Appendix_Simulation

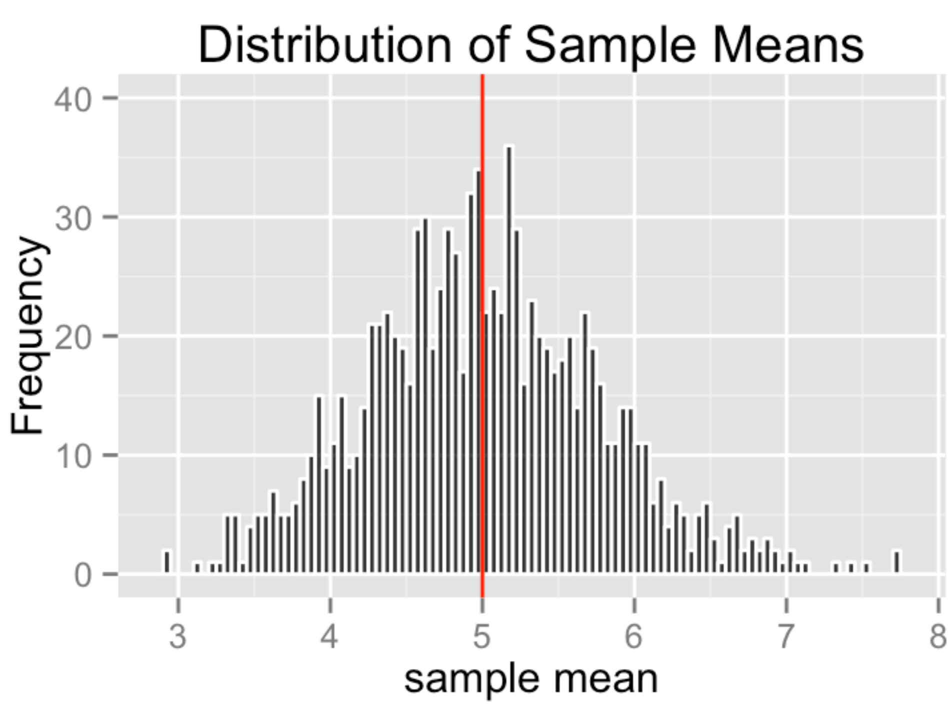
Ning Zhang

August 21, 2015

```
library(ggplot2)
library(knitr)
opts_chunk$set(warning = FALSE, message = FALSE, fig.width = 4, fig.height = 3, resul
ts="hide")
```

```
set.seed(123)
simulation <- matrix(nrow = 1000, ncol=40)
for (i in 1:1000){
    simulation[i,] <- rexp(40, rate = 0.2)
}
```

