How many number of the observations and sample you will need, according to the Central-limit theorem?

Md.Shahrear Zaman

March 2020

Introduction:

We can find out the answer of the relevant question by checking the probability density function of different number of the observations and sample. According to the central-limit theorem the desired number of the observations and sample should lead us to the normal distribution with mean 0 and standard deviation 1.

Central-limit theorem:

Let, Z_n is the random variable which is defined by

$$Z_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{var[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

$\bar{X}_n$ is the sample mean

Let $f(X)$ be a density with mean μ and finite variance σ^2 and n is the sample size.

Let, $E(\bar{X}_n) = \mu = 0$ and

$\sigma^2 = 1$

Therefore,

$$Z_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{var[\bar{X}_n]}}$$

$$\Rightarrow \quad Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

$$\Rightarrow \quad Z_n = \frac{\bar{X}_n - 0}{1/\sqrt{n}} \quad [\mu = 0 \; and \; \sigma^2 = 1 \; at \; the \; limit]$$

$$\Rightarrow \qquad Z_n = \sqrt{n}\,\overline{X}_n \quad \left[scaled\ by\ \sqrt{n}\right]$$

Now If we increase n , the probability distribution of Z should go towards the normal distribution which has 0 mean and 1 standard deviation.

(p.234; Graybill, Franklin A. ; Boes, Duane C. (1963))

