Investigation of the exponential distribution

Rinnette Ramdhanie

8 November 2019

Overview

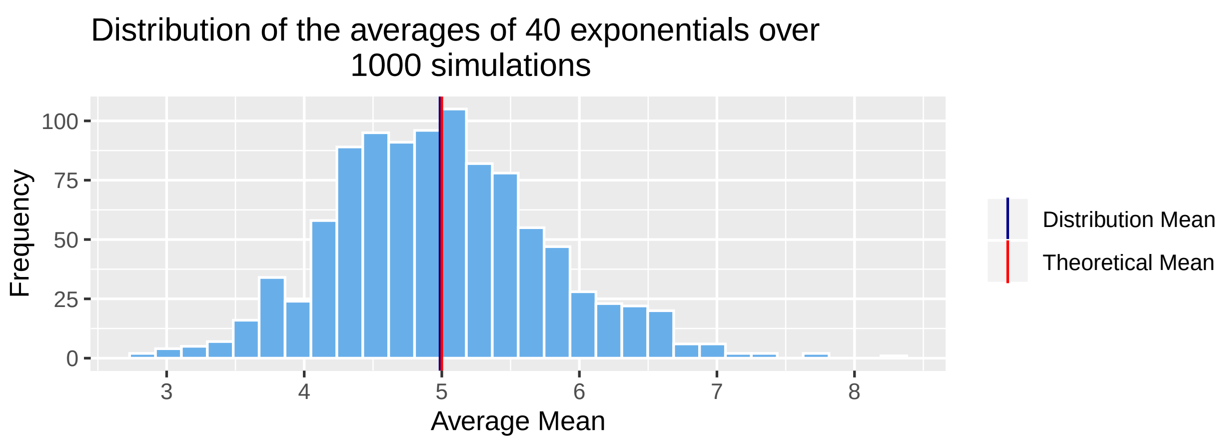
This report investigates the exponential distribution in R. The average of 40 random simulations was found one thousand times and the sample mean and variance were compared to the theoretical mean and variance. It was found that the sample statistics very closely matched the theoretical values. These averages were then normalised and plotted on a histogram and it was found that it approximated the standard normal distribution which conforms with the Central Limit Theorem.

Simulations

Forty simulations were generated one thousand times using the rexp() function in R. The rate parameter for the exponential distribution, called lambda in this report, was set to 0.2 in this investigation. For each group of 40 variates, the mean and variance was calculated. Both the mean and standard deviation is equal to $\frac{1}{lambda}$.

Sample Mean versus Theoretical Mean

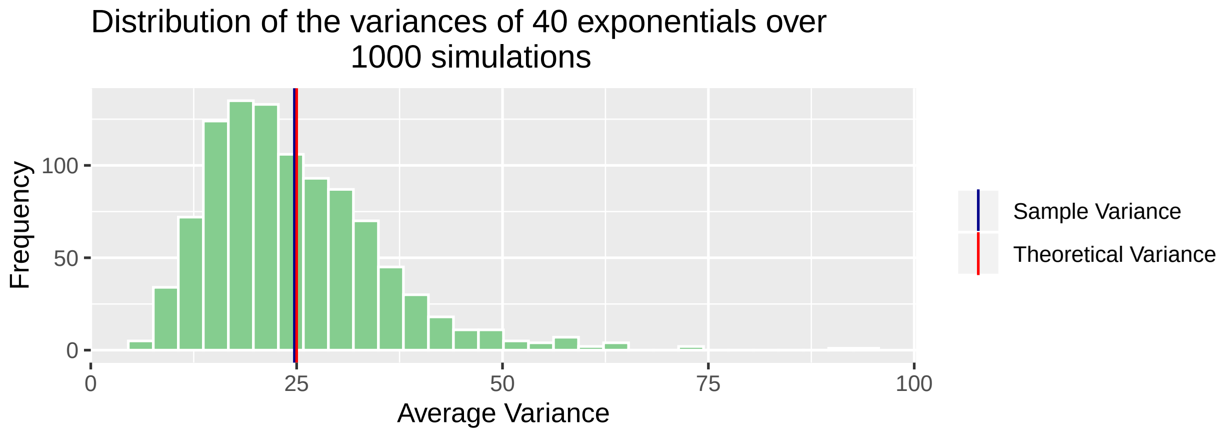
The theoretical mean of the exponential distribution is $\frac{1}{lambda}$ or $\frac{1}{0.2} = 5$. A histogram was used to show the distribution of the average means and how it compares to the theoretical mean.



The distribution is centered at about 4.9 which is very close to the theoretical mean of 5.

Sample Variance versus Theoretical Variance

The theoretical variance of the exponential distribution is $\frac{1}{lambda}^2$. In this case $\frac{1}{0.2}^2$ or 25. A histogram was used to show the distribution of the average variances and how it compares to the theoretical variance.



The histogram shows the variability of the sample means. The centre of the distribution of the variances is about 24.9 which is almost the same as the theoretical mean of 25. The lines showing the x intercept of these figures are almost on top of each other as the values are very close.

