Statistical Inference Report

Anandu R

9/20/2020

Within this exercise using statistical inference and knowledge on various distributions, I investigate the exponential distribution and compare it with the Central Limit Theorem. I will attempt to relate the two distributions - Exponential distribution and Uniform distribution.

Creating an exponential distribution

```
set.seed(13)
lambda = 0.2
n = 40
simulations = 1000
dstbn = rexp(n, lambda)
```

We will now use bootstrap sampling to create a sample distribution

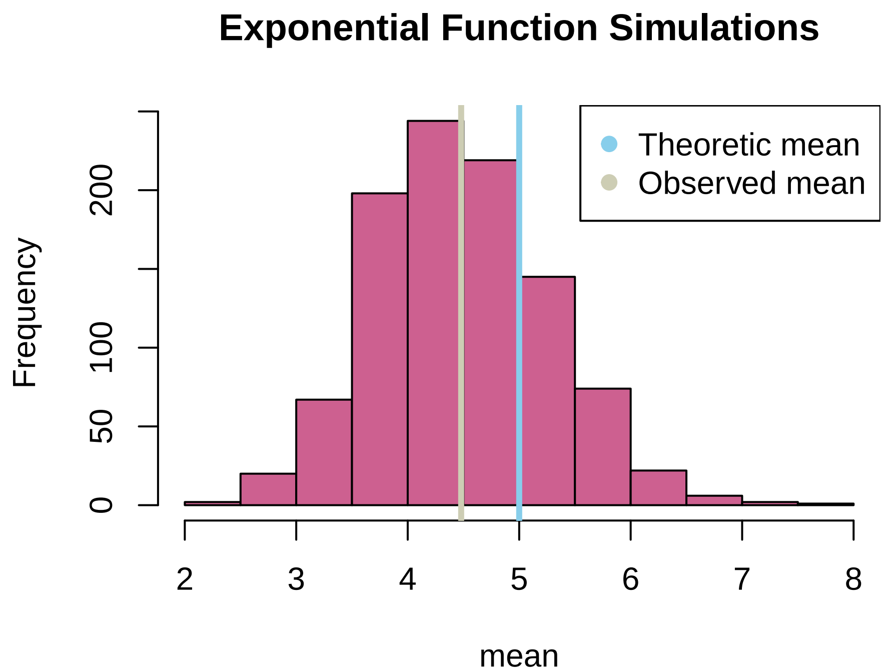
```
dstbn_sampled = matrix(sample(dstbn,simulations*n,replace = T),simulations,n)
dstbn_resampled = apply(dstbn_sampled, 1, mean)
```

Comparing the means

```
##   distribution    means
## 1  Theoretical 5.000000
## 2     Observed 4.479315
```

Visualization

```
hist(dstbn_resampled, xlab = "mean", main = "Exponential Function Simulations", col = "hotpink3")
abline(v = theory_mean, col = "skyblue", lwd = 3)
abline(v = observed_mean, col = "lightyellow3", lwd = 3)
legend("topright", col = c("skyblue","lightyellow3"), legend = c("Theoretic mean", "Observed mean"), pc
```



The theoretical mean is 5 whereas the observed mean of the 40 averages from the 1000 sampled distributions is 4.48

Comparing variability

Comparing variability in the sampled distribution compared to the theoretical distribution variance.

```
##   distribution  variance
## 1  Theoretical 0.6250000
## 2     Observed 0.6238251
```

The theoretical variance is 0.625 whereas the observed variance of the 40 averages from the 1000 sampled distributions is 0.6238

