

Simulate CLT

Stanislav Prikhodko

Sunday, May 10, 2015

Part 1. Sampling Exponential Distribution

Overview

In this project I will investigate the exponential distribution and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. I will investigate the distribution of averages of 40 exponentials.

Project

The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. I will use $\lambda = 0.2$ for all of the simulations. In this study I will investigate the distribution of averages of 40 exponentials.

Theoretical Mean of the distribution

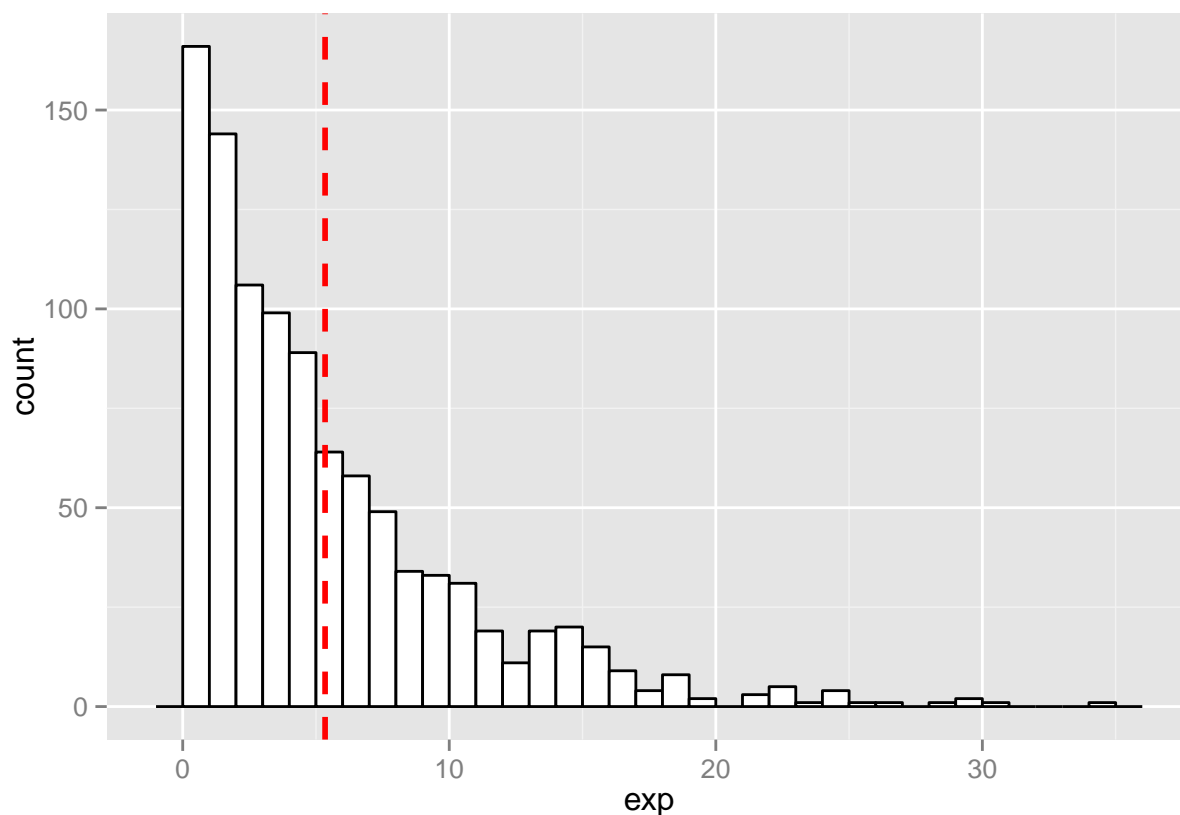
Creating 1000 values from the exponential distribution

```
lambda <- 0.2
exp <- rexp(n=1000, rate = lambda)
exp_mean <- mean(exp)
theoretical_mean <- (1/lambda)
```

The sample mean is **5.3411229**. Theoretical mean is **5** (calculated as $1/\lambda$).

Let's plot a histogram to show the sample mean and compare it to the theoretical mean of the distribution.

```
ggplot() + aes(exp)+
  geom_histogram( binwidth=1, colour="black", fill="white") +
  geom_vline(aes(xintercept=exp_mean), color="red", linetype="dashed", size=1)
```



Theoretical Variance of the Sample

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
sample_size <- 40
mns <- rep(0,1000)
for (i in 1:1000) {
  mns[i] <- mean(exp[sample(1:1000, sample_size, replace=FALSE)])
}

sample_var <- var(mns)
theoretical_var <- (1/lambda)^2 / sample_size
```

3. Show that the distribution is approximately normal.

```
#distribution of 1000 averages of 40 random uniforms
```

```
p1 <-
  ggplot() + aes(mns)+
```

```
geom_histogram(aes(y=..density..), binwidth=0.1, colour="black", fill="white") +
geom_vline(aes(xintercept=mean(mns, na.rm=T)), color="red", linetype="dashed", size=1) +
geom_density(alpha=.2)
```

```
p2 <-
ggplot() + aes(mns)+
geom_histogram( binwidth=0.1, colour="black", fill="white") +
geom_vline(aes(xintercept=mean(mns, na.rm=T)), color="red", linetype="dashed", size=1)

multiplot(p1,p2,cols=2)
```

