

# Practice 14: Central Limit Theorem.

Mayra Cristina Berrones Reyes 6291

December 8, 2020

## 1 Introduction

The Central Limit Theorem (CLT) tells us that in many situations, when independent random variables are added with unknown distribution (it could be uniform, binomial or completely random), the sample means will approximate the normal distribution, such as a bell curve.

Often, the CLT is confused with “the law of large numbers”. This law states that as the size of a sample increases, the sample mean will become a more accurate estimate of the population mean. The main difference between the two theorems is that the law of large numbers pertains to a single sample, meanwhile, the CLT pertains to the distribution of sample means [2].

As statistical significance goes, the CLT [1]:

- Helps us analyze data like hypothesis testing and constructing confidence intervals, because these methods assume the population is normally distributed, so we can treat the sampling distribution as normal according to the CLT.
- When we increase the samples from the population, the standard deviation of sample means will decrease. This helps us estimate the population mean much more accurately.

The mean of the sample means is denoted as Equation 1 and the standard deviation of the sample mean is denoted as Equation 2

$$\mu_{\bar{x}} = \mu, \tag{1}$$

where,

- $\mu_{\bar{x}}$  is the mean of the sample means.
- $\mu$  is the population mean.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \tag{2}$$

where,

- $\sigma_{\bar{x}}$  is the standard deviation of the sample mean.
- $\sigma$  is the population standard deviation.
- $n$  is the sample size.

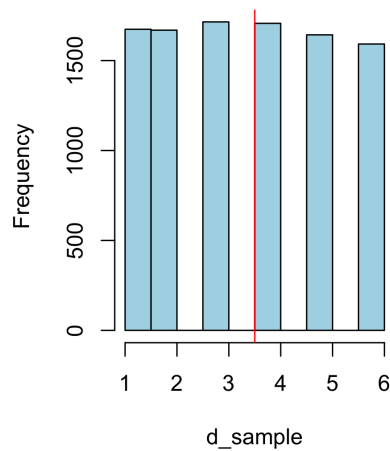


Figure 1: Histogram of the frequency of each outcome.

## 1.1 Example: Die roll

Following similar steps as the practice of Law of Large Numbers (LLN), to further analyze and understand the CLT, we use the example of the dice and with code in R 1.

A fair die can be modeled with a discrete random variable with outcome 1 through 6, each with the equal probability of  $\frac{1}{6}$ . The expected value is  $\frac{1+2+3+4+5+6}{6} = 3.5$ . In this experiment we are going to simulate throwing a fair die 10,000 times where we first have as a result Figure 1.

```

1 png("Ej14_dice.png", width = 1000, height = 1300, res = 300)
2 d_sample <- sample(1:6,10000, replace= TRUE)
3 hist(d_sample, col = "light blue",main="")
4 abline(v=3.5, col = "red",lty=1)
5 dev.off()
6
7 x30 <- c()
8 x100 <- c()
9 x1000 <- c()
10 k =10000
11 for ( i in 1:k){
12   x30[i] = mean(sample(1:6,30, replace = TRUE))
13   x100[i] = mean(sample(1:6,100, replace = TRUE))
14   x1000[i] = mean(sample(1:6,1000, replace = TRUE))
15 }
16
17 png("Ej14_dice1.png", width = 1000, height = 1300, res = 300)
18 hist(x30, col = "light blue",main="",xlab = "die roll")
19 abline(v = mean(x30), col = "blue")
20 dev.off()
21
22 png("Ej14_dice2.png", width = 1000, height = 1300, res = 300)
23 hist(x100, col = "pink", main="",xlab = "die roll")

```

```

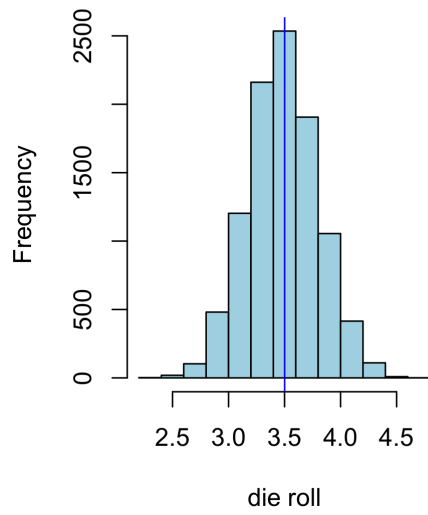
24 abline(v = mean(x100), col = "red")
25 dev.off()
26
27 png("Ej14_dice3.png", width = 1000, height = 1300, res = 300)
28 hist(x1000, col = "orange", main = "", xlab = "die roll")
29 abline(v = mean(x1000), col = "red")
30 dev.off()

