Statistical Inference Assignment: Simulation

Francisco Marco-Serrano

Overview

The exponential distribution is simulated by generating 1000 series of 40 data points.Differences of the mean and variance are assessed copmaring them to the theoretical values. Finally, the normality of the simulated distribution is evaluated using graphical and numerical tests.

Simulations

Generate one thousand series of length forty entries to compare the results with the theoretical ones.

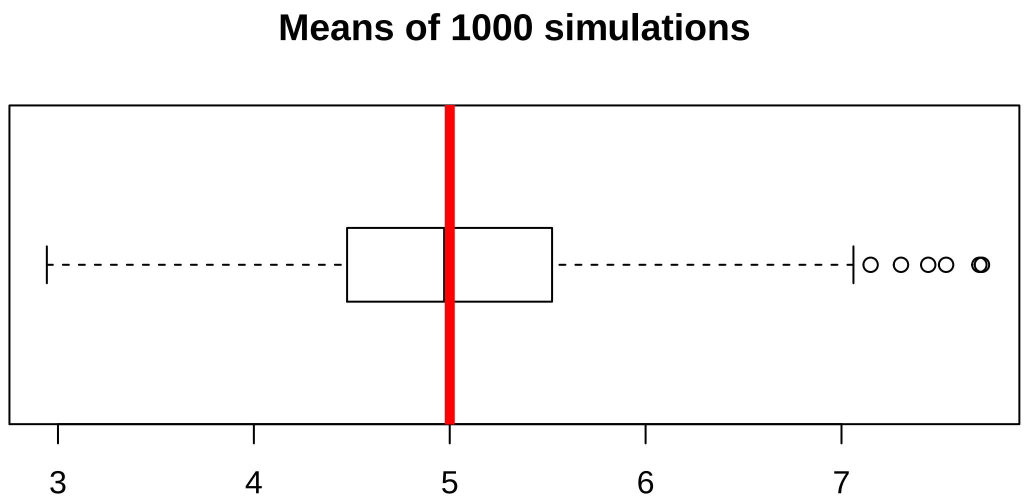
```
# define parameters

n <- 40         # number of data points
lambda <- 0.2   # value of exponential function parameter
set.seed(123)   # set seed for reproducibility purposes

# initialise counters
mns = NULL
vns = NULL

# run simulation
for (i in 1 : 1000){
  x = rexp(n, lambda)     # simulate exponential distribution
  mns = c(mns, mean(x))  # obtain mean
  vns = c(vns, var(x))   # obtain variance
}
```

Sample Mean vs Theoretical Mean

The mns variable gathers the means of the 1000 simulations. Find below the theoretical mean and the actual simulated mean. In the boxplot we represent the distribution of the calculated means. The vertical red line marks the theoretical mean, calculated as 1/lambda.

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

```
## [1] 5.011911
```

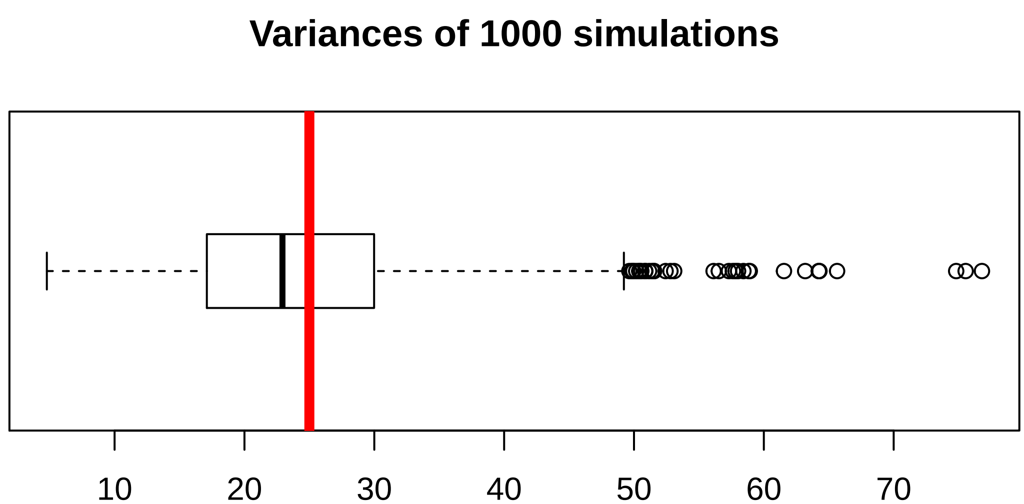


Sample Variance vs Theoretical Variance

The vns variable gathers the variances of the 1000 simulations. Find below the theoretical variance and the actual simulated variance. In the boxplot we represent the distribution of the calculated variances. The vertical red line marks the theoretical variance, calculated as (1/lambda)^2.

