

Statistic inference report - Part 1

Tom

Saturday, June 13, 2015

Overview

In this report, we are going to explore the following topics:

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

In this study, we generate 1000 sets of random numbers from a selected exponential distribution ($\lambda = 0.2$). Each set contains 40 random numbers. The means and variance of the 1000 sample means are studied to see what kind of distribution these sample means follow. For reproducibility, 942387 is selected as random seed.

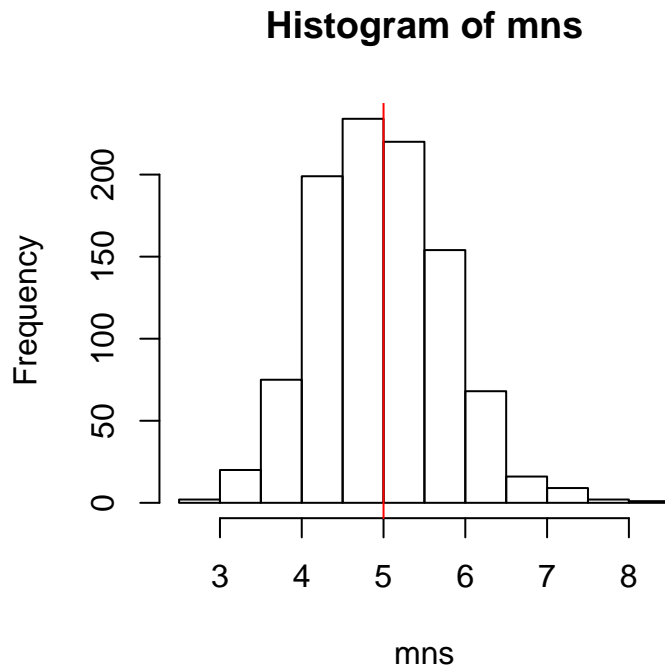
Sample Mean versus Theoretical Mean

The theoretical mean is $= 1/\lambda = 1/0.2 = 5$.

The following shows a histogram for the 1000 sample means and a vertical line at 5.

We can see that most of the sample means are around 5.

```
mns = NULL
set.seed(942387)
for (i in 1 : 1000) mns = c(mns, mean(rexp(40, 0.2)))
hist(mns)
abline(v=5, col='RED')
```



Below shows the average and median of the 1000 sample means.

```
mean(mns)
```

```
## [1] 4.969908
```

```
median(mns)
```

```
## [1] 4.943481
```

Conclusion: The sample mean at most of the time follow the theoretical mean, in the case of extreme case, the sample means can be as low as 3 or as high as 8.

Sample Variance versus Theoretical Variance

The theoretical standard deviation of the random numbers is $= (1/\lambda)^2 = (1/0.2)^2 = 25$.

According to the Central limit theorem, The standard error of sample mean is $\sigma/\sqrt{n} = 25/\sqrt{1000} = 0.791$.

The theoretical variance of 1000 sample mean is $0.791^2 = 0.625$.

The sample variance of 1000 sample means is

```
var(mns)
```

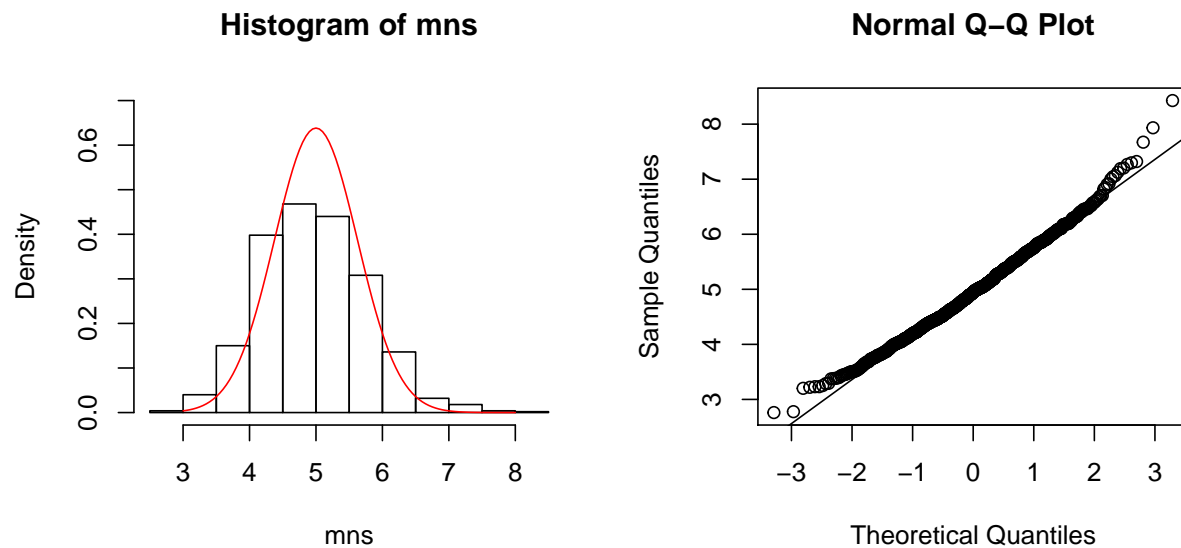
```
## [1] 0.6196529
```

Conclusion: The sample variance of the 1000 sample means is closed to the theoretical value.

Distribution

To verify that the distribution of 1000 sample means is approximately normal, we are going to plot 2 graphs. First, we plot the histogram of the sample mean and draw a red line representing the theoretical distribution. To check if the histogram follows the theoretical one.

```
par(mfrow=c(1,2))
hist(mns, freq=FALSE, ylim=c(0,0.7))
curve(dnorm(x, 5, 0.625), from=3, to=8, col=2, add=TRUE)
qqnorm(mns)
qqline(mns)
```



Secondly, we draw the QQ-plot for the 1000 sample means.

From the QQ-plot, it shows that there is a slightly fat negative tail and a slightly thin positive tail and hence the distribution of sample means is a negative skewed distribution.

From the above 2 graphs, we see that the distribution follows most of the feature of a normal distribution. We can conclude that the distribution of 1000 sample means is approximately normal.