Statistical Inference: The Study of the Exponential Distribution, A Simulation Exercise

Alexander Tuzhikov

September 14, 2015

Contents

1 Overview: Exponential Distribution

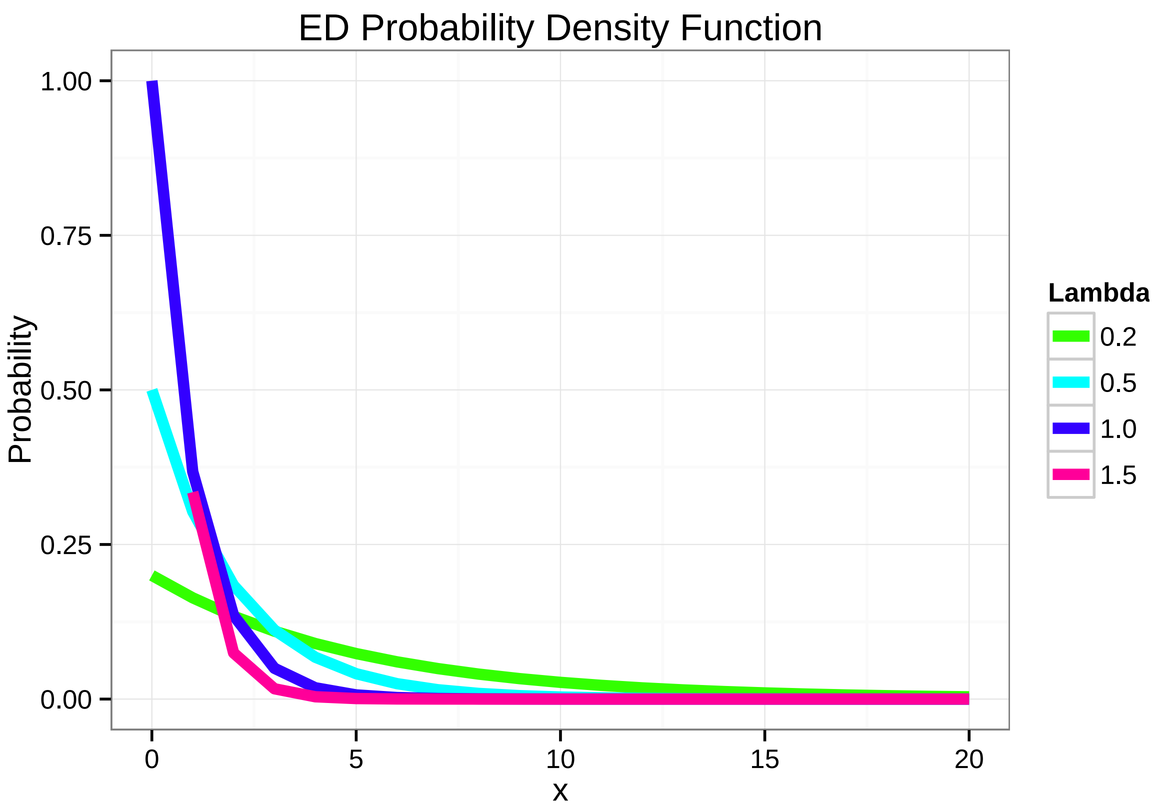
In accordance with Wikipedia, exponential distribution (ED) is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. Both mean and standard deviation of the ED is 1/lambda. As suggested in the study objective, here we will use lambda = 0.2. However, for the purpose of introduction, let's reconstruct the wiki plots of the ED with different lambda:

```
lambdas<- c(0.2, 0.5, 1, 1.5)#the given in the task + those from wikipedia
n<- 40 #given by ".. you will investigate the distribution of
       #averages of 40 exponential(0.2)s" in the task
sampling.count<- 1000  #given by ".. you will need to do a
       #thousand or so simulated averages of 40 exponentials" in the task
set.seed(2015) #seed for reproducibility
ed.mean<- mean(rexp(1e6, 0.2))
ed.sd<- sd(rexp(1e6, 0.2))
all.equal(ed.mean, ed.sd, tolerance = 1e-2) #mean and standard deviation are equal up to 1e-2 level of p
```

```
## [1] TRUE
```

```
#prepare a data.frame for the plot, melt by x, plot as line
ed.plot.df<- as.data.frame(cbind(
        x=0:40,
        la.0.2=dexp(x=0:40, lambdas[1]),
        la.0.5=dexp(x=0:40, lambdas[2]),
        la.1=dexp(x=0:40, lambdas[3]),
        la.1.5=dexp(x=0:40, lambdas[4])
)) %>%
        melt(id.vars="x") %>%
        ggplot(data=., mapping=aes(x=x, group=variable, y=value, color=variable)) +
```

```
        geom_line(size=2) + theme_bw() + xlim(0,20) + ylim(0,1) +
        labs(title="ED Probability Density Function") + ylab("Probability") +
        scale_color_manual(values=rainbow(4, start = 0.3, end = 0.9), labels=c("0.2","0.5", "1.0", "1.5
plot(ed.plot.df)
```



We just checked if the mean and the standard deviation are equal and plotted the probability density function at different levels of lambda.

2 Simulations: Sampling the ED

Now, let's move to the first task:
1. Show the sample mean and compare it to the theoretical mean of the distribution. First we gonna need 1000 samples of size 40 form ED.

```
#sampling
samples<- as.data.frame(do.call(what = "cbind", args = lapply(1:sampling.count, function(x){return(rexp
samples.colMeans<- colMeans(samples) #calculate the column means
```

Ok, now let's compare the theoretical mean, which is:

```
1/lambdas[1]
```

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

to that of the simulation procedure:

```r
mean(samples.colMeans) #calculate the total mean
```

```
## [1] 5.016687
```

```r
#compare
all.equal(1/lambdas[1],mean(samples.colMeans))
```

```
## [1] "Mean relative difference: 0.003337434"
```

The difference is neglegible in our case.

3 Sample Mean versus Theoretical Mean

4 Sample Variance versus Theoretical Variance

5 Distribution