```

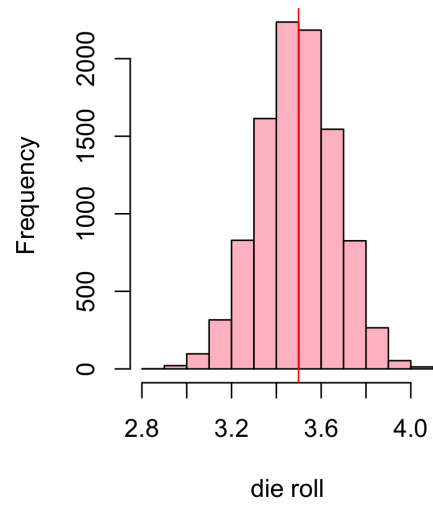
Listing 1: R extract of the code used to perform the dice experiment

We will take samples of size 10, from the above 10,000 observations of the outcome of die roll. From this, we will take the arithmetic mean and try to plot the mean of sample. This procedure will be reproduced k times (in this case k= 10,000).

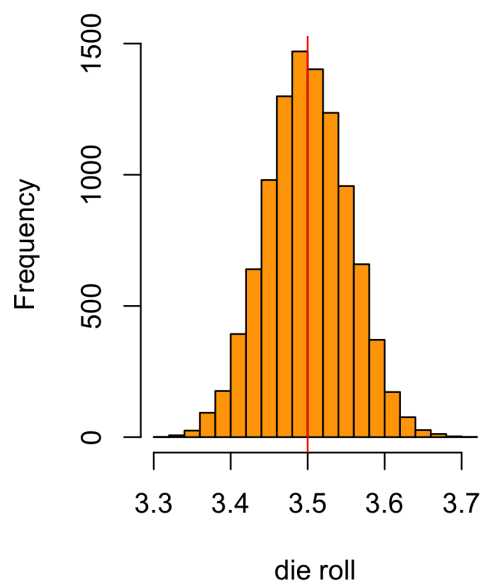
By theory , we know as the sample increases, we get better bell shaped curve. As the n approaches infinity , we get a normal distribution. We will achieve this by increasing the sample size to 30, 100 and 1,000, and the results can be seen in Figure 2.



(a)  $n = 30$



(b)  $n = 100$



(c)  $n = 1,000$

Figure 2: Different parameters for the dice experiment performed in R.

Table 1: Calculation of the standard error of the dice example.

Sample size		
10	100	1,000
$\sigma_{\bar{x}} = \frac{1.71}{\sqrt{10}} = 0.54$	$\sigma_{\bar{x}} = \frac{1.71}{\sqrt{100}} = 0.17$	$\sigma_{\bar{x}} = \frac{1.71}{\sqrt{1000}} = 0.05$

## 2 Applications of Central Limit Theorem in Machine Learning

The CLT has important implications in applied machine learning. The theorem does inform the solution to linear algorithms such as linear regression, but not exotic methods like artificial neural networks that are solved using numerical optimization methods. Instead for these type of methods, we must use experiments to observe and record the behavior of the algorithms and use statistical methods to interpret their results.

For the significance test, we can make comparisons of the performance of one model against another model. In this case, we can use the tools of statistical significance to estimate the likelihood that the two samples of model skill scores were taken from the same or a different distribution of model scores. If it looks like the samples were drawn from the same population, then we assume there is no difference between the models, and any differences are due to statistical noise.

Another example is when we have trained a final model in machine learning, we want to make a guess about how good the model is expected to be in practice. This can be called a confidence interval. We can develop multiple independent evaluations of a model accuracy to result in a population of model accuracy estimates. The mean of these estimates will is regarded as the error of the true estimate of the model accuracy on the problem.

To illustrate this, we can actually take the example of the dice, and verify the standard error on each of the experiments seen on Figure 2.

