Part 1: Simulation Exercise

Olha Klishchuk

2023-10-18

Overview

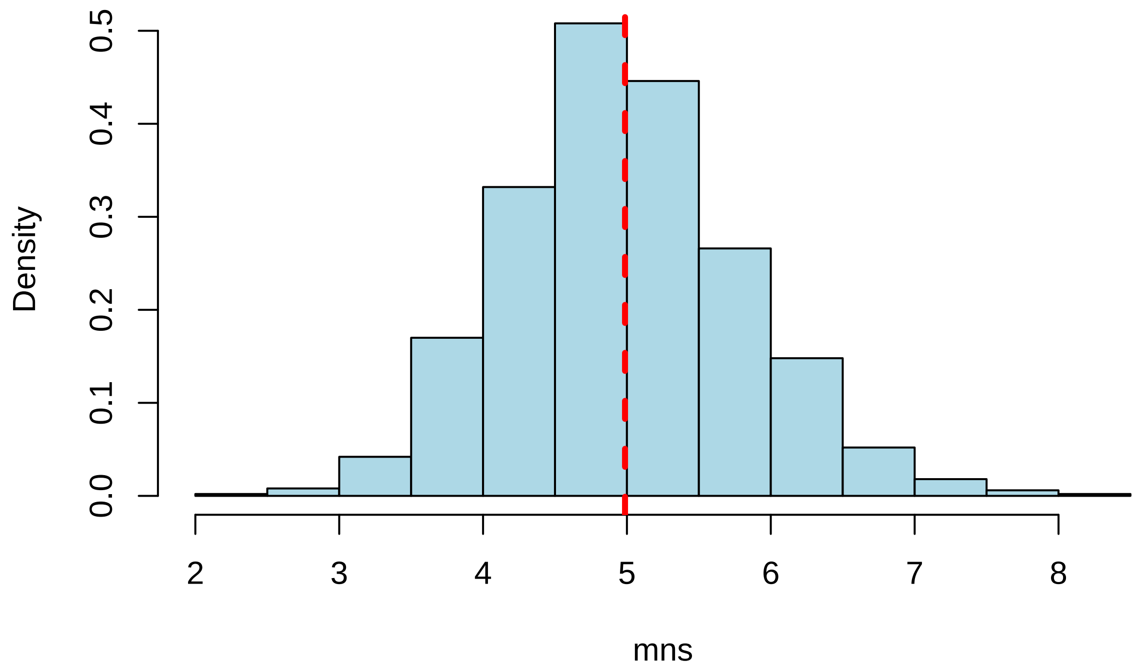
We have some distribution function $f = e^{\lambda} * n$. We need to analyze does it converges to st.normal distribution. Let's play with this distribution and plot density histogram.

```r
library(ggplot2)
```

```r
# Create simulation
lambda <- .2
n = 40

invisible(rexp(n, lambda))

#Redraw
mns = NULL
for (i in 1 : 1000) {mns = c(mns, mean(rexp(n, lambda)))}
options(repr.plot.width = 3.5, repr.plot.height = 3.5)
hist1<-invisible(hist(mns, col = 'lightblue', freq = F, breaks = 20, main = '',plot = T))
abline(v = mean(mns), col = 'red', lwd = 3, lty = 2)
```



Plot 1. Histogram of mns distribution

Take a look at the descriptives.

```
#Descriptive statistics
summary(mns);cat("  Sd\n",round(sd_e<-sd(mns), 3),"\n")
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.438   4.421   4.940   4.987   5.490   8.220
```

```
##   Sd
##  0.826
```

```
#Kurtosis and skewness
# suppressPackageStartupMessages(install.packages('PerformanceAnalytics'))
suppressPackageStartupMessages(library(PerformanceAnalytics))
```

```
## Warning: package 'PerformanceAnalytics' was built under R version 4.2.3
```

```
cat('Kurtosis:',round(kurtosis(mns), 2), '\n')
```

```
## Kurtosis: 0.24
```

```r
cat('Skewnwss:', round(skewness(mns), 2))
```

```
## Skewnwss: 0.3
```

Consider mean interval at the 5% quantile:

```r
round(mean(mns) + c(-1,1)*qnorm(.975)*sd(mns)/sqrt(length(mns)), 5)
```

```
## [1] 4.93561 5.03803
```

Hypothesis testing

Now we will compare the standard deviation of empirical and theoretical distributions. Lets take into account that the theoretical mean and standard deviation for this distribution are similar as for Poisson distribution, we will have:

$$mean_t = \frac{1}{\lambda}$$
$$sd_t = \frac{mean_t}{\sqrt{n}}$$

```r
#Theoretical and empirical mean
mean_t = 1/lambda
mean_e = mean(mns)
```

```r
# Standard deviation from analytic expression
sd_t <- (1/lambda)/sqrt(n)
sd_t
```

```
## [1] 0.7905694
```

```r
sd_e^2
```

```
## [1] 0.6825977
```

We see that empirical standard deviation is less volatile than theoretical one. Let's test hypothesis on closeness to normality according to the *Central Limit Theorem*.

```r
# Shapiro-Wilk-test
shapiro.test(mns)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mns
## W = 0.99448, p-value = 0.001011
```

Therefore, in the Appendix 1.1 you can see the plot that compares actual density of distribution with theoretical, where with dashed lines outlined theoretical and empirical means. We see that empirical distribution is slightly left skewed, but the hypothesis testing approves hypothesis on normality.