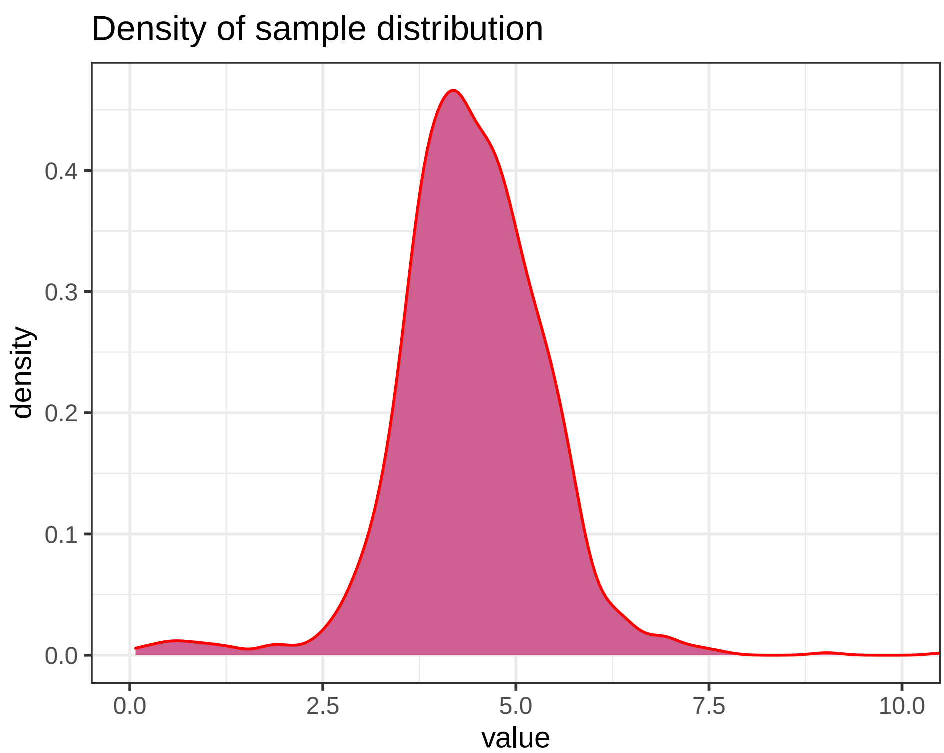
Visualising the sampled distribution

Representation of the sample distribution as a density plot to investigate the normality of the distribution

```r
## Creating dataframe for visualization
df = data.frame(obs=c(rep("observed",40),rep("sampled",40)),value=c(dstbn,dstbn_resampled))

## Visualization
library(ggplot2)
ggplot(
  df,
  aes(
    x = value,
```

```
  )
) + geom_density(fill = "hotpink3", color = "red") +
  theme_bw() +
  coord_cartesian(xlim = c(0,10)) +
  ggtitle("Density of sample distribution")
```



We can observe that the distribution has a bell-shaped curve, a classical feature of normal distribution.

Approximating the sample distribution to a standard normal

Normalizing the sample means using z-scores

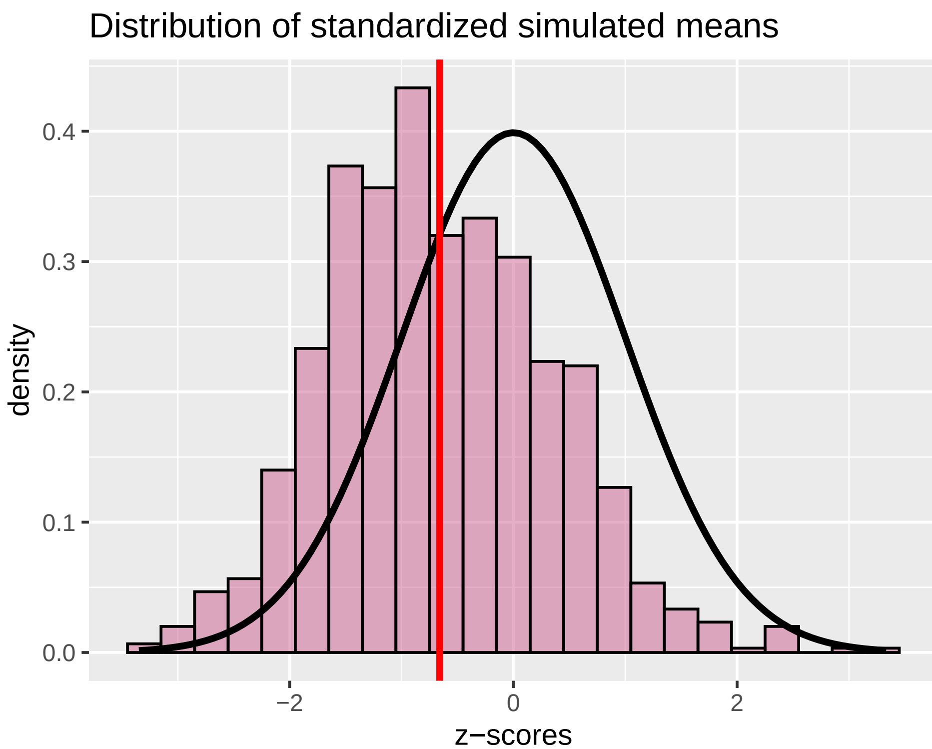
According to the central limit theorem the z-scores should be centered around the mean of the standard normal, 0.

```
dstbn_normalized = data.frame(z=as.numeric((dstbn_resampled-theory_mean)/sqrt(theory_var)))
z_mean = mean(dstbn_normalized$z)
```

Visualizing

```
ggplot(
  dstbn_normalized,
  aes(
    x = z
  )
) + geom_histogram(alpha=0.5,  binwidth = 0.3, fill="hotpink3", color="black", aes(y = ..density..)) +
   stat_function(fun = dnorm, size = 1.2) +
```

```
  geom_vline(xintercept = z_mean, color="red", size = 1.2) +
  ggtitle("Distribution of standardized simulated means") +
  xlab("z-scores")
```



We see that the distribution falls almost within the standard normal distribution, and hence approximated.