# Exponential Distribution is Normal?

Giulio Mario Martena

6/3/2020

## Overview

The purpose of this analysis is to verify the similarities between a normal an exponential distribution.

- *The exponential distribution is closely related to the Poisson distribution, and can be simulated in R with `rexp(n,lambda)`, where lambda is the rate parameter.*

The mean of the exponential distribution is `1/lambda`, and the standard deviation is also `1/lambda`.

In this study lambda will be fixed to **0.2**.

*The distribution of averages across 40 exponential* will be investigated in this study, across 1000 simulations.

## Exponential distribution simulation

```r
set.seed(1234) # For reproducibility
lambda <- 0.2 ; n <- 40 # Number of distributions
sims <- 1000 # Number of simulations per distribution
simulation <- replicate(sims, rexp(n,lambda)) #replicate returns a matrix
meanExp <- apply(simulation,2,mean)
```

### Sample Mean versus Theoretical Mean

```r
meanSample <- mean(meanExp)
meanTheoretical <- 1 / lambda
```

The mean sample is 4.9742388 and the theoretical is 5. The difference between the two is 0.03, so they are indeed similar. The distribution is plotted in *figure 1*, where a normal trend is visible.

The natural next step is to repeat the steps for the **variance**.

### Comparison of sample and theoretical variance of the distribution

```r
varianceSample <- var(meanExp)
varianceTheoretical <- (1 / lambda * 1 / sqrt(n)) ^ 2
```

The mean std. deviation is 0.7554171 and the theoretical is 0.7905694. The difference between the two std. deviations is 0.03.

### Check if distribution tends to a normal

*If theory checks out, the distribution should tend to a normal distribution with mean and standard deviation equal to the sample ones.*

```r
xfit <- seq(min(meanExp), max(meanExp), length = 2*n)
yfit <- dnorm(xfit, mean = meanSample, sd = sqrt(varianceSample))
```

The code above constructs a normal fit from the exponential means. *Figure 2* shows such similarity.

```r
mean(meanSample) + c(-1, 1) * qnorm(.975) * sqrt(varianceSample)/sqrt(n)
```

```
## [1] 4.740137 5.208341
```

The code above calculates the 95% C.I. for the sample mean.

- In terms of hypothesis this means that $H_0 : \mu - \mu_s = 0$ **cannot** be rejected.

Furthermore the similarity was expected from the CLT, of course, but how do we know we are not dealing with an actual normal distribution? Performing a normality check. In this case I will use Kolmogorov-Smirnov test.

```r
ks.test(xfit, meanExp)$p.value
```

```
## [1] 0.0001360558
```

Since $p_{ks} < \alpha$ we reject $H_0 : X \sim N(\mu_s, \sigma^2/n)$ and conclude that the two distributions are not, in fact, equally distributed.

## Appendix

```r
hist(meanExp, breaks = n,
     xlab="",ylab="", main="Distribution means",
     col = rgb(0,0,0,.15), xlim = c(4,6));
abline(v = meanSample, col='blue');
abline(v = meanTheoretical, col='red', lty=2);
legend('topright', col=c('blue','red'),
       legend = c('Sample Mean','Theoretical mean'),
       pch='-', cex = .5)
```

```r
hist(meanExp,breaks=n,prob=T,xlab = "",main="",ylab="",col=rgb(.8,.6,.4,.7));
lines(xfit, yfit, col='red', lwd=5)
```

