Statistical Inference Course Project

Ehson Ghandehari

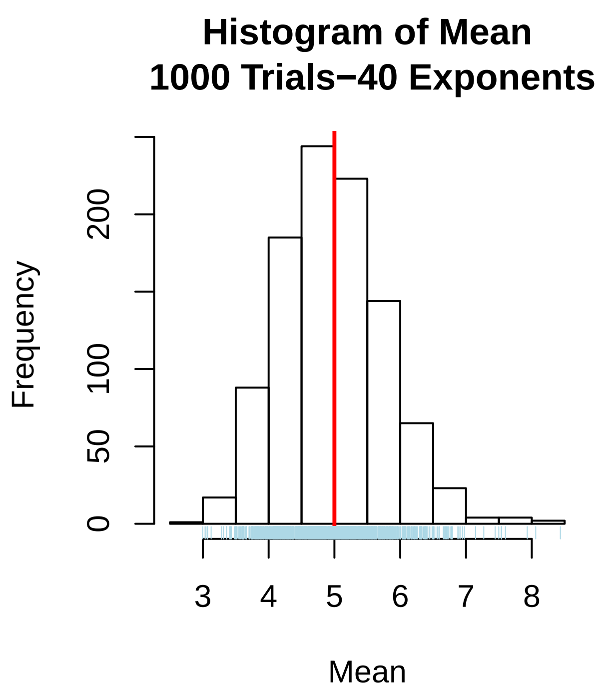
January 13, 2016

Overview:

The purpose of this manuscript is to perform 1000 simulations, using R exponential distribution function. Each simulation was conducted with two pre-defined paramters: 1) fourty observations and 2) a constant theoritcal rate (rate=1/lambda) by setting lambda=0.2. The mean and variance of experimental simulations was compared to the theoritical values (i.e. mean of 1/lambda=5 and standard deviation of 1/lambda=5). The shape of the distrubution was identified by Kurtosis and Tailing of the distribution. At last, it was confirmed that the distibution behaved as a normal distribution by consucting simulations with thousand eqponentials. The code for all calculations and visualization can be found in ** Appendix A**.

Simulations:

At first, we store the outcome of the simulation into a dataframe called **Results**. The outcome of Nth simulation can be found in Nth row of this dataframe. The mean and variance of each simulation were calculated and stored in **Results.mean** and **Results.var** subsequently. The histogram (including the rugged frequency) of mean and variance after thoudans simulations, averaging over 40 randomly created numbers (followinf exponential distribution) is presented in Figure 1. The red vertical lines, as refrence, show the population mean of 5 and variance of 25.



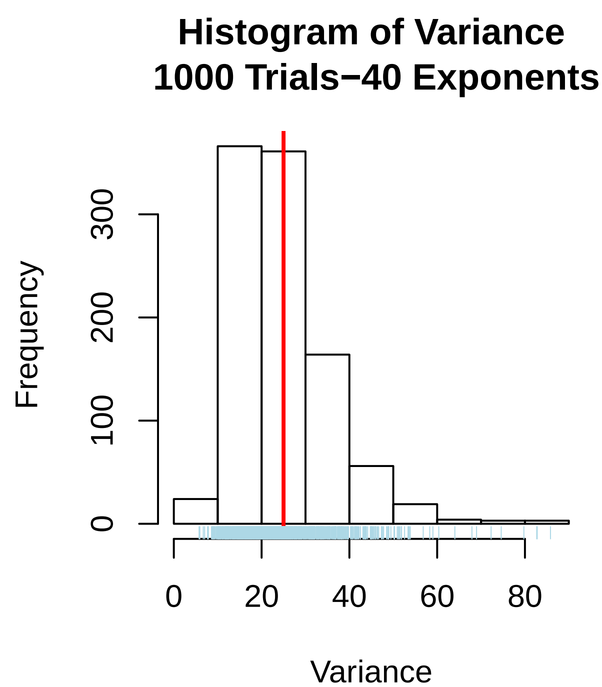


Figure 1. The distribution of mean and its variance after thousand simulation, each calculating the mean of 40 exponential-distribution randomly created numbers.

The **mean** distribution had a kurtosis of 0.528 and skewness of 0.433. Therefore, the peak of distribution is far from value 5 and there is a slight right skewness in distribution. The **variance** distribution had a kurtosis of 3.854 and skewness of 1.469. Therefore, the peak of distribution is close to population value of 25 and there is right skewness in the distribution.

The histogram (including the rugged frequency) of mean and variance after thoudans simulations, averaging over 1000 randomly created numbers (followinf exponential distribution) is presented in Figure 1. The red vertical lines, as refrence, show the population mean of 5 and variance of 25.

