StatisticalInferenceProject1

Deepthi Murthy

July 25, 2015

{r global_options, include=FALSE} library(knitr) opts_chunk$set(fig.width=7, fig.height=4, warning=FALSE, message=FALSE)

Executive summary

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. I will set lambda = 0.2 for all of the simulations. I will investigate the distribution of averages of 40 exponentials. Note that I will need to do a thousand simulations.

Comparing sample mean and the theoretical mean of the distribution

Draw 1000 samples of size 40 from an $Exp(\frac{1}{0.2}, \frac{1}{0.2})$ distribution. For each of the 1000 samples calculate the mean. This is the same as drawing a single sample of size 1000 from the corresponding sampling distribution with $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$.

Based on CLT you expect that each single mean of those 1000 means is already approximately $\frac{1}{\lambda} = \frac{1}{0.2} = 5$. Since we now caluculate the mean of 1000 sampled means we expect the output to be very close to 5.

check if this is the case.

Simulations

```
# load neccesary libraries
library(ggplot2)

# set constants
lambda <- 0.2 # lambda for rexp
n <- 40 # number of exponetials
numberOfSimulations <- 1000 # number of tests

# set the seed to create reproducability
set.seed(11081979)

# run the test resulting in n x numberOfSimulations matrix
exponentialDistributions <- matrix(data=rexp(n * numberOfSimulations, lambda), nrow=numberOfSimulations)
exponentialDistributionMeans <- data.frame(means=apply(exponentialDistributions, 1, mean))
```

{r echo=FALSE} # plot the means ggplot(data = exponentialDistributionMeans, aes(x = means)) + geom_histogram(binwidth=0.1) + scale_x_continuous(breaks=round(seq(min(exponentialDistributio
max(exponentialDistributionMeans$means), by=1)))

Sample Mean versus Theoretical Mean

The expected mean μ of a exponential distribution of rate λ is

$\mu = \frac{1}{\lambda}$

```
mu <- 1/lambda
mu
```

Let $\bar{X}$ be the average sample mean of 1000 simulations of 40 randomly sampled exponential distributions.

```
meanOfMeans <- mean(exponentialDistributionMeans$means)
meanOfMeans
```

As you can see the expected mean and the avarage sample mean are very close

Sample Variance versus Theoretical Variance

The expected standard deviation σ of a exponential distribution of rate λ is

$\sigma = \frac{1/\lambda}{\sqrt{n}}$

The e

```
sd <- 1/lambda/sqrt(n)
sd
```

The variance Var of standard deviation σ is

$Var = \sigma^2$

```
Var <- sd^2
Var
```

Let Var_x be the variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution, and σ_x the corresponding standard deviation.

```
sd_x <- sd(exponentialDistributionMeans$means)
sd_x

Var_x <- var(exponentialDistributionMeans$means)
Var_x
```

As you can see the standard deviations are very close Since variance is the square of the standard deviations, minor differnces will we enhanced, but are still pretty close.

Distribution

Comparing the population means & standard deviation with a normal distribution of the expected values. Added lines for the calculated and expected means

```
{r echo=FALSE} # plot the means ggplot(data = exponentialDistributionMeans, aes(x =
means)) + geom_histogram(binwidth=0.1, aes(y=..density..), alpha=0.2) + stat_function(fun
= dnorm, arg = list(mean = mu , sd = sd), colour = "red", size=1) + geom_vline(xintercept
= mu, size=1, colour="#CC0000") + geom_density(colour="blue", size=1) + geom_vline(xintercept
= meanOfMeans, size=1, colour="#0000CC") + scale_x_continuous(breaks=seq(mu-3,mu+3,1),
limits=c(mu-3,mu+3))
```

Conclusions

In this analysis it is shown that the sampling distribution of the mean of an exponential distribution with $n = 40$ observations and $\lambda = 0.2$ is approximately $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$ distributed.