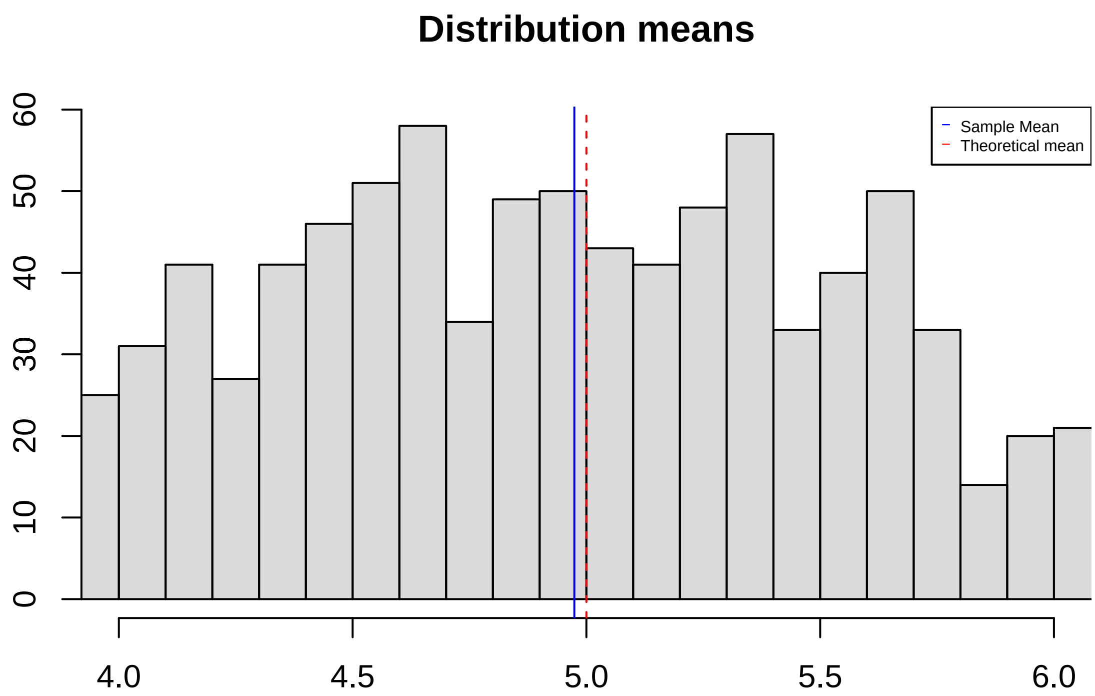


Figure 1: Histogram showing the sample (in blue) and theoretical (in red) mean of the distributions

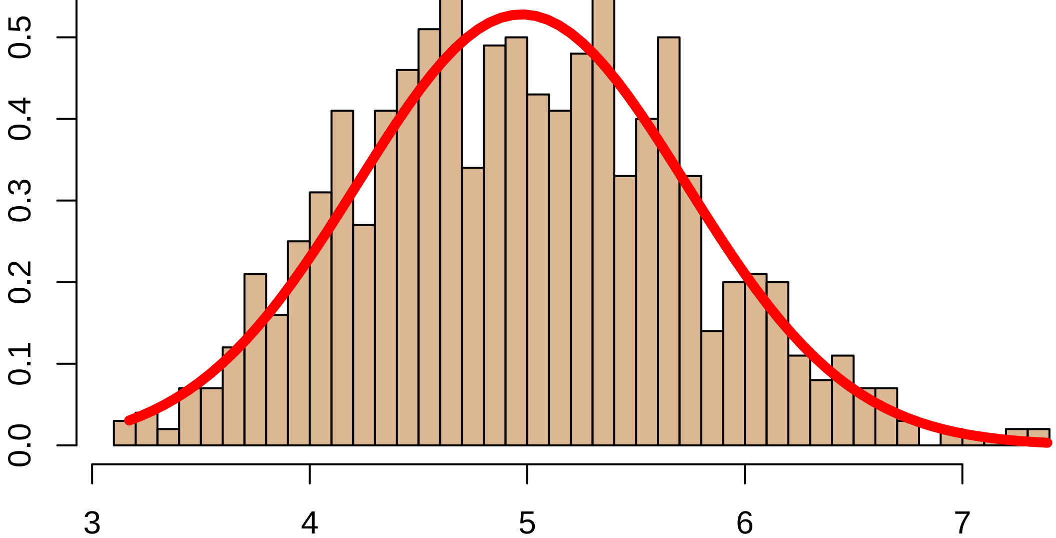


Figure 2: Exponential vs Normal Distribution