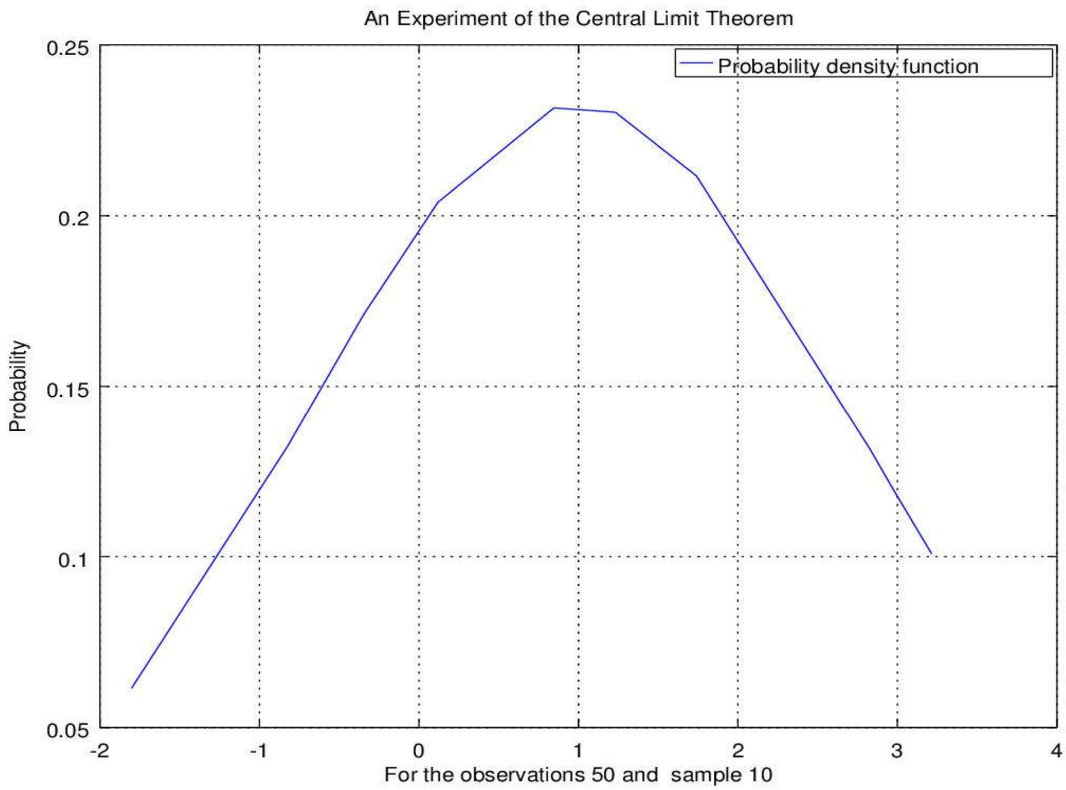
<u>Code:</u>

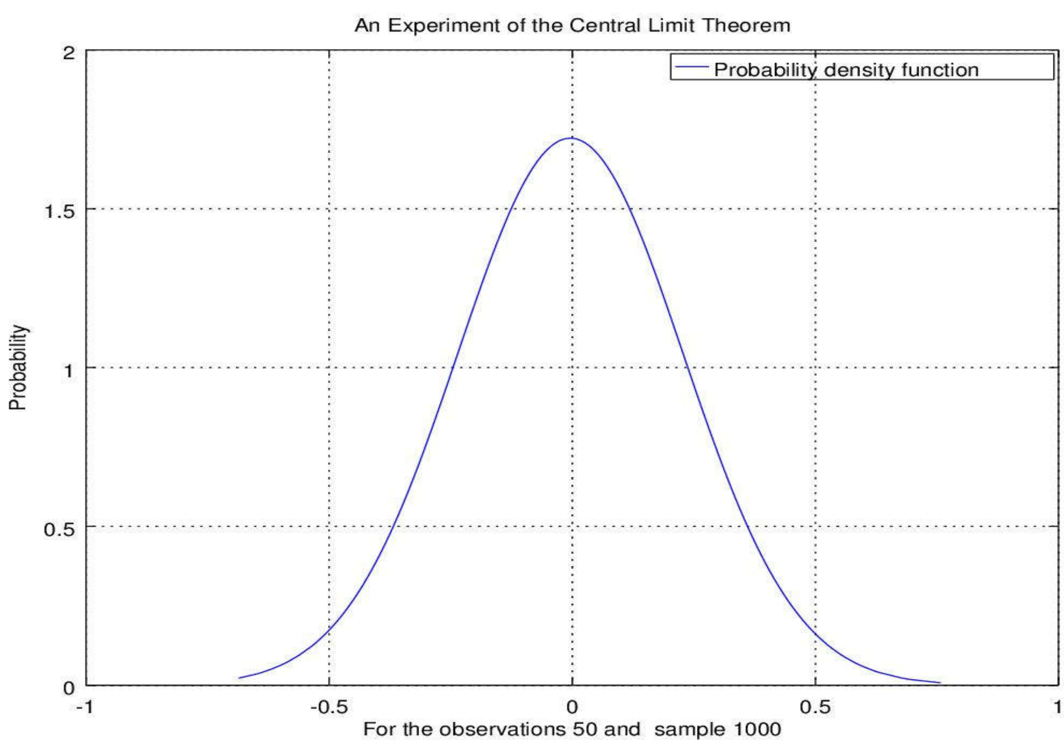
Let's check this with octave,

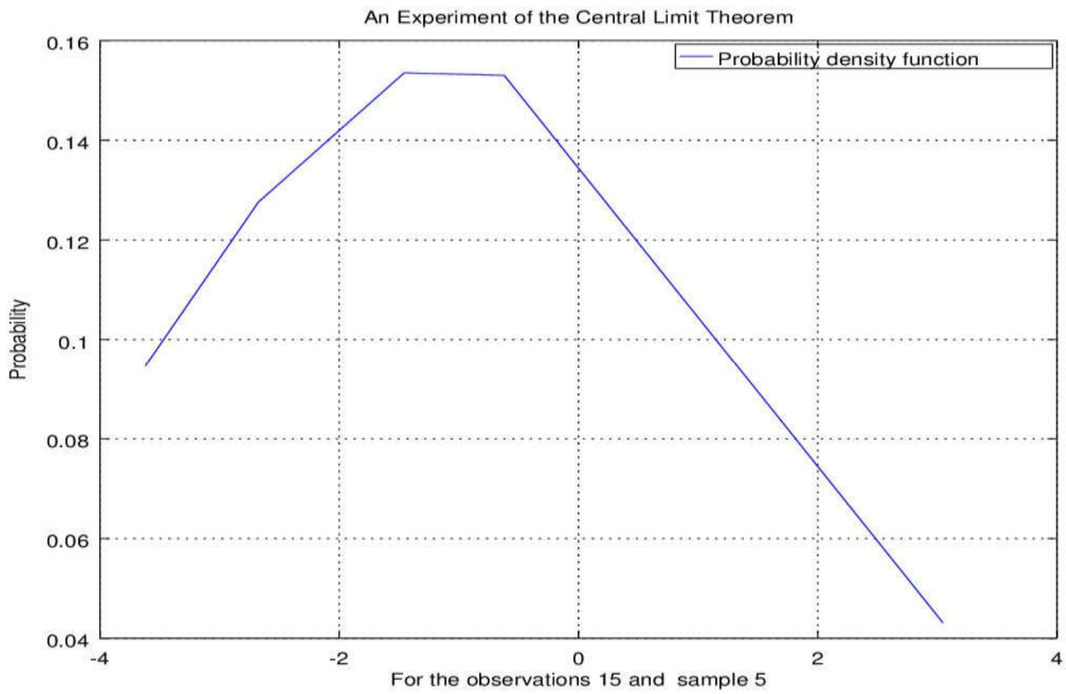
```
% created by : Md. Shahrear Zaman

%(c) Md. Shahrear Zaman

% shahrear.zaman1971@gmail.com

% student.eco86@gmail.com

% date 18/03/2020

% This code will help you to understand that, under central-
%limit theorem(CLT), the sum of a %large number of random
%variables will have an approximately normal distribution

clear all

a= 5;

b= 5;

for i = 1 : a

    for k = 1 : b

              x = randn(1,i,k); % random number generator. We can
                                            %also take the code rand()

            z = sqrt(k)*(sum(x)/k);

    end
```

