Part 1: Simulation Exercise

Hasan Shojaei

20 June 2018

Overview

In Part 1, we investigate the exponential distribution and compare it with Central Limit Theorem. We show that the distribution of sample means is centered at theoretical mean of the popilation, and has a variance that is equal to the variance of the population divided by sample size.

Simulating Exponential Distribution

We use rexp(n, lambda) to simulate exponential distribution in R. We investigate the distribution of averages of 40 exponentials by performing a thousand simulations.

```
n1 <- 1000       # number of simulations
n2 <- 40         # sample size
lambda <- 0.2    # exponential rate, also 1/mean and 1/sd
set.seed(1)      # set seed to ensure reproducibility
# drwa 40 random exponentials and take their average. repeat 1000 times
mns = NULL
for (i in 1 : n1) mns = c(mns, mean(rexp(n2,lambda)))
# calculate mean and variance of averages
avg <- mean(mns); var <- var(mns)
print(avg)
```

```
## [1] 4.990025
```

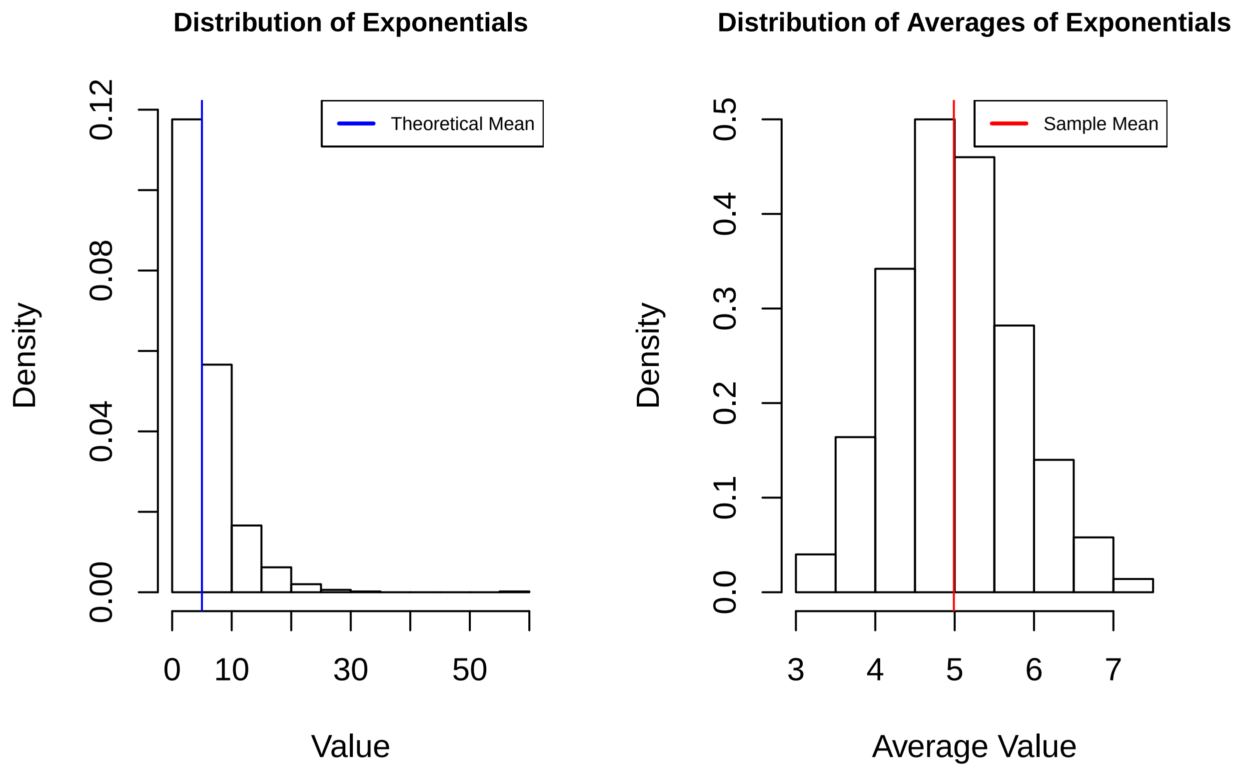
```
print(var)
```

```
## [1] 0.6111165
```

Sample Mean vs. Theoretical Mean

We compare sample mean and theoretical mean in this section. Let's plot the distribution of a large collection of random exponentials (left figure below) and the distribution of a large collection of averages of 40 exponentials (right figure below) first. The vertical lines denote mean values.

```
par(mfrow=c(1,2))
# draw 1000 random exponentials and plot their distribution
hist(rexp(n1,lambda), prob=T,
     main="Distribution of Exponentials", xlab="Value", cex.main=0.85)
abline(v=1/lambda, col="blue")
legend("topright", legend="Theoretical Mean", col="blue", lwd=2, cex=0.6)
# plot the distribution of sample means
hist(mns, prob=T,
     main="Distribution of Averages of Exponentials",
     xlab="Average Value", cex.main=0.85)
abline(v=avg, col="red")
legend("topright", legend="Sample Mean", col="red", lwd=2, cex=0.6)
```



Sample mean is **4.99**, which is very similar to the theoretical mean of the distribution: $\mu = 1/\lambda = \mathbf{5}$.

Sample Variance vs. Theoretical Variance

Sample variance is **0.61**, which is much smaller than the theoretical variance of the distribution: $\sigma^2 = 1/\lambda^2 = \mathbf{25}$. This is evident from the figure above too. According to Central Limit Theorem, if we multiply sample variance (**0.61**) by sample size (**40**), we obtain a good estimate of the distribution variance: **24.44**.

Sampling Distribution of the Mean vs. Population Distribution

In this section we compare sampling distribution of the means (right figure below) with population distribution (left figure below). The sampling distribution of the mean looks far more Gaussian than the original exponential distribution! For reference, we have added Exponential (with rate=lambda) and Gaussian (with mean=1/lambda and sd=1/lambda/sqrt(n2)) probability density plots to the left and right figures respectively.

```
par(mfrow=c(1,2))
# draw 1000 random exponentials and plot their distribution
exp_distr <- rexp(n1,lambda)
hist(exp_distr, prob=T,
     main="Distribution of Exponentials", xlab="Value", cex.main=0.85)
x <- seq(min(exp_distr), max(exp_distr), length = 100)
lines(x, dexp(x, rate=lambda), pch = 25, col = "green3")
legend("topright", legend="Exponential distribution", col="green3", lwd=2, cex=0.6)
```

```r
# plot the distribution of sample means
hist(mns, prob=T,
     main="Distribution of Averages of Exponentials",
     xlab="Average Value", cex.main=0.85)
x <- seq(min(mns), max(mns), length = 100)
lines(x, dnorm(x, mean = 1/lambda, sd = (1/lambda/sqrt(n2))), pch = 25, col = "green3")
legend("topright", legend="Normal distribution", col="green3", lwd=2, cex=0.6)
```