Investigating the exponential distribution in R

Kiattisak Chaisomboon

07/06/2020

Overview

We will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with "rexp(n, λ)" where n is number of observations and λ is the rate parameter.

Simulations

Creating the list of sample mean (named 'mns') and the list of sample variance (named 'vrs') from 1000 samples which simulated using the function of the exponential distribution "rexp(n, λ)" where n = 40 and λ = 0.2

```
library(ggplot2)

set.seed(23)
sim <- 1000
n <- 40
lambda = 0.2

mns = NULL
vrs = NULL
for (i in 1 : sim) {
  rnd = rexp(n, lambda)
  mns = c(mns, mean(rnd))
  vrs = c(vrs, var(rnd))
}
```

Sample Mean versus Theoretical Mean

Calculating the sample mean.

```
sample.mean <- mean(mns)
sample.mean
```

```
## [1] 5.01425
```

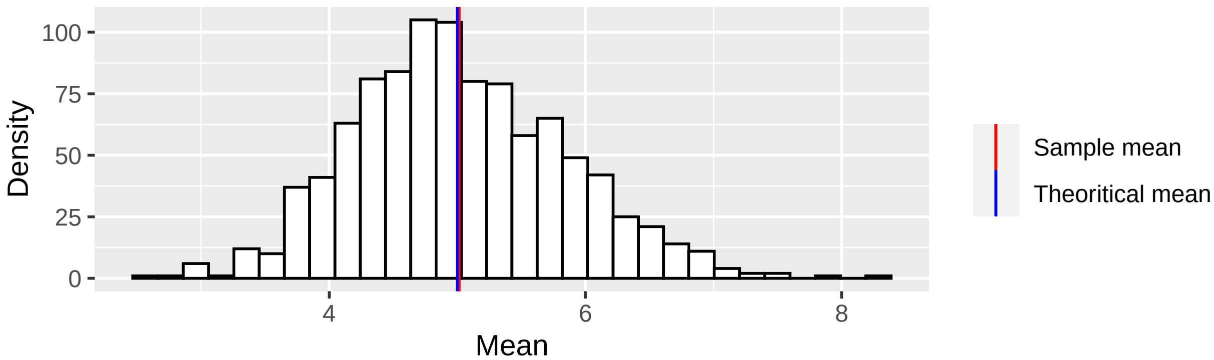
The mean of exponential distribution is $1/\lambda$, can be used to calculate the theoritical mean as the follow:

```
theoritical.mean <- 1/lambda
theoritical.mean
```

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

The sample mean (5.01425) is very close to the theoritical mean (5), as shown in the following figure:

```
dfm <- data.frame(mns)
ggplot(dfm, aes(mns)) +
  geom_histogram(bins=30, colour="black", fill="white") +
  geom_vline(aes(xintercept = sample.mean, colour='Sample mean')) +
  geom_vline(aes(xintercept = theoritical.mean, colour='Theoritical mean')) +
  scale_color_manual(name="", values = c('Sample mean' = 'red',
                                         'Theoritical mean' = 'blue')) +
  xlab("Mean") + ylab("Density")
```



Sample Variance versus Theoretical Variance

Calculating the sample variance.

```
sample.variance <- mean(vrs)
sample.variance
```

```
## [1] 24.92895
```

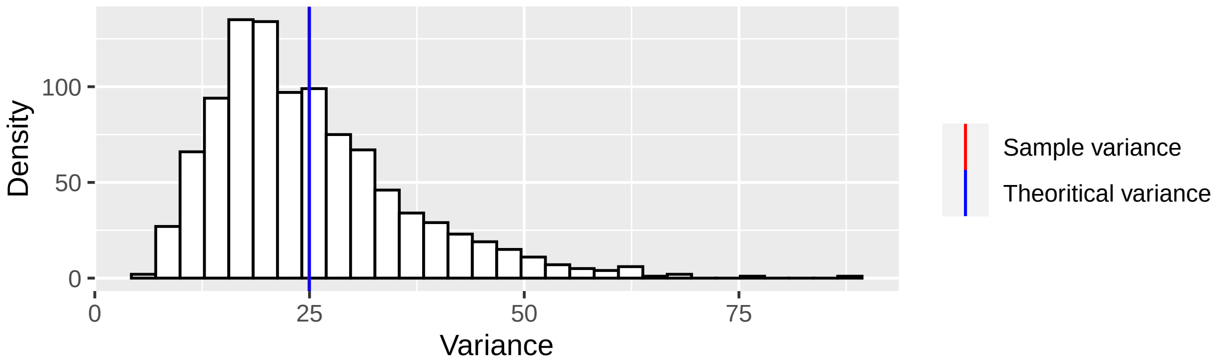
The variance of exponential distribution is $(1/\lambda)^2$, can be used to calculate the theoritical variance as the follow:

```
theoritical.variance <- (1/lambda)^2
theoritical.variance
```

```
## [1] 25
```

The sample variance (24.92895) is very close to the theoritical variance (25), as shown in the following figure:

```
dfv <- data.frame(vrs)
ggplot(dfv, aes(vrs)) +
  geom_histogram(bins=30, colour="black", fill="white") +
  geom_vline(aes(xintercept = sample.variance, colour='Sample variance')) +
  geom_vline(aes(xintercept = theoritical.variance, colour='Theoritical variance')) +
  scale_color_manual(name="", values = c('Sample variance' = 'red',
                                         'Theoritical variance' = 'blue')) +
  xlab("Variance") + ylab("Density")
```



Distribution

The CLT states that the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases. The result is that $Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ has a distribution like that of a standard normal for large n.

```
z <- ((mns - sample.mean)/(sample.variance / sqrt(sim)))

df <- data.frame(z)
ggplot(df, aes(x=z)) +
  geom_histogram(aes(y=..density..), bins=30, colour="black", fill="white") +
  geom_density(aes(colour='Sampling Distribution')) +
  geom_line(stat = "function", fun = "dnorm", args = list(mean = 0, sd = 1),
            aes(colour='Standard Normal Distribution')) +
  scale_color_manual(name="", values = c('Sampling Distribution' = 'red',
                                         'Standard Normal Distribution' = 'blue')) +
  xlab("Z") + ylab("Density")
```