Exponential distribution and CLT

Paolo Saracco

2024-03-21

Overview

In this project we explore the exponential distribution in R and compare it with the Central Limit Theorem. Our aim is to show that the empirical distribution of the mean of 40 iid exponentials with parameter $\lambda = 0.2$ is approximately a normal of mean $1/\lambda$ and variance $1/40\lambda^2$.

Introduction

According to Wikipedia "Exponential distribution":

> The *exponential distribution* or *negative exponential distribution* is the probability distribution of the distance between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate; the distance parameter could be any meaningful mono-dimensional measure of the process, such as time between production errors, or length along a roll of fabric in the weaving manufacturing process. The probability density function (pdf) of an exponential distribution is
>
> $$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$
>
> Here $\lambda > 0$ is the parameter of the distribution, often called the rate parameter. The distribution is supported on the interval $[0, \infty)$. If a random variable X has this distribution, we write $X \sim Exp(\lambda)$.

The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. For the sake of simplicity, we set $\lambda = 0.2$ for all of the simulations and we investigate the distribution of averages of 40 exponentials of parameter λ.

The following packages will be used throughout this report

```
library(ggplot2);
```

Simulations

Empirical density of the exponential

We begin by simulating 1000 exponentials with $\lambda = 0.2$

```
set.seed(0);
lambda <- 0.2;
m <- 1000;
simExp <- data.frame(rexp(m,lambda));
names(simExp) <- "Exp.2";
```

and by plotting the resulting empirical density

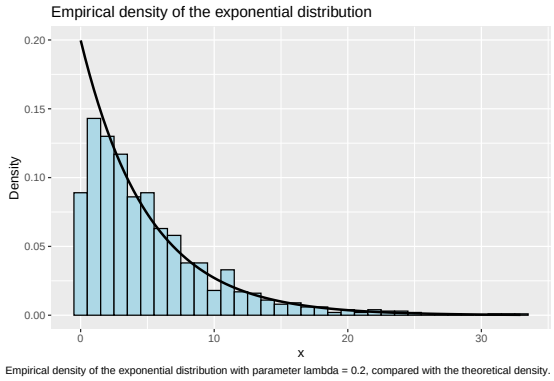

Figure 1: Empirical Density

Clearly this is very different from the density of a normal distribution.

Now, let us simulate 1000 means of 40 exponentials with parameter $\lambda = 0.2$

```
n <- 40;
m <- 1000;
mn <- data.frame(Sample.mean = apply(matrix(rexp(n*m,lambda),m,n),1,mean));
```

Sample Mean versus Theoretical Mean

We can compare the sample mean of our 1000 means

```
mean(mn$Sample.mean);
```

```
## [1] 4.994067
```

with the theoretical mean of the mean, which coincides with the mean of one of the iid exponentials

```
1/lambda;
```

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

Figure 2 shows the sample mean (in red) of our empirical density, which is clearly very close to the mean of the theoretical normal distribution (represented with a dashed blue curve).

Sample Variance versus Theoretical Variance

Analogously, we can compare the sample variance of our 1000 means

```
sd(mn$Sample.mean)^2;
```

```
## [1] 0.6670455
```

with the theoretical variance of the mean, which coincides with the variance of one of the iid exponentials divided by the number of random variables considered

```
1/(n*lambda^2);
```

```
## [1] 0.625
```

Figure 2 again shows one sample standard deviation (in orange) from the sample mean of our empirical density. For comparison purposes, the shaded blue area under the normal density represents one theoretical standard deviation around the theoretical mean.

Distribution

In order to verify that the distribution of our average of 40 iid exponentials is approximately normal, let us plot the empirical distribution in comparison with the theoretical one. Let us also highlight a few parameters like empirical and theoretical means and standard deviations.

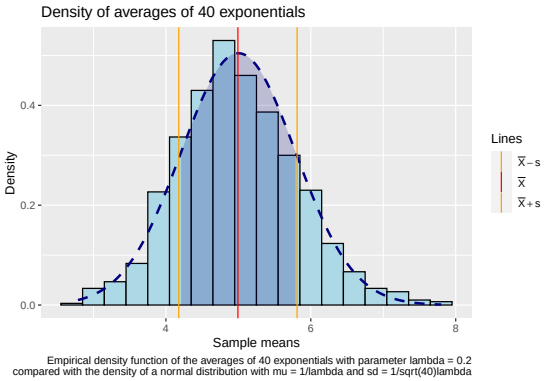

Figure 2: Density Comparison

As we expect, in view of the Central Limit Theorem, the empirical distribution of the mean of 40 iid exponentials of parameter λ is approximately normal with mean $1/\lambda$ and standard deviation $1/\sqrt{40}\lambda$. In order to find additional evidence, let us plot a q-q plot of the theoretical quantiles against the empirical ones.

