Investigate the exponential distribution in R and compare it with the Central Limit Theorem

Glenn Kerbein

April 1, 2017

Overview

The purpose of this analysis is to investigate the exponential distribution and compare it to the Central Limit Theorem. For this analysis, set `lambda` equal to `0.2` for all simulations. The investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.

Simulations

Set the simulation variables lambda, exponentials, and seed.

```
ECHO=TRUE
set.seed(9001)
lambda <- 0.2
exponentials <- 40
```

Run the simulations with variables.

```
simMeans = NULL
for (i in 1 : 1000) {
    simMeans <- c(
        simMeans,
        mean(
            rexp(
                exponentials,
                lambda
            )
        )
    )
}
```

Sample Mean versus Theoretical Mean

Sample Mean

Calculating the mean from the simulations gives the sample mean.

```
mean(simMeans)
```

```
## [1] 5.004444
```

Theoretical Mean

The theoretical mean of an exponential distribution is `lambda^(-1)`.

```
lambda ^ (-1)
```

```
## [1] 5
```

Comparison

```
abs(mean(simMeans)-lambda^(-1))
```

```
## [1] 0.004443579
```

There is only a slight difference between the simulations sample mean and the exponential distribution theoretical mean.

Sample Variance versus Theoretical Variance

Sample Variance

Calculating the variance from the simulation means gives the sample variance.

```
var(simMeans)
```

```
## [1] 0.6335885
```

Theoretical Variance

The theoretical variance of an exponential distribution is `( lambda * sqrt(n) ) ^ (-2)`.

```
( lambda * sqrt(exponentials) ) ^ (-2)
```

```
## [1] 0.625
```

Comparison

There is only a small difference between the simulations sample variance and the exponential distribution theoretical variance.

```
abs(var(simMeans)-(lambda * sqrt(exponentials))^(-2))
```

```
## [1] 0.008588534
```

Distribution

Below is a density histogram of the 1000 simulations. There is an overlay with a normal distribution that has a mean of `lambda ^ (-1)` and standard deviation of `( lambda * sqrt(n) ) ^ (-1)`, the theoretical normal distribution for the simulations.

```
hist(
    simMeans,
    prob=TRUE,
    col="#0054a6",
    main="Mean distribution for rexp()",
    xlab = "Simulated means",
    breaks = 40
```

```
)
lines(
    density(simMeans),
    lwd=3,
    col="#e36209"
)
```