Exploring Variances of the Exponential Distribution

ww44ss

October 2014

Synopsis

This analysis shows how the mean and standard deviation of a finite random sample of a uniform distribution converge to population mean with increasing sample size.

Here is a sample of the data.

```
##create data for analysis

        ##Set seed to make results reproducible
        set.seed(8675309)

        ##makes sample sizes variable
        rowsofsamples <- 40
        columnsofdistributions <- 1001

        ##set lambda
        lambda <- .2

        ##create data frame using slightly different method than in class
        ##c rows is the sample index and column indexes the distribution.
        rdata <- replicate(columnsofdistributions, rexp(rowsofsamples, lambda))
```

This analysis looks at a set of 1001 distributions of 40 samples.

Here is a sample of some data points

```
        rdata[1:10,1]
```

```
##  [1] 8.3108 6.1163 3.8367 1.7305 1.5428 3.3823 0.8332 0.1095 3.6691 2.4949
```

Question1: Show where the distribution is centered and compare it to the theoretical center of the distribution.

```
##Take means of columns and plot histrogram with mean and theoretical mean

        ##Compute Column means
        expmeans <- colMeans(rdata)

        ##Calculate "Mean of Means"
        meanofdistmeans<-mean(expmeans)
```

The mean of the distribution 5.0123 differs by 0.2453% from the theoretical population mean of 5.0

Variance: Show how variable it is and compare it to the theoretical variance of the distribution.

```
## Calculate the Variance
varofdistmeans = var(expmeans)

expectedvariance = 1/lambda/(lambda*rowsofsamples)
```

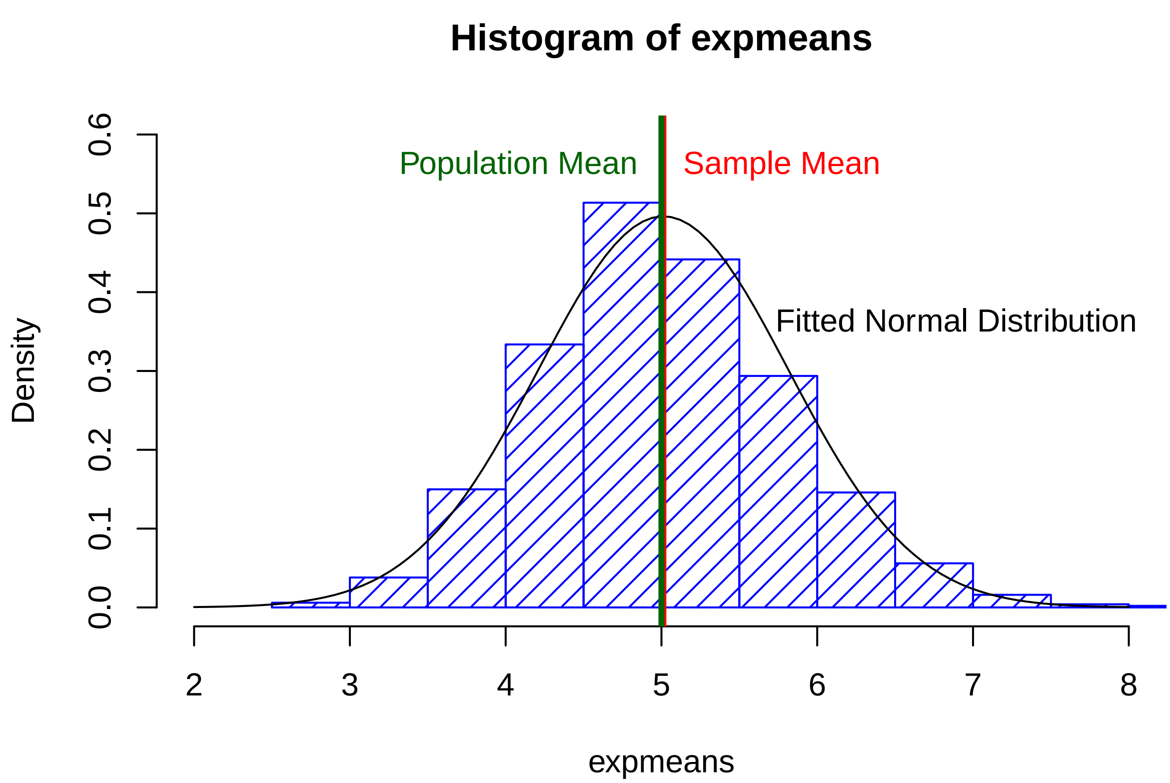
The theoretical variance of the distribution is lambda = 0.625 versus the standard deviation the data 0.6463, which differs by 3.41% .