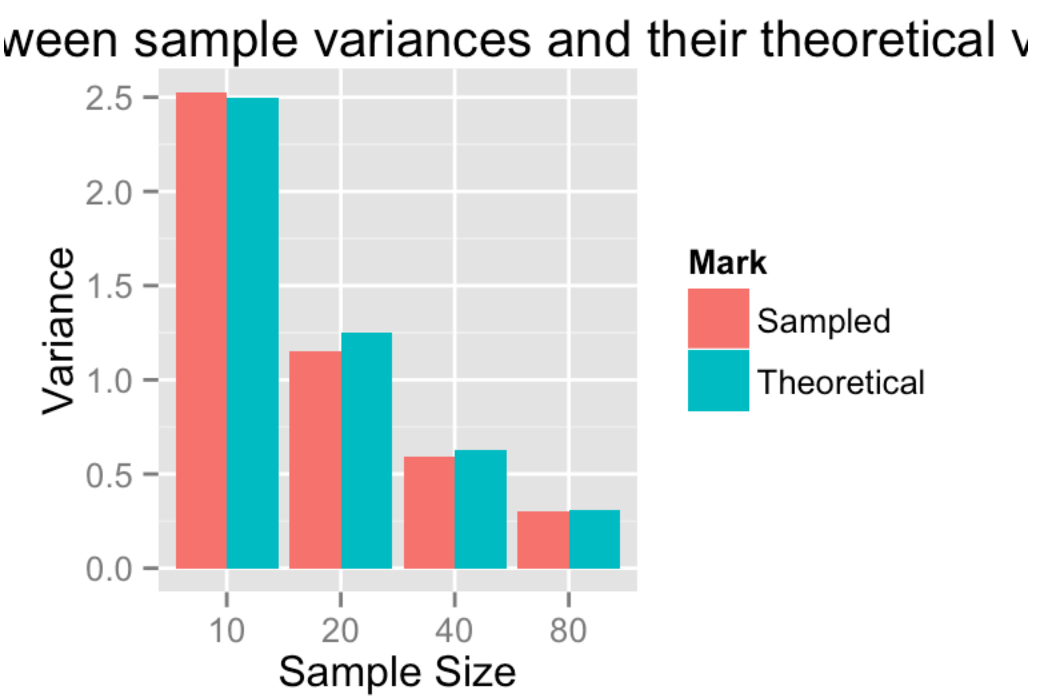
```
# a histogram showing the distribution of 1,000 samples
#hist(rowMeans(simulation), plot = TRUE, xlab = "sample mean", main = "Distribution o
f Sample Means")
sample_means <- data.frame(means = rowMeans(simulation))
population_mean <- 1/0.2
means_hist <- ggplot(data=sample_means, aes(x=means))+ ylim(0,40)
means_hist <- means_hist + geom_histogram(binwidth=0.05,colour = "white",fill = "blac
k",alpha=0.8) + xlab("sample mean") + ylab("Frequency") + ggtitle("Distribution of Sa
mple Means")
means_hist <- means_hist + geom_vline(xintercept=population_mean, colour="red", size=
0.5) + scale_x_continuous(breaks = c(2, 3, 4, 5, 6, 7, 8))
means_hist
```



```
theoretical_variance <- 1/(0.2^2)
set.seed(124)
simulation_10 <- matrix(nrow=1000, ncol=10)
for (i in 1:1000){ simulation_10[i,] <- rexp(10, rate = 0.2)}
simulation_20 <- matrix(nrow=1000, ncol=20)
for (i in 1:1000){simulation_20[i,] <- rexp(20, rate = 0.2)}
simulation_40 <- matrix(nrow = 1000, ncol=40)
for (i in 1:1000){simulation_40[i,] <- rexp(40, rate = 0.2)}
simulation_80 <- matrix(nrow=1000, ncol=80)
for (i in 1:1000){simulation_80[i,] <- rexp(80, rate = 0.2)}
```

```
variance_10 <- var(rowMeans(simulation_10))
variance_20 <- var(rowMeans(simulation_20))
variance_40 <- var(rowMeans(simulation_40))
variance_80 <- var(rowMeans(simulation_80))
sample_variances <- c(variance_10, variance_20, variance_40, variance_80)
theoretical_sample_variances <- c(theoretical_variance/10, theoretical_variance/20, theoretical_variance/40, theoretical_variance/80)
variance_mark = c(replicate(4, "Sampled"), replicate(4, "Theoretical"))
group = as.vector(replicate(2, c(10,20,40,80)))
```

```
#drawing a barplot to compare variances of sample means of various sample size with their respective standard error.
variances <- data.frame(Group = as.factor(group), Variance = c(sample_variances, theoretical_sample_variances), Mark = variance_mark)
comparison_var <- ggplot(variances, aes(x=Group, y=Variance)) + geom_bar(aes(fill = Mark), position = "dodge", stat= "identity") + xlab("Sample Size") + ggtitle("Comparison between sample variances and their theoretical values")
comparison_var
```



```r
#calculate mean and variance
mean_of_samples <- mean(sample_means$means)
variance_of_samples <- var(sample_means$means)
mean_of_samples
variance_of_samples
```

```r
# draw a plot showing density distribution of sample means and a normal distribution
with the same mean and standard deviation.
means_density <- ggplot(data=sample_means, aes(x=means)) +
    geom_density(stat="density", position = "identity")

cols <- c("Sample_Mean"="#f04546","Normal"="#3591d1")
means_density <- ggplot(data=sample_means, aes(x=means)) + geom_density(aes(y=..densi
ty.., position="stack", colour = "Sample_Mean")) + stat_function(fun = dnorm, aes(col
our = "Normal"), args = list(mean = mean_of_samples, sd = sqrt(variance_of_samples)))
+ scale_colour_manual(name="Distribution",values=cols) + geom_vline(xintercept=mean_o
f_samples, colour="#f04546", size = 0.5) + xlab("Means") + ylab("Density") +  ggtitle
("Distribution Of Sample Means of The Exponential Distribution")
means_density
```