Statistical Inference: Exponential Distribution

Ariel Lev

26. April 2015

Overview

In this report I would like to emperically show the properties of the Central Limit Theorem given samples of exponential distrubted iid variables. I will show that the estimates of the mean and variance converge respectively to the population's mean and variance, and that the distribution of the sample converge to a normal distribution.

Simulations

```
# number of exponentials in each simulation
n <- 40
# number of simulations
nsim <- 1000
# lambda
lambda <- 0.2

# mean and variance vectors
mns = NULL
vars = NULL

# theoretical mean and variance
data <- rexp(n*nsim, lambda)
data.frame( mean = mean(data), var = var(data))
```

```
##      mean      var
## 1 4.96664 24.62318
```

```
## simulation begins
for (i in 1:nsim) {
  r_expos <- rexp(n, lambda)
  mns <- c(mns,mean(r_expos))
  vars <- c(vars, var(r_expos))
  }
```

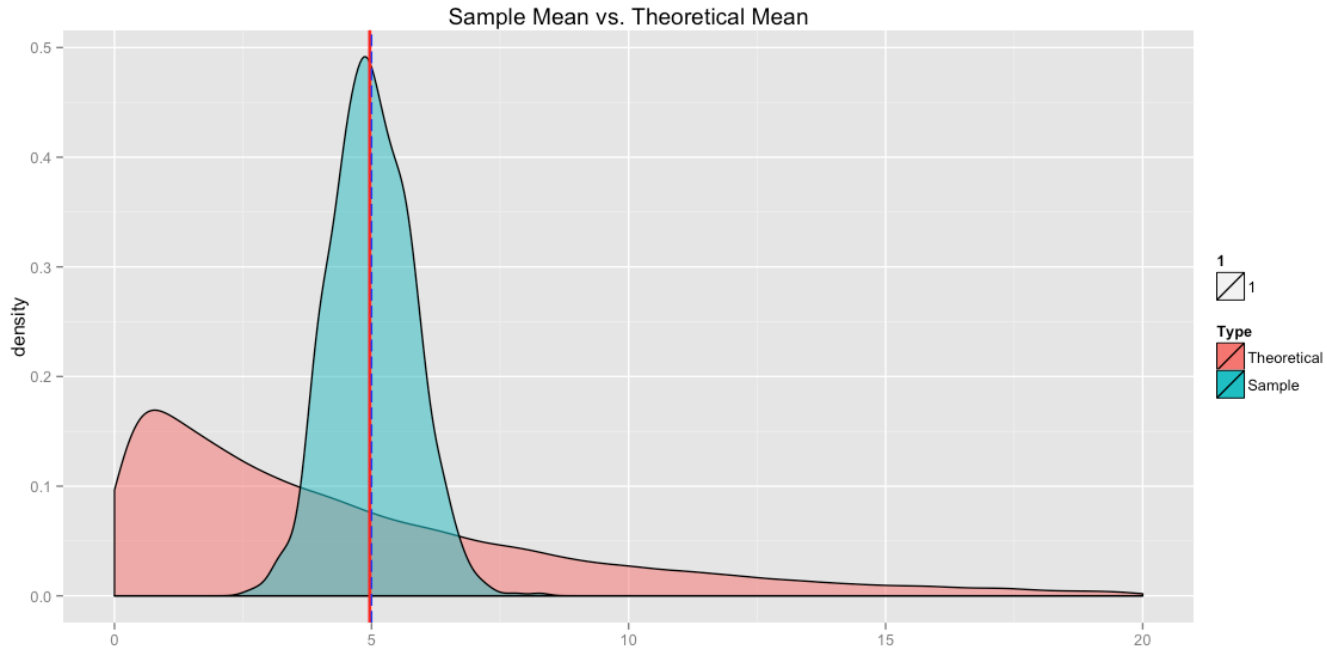
Sample Mean vs. Theoretical Mean

```
expo_df <-data.frame( value = data, type = "Theoretical")
mns_df <-data.frame( value = mns, type = "Sample")
plot_df <- rbind(expo_df, mns_df)

data.frame( mean = c(mean(data), mean(mns)), row.names = c("theoretical", "samp
le"))
```

```
##                mean
## theoretical 4.966640
## sample      5.000913
```

```
ggplot(as.data.frame(plot_df$value),aes(x=plot_df$value)) + geom_density(aes(fi
ll=plot_df$type, alpha = 1)) + geom_vline(xintercept = mean(expo_df$value), col
or = "red", size = 1) + geom_vline(xintercept = mean(mns_df$value), linetype =
"longdash", color = "blue", size = 0.5) + scale_x_continuous(breaks = sort(se
q(0,20,5))) + ggtitle(label = "Sample Mean vs. Theoretical Mean") + theme(axi
s.title.x=element_blank()) + guides(fill=guide_legend(title="Type")) + xlim(0,2
0)
```



Sample Variance vs. Theoretical Variance

```
expo_df <-data.frame( value = data, type = "Theoretical")
mns_df <-data.frame( value = mns, type = "Sample")
plot_df <- rbind(expo_df, mns_df)

# theoretical variance and mean of sample variance
data.frame( var = c(var(data), mean(vars)), row.names = c("theoretical", "sampl
e"))
```

```
##                var
## theoretical 24.62318
## sample      24.55614
```

```
# theoretical variance of samples
var(data)/(n-1)
```

```
## [1] 0.6313636
```

```
# empirical variance of samples
var(mns)
```

```
## [1] 0.6183946
```

Distribution

```
normal_df <-data.frame( value = rnorm(nsim, mean = mean(mns), sd = sd(mns)), ty
pe = "Normal")
mns_df <-data.frame( value = mns, type = "Sample")
plot_df <- rbind(normal_df, mns_df)

ggplot(as.data.frame(plot_df$value),aes(x=plot_df$value)) + geom_density(aes(fi
ll=plot_df$type, alpha = 1)) + geom_vline(xintercept = mean(normal_df$value), c
olor = "red", size = 1) + geom_vline(xintercept = mean(mns_df$value), linetype
= "longdash", color = "blue", size = 0.5) + scale_x_continuous(breaks = sort(se
q(0,20,5))) + ggtitle(label = "Sample vs. Normal Distribution") + theme(axis.ti
tle.x=element_blank()) + guides(fill=guide_legend(title="Type")) + xlim(0,20)
```