Statistical Inference

Neel

9/17/2020

Overview

The purpose of this report is to be able to take the 1000 means of 40 randomly generated exonential distributions. By doing so many simulations we expect to see a nearly normal distribution with means and variances that are quite close to the theoretical values.

Sample mean vs Theoretical mean

Exponential means dataframe

We create a dataframe of 1000 means from an exponential distribution of lambda = 0.2.

```
n <- 40
n_trials <- 1000
lambda <- 0.2

mean_df <- data.frame(trial = numeric(), mean = numeric())
for (i in 1:n_trials) {
    mean_df[i,] <- c(i,mean(rexp(n,lambda)))
}
```

Sample mean

We calculate the sample mean by using the mean function.

```
sample_mean <- mean(mean_df$mean)
sample_mean
```

```
## [1] 5.042706
```

Theoretical mean

We calculate the theoretical mean by 1/lambda as the equation.

```
theoretical_mean <- 1/lambda
theoretical_mean
```

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

Sample vs theoretical mean histogram

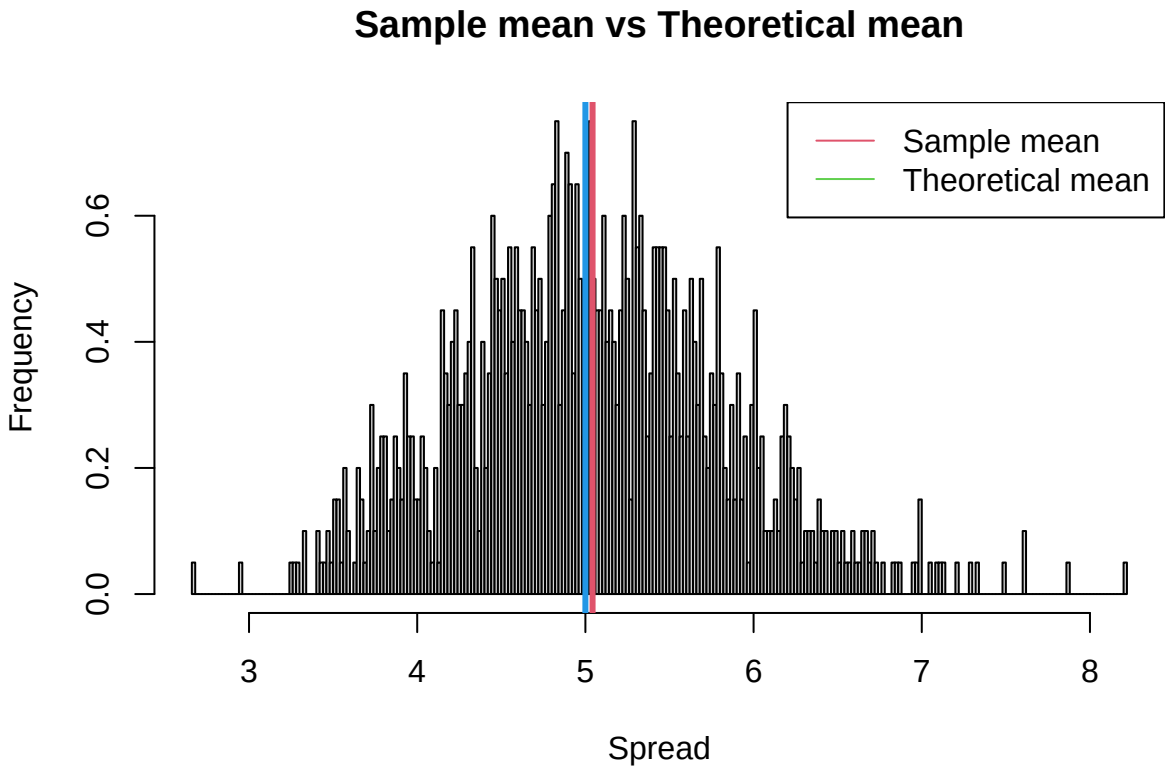
We create a histogram with 200 breaks to see a greater spread of the distribution and draw a vertical red line to denote the sample mean and a green line to deonte the theoretical mean.

```
hist(mean_df$mean, breaks = 200, prob = T, main = "Sample mean vs Theoretical mean", xlab = "Spread", yl

#vertical line for sample mean
abline(v = sample_mean, col = 2, lwd = 3)

#vertical line for theoretical mean
abline(v = theoretical_mean, col = 4, lwd = 3)

legend("topright", c("Sample mean", "Theoretical mean"), col = c(col = 2, col = 3), lty = c(1,1))
```



From the histogram above we can see that the sample mean is very close to that of the theoretical mean which is what we expected to see from a nearly normal distribution due to the 1000 trials.

