Statistical Inference Course Project Part 1

Tulin Varol

04/06/2018

Simulation Exercise

In this project, the Central Limit Theorem (CLT) will be explored. In order to do this, 1000 simulations will be generated using an exponential distribution with a rate parameter of 0.2. In each simulation, 40 random numbers will be generated.

Sample Mean versus Theoretical Mean

Let's sample 40 exponentials 1000 times and get the distribution of means. For the exponential distribution, the theoretical mean is 1/lambda which is 1/0.2=5 in this case. The mean of the means from the simulation should be close to the theoretical mean. We'll check if this is really the case.

```
# Set the parameters as given in the assignment
n <- 40
lambda <- 0.2
nSim <- 1000

# First compute the theoretical mean
true.mean <- 1/0.2

# Set the seed for reproducibility
set.seed(1234)

# Generate 1000 simulations with 40 exponentials and with a rate of 0.2
means <- replicate(nSim, {
  mm <- rexp(n, lambda)
  mean(mm)
  })

# Print the mean of means
sample.mean <- mean(means)
print(sample.mean)
```

```
## [1] 4.974239
```

The mean of means from this distribution is 4.97 which very close to the theoretical value of 5. This can be seen on the plot (last page) as well. Please note that it is a density distribution, so it doesn't show the frequency of means but density of means. In other words, the distribution is normalized to unity. Theoretical and sample means are shown by blue and red vertical lines, respectively.

Sample Variance versus Theoretical Variance

The theoretical standard deviation should be the standard deviation of the exponential distribution divided by the square root of the sample size: $\frac{\sigma}{\sqrt{n}}$ so $5/\sqrt{40} = 0.79$ in our case. The theoretical variance, in this case, is expected to be $0.79^2 = 0.62$. We'll check if the sample variation is close to the theoretical variance.

```
# First compute the theoretical standard deviation and variance
true.sd <- (1/0.2)/sqrt(40)
true.variance <- (true.sd)^2
print(true.variance)
```

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

```
# Print the variance of means
sample.sd <- sd(means)
sample.variance <- (sample.sd)^2
print(sample.variance)
```

