Statistical inference Project

Mario Segal

August 9, 2014

This is my Class Project for the Coursera Class of Statistical Inference, August 2014

Question A: Explore Exponential Distribution with Lambda = 0.2

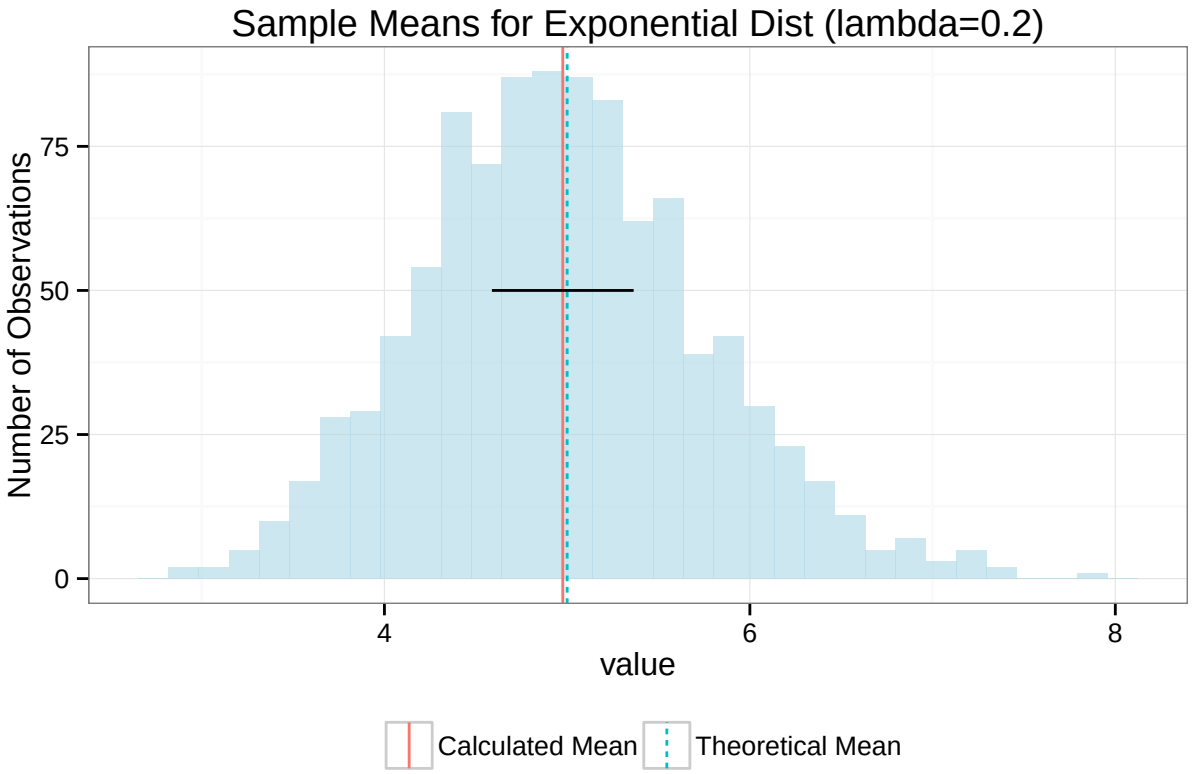
1. Repeatedly sample 40 exponential random numbers and plot their averages

```
require(plyr);require(ggplot2);require(scales)
```

```
## Loading required package: plyr
## Loading required package: ggplot2
## Loading required package: scales
```

```
n=40
lambda=0.2
samples=1000
data <- data.frame(Measure="Sample Means",value=sapply(1:1000,function(x) mean(rexp(n,lambda))),strings
#data <- ddply(data,.(),mutate,calc=mean(value),theory=1/lambda)
means<- data.frame(Type=c("Calculated Mean","Theoretical Mean"),mean=c(mean(data$value),1/lambda))
sd_calc <- sd(data$value)
ggplot(data,aes(x=value))+geom_histogram(alpha=0.6,fill="lightblue")+theme_bw()+
  theme(legend.position="bottom")+ggtitle("Sample Means for Exponential Dist (lambda=0.2)")+
  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("Number of 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.
```



The figure below whows the distribution of means from 100 repetitons of 40 exponential samples each. The red line is the calculated average of the 100 sample means, while the blue line denotes the theoretical mean of an exponential distribution with lambda = 0.2, which is 1/lambda = 5 The black horizontal line shows the calculated standard deviation of the sample means centered around the calculated sample mean which is equal to `round(sd_calc,2)`*. According to the central limit theorem the standard deviation of the sample means is equal to the standard deviation of the actual distribution divided by sqrt(n), where n is the sample size (40 in this case). For comparison purposes the estimated population standard deviatio would be* `round(sd_calc*sqrt(4),2)` *which is close to the theoretical value of 5*