Project Statistical Inference

Summary

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should 1. Show where the distribution is centered at and compare it to the theoretical center of the distribution. 2. Show how variable it is and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal. 4. Evaluate the coverage of the confidence interval for 1/lambda: X¯±1.96Sn???.

Load Packages required

```
require(plyr)
```

```
## Loading required package: plyr
```

```
require(ggplot2)
```

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

```
require(scales)
```

```
## Loading required package: scales
```

Question A: Explore Exponential Distribution with Lambda = 0.2

```
## create data frame and boundary conditions

set.seed(12345)
n = 40
lambda = 0.2
samples = 1000
data <- data.frame(Measure = "Sample Means", value = sapply(1:samples, function(x) mean(rexp(n,
    lambda))), stringsAsFactors = F)
# data <- ddply(data,.(),mutate,calc=mean(value),theory=1/lambda)
```

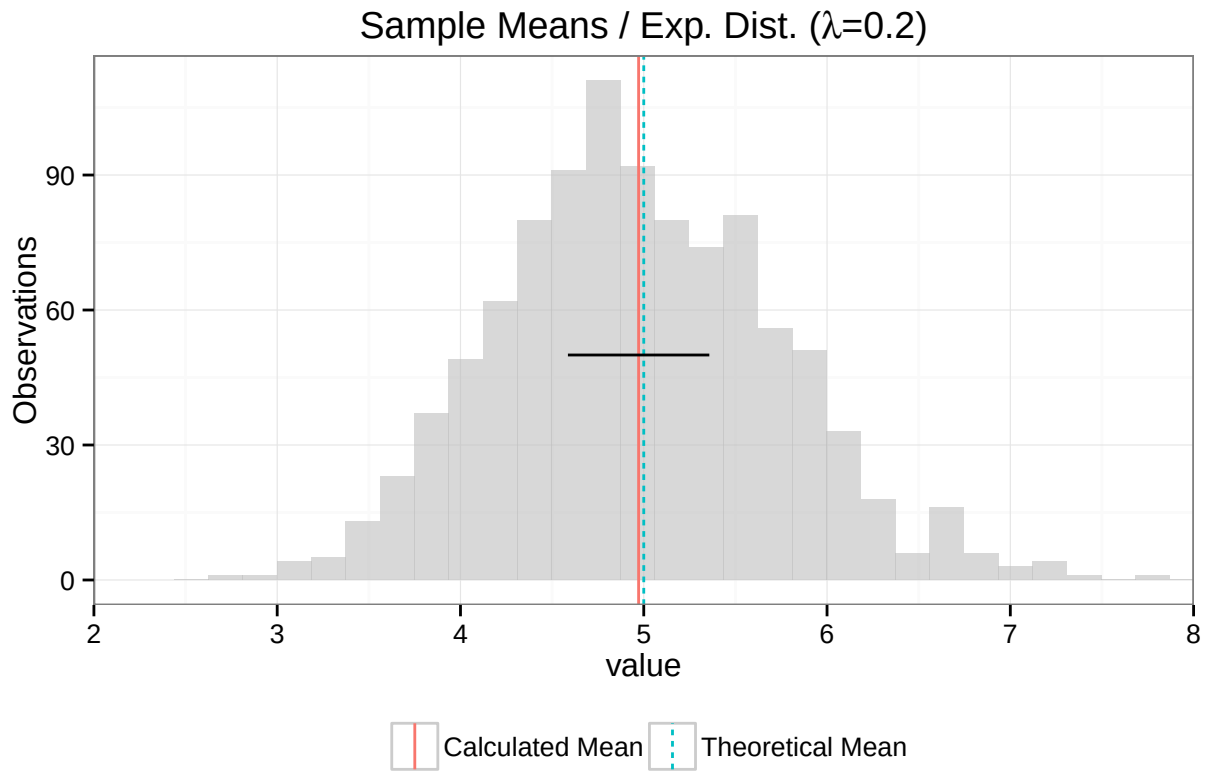
Compute mean & standar deviation

```
means<- data.frame(Type=c("Calculated Mean","Theoretical Mean"),mean=c(mean(data$value),1/lambda))
sd_calc <- sd(data$value)
```

Create plot histogram

```
title <- expression(paste("Sample Means / Exp. Dist. (",lambda, "=", "0.2",")",sep=""))
ggplot(data,aes(x=value))+geom_histogram(alpha=0.6,fill="grey")+
  theme_bw()+theme(legend.position="bottom")+ggtitle(title)+coord_cartesian(xlim=c(2,8))+
  geom_vline(data=means,aes(xintercept=mean,color=Type,linetype=Type), show_guide = TRUE)+
  guides(color=guide_legend(title=NULL),linetype=guide_legend(title=NULL))+
  scale_y_continuous("Observations",labels=comma)+
  geom_line(data=NULL,aes(x=c(mean(data$value)-sd_calc/2,mean(data$value)+sd_calc/2),y=c(50,50)),color=
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



Plot: 1,000 repeated samples of 40 exponential random numbers with λ =0.2. Red line = Average of the 1,000 sample means, Blue line = Theoretical mean of an exponential distribution which is $\frac{1}{\lambda} = 5$. The Caluclatd Mean is 4.972 with a 95% confidence interval of 4.9241 to 5.0198. Black line = Standard deviation of the sample means centered around the calculated sample mean

Result: The distribution of sample means appears to be normal and centered around the distribution mean Evaluate the coverage of the confidence interval for 1/lambda: $\bar{X} \pm 1.96 Sn$???.

The plot shows the sensitivity of the Confidence Interval from the sampled exponential distribution. Left panel: Simulates different values or lambda (1,000 simulations of 40 exponential samples, 95 threshold) Right panel: simulates different exponential samples (*lambda* = 0.21,000 simulations) The rigth panel show better predictability vs the left panel.