Simulation Exponential Distribution

Yato

March 15, 2017

Overview

1. Sample mean is consistent to the theoretical mean.
2. Sample variance is consistent to the theoretical variance.
3. According to the Central Limit Theory, the distribution of 1000 averages of 40 exponentials is approximately normal.

Simulations

- simulate a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.
- Set lambda = 0.2 for all of the simulations. The mean of exponential distribution is 1/lambda, 5, and the Variance is 1/lambda^2, 25.

```
# generate 1000 random exponentials
set.seed(17); lambda <- 0.2; simexp1 <- rexp(1000, 0.2)
# generate 1000 averages of 40 random exponentials
simexp2 = NULL
for(i in 1:1000) simexp2 = c(simexp2, mean(rexp(40, 0.2)))
```

Sample Means versus Theoretical Mean

```
mean_sample1 <- mean(simexp1); mean_sample2 <- mean(simexp2); mean_theo <- 1/lambda;
mean_sample1
```

```
## [1] 5.035278
```

```
mean_sample2
```

```
## [1] 4.958807
```

```
mean_theo
```

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

Sample Variance versus Theoretical Variance

```
var_sample1 <- var(simexp1);var_sample2 <- var(simexp2);var_theo <- (1/lambda^2/40);
var_sample1
```

```
## [1] 27.2214
```

```
var_sample2
```

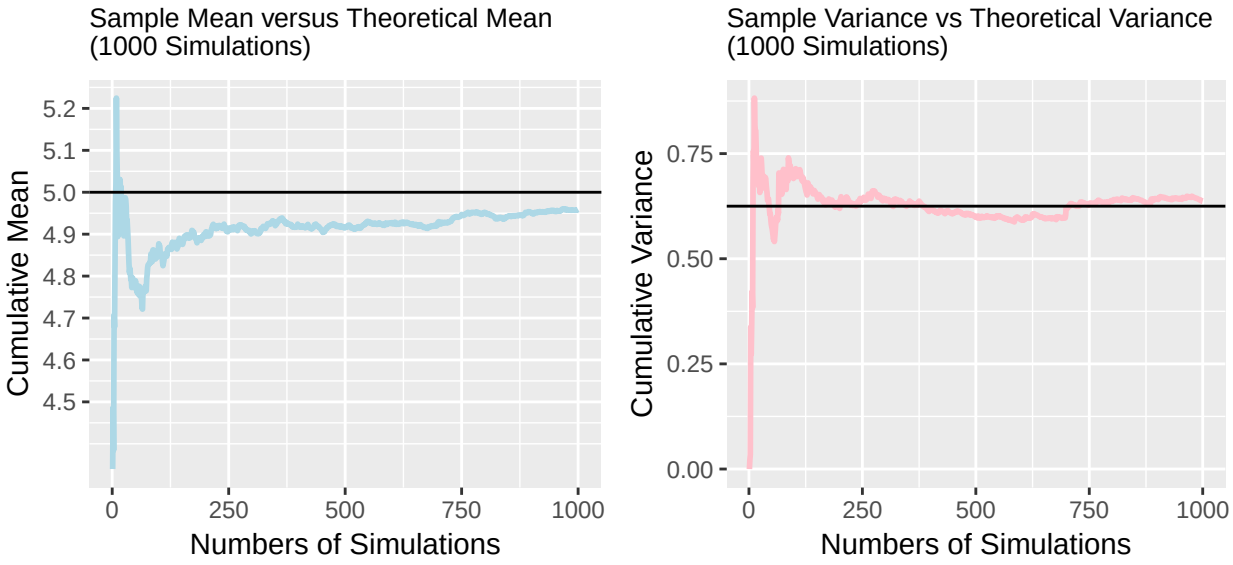
```
## [1] 0.6379003
```

```
var_theo
```

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

```
megg2 <- ggplot(data.frame(x = 1:1000, y = cumsum(simexp2)/(1:1000)), aes(x = x, y=y)) +
        geom_line(col = "lightblue", size = 1) + geom_hline(yintercept = 5) +
        scale_y_continuous(breaks = seq(4.5, 5.3, by = 0.1)) +
        ggtitle("Sample Mean versus Theoretical Mean\n(1000 Simulations)") +
        theme(plot.title=element_text(size = 10)) +
        xlab("Numbers of Simulations") + ylab("Cumulative Mean")

vargg2 <- ggplot(data.frame(x = 1:1000,
       y = cumsum(simexp2^2)/(1:1000)-(cumsum(simexp2)/(1:1000))^2), aes(x = x, y=y)) +
        geom_line(col = "pink", size = 1) + geom_hline(yintercept = 0.625) +
        ggtitle("Sample Variance vs Theoretical Variance\n(1000 Simulations)") +
        theme(plot.title=element_text(size = 10)) +
        xlab("Numbers of Simulations") + ylab("Cumulative Variance")
grid.arrange(megg2, vargg2, ncol=2)
```



