Comparison of Distributions of 1,000 Simuations of 40 samples to Theoritical Distribution

JJanzen

March 18, 2015

This report will summarize with R code and plots the mean of 1,000 simulations of samples of 40 compared the the theoritical normal distribution. The plots will include a distribution of simulations vs therotical, variance comparison, and also a confidence interval summary.

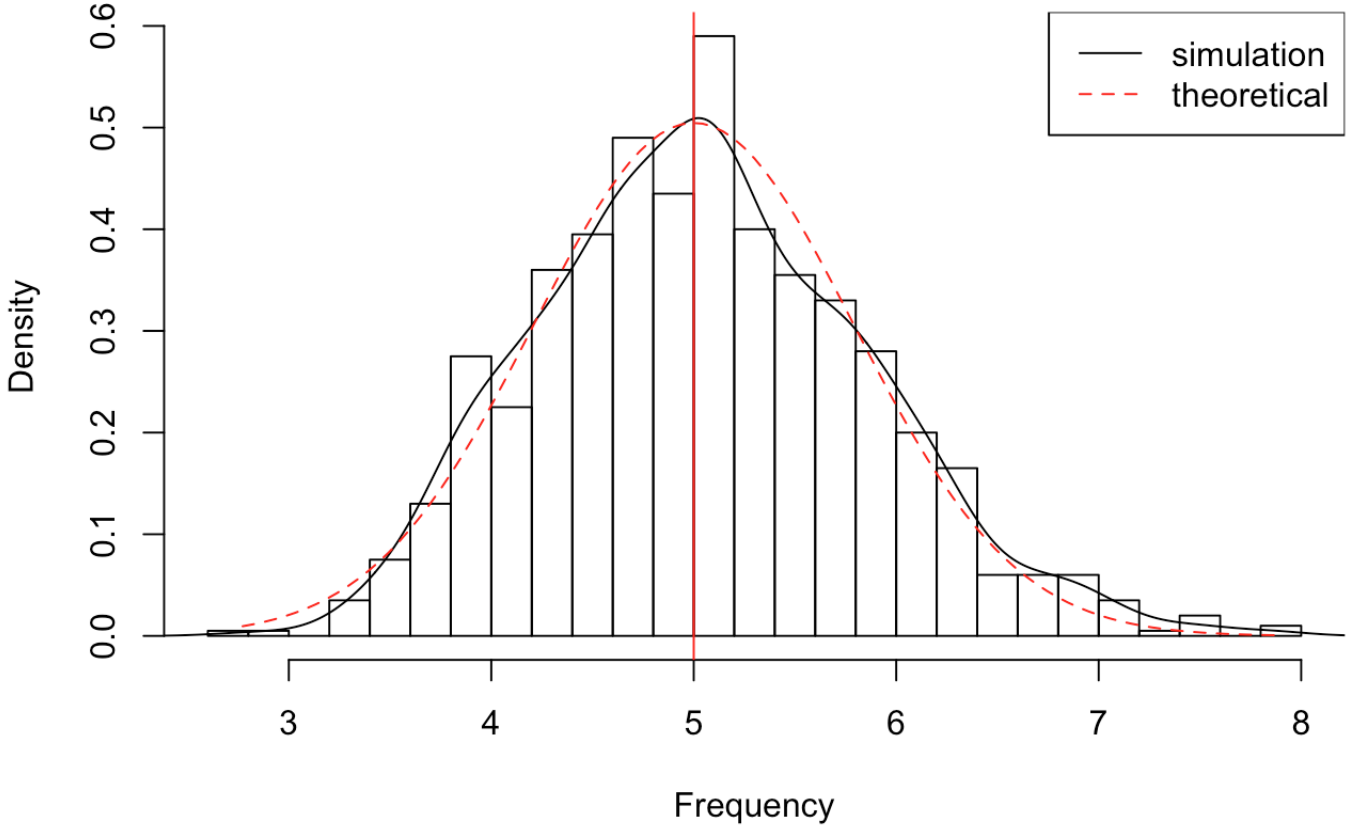
```
# give rate parameter
lambda <- 0.2
# number of exponentials
n <- 40
# number of simulations to invensigate distribution
nosim <- 1000
set.seed(10)
```

Create simulations (samples) using 1000 simulations (rows) and values of each (columns).

```
## matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)
simulation <- matrix(rexp(nosim*n, rate=lambda), nosim, n)
# get mean of each of the 1,000 simulations
row_means <- rowMeans(simulation)
```

Create historgram of means of 1,000 simuations.

```
hist(row_means, breaks = 25, prob=T, main = "Distribution of Means of 1k Simulations", xla
b="Frequency")
# add density line of best fit for mean of simulations
lines(density(row_means))
# center of theoretical distribution
abline(v=1/lambda, col="red")
# theoretical density of mean of simulations
xfit <- seq(min(row_means), max(row_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
lines(xfit, yfit, pch=22, col="red", lty=2)
# add legend
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "red"))
```



Simulation mean and variance versus Theoritcal mean and variance.

```
# simulation mean
mean(row_means)
```

```
## [1] 5.04506
```

```
# theoritical mean
lambda^-1
```