For that example we have the variance of rolling a fair die equal to 2.92 and the standard deviation to 1.71, so we have the results in Table 1. In that table we can see, as the sample increases, the error gets smaller and smaller.

## 3 Conclusion

In this case with the CLT definition, a part of the process of training a Neural Network comes to mind. When training the models used for our investigation of the thesis, we focused on comparing the efficiency and accuracy of the models. In the training process, there is a sort of monitoring of the results and development of the model. Similar to what we did here, we chose a set number of iterations in which our models are to be trained. This iterations are also divided in epochs, that are a form of larger way to see the iterations and the slow process of improving the accuracy of your model.

The way it improves is dependent of the loss function you choose. In our case, we chose the

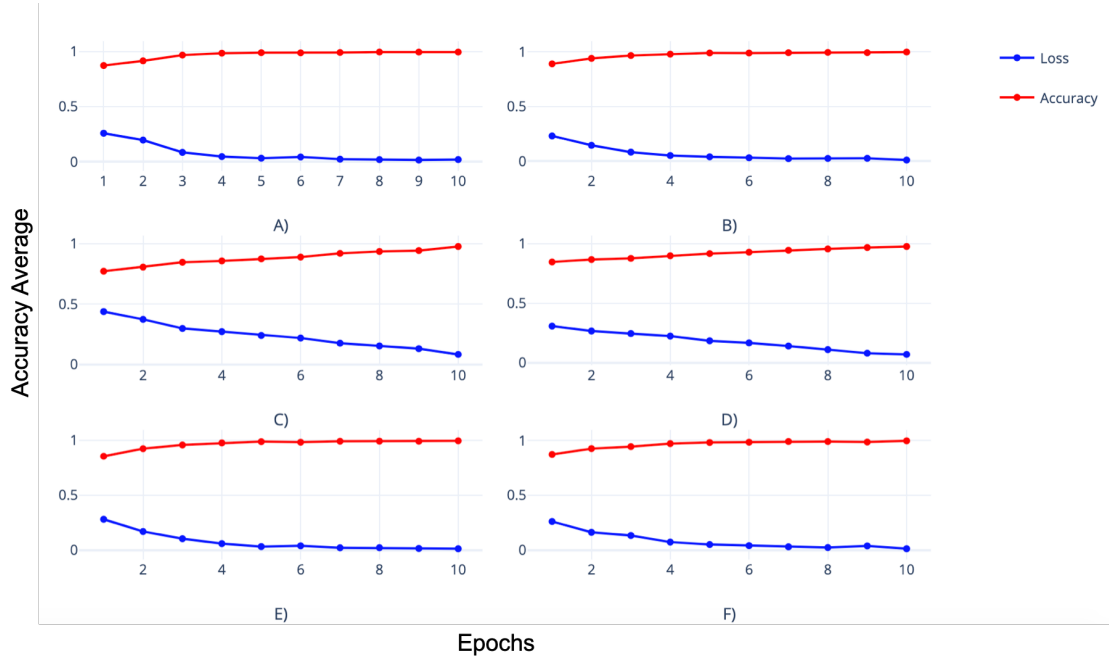


Figure 3: Loss and accuracy of the best models.

binary cross-entropy (specifically for the type of data we had) and the main goal is to reduce the loss function as close to zero as we can, and have the accuracy as close to 1. This is the part of the process that most resonated while reading for examples, because of the way we look for one part of the process to be as close as zero and the other to converge to 1 (this part more so in the Law of Large Numbers).

In Figure 3 we have a plot of the loss and accuracy of the best models trained for the experimentation of our thesis. As we can see, in the best models, the accuracy is close to one in the last epoch, and the loss is closest to zero.

In conclusion, understanding the central limit theorem is important when it comes to trusting the validity of the results of the models we train and assessing the precision of the estimated accuracy it gives.

## References

- [1] Statistics for data science: Introduction to the central limit theorem. <https://www.analyticsvidhya.com/blog/2019/05/statistics-101-introduction-central-limit-theorem/>. Accessed: 2020-12-08.
- [2] Why the central limit theorem in data science. <https://towardsdatascience.com/why-the-central-limit-theorem-in-data-science-be8997b95f3>. Accessed: 2020-12-08.