Central Limit Theorem and the Exponential Distribution

Derwin McGeary

September 18, 2015

Overview

We are going to demonstrate how the central limit theorem applies by using a simulation from the exponential distribution.

Simulations

So we start by setting the parameters and loading the ggplot2 library. We also set the seed here to make the results fully reproducible. They don't need to be in this case, but it's a good habit.

```
library(ggplot2)
lambda <- 0.2
nosim <- 1000
n <- 40
set.seed(1477)
```

Now we take 1000 averages from 40 samples from `rexp` with $\lambda = 0.2$.

```
averages=NULL
for(i in 1:nosim) averages=c(averages, mean(rexp(n,lambda)))
```

Sample Mean versus Theoretical Mean

Let us take the mean of our samples.

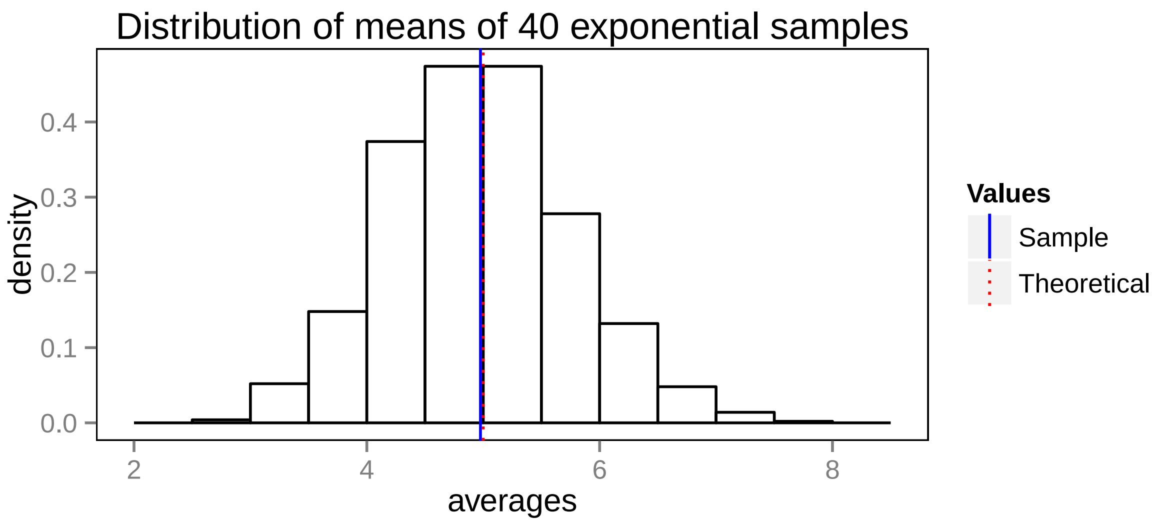
```
avmean <- mean(averages)
emean <- 1/lambda
```

The mean of the averages is 4.9761391. The expected mean of the exponential distribution is $\mu = \frac{1}{\lambda} = 5$. That's a difference of 0.48%.

Let's look at that graphically

```
histo <- ggplot(as.data.frame(averages), aes(x=averages)) +
  geom_histogram(binwidth=.5, colour="black", fill="white", aes(y=..density..)) +
  scale_colour_manual(name="Values", values=c(Sample="blue",Theoretical="red")) +
  scale_linetype_manual(name="Values", values=c(Sample="solid",Theoretical="dotted")) +
  theme(legend.background=element_blank(),
        panel.grid.major=element_blank(),
        panel.grid.minor=element_blank(),
        panel.background=element_rect(colour="black", fill="white")) +
  ggtitle("Distribution of means of 40 exponential samples")
histo +
```

```
  geom_vline(aes(xintercept=mean(averages),
                 color="Sample", linetype="Sample"), show_guide = TRUE) +
  geom_vline(aes(xintercept=1/lambda,
                 colour="Theoretical", linetype="Theoretical"), show_guide = TRUE)
```



We can see that the sample and theoretical means are practically on top of each other.

Sample Variance versus Theoretical Variance

Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

The theoretical variance of the exponential function is λ^{-2} which is 25 in this case. The expected sample variance is $\frac{\lambda^{-2}}{N}$ which is 0.625. In reality, our samples give us

```
esampvar <- (lambda^-2)/n # this is the expected sample variance
print(avvar <- var(averages)) # this is our actual sample variance
```

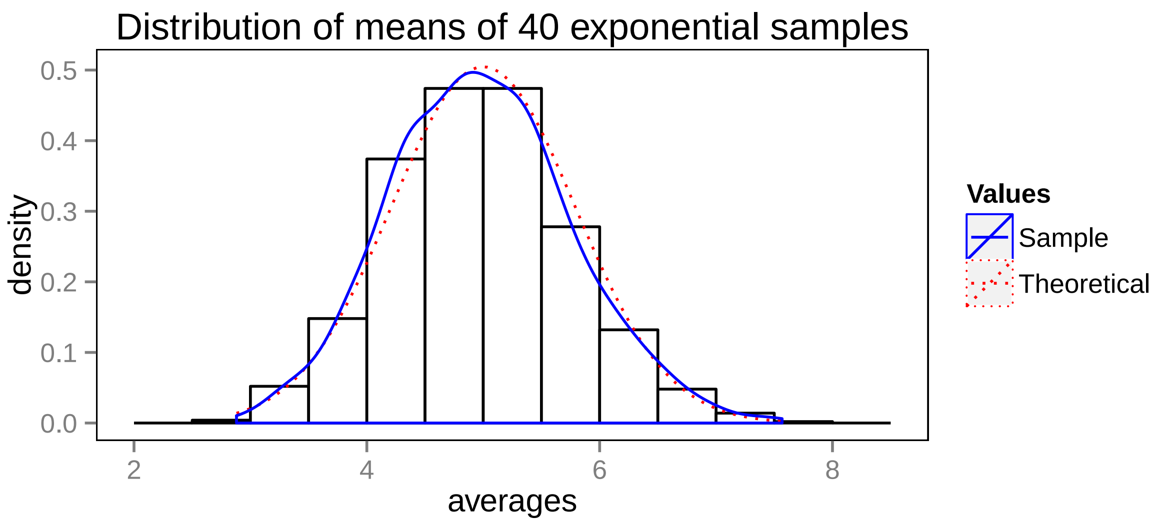
```
## [1] 0.6043701
```

So the difference is 3.3%.

Distribution

The following chart displays the density function and the normal curve.

```
histo +
  stat_function(fun = dnorm, args=list(mean=1/lambda, sd=sqrt(esampvar)),
                aes(colour="Theoretical", linetype="Theoretical")) +
  geom_density(alpha=.2, aes(colour="Sample",linetype="Sample"))
```



So it can be seen that our distribution is indeed approximately normal.