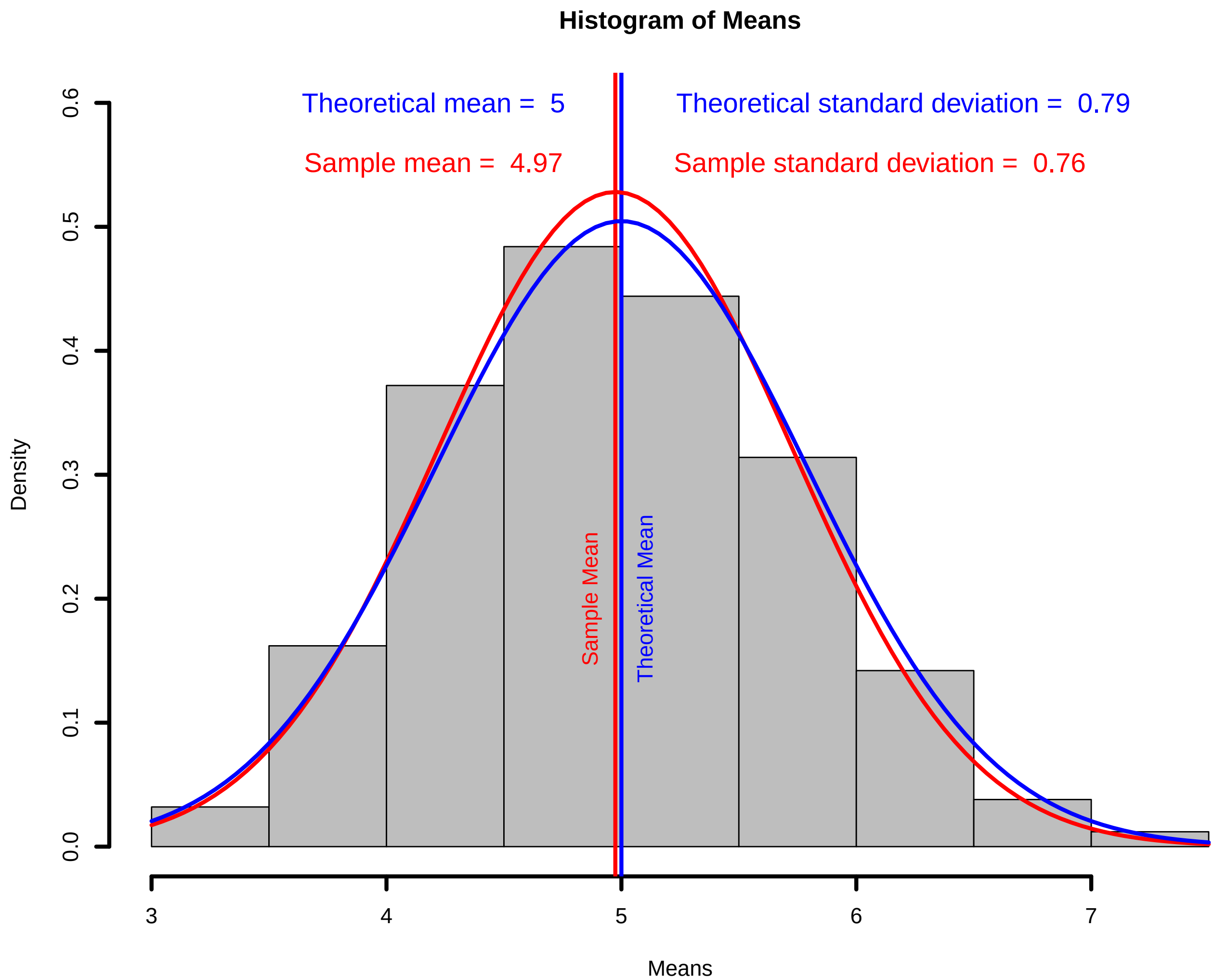
```
## [1] 0.5706551
```

The sample variance is calculated as 0.57 which is close to 0.62 as expected.

Distribution

The CLT states that the distribution of averages of iid random variables becomes that of a normal as the sample size increases. Thus, we'll investigate to see if the distribution of 1000 averages of 40 exponentials looks like a normal distribution. It is easier to check this by fitting the distribution.

```
# Plot the density distribution of means from 1000 simulations sampling 40 exponentials
hist(means, breaks = 15,
    main = "Histogram of Means",
    xlab = "Means",
    ylab = "Density",
    ylim=c(0,0.6),
    prob=TRUE,
    lwd = 3,
    col = "gray")
abline(v=mean(means),col="red", lwd=3)
abline(v=true.mean,col="blue", lwd=3)
text(x = mean(means) - 0.1, y = 0.2, "Sample Mean", srt = 90., col = "red")
text(x = 1/lambda + 0.1, y = 0.2, "Theoretical Mean", srt = 90., col = "blue")
curve(dnorm(x, sample.mean, sample.sd),
      add=TRUE,
      col = "red",
      lwd = 3)
text(x = 4.2, y = 0.55, paste("Sample mean = ", round(sample.mean,2)),
     col = "red", cex = 1.25)
text(x = 6.1, y = 0.55, paste("Sample standard deviation = ", round(sample.sd,2)),
     col = "red", cex = 1.25)
curve(dnorm(x,true.mean,true.sd),
    add = TRUE,
    col = "blue",
    lwd = 3)
text(x = 4.2, y = 0.60, paste("Theoretical mean = ", round(true.mean,2)),
     col = "blue", cex = 1.25)
text(x = 6.2, y = 0.60, paste("Theoretical standard deviation = ", round(true.sd,2)),
     col = "blue", cex = 1.25)
```



The blue curve is the fitted normal distribution with mean of 4.97 and standard deviation of 0.76, while the red curve shows a normal distribution with the theoretical mean of 5 and theoretical standard deviation of 0.79. There is a good agreement between these two curves, meaning that the distribution of means of iid random variables tends toward a normal distribution as CLT states.