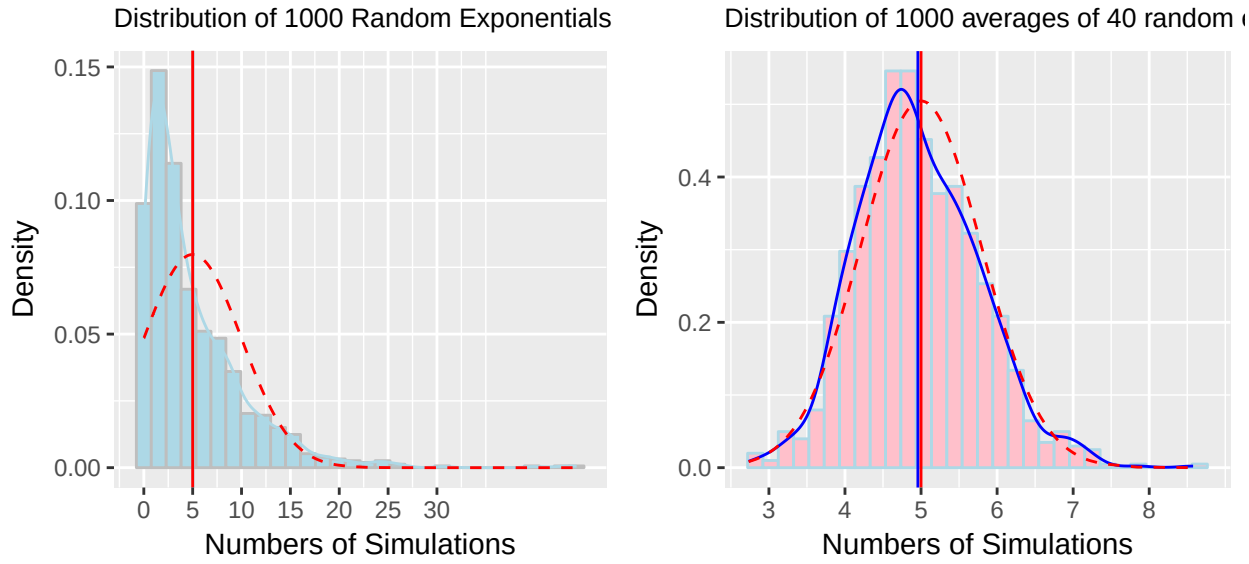
Comparing the Two Distributions and Testing the normality

Notice the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

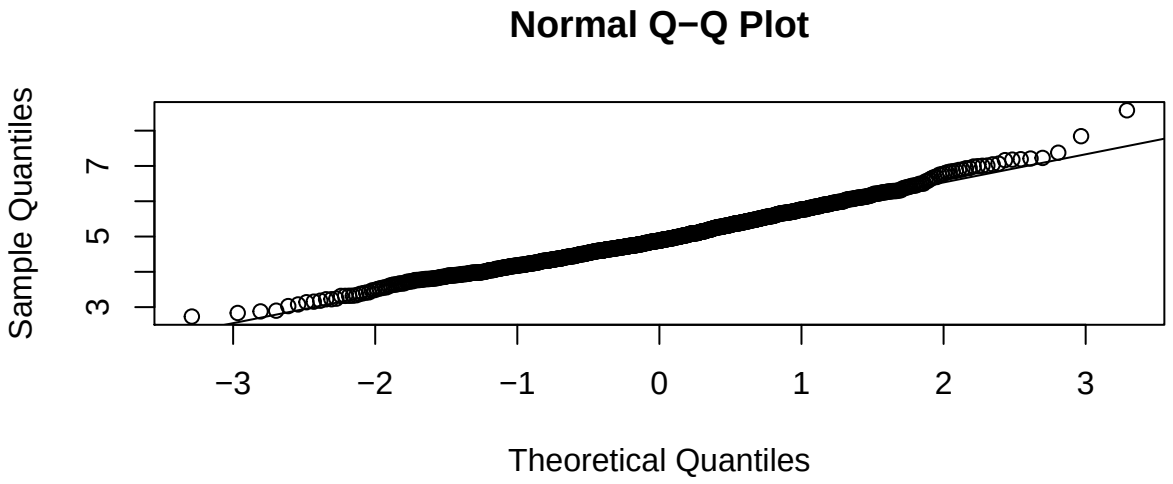
```
gg1 <- ggplot(NULL, aes(x = simexp1, y = ..density..)) +
        geom_histogram(fill = "lightblue", colour= "grey", bins = 30) +
        geom_line(stat = "density", col="lightblue") +
        geom_line(aes(x=simexp1, y= dnorm(simexp1, mean=5, sd=5)), col = "red", linetype="dashed") +
        geom_vline(xintercept = c(mean_sample1, mean_theo), col=c("lightblue", "red")) +
        scale_x_continuous(breaks = seq(0, 30, 5)) +
        ggtitle("Distribution of 1000 Random Exponentials") +
        theme(plot.title=element_text(size = 10)) +
        xlab("Numbers of Simulations") + ylab("Density")

gg2 <- ggplot(NULL, aes(x = simexp2, y = ..density..)) +
        geom_histogram(fill = "pink", colour="lightblue", bins = 30) +
        geom_line(stat = "density", col = "blue") +
```

```
        geom_line(aes(x=simexp2, y= dnorm(simexp2, mean=5, sd=5/sqrt(40))), col = "red", linetype="dash
        geom_vline(xintercept = c(mean_sample2, mean_theo), col=c("blue", "red")) +
        scale_x_continuous(breaks = seq(1, 8, 1)) +
        ggtitle("Distribution of 1000 averages of 40 random exponentials") +
        theme(plot.title=element_text(size = 10)) +
        xlab("Numbers of Simulations") + ylab("Density")
grid.arrange(gg1, gg2, ncol=2)
```




```
qqnorm(simexp2); qqline(simexp2)
```



The distribution of sample means is centered at 4.96 and the theoretical center of the distribution is 5. The variance of sample means is 0.6379 where the theoretical variance of the distribution is 0.625. Due to the central limit theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. The q-q plot suggests the normality.