

Statistical Inference - Simulation Exercise

Markus Divjak

Overview

The following report investigates the properties of a sample of 40 random exponential variables each. In order, the report compares the mean of the random variables with the expected mean, the variance of the random variables with the variance and, lastly, investigates whether the distribution means can be understood as being approximately standard normal. For this report, the parameter lambda was chosen to be 0.2.

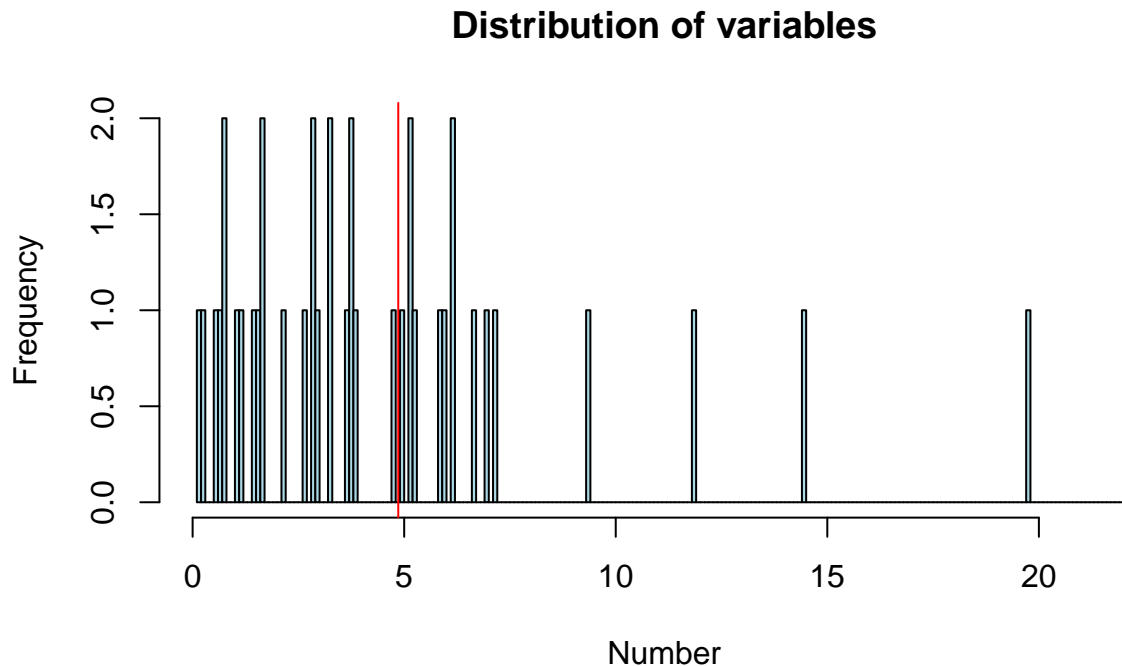
Sample Mean vs. Theoretical Mean

The theoretical mean of an exponential distribution is $1 / \text{lambda} = 5$. We can compare this to the mean of our experiment. The plot below shows the distribution of the variables with the sample mean being marked by a red vertical line.

```
set.seed(1)
x<-rexp(40,0.2)
mean(x)

## [1] 4.860372

hist(x,col="lightblue",breaks=200,main="Distribution of variables",xlab="Number")
abline(v=mean(x),col="red")
```



Note that the mean is approximately 5 (as expected).

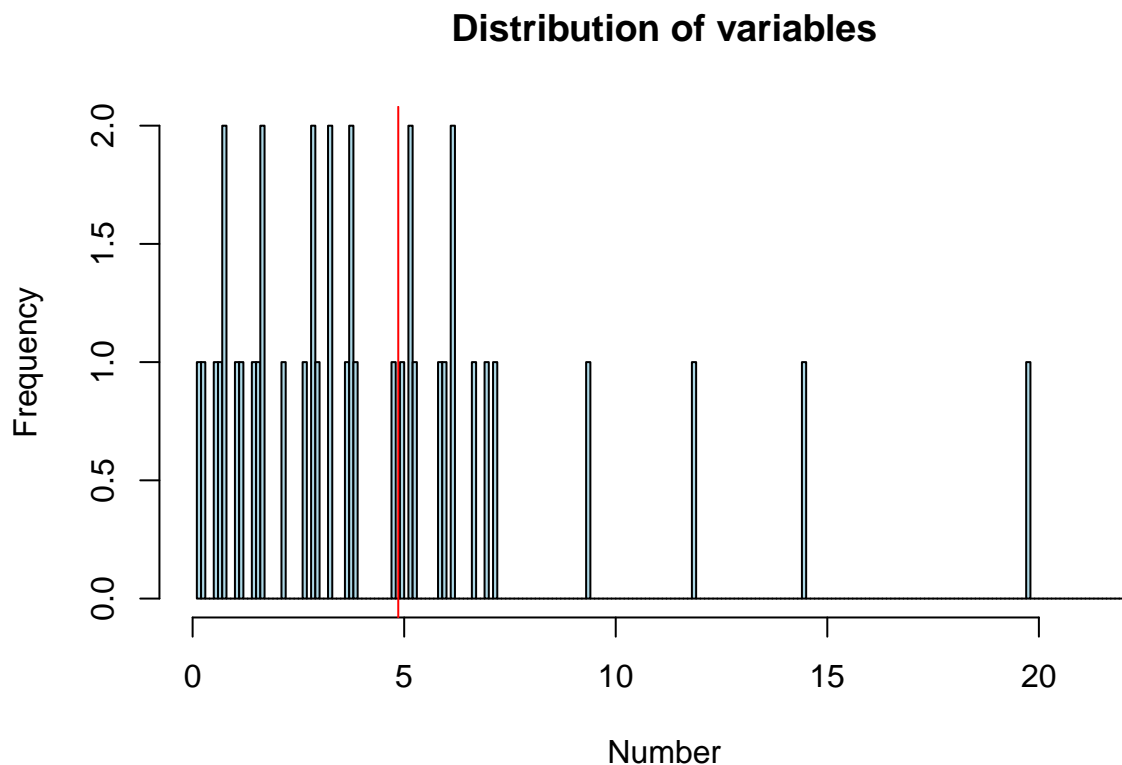
Sample Variance vs. Theoretical Variance

The theoretical variance of an exponential distribution with $\lambda = 0.2$ is 25. We can compare this to the variance of our experiment. The plot below shows the distribution of the variables with the sample variance being marked by a green vertical line.

```
set.seed(1)
var(rexp(40,0.2))
```

```
## [1] 23.56906
```

```
hist(x,col="lightblue",breaks=200,main="Distribution of variables",xlab="Number")
abline(v=mean(x),col="red")
abline(v=var(x),col="green")
```



Note that the sample variance is approximately 25 (As expected)

Distribution

According to the Central Limit Theorem (CLT) the distribution of means of iid variables becomes that of a standard normal as the sample size increases. In this case we run 1000 experiments of 40 random variables each by structuring them in a matrix of 1000 rows and 40 columns. We then take the mean of each row and plot those means in the histogram you will find below, which shows that the distribution is approximately standard normal (note that typical bell shape).

```
set.seed(1)
sim <- matrix(rexp(40000,0.2),ncol=40)
hist(apply(sim,1,mean),col="red",breaks=20, main="Distribution of means of 40 variables", xlab="Mean")
```