Distribution: Show that the distribution is approximately normal.

```
##Generate Historgram
hist(expmeans, prob=TRUE, density=12, angle=45, xlim=c(2,8), ylim=c(0,0.6),col="blue")
curve(dnorm(x, mean=mean(expmeans), sd=sd(expmeans)), add=TRUE)


##Draw lines at mean and theoretical mean = 5
abline(v=meanofdistmeans, lwd=3, col="red")
abline(v=5, lwd=3, col="darkgreen")

##Label means on graph
text(meanofdistmeans,0.55, "Sample Mean", col="red", adj=c(-0.1,0))
text(5,0.55, "Population Mean", col="darkgreen", adj=c(1.10,0))
text(5.5,0.35, "Fitted Normal Distribution", col="black", adj=c(-.10,0))
```

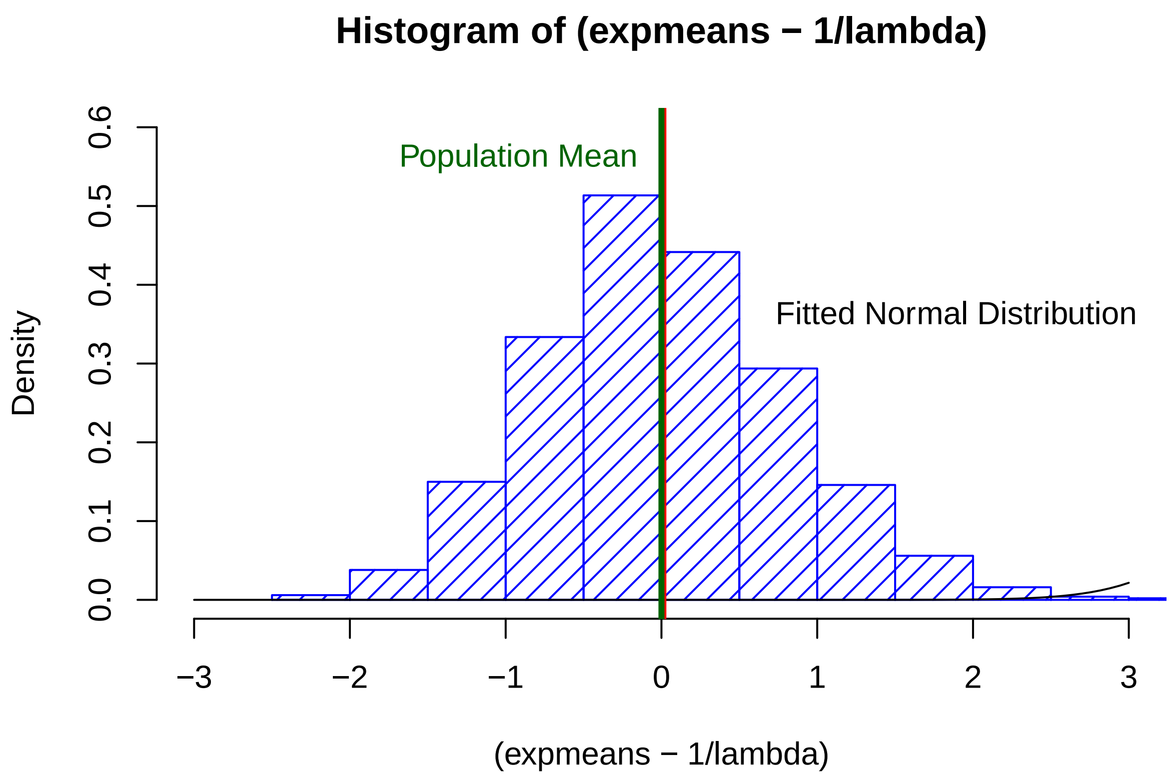


```
###Draw Std Deviation Lines
```

```
##Generate Historgram
hist((expmeans-1/lambda), prob=TRUE, density=12, angle=45, xlim=c(-3,3), ylim=c(0,0.6),col="blue
curve(dnorm(x, mean=mean(expmeans), sd=sd(expmeans)), add=TRUE)


##Draw lines at mean and theoretical mean = 5
abline(v=meanofdistmeans-1/lambda, lwd=3, col="red")
abline(v=0, lwd=3, col="darkgreen")

##Label means on graph
text(meanofdistmeans,0.55, "Sample Mean", col="red", adj=c(-0.1,0))
text(0,0.55, "Population Mean", col="darkgreen", adj=c(1.10,0))
text(0.5,0.35, "Fitted Normal Distribution", col="black", adj=c(-.10,0))
```




```
###Draw Std Deviation Lines
```

Above is a plot of a histogram of the exponentials. The sample mean of the data, shown in the graph by the red vertical line at 5.0123 is barely distinguishable from the popultion mean at 5.000. The fit of the normal distribution, from the sample mean and standard deviation 0.8039 follows the data well.

Convergence: Evaluate the coverage of the confidence interval

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
        lambda<-.2
        noofreps<-5000
        samples<-c(10, 20, 50, 100, 200, 500, 1000, 2000)

        pmean=1/lambda
        psd=1/lambda
        ssd=psd/sqrt(samples)

coverage <- sapply(samples, function(samples) {
        rdata <- replicate(noofreps, rexp(samples, lambda))
        expmeans <- (colMeans(rdata)-pmean)/ssd
        mean<-mean(expmeans)
        ll <- expmeans - qnorm(0.975) * ssd
        ul <- expmeans + qnorm(0.975) * ssd
        mean(ll < 0 & ul > 0)
})

ggplot(data.frame(samples, coverage), aes(x = samples, y = coverage)) + geom_line(size = 2) + geom_hline
```

```
## Warning: Removed 3 rows containing missing values (geom_path).
```