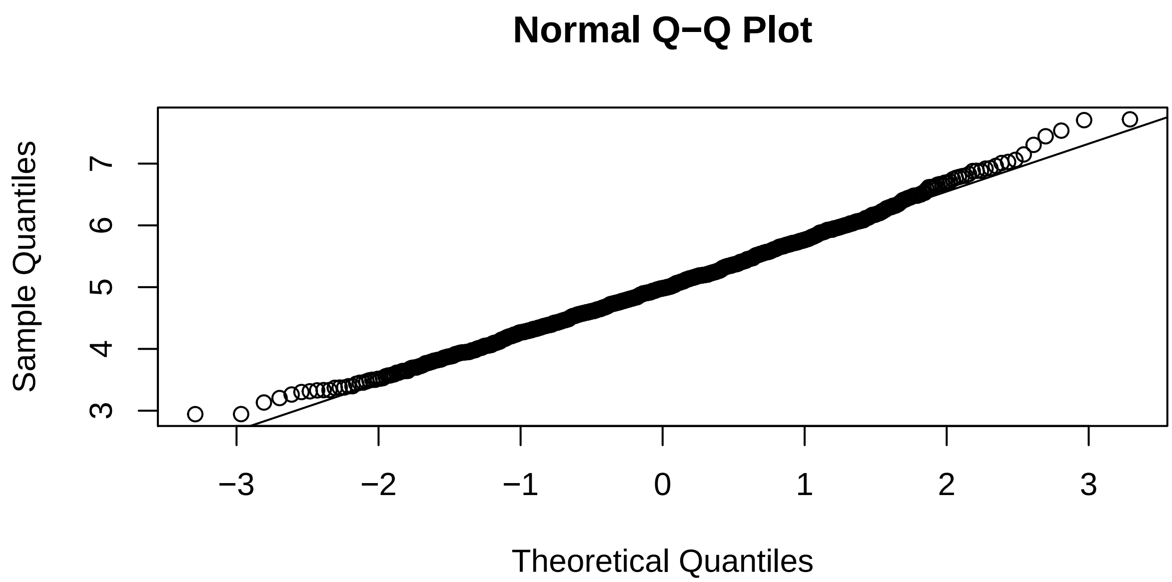
```
## [1] 25
```

```
## [1] 24.84317
```



Distribution

In order to test the normality of the simulated data we will be using the Q-Q plot (see this article) and the Shapiro-Wilk test (see this reference).




```
##
##  Shapiro-Wilk normality test
##
## data:  mns
## W = 0.9956, p-value = 0.005322
```

The Q-Q plot seems to reflect a Normal distribution. However, when analysisng the results of the Shapiro-Wilk test we see it is not. The null hypothesis of the test is that the population is normally distributed. Since the p-value is almost 0 (nil), we cannot accept the data is generated from a Normal probability distribution.

Appendix

Means code

```
# theoretical mean
tmean = 1/lambda; tmean

# simulated mean
smean = mean(mns); smean

boxplot(mns, horizontal = TRUE, main = "Means of 1000 simulations", boxwex = 0.5)
abline(v = tmean, col = "red", lwd = "5")
```

Variances code

```
# theoretical variance
tvar = (1/lambda)^2; tvar

# simulated variance
svar = mean(vns); svar

boxplot(vns, horizontal = TRUE, main = "Variances of 1000 simulations", boxwex = 0.5)
abline(v = tvar, col = "red", lwd = "5")
```

Normality code

```
# The normal Q-Q plot, including the fit line
qqnorm(mns); qqline(mns)

# Perform the Shapiro test
shapiro.test(mns)
```