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

Anshuman Swaroop

4th October 2024

Overview

The purpose of this analysis is to understand exponential distribution and compare it to the Central Limit Theorem. We'll set lambda = 0.2, simulate 1000 averages of 40 exponentials, and perform the following tasks:

```
1) Sample Mean vs Theoretical Mean
2) Sample Variance vs Theoretical Variance
3) Distribution Approximation to Normal
```

Simulations

Set the simulation variables lambda, exponentials, and seed.

```
ECHO=TRUE
set.seed(1337)
lambda = 0.2
exponentials = 40
```

Run 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 with give the sample mean.

```
mean(simMeans)
```

```
## [1] 5.055995
```

Theoretical Mean The theoretical mean of an exponential distribution is lambda^-1.

```
lambda^-1
```

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

Comparison There is a difference in the simulations sample mean and the exponential distribution's theoretical mean.

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

```
## [1] 0.05599526
```

Sample Variance versus Theoretical Variance

Sample Variance Calculating the variance from the simulation means with give the sample variance.

```
var(simMeans)
```

```
## [1] 0.6543703
```

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

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

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

Comparison There is a slight difference between the simulations sample variance and the exponential distribution's theoretical variance.

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

```
## [1] 0.0293703
```

Distribution

This 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.

```
library(ggplot2)
ggplot(data.frame(y=simMeans), aes(x=y)) +
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="#0072B2",
                 color="black") +
  stat_function(fun=dnorm, arg=list(mean=lambda^-1,
                                    sd=(lambda*sqrt(exponentials))^-1),
                size=2) +
  labs(title="Plot of the Simulations", x="Simulation Mean")
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## Warning in stat_function(fun = dnorm, arg = list(mean = lambda^-1, sd = (lambda
## * : Ignoring unknown parameters: ‘arg‘
```

```
## Warning: The dot-dot notation (‘..density..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(density)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```