Simulation Exercise [Peer-graded Assignment: Statistical Inference Course Project]

Dr.Chamika Senanayake

9/26/2020

Overview:

the main objectives of the given report is to 1. Show the sample mean and compare it to the theoretical mean of the distribution. 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal.

Simulations:

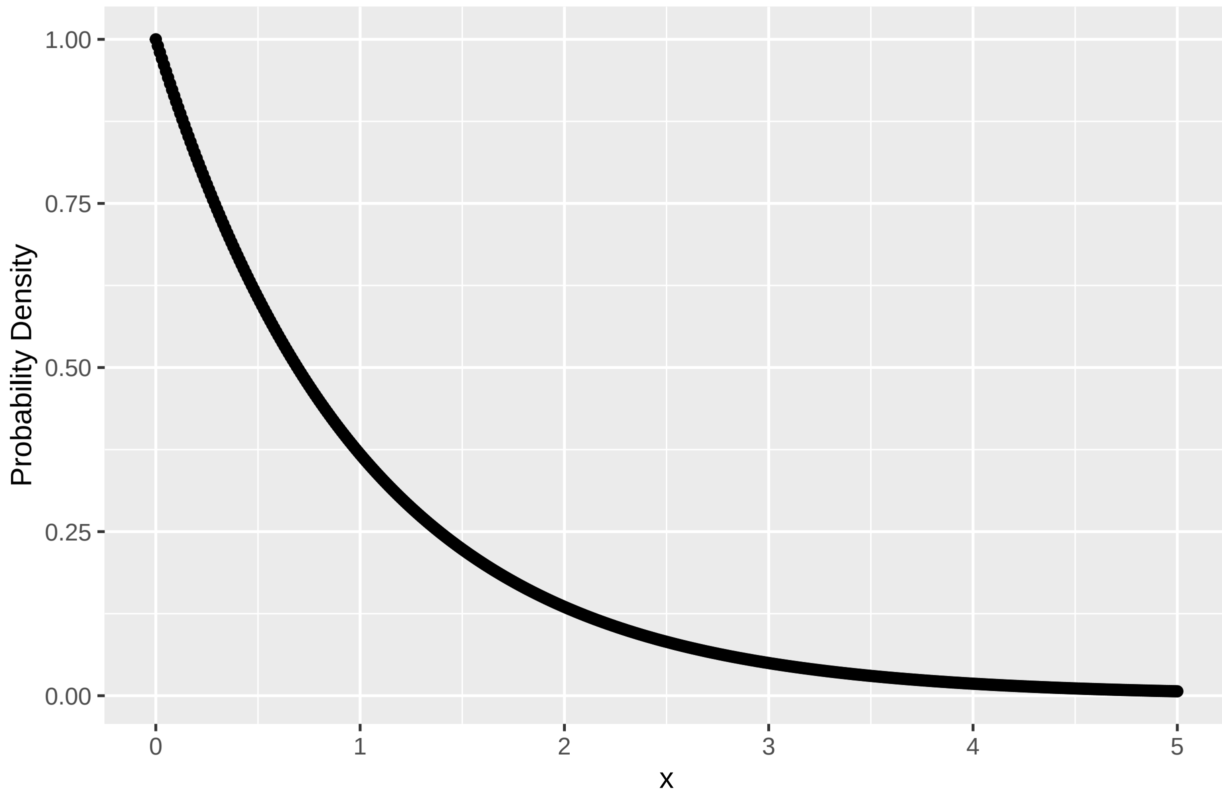
The Exponential Distribution

The exponential distribution describes the arrival time of a randomly recurring independent event sequence. If `mu` is the mean waiting time for the next event recurrence, its probability density function is `f(x) = (1/mu * e ^ (-x/mu)) when x >= 0; 0 when x < 0`. If we call `lambda (rate) = 1 / mu`, then the **exponential probability density function (dexp)** becomes `f(x) = lambda * e ^ (-lambda * x)`. The following is the plot of the exponential probability density function.

```
x <- seq(0.0, 5.0, 0.01)
y <- dexp(x)
qplot(x, y, type="l", xlab = "x",
      ylab = "Probability Density",
      main = "Figure 1: Exponential Probability Density Function")
```

```
## Warning: Ignoring unknown parameters: type
```

Figure 1: Exponential Probability Density Function



the parameters for this simulations are as follows

```
set.seed(1)
sims <- 1000
lambda <- 0.2
n <- 40
```

1000 random samples of size 40 were generated, with rate (lambda) equals 0.2.calculated mean of these several samples were stored into the `means` vector

```
exp_sim <- matrix(rexp(sims * n, lambda), sims, n)
means <- rowMeans(exp_sim)
```

##Sample Mean versus Theoretical Mean:

The sample mean, is obtained in what follows.

```
sample_mean <- mean(means)
theoretical_mean <- 1 / lambda
```

Sample Variance versus Theoretical Variance:

```
sample_variance <- var(means)
theoretical_variance <- (((1 / lambda)^2) / 40)
```

The sample variance is **0.6177072** and the theoretical variance is **0.625**. When comparing the estimated value and the theoretical one, there is a small difference between then, -0.0072928, to be exact.

Distribution:

This plot contains the histogram (light blue) of the means of 1000 random samples of size 40. It also contains the density curve (blue) to demonstrate how it approximates the normal distribution. Additionally, the dashed red line representes the mean of the sample.

```
g <- ggplot(data.frame(means), aes(x = means))
g <- g + geom_histogram(aes(y=..density..), colour = "black", fill = "white", binwidth = lambda)
g <- g + geom_density(colour = "skyblue1", fill="skyblue1", alpha = .3)
g <- g + geom_vline(aes(xintercept = sample_mean), linetype = "dashed", size = 2, colour = "red")
g <- g + scale_x_continuous(breaks = c(1:10))
g <- g + ggtitle("Means of 1000 Random Samples of Size 40"); g
```