Sample variance vs Theoretical variance

Sample variance

We calculate the sample variance by using the var function on the means.

```r
sample_variance <- var(mean_df$mean)
sample_variance
```

```
## [1] 0.6240124
```

Theoretical variance

We calculate the theoretical variance by using the formula (1/lambda)^2/(number of means per trial).

```r
theoretical_variance <- (((1/lambda)^2)/(n))
theoretical_variance
```

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

Sample vs theoretical variance histogram

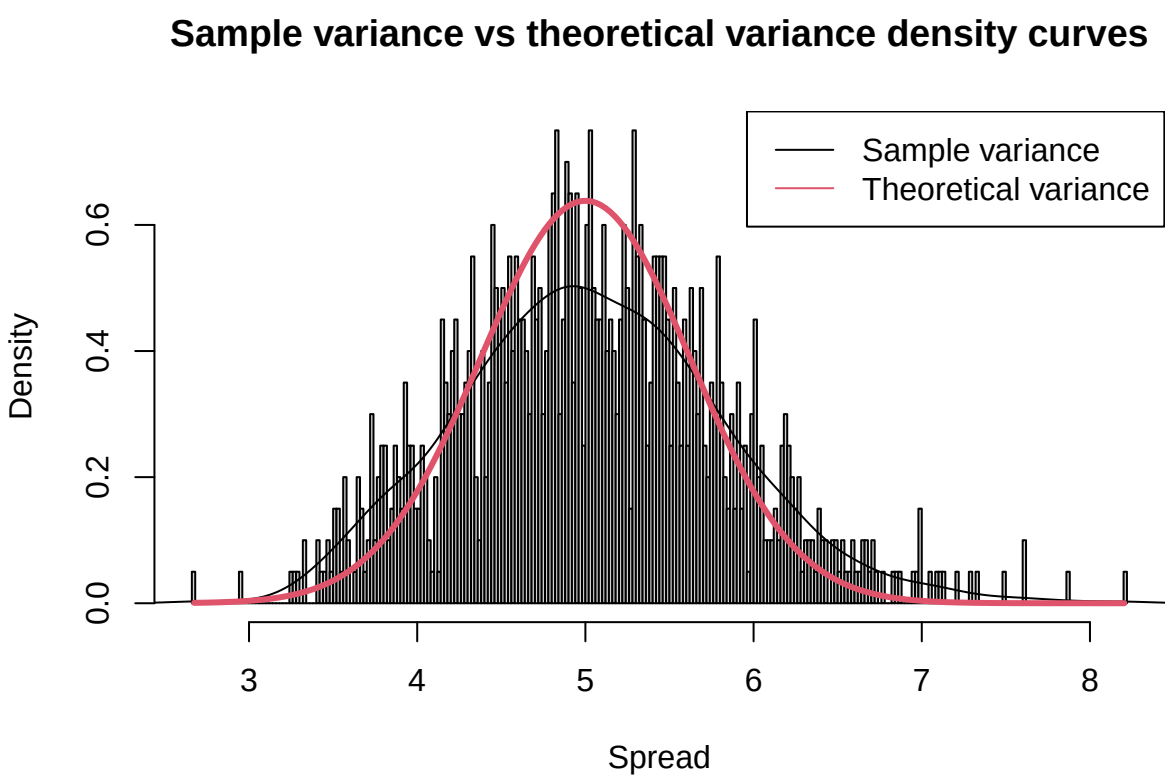
We create a histogram with 200 breaks to see a greater spread of the distribution. We then add a density curve using the density function to plot the sample variance spread through the data. We add another density curve of the theoretical variance by using x and y values based on the 200 breaks such that we find the dnorm value at the specific x value based on the theoretical mean and variance.

```r
hist(mean_df$mean, breaks = 200, prob = T, main = "Sample variance vs theoretical variance density curve

#density curve of sample variance
lines(density(mean_df$mean))

#density curve of theoretical variance
minimum_mean <- min(mean_df$mean)
maximum_mean <- max(mean_df$mean)
x_vals <- seq(minimum_mean, maximum_mean, length.out = 200)
y_vals <- dnorm(x_vals, mean = theoretical_mean, sd = theoretical_variance)
lines(x_vals, y_vals, col = 2, lwd =3)

legend("topright", c("Sample variance", "Theoretical variance"), col = c(col = 1, col = 2), lty = c(1,1
```