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

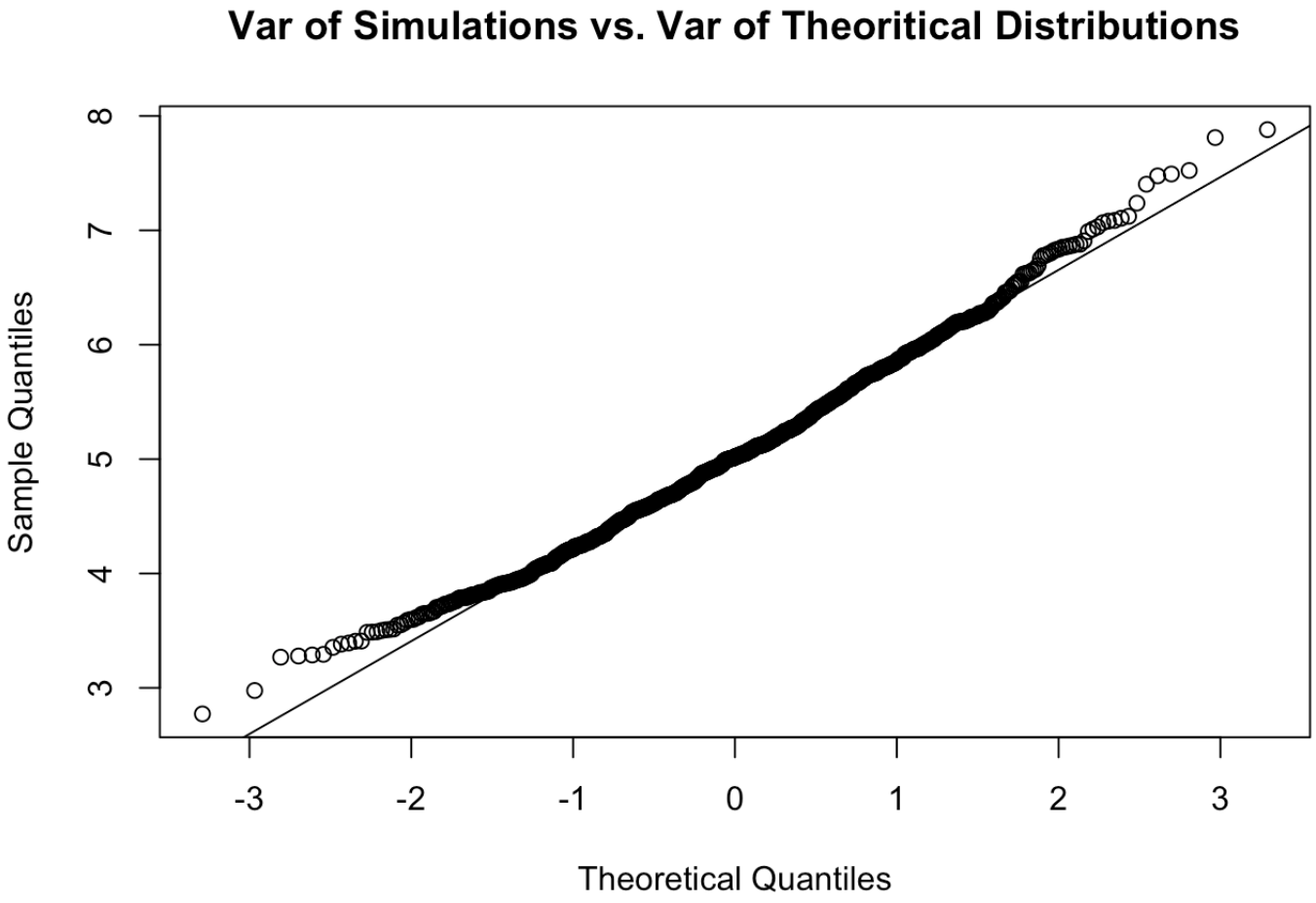
```
# simulation variance
var(row_means)
```

```
## [1] 0.6541736
```

```
# theoritcal variance
1/(lambda^2*n)
```

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

```
# plot of variance of simulation to variance of theoritical distribution
qqnorm(row_means,main = "Var of Simulations vs. Var of Theoritical Distributions"); qqline(ro
w_means)
```



The distribution below is relativily normal. The mean of the simulation is right about at 5, which is within the 95% confidence interval.

```
# confidence interval
?seq # seq(from = 1, to = 1, by = ((to - from)/(length.out - 1)), length.out = NULL, along.wi
th = NULL, ...)
# generate values from 4 to 6 by .01
lambda_vals <- seq(4, 6, by=0.01)
coverage <- sapply(lambda_vals, function(lamb) {
    mu_hats <- rowMeans(simulation)
    ll <- mu_hats - qnorm(0.975) * sqrt(1/lambda**2/n)
    ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda**2/n)
    mean(ll < lamb & ul > lamb)
})

library(ggplot2)
?qplot
qplot(lambda_vals, coverage, main = "Confidence Interval of Mean Simulations") + geom_hline(y
intercept=0.95)
```