Investigating CLT for IID Exponential Random Variables

Aseem Rohatgi

Overview

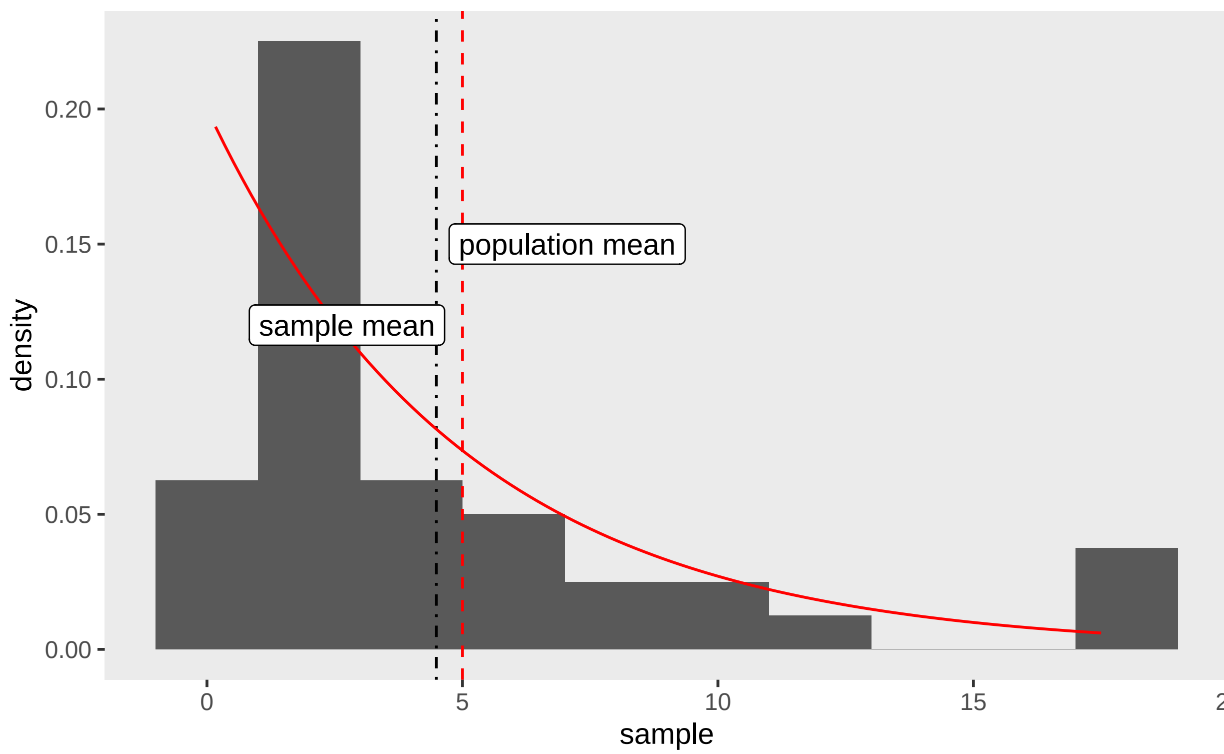
This report studies if a set of 40 independent and identically distributed $Exp(0.2)$ variables, behave according to the Central Limit Theorem. To perform this analysis, we will simulate random samples from the assumed distribution and then check of the means of those samples are normally distributed.

Simulations

We begin by obtaining 40 random values from a $Exp(0.2)$ distribution. We run this simulation 5000 times and then calculate the mean of each simulation.

```
exp_means<-sapply(1:5000, function(i) mean(rexp(40,0.2)))
```

The results are stored in a variable `exp_means`. The following plot shows the histogram of one such sample, along with the parent distribution density.



Central Limit Theorem

The theorem states that the mean of set of IID random variables is approximately normally distributed, and that this approximation gets better with increasing sample sizes.