From the above histogram we see that the sample spread is not symmetic but is nearly symmetric to that of the theoretical spread which once again shows that the data is indeed tending towards the side of a nearly normal distribution.

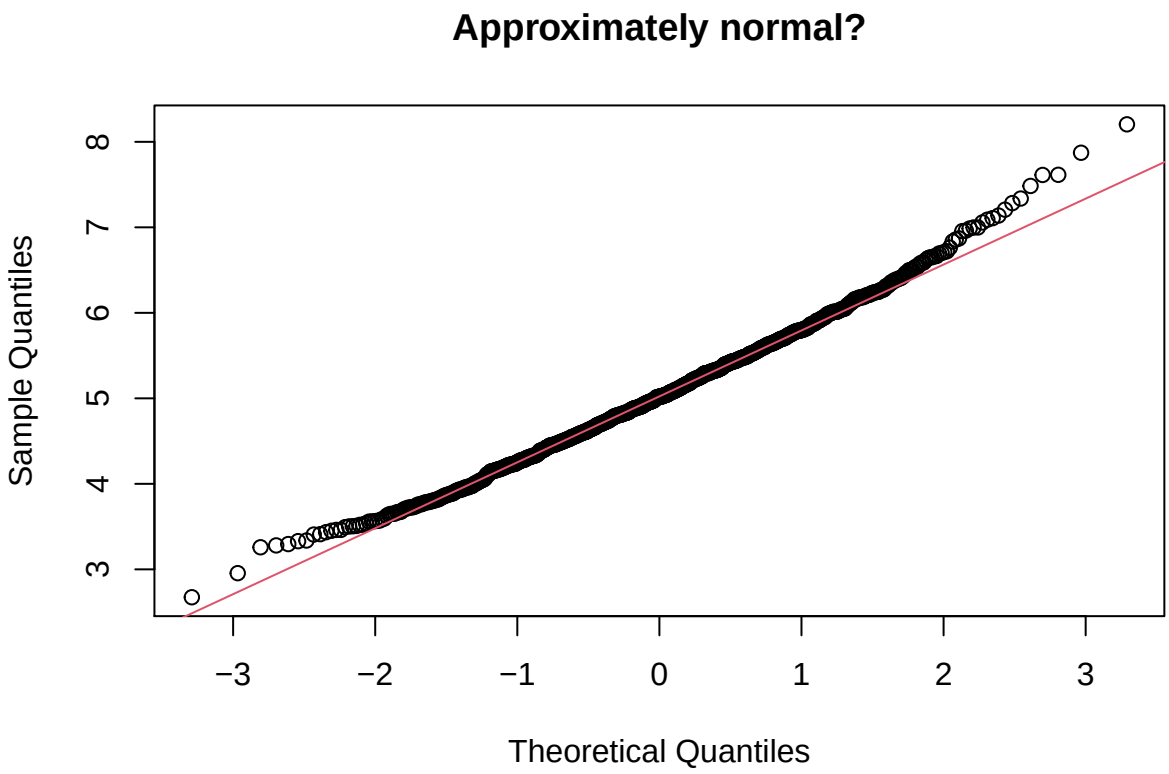
Approximately normal

In order to check the entire dataframe of means for normality we can use a qqplot to show how the data follows the theoretical values for the central quantiles.

QQPLOT of exponential means

We plot a qqplot with the means from the dataframe and draw a theoretical line to see how the data models the theoretical values.

```
qqnorm(mean_df$mean, main = "Approximately normal?")
qqline(mean_df$mean, col = 2)
```



Fron the above plot we can once again see that the data is nearly normal as it quite closely adheres to the line for the central quantiles.