Summary

Therefore, the mns distribution is approximately converges to a standard normal distribution due to the sufficiently large number of observations.

Appendix

A Empirical VS Theoretical Distrubution

```
suppressMessages(library(RNifti))
```

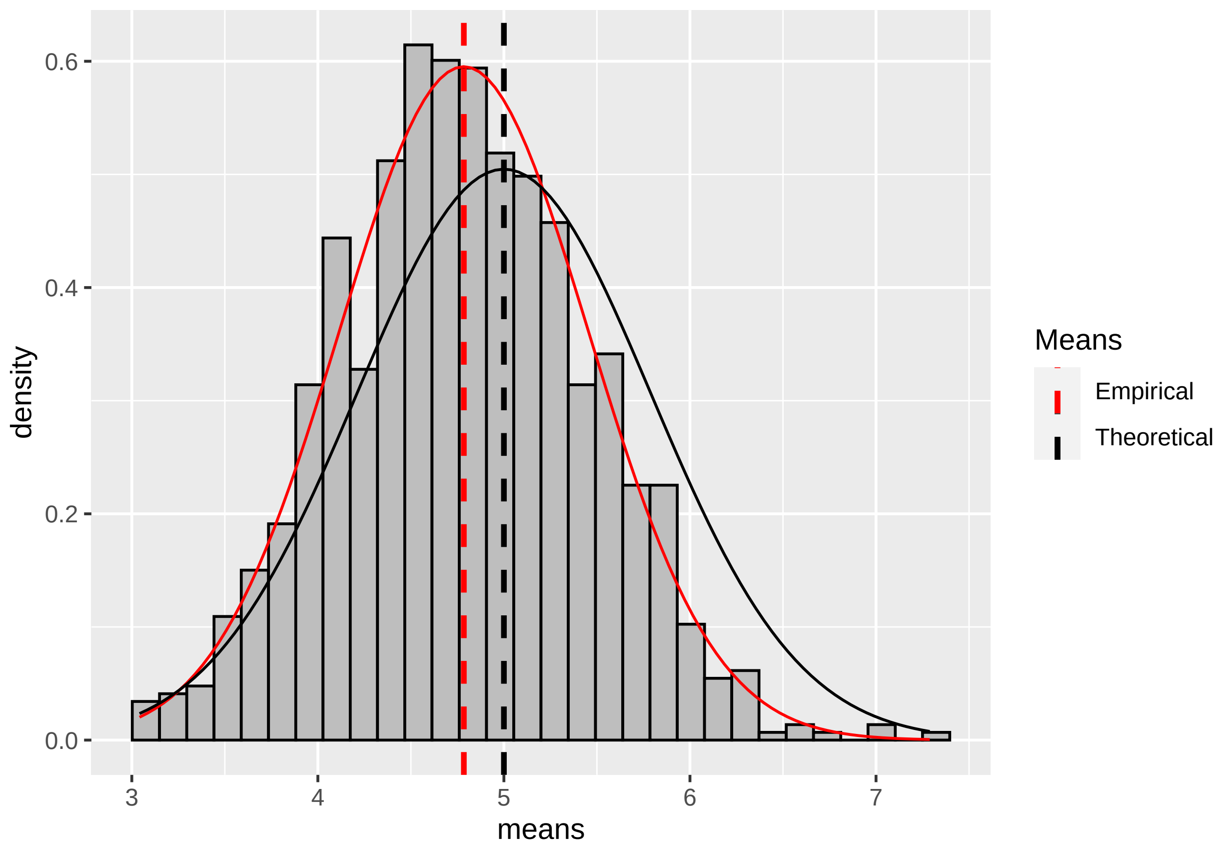
```
## Warning: package 'RNifti' was built under R version 4.2.3
```

```
#Adding distributions to the previous graph.
#Lets create 1000 draws
sim = 1000
resamples <- matrix(sample(rexp(n, lambda), n*sim, replace = T), sim, n)
means <-apply(resamples, 1, mean)
df<-data.frame(means)
```

```
#Plotting
plot_hist<- ggplot(df, aes(means)) +
    geom_histogram(bins = 30, fill="grey",color="black", aes(y=..density..))

hist2<-plot_hist+stat_function(fun=dnorm,args=list(mean=mean(df$means), sd=sd(df
    $means)),color = "red") +
    geom_vline(aes(xintercept = mean(means), color="Empirical"), size = 1,
              linetype = 'dashed')

hist2+stat_function(fun=dnorm,args=list(mean=mean_t, sd=sd_t), color = "black")+
    geom_vline(aes(xintercept = mean_t, color = 'Theoretical'), size = 1,
              linetype= 'dashed') +
    scale_color_manual(name = "Means", values = c(Empirical = "red", Theoretical
              = "black"))
```



0.6
0.4
0.2
0.0
density
3
4
5
6
7
means
Means
Empirical
Theoretical