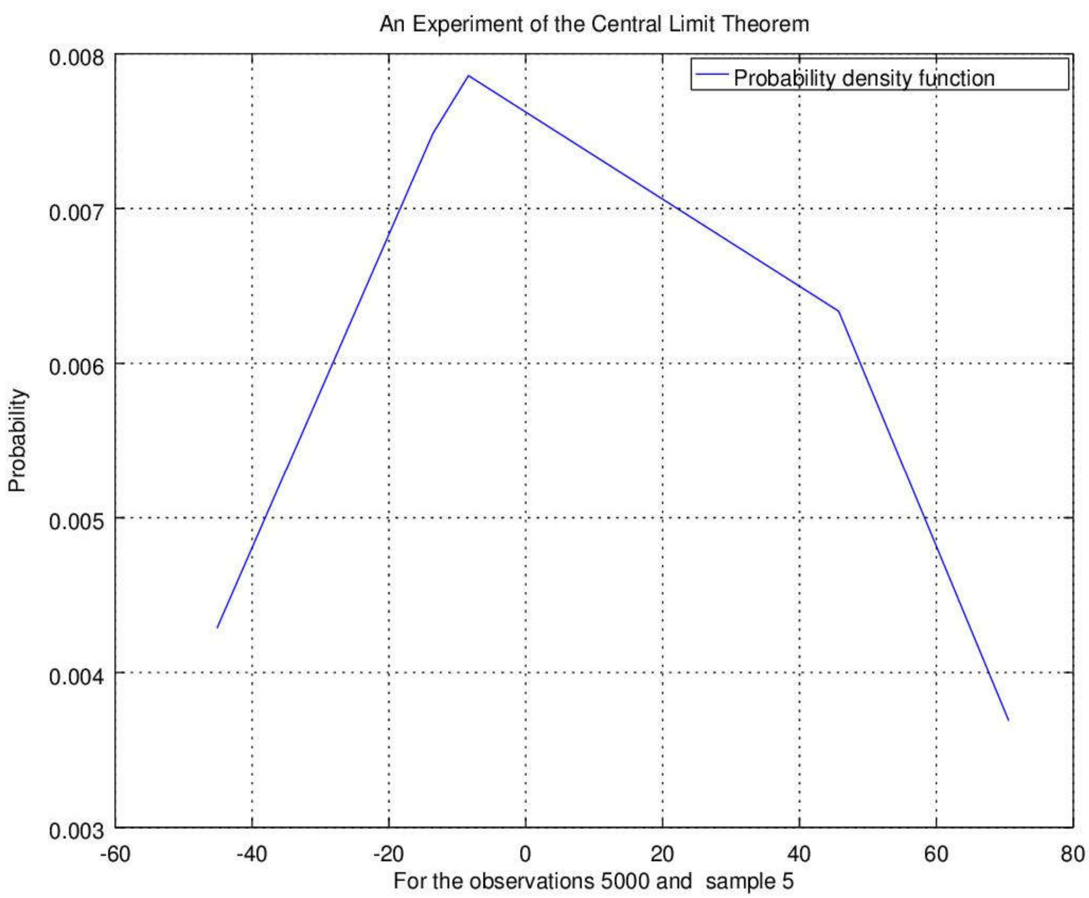
```
end

 [r c1 r1] = size(z);

  N1 = r1;

  mu = sum(z)/N1;

  sigma2 = (sum((z - (sum(z)/N1)).^2)/(N1-1));


zsort = sort(z);

zsort1 = zsort(1:b);


f = ( (1 / sqrt(2*pi*(sigma2))) * exp(-(1/2)*(((zsort1 -
mu).^2)/sigma2)));

figure (); plot (zsort1,f);

grid on;

ylabel('Probability')

legend ({"probability density function"})

title('An Experiment of the Central Limit Theorem');
```

Performace of the Code:






An Experiment of the Central Limit Theorem
Probability density function
Probability
0.16
0.14
0.12
0.1
0.08
0.06
0.04
-4
-2
0
2
4
For the observations 15 and sample 5


An Experiment of the Central Limit Theorem
Probability density function
Probability
0.008
0.007
0.006
0.005
0.004
0.003
-60
-40
-20
0
20
40
60
80
For the observations 5000 and sample 5

According to the eye ball inspection of the above graph, it is better to take 1000 sample with 50 observations for each.

References :

1. Introduction to The Theory of Statistics; Graybill, Franklin A. ; Boes, Duane C.
2. Mathematical Statistics for Economics and Business; Mittelhammer, Ron C.
3. Wikipedia; Normal Distribution

Note: