Statistical Inference Project - Part I

Antoni Parellada

Saturday, January 17, 2015

Overview:

The central limit theorem (CLT) states that the mean of the distribution of sample means converges towards the population mean μ. The distribution of the means of the sample assumes normality with decreasing variance as the size of the samples increases: $var = \sigma^2/n$.

The theoretical mean of an exponential distribution with $\lambda = 0.2$ equals $1/\lambda = 5$.

Simulations:

Sample Mean versus Theoretical Mean:

A Monte Carlo simulation of 1,000 random draws from an exponential distribution of this rate parameter (0.2) will empirically yield a mean of:

```
set.seed(0)
lambda =0.2
samples = 40
x_bar <- NULL
for (i in 1:1000) x_bar[i] = mean(rexp(samples,rate = lambda))
mean(x_bar)
```

```
## [1] 4.989678
```

Sample Variance versus Theoretical Variance:

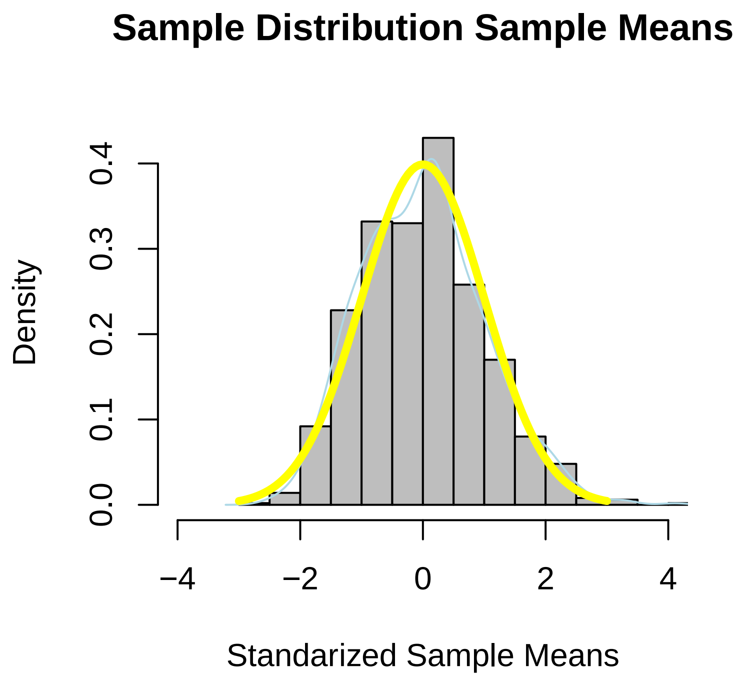
The variance of the sample distribution of the sample means is calculated as $\sigma^2/n = (1/0.2)^2/40 = 0.625$, while the empirical calculation is:

```
var(x_bar)
```

```
## [1] 0.6181582
```

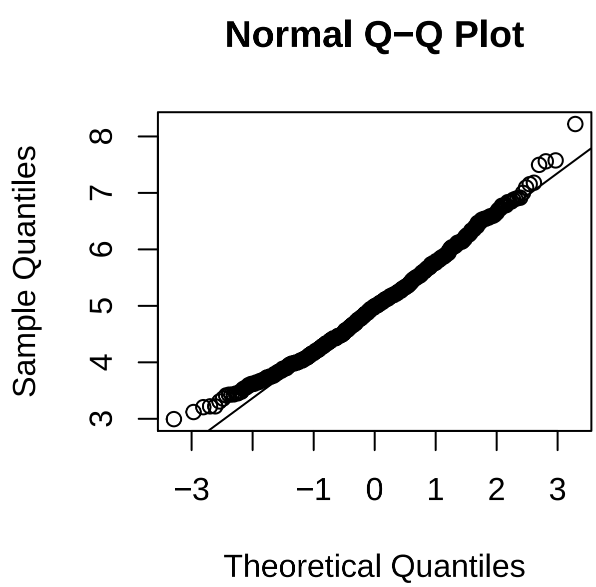
Distribution:

The overlap of the normal distribution curve to the samples means is shown with normalized values:



Further a QQ Plot of sample means visually suggests normality:

```
qqnorm(x_bar)
qqline(x_bar)
```



The fact that the sample means align themselves following a normal distribution with mean = μ should also guarantee that the confidence interval in our Monte Carlo simulation contains the theoretical mean in 95% of the cases: 5. Let's prove it:

```r
lower <- x_bar - qnorm(0.975) * 5/sqrt(40)
upper <- x_bar + qnorm(0.975) * 5/sqrt(40)
frac_mean <- mean(lower< 5 & upper > 5)
frac_mean
```

```
## [1] 0.954
```