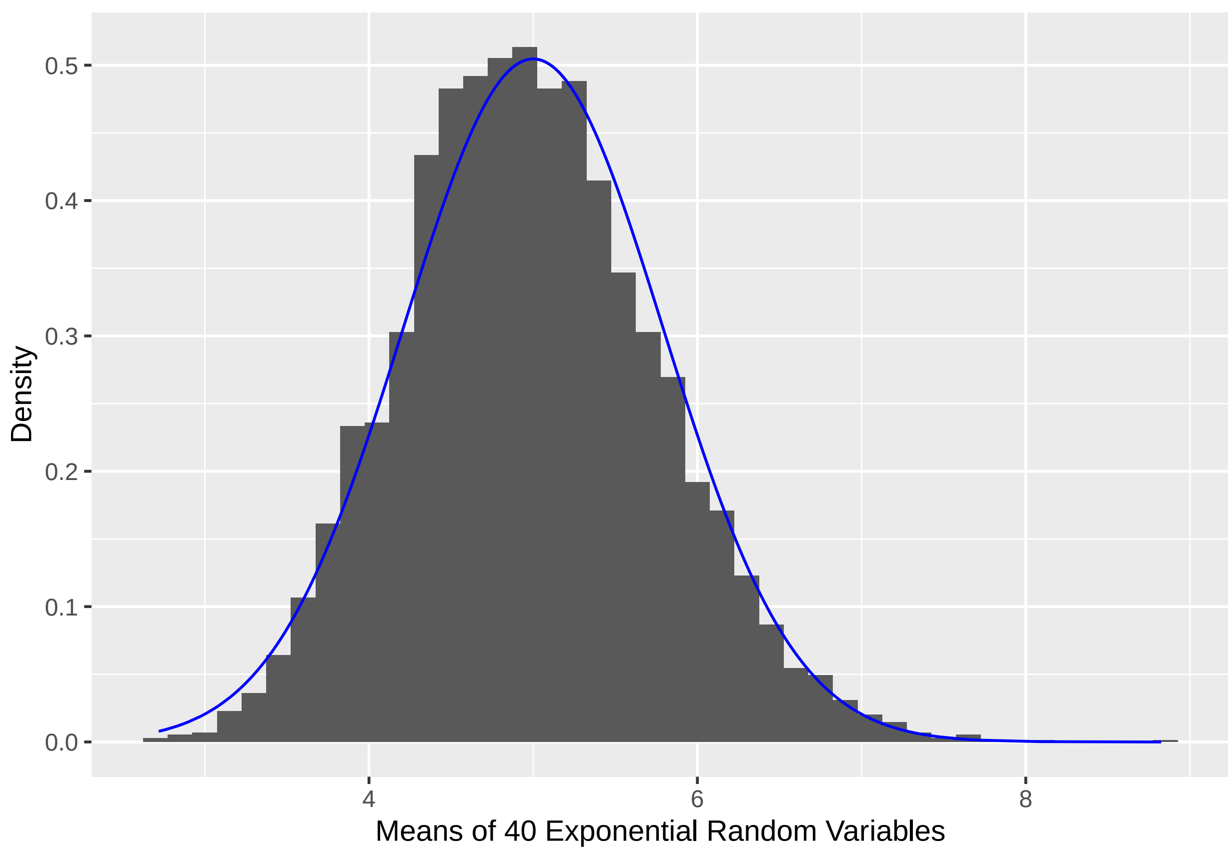
$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

This implies that for our sample of a set of 40 $Exp(0.2)$ variables, we have that,

$$\frac{\bar{X} - 5}{\frac{5}{\sqrt{40}}} \sim N(0, 1)$$

Observed sample means

As can be seen from the plot, the distribution of sample means is approximately normal. The blue curve is the density of a $N(5, \frac{5}{\sqrt{40}})$ variable.




```
5; mean(exp_means)
```

```
## [1] 5
```

```
## [1] 4.978541
```

The mean of the simulated distribution and that of the theoretical distribution are extremely close. The difference is 0.4291832% of the theoretical value.

```r
5/sqrt(40); sd(exp_means)
```

```
## [1] 0.7905694
```

```
## [1] 0.7704208
```

The standard deviation of the simulated distribution is also pretty close to the theoretical standard deviation. The difference is 0.4029722% of the theoretical value.

Conclusion

We randomly simulated numerous sets of IID exponential random variables and plotted the density of their means on a histogram. We saw that the distribution was very close to a normal distribution with the appropriate parameters. Thus, we were successful in confirming the validity of the Central Limit Theorem for this particular exponential distribution.

Appendix

Load `ggplot2` package

```r
library(ggplot2)
```

Seed for getting means

A seed was set for simulating the exponential variables. This was used again in extracting the first simulation to plot the sample data

```r
set.seed(137)
```

Sample data plot

```r
set.seed(137)

sample<-rexp(40,0.2)

x<-seq(min(sample),max(sample),length.out=100)

p<-ggplot(as.data.frame(sample), aes(x=sample, y=..density..)) +
  geom_histogram(binwidth=2) +
  geom_vline(xintercept=mean(sample), colour="black", linetype="dotdash")

p<-p + geom_line(data=data.frame(x=x, y=dexp(x,0.2)), aes(x=x, y=y), colour="red") +
```

```
  theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank()) +
  geom_vline(xintercept=5, colour="red", linetype="dashed")+
  annotate(geom="label", x=c(mean(sample)-1.75,7.05), y=c(0.12,0.15),
           label=c("sample mean", "population mean"))

print(p)
```

Sample means plot

```
q<-ggplot(as.data.frame(exp_means), aes(x=exp_means, y=..density..)) +
  geom_histogram(binwidth=0.15) +
  labs(x="Means of 40 Exponential Random Variables", y="Density")

q<-q + geom_line(data=data.frame(x=exp_means ,y=dnorm(exp_means,5,5/sqrt(40))),
                 aes(x, y), col="blue")

print(q)
```

Sample mean vs theoretical mean

```
meandiff<-abs((5-mean(exp_means)))*100/5
```

Sample standard deviation vs theoretical standard deviation

```
sddiff<-abs((5/sqrt(40)-sd(exp_means)))*100/5
```