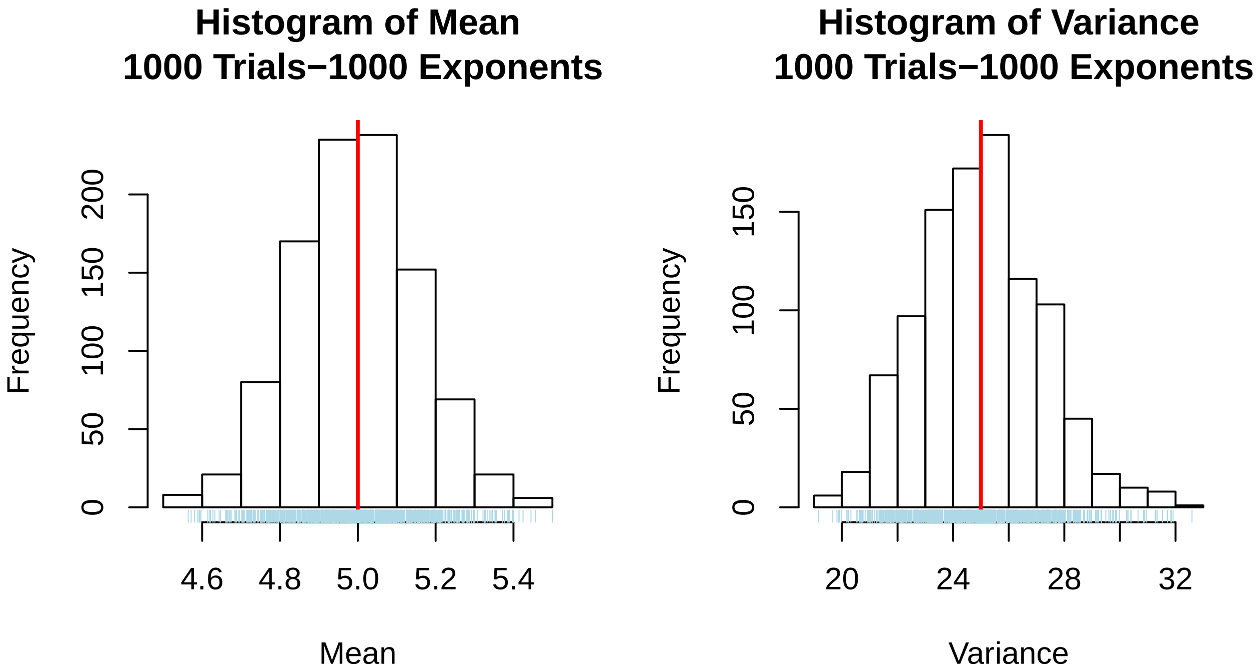


Figure 2. The distribution of mean and its variance after thousand simulation, each calculating the mean of 1000 exponential-distribution randomly created numbers.

The **mean** distribution had a kurtosis of -0.135 and skewness of 0.039. Therefore, the peak of distribution became closer to value of 5 (and symmetric) and the skewness to right was reduced. The **variance** distribution had a kurtosis of 0.076 and skewness of 0.287. Therefore, the peak of distribution got closer to the population value of 25 and the right skewness was significantly reduced.

Appendix A:

```
# Creating 1000 simulations with 40 exponentials
# require(e1071)  # this package is installed for Kurtosis calcualtions.
# library(e1071)
# Results<-data.frame()
# for (i in 1:1000) {sim<-rexp(n=40,rate=0.2)
# Results<-rbind(Results,sim)}
# colnames(Results)<-paste("Obs-",as.character(1:40), sep="")
# Results.mean<-apply(Results,1,mean) # this calculates the mean of each row.
# Results.var<-apply(Results,1,var) # this calculated the variance of each row.


#{r Fig1, echo=FALSE, warning=FALSE, error=FALSE, message=FALSE, fig.height=4,
# fig.width=7, fig.cap="Figure 1:"}

# Graphing the distribution of Means
# par(mfrow=c(1,2))
# hist(Results.mean, xlab="Mean", main="Histogram of Mean\n 1000 Trials-40 Exponents")
# abline(v=5,col="red",lwd="2",)
# rug(Results.mean, col="light blue")
# Graphing the distribution of Variance
```

```r
# hist(Results.var, xlab="Variance", main="Histogram of Variance\n 1000 Trials-40 Exponents")
# abline(v=25,col="red",lwd="2")
# rug(Results.var, col="light blue")

#{r data.creation2, echo=FALSE, warning=FALSE, error=FALSE, message=FALSE}
# Creating 1000 simulations with 1000 exponentials
# Results<-data.frame()
# for (i in 1:1000) {sim<-rexp(n=1000,rate=0.2)
# Results<-rbind(Results,sim)}
# colnames(Results)<-paste("Obs-",as.character(1:1000), sep="")
# Results.mean<-apply(Results,1,mean) # this calculates the mean of each row.
# Results.var<-apply(Results,1,var) # this calculated the variance of each row.

# {r Fig2, echo=FALSE, warning=FALSE, error=FALSE, message=FALSE, fig.height=4,
# fig.width=7, fig.cap="Figure 2:"}
# Graphing

# par(mfrow=c(1,2))
# hist(Results.mean, xlab="Mean", main="Histogram of Mean\n 1000 Trials-1000 Exponents")
# abline(v=5,col="red",lwd="2",)
# rug(Results.mean, col="light blue")
# Graphing the distribution of Variance

# hist(Results.var, xlab="Variance", main="Histogram of Variance\n 1000 Trials-1000 Exponents")
# abline(v=25,col="red",lwd="2")
# rug(Results.var, col="light blue")
```