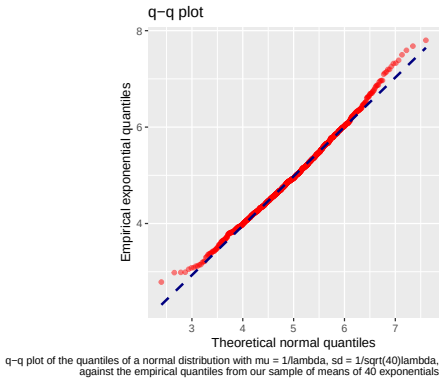

Figure 3: q-q plot

For the convenience of the interested reader, the *quantile-quantile (q-q) plot* is a graphical technique for determining if two data sets come from populations with a common distribution. A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. A 45-degree reference line is also plotted. If the two sets come from a population with the same distribution, the points should fall approximately along this reference line. The greater the departure from this reference line, the greater the evidence for the conclusion that the two data sets have come from populations with different distributions.

Appendices

Code for the empirical density

Code for Figure 1.

```
ggplot(data = simExp, aes(x = Exp.2)) +
        geom_histogram(binwidth = 1,
                       colour = "black",
                       fill = "lightblue",
                       aes(y = after_stat(density))) +
        stat_function(fun = dexp, args = list(rate = lambda), linewidth = 1) +
        labs(x = "x",
             y = "Density",
             title = "Empirical density of the exponential distribution",
             caption = paste("Empirical density of the exponential",
                             "distribution with parameter lambda = 0.2,",
                             "compared with the theoretical density.",
                             sep = " "));
```

Code for the density comparison

Code for Figure 2.

```
line.data <- data.frame(xintercept = c(mean(mn$Sample.mean)-sd(mn$Sample.mean),
                                       mean(mn$Sample.mean),
                                       mean(mn$Sample.mean)+sd(mn$Sample.mean)),
                        Lines = c("mu - sigma",
                                  "mu",
                                  "mu + sigma"),
                        colour = c("mu - sigma" = "orange",
                                   "mu" = "red",
                                   "mu + sigma" = "orange"),
                        stringsAsFactors = F);

my.labs <- list(bquote(bar(X) - s),
                bquote(bar(X)),
                bquote(bar(X) + s));

ggplot(data = mn, aes(x = Sample.mean)) +
        geom_histogram(binwidth = .3,
                       colour = "black",
                       fill = "lightblue",
                       aes(y = after_stat(density))) +
        stat_function(fun = dnorm,
                      args = list(mean = 1/lambda, sd = 1/(sqrt(n)*lambda)),
                      linewidth = 1,
                      colour = "navy",
                      linetype = 2) +
        stat_function(fun = dnorm,
                      args = list(mean = 1/lambda, sd = 1/(sqrt(n)*lambda)),
                      geom = "area",
                      fill = "navy",
                      alpha = 0.2,
                      xlim = 1/lambda + c(-1,1)*1/(sqrt(n)*lambda)) +
```

```
        geom_vline(aes(xintercept = xintercept, colour = Lines),
                   data = line.data,
                   linewidth = 0.5)  +
        scale_colour_manual(values = line.data$colour,
                            breaks = line.data$Lines,
                            labels = my.labs) +
        labs(x = "Sample means",
             y = "Density",
             title = "Density of averages of 40 exponentials",
             caption = paste("Empirical density function of the averages of",
                             "40 exponentials with parameter lambda = 0.2\n",
                             "in comparison with the density of a normal",
                             "distribution with mu = 1/lambda and",
                             "sd = 1/sqrt(40)lambda",
                             sep = " "));
```

Code for the q-q plot

Code for Figure 3.

```
ggplot(data = mn, aes(sample = Sample.mean)) +
        stat_qq(distribution = qnorm,
                dparams = list(mean = 1/lambda, sd = 1/(sqrt(n)*lambda)),
                colour = "red",
                alpha = 0.5) +
        stat_qq_line(distribution = qnorm,
                     dparams = list(mean = 1/lambda, sd = 1/(sqrt(n)*lambda)),
                     colour = "navy",
                     linewidth = 1,
                     linetype = 2) +
        labs(x = "Theoretical normal quantiles",
             y = "Empirical exponential quantiles",
             title = "q-q plot",
             caption = paste("q-q plot of the quantiles of a normal",
                             "distribution with mu = 1/lambda",
                             "and sd = 1/sqrt(40)lambda\n",
                             "against the empirical quantiles from",
                             "our sample of means of 40 exponentials",
                             sep = " ")) +
        theme(aspect.ratio=1);
```