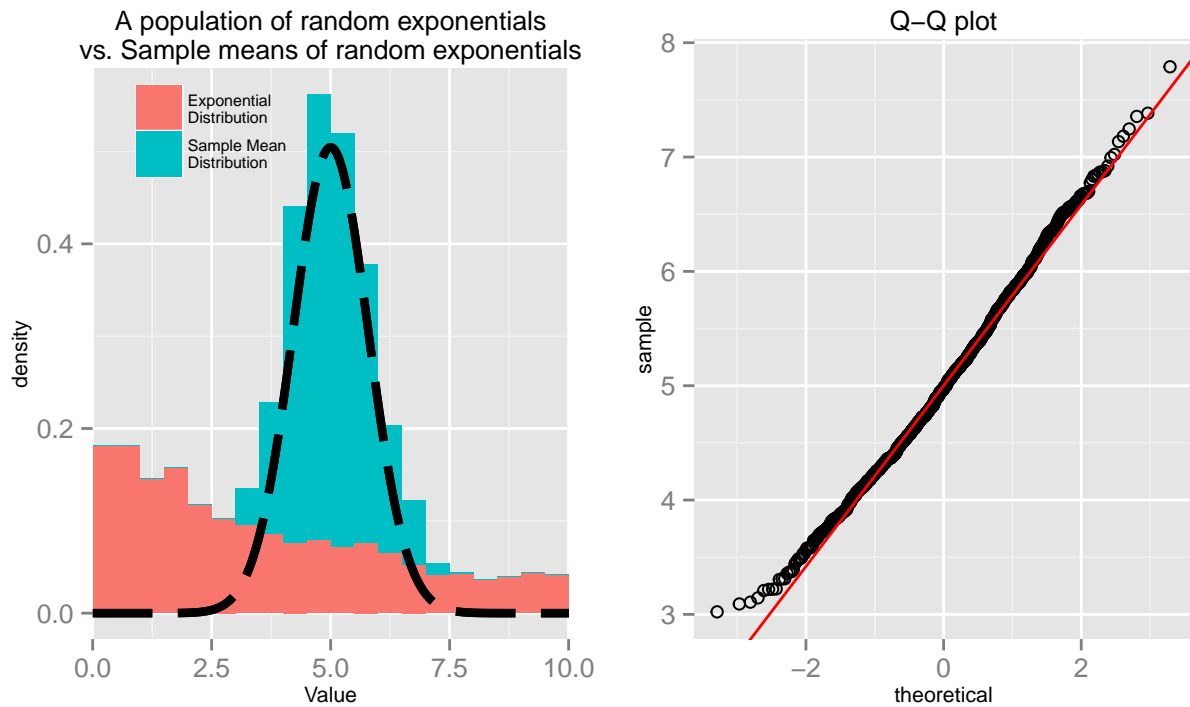


Figure 1: Demonstrating that the Distribution of Sample Means is Normal

Conclusions

1. We simulated sampling the mean of an exponential distribution. The simulated mean was 5.015, close to the theoretical mean, 5.
2. The simulated variance, 0.625, was close to the variance of the sampled means, 0.627.
3. The two endpoints of the 95% CI of the distribution of sample means, 4.762, 5.269, match closely the 95% CI of a normal distribution centered at μ , 4.755, 5.245.
4. We began with a population of exponentials that were not normally distributed. Then, we took the means of many samples of size n . The distribution of these sample means *are normally distributed*. This is the main prediction of the Central Limit Theorem.

Note: Code for graphs not included in this report due to space limits, but markdown file can be found [here](#).