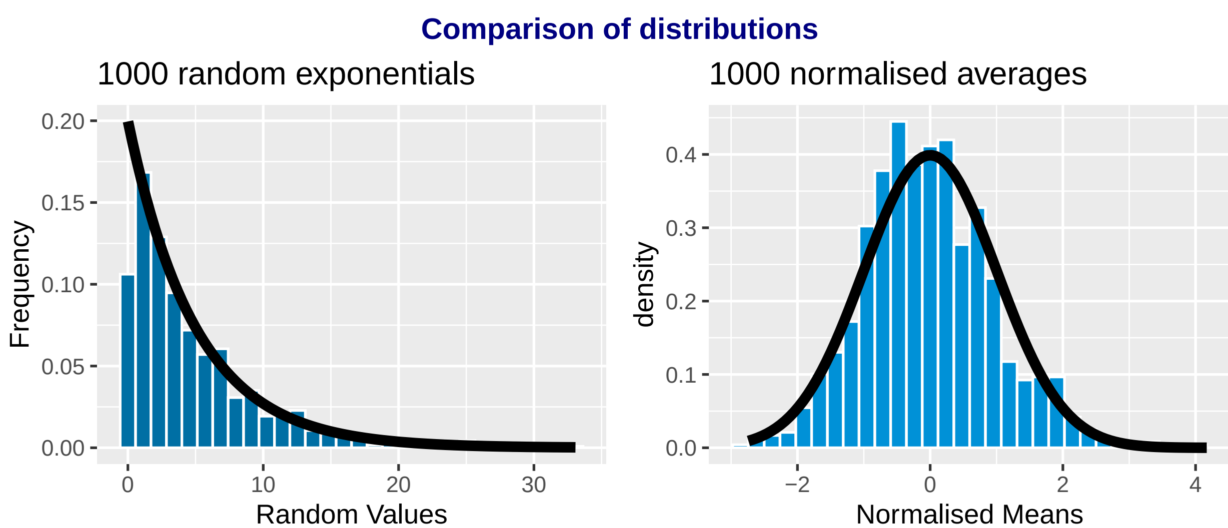
Distribution

In order to show whether the distribution of the one thousand averages of 40 random exponentials is approximately normal, the normalised averages were found and plotted on a histogram. This was done by subtracting the theoretical mean from each average and the result was divided by the standard error of the mean.

The standard error is $\sqrt{\frac{variance}{n}}$ where variance is 25 and n is 40 in this case.

The second histogram displays the distribution of these normalized averages.

The first histogram in the panel shows the distribution of one thousand random exponentials for comparison.



The first plot shows the probability density function of the exponential distribution which is convex and decreases in frequency as the random values increase.

Since the shape of the second plot looks like a bell curve, this means that the distribution is approximately standard normal which complies with the Central Limit Theorem.

Appendix

```r
knitr::opts_chunk$set(echo = FALSE)
        library(ggplot2)
        library(ggpubr)
        ## Set up variables to be used in the report
        lambda <- 0.2
        numexp <- 40
        meanexp <- 1/lambda
        varexp <- (1/lambda)^2
        ## Simulate 40 exponentials one thousand times and find the
        ## average means and variances.
                set.seed(11)
                means <- NULL
                for (i in 1:1000) means = c(means, mean(rexp(40, 0.2)))

                set.seed(12)
                vars <- NULL
                for (i in 1:1000) vars = c(vars, var(rexp(40, 0.2)))
        ## Calculate the distribution and theoretical means and store in a dataframe
                distMean <- mean(means)
                theoMean <- 1/lambda
                central_means <- data.frame(statistic = c("Distribution Mean",
                                                          "Theoretical Mean"),
                             value = c(distMean, theoMean))

        ## Plot the average means on a histogram and compare to the theoretical mean
                ggplot(data.frame(means), aes(x=means)) +
                        geom_histogram(col = "white", fill = "#68AFEA") +
                        scale_x_continuous(breaks=seq(0, 10, 1)) +
                        geom_vline(data = central_means, aes(xintercept = value,
                                                             color = statistic)) +
                        scale_color_manual(values = c("dark blue", "red"), name = NULL) +
                        labs(title="Distribution of the averages of 40 exponentials over
                             1000 simulations", x = "Average Mean", y = "Frequency")
        ## Calculate the distribution and theoretical variances and store in a dataframe
                distVar <- mean(vars)
                theoVar <- (1/lambda)^2
                central_mean_var <- data.frame(statistic = c("Sample Variance",
                            "Theoretical Variance"), value = c(distVar, theoVar))

        ## Plot the average variances and compare to the theoretical variance
                ggplot(data.frame(vars), aes(x=vars)) +
                        geom_histogram(col = "white", fill = "#85CD8F") +
                        geom_vline(data = central_mean_var, aes(xintercept = value,
                                                                color = statistic)) +
                        scale_color_manual(values = c("dark blue", "red"), name = NULL) +
                        labs(title="Distribution of the variances of 40 exponentials over
                             1000 simulations", x = "Average Variance", y = "Frequency")
        se <- sqrt(varexp/numexp)  #standard error
        norm <- (means-meanexp)/se #normalised averages

        # Plot the distribution of 1000 random exponentials
```

```r
        set.seed(13)
        distr1000 <- data.frame(outcome = rexp(1000, lambda))
        distr1000_plot <- ggplot(distr1000, aes(x=outcome)) +
                geom_histogram(aes(y=..density..), col = "white",
                               fill = "#006FA4") +
                stat_function(fun = dexp, args = list(rate=lambda), size = 2) +
                labs(title = "1000 random exponentials", x = "Random Values",
                     y = "Frequency")

# Plot the distribution of the 1000 normalised averages
        distr1000_40_plot <- ggplot(data.frame(norm), aes(x=norm)) +
                geom_histogram(aes(y=..density..), col = "white",
                               fill = "#0091D7") +
                stat_function(fun = dnorm, size = 2) +
                labs(title = "1000 normalised averages", x = "Normalised Means")

# Display in one panel
        distr_panel <- ggarrange(distr1000_plot, distr1000_40_plot, align = "h")
        annotate_figure(distr_panel,
                top = text_grob("Comparison of distributions",
                                color = "navy blue", face = "bold", size = 12))
```