5 week Law of large numbers and central limit theorem

Learning Objectives:

5.1 Know statement of law of large numbers.
5.2 Know statement of central limit theorem.
5.3 Know properties of normal distribution.
5.4 Know and apply Chebyshev's inequality

Оглавление:

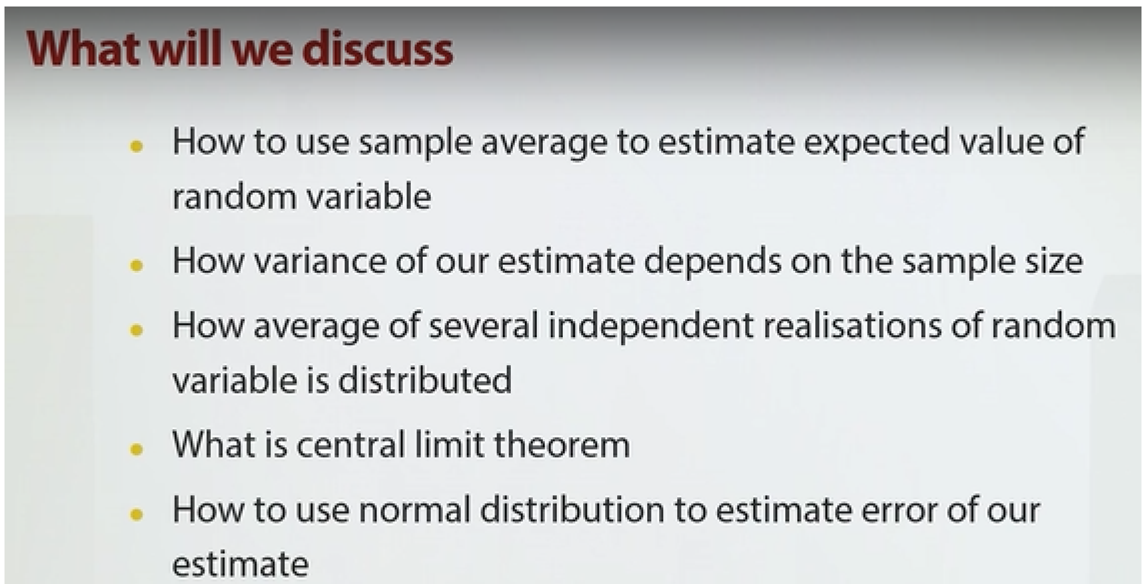
Introduction to the 5th week

In this week, we will study law of large numbers and central limit theorem. These are crucial facts and they play a very important role in probability theory and statistics. How can we use them?
Assume that we want to use data to answer some questions about the real world. For example, we have Africa. And we want to know what is the average age or average income of people in Africa? To answer this question we can conduct a survey. We can pick several persons from Africa and ask them questions about their age and income.We get a table:

person id	age	income
AA1	23	12,34
AB2	45	34,53
ZX3	37	23,45
Average	35	23,44

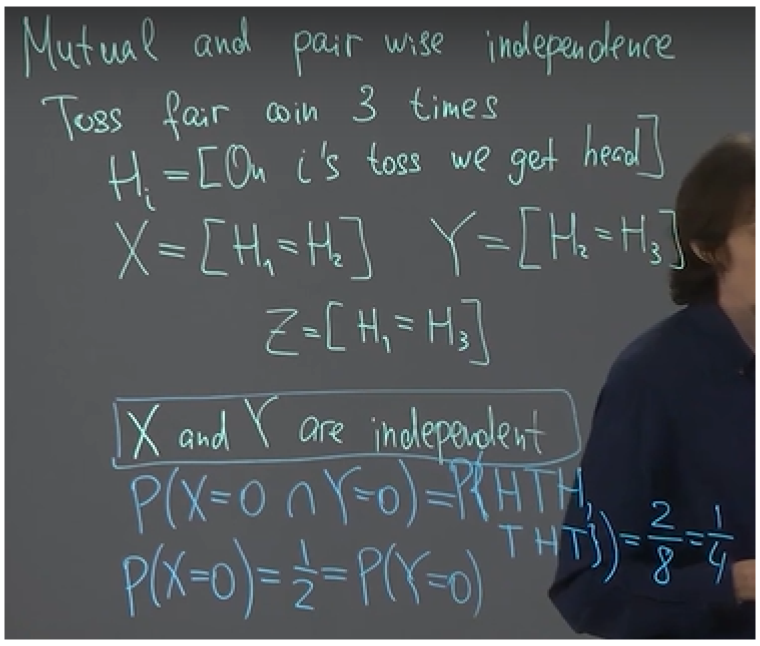
In this case, we have only 3 rows in this table, meaning that we surveyed only 3 persons. But, of course, in actual research we have hundreds or thousands, or even millions of rows in our table. Anyway, now we want to use this data to answer the question. To do so, we'll find so-called sample average. It means that we just find average of numbers that in our table: average age an daverage income. Then we will use these averages as estimates of the corresponding quantites in real world. However, we have a question.
How accurate our estimates of average?
Indeed, if we have a study with only 3 participants, then if we perform another study, we will get different numbers as our estimates. It means that these estimates cannot be too accurate. **How many people should we include into our study to make these estimates accurate? How can we estimate the inaccuracy of our estimates?** This is what statistics do based on law of large numbers and central

limit theorem.

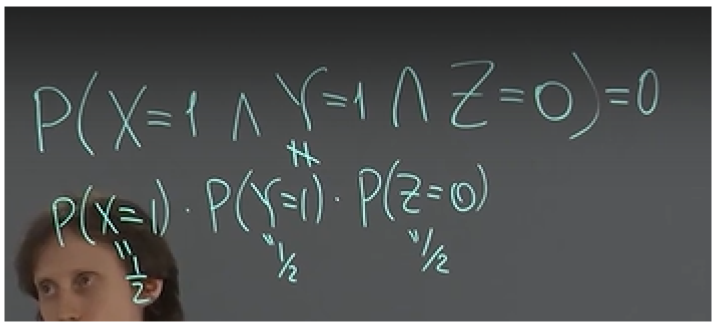


Mutual and pairwise independence

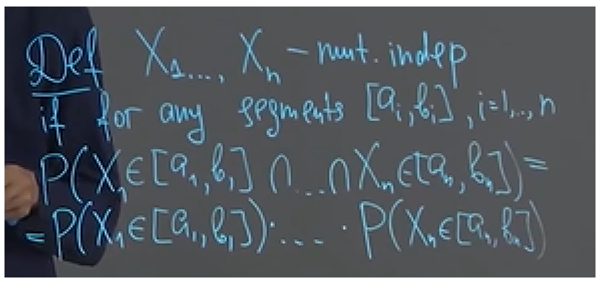
Before we start discussing law of large numbers and central limit theorem, we have to refine(уточнить) our notion of independence for random varibles. In case of 2 r.v. things are simple. But if you have several r.v.(s), we have to introduce a new notion, mutual independence. This is in close relation to the notion of mutual independence of events that we discussed previously.
Example:



Let us consider all 3 variables together.



In this case, we say that these 3 variables are **pairwise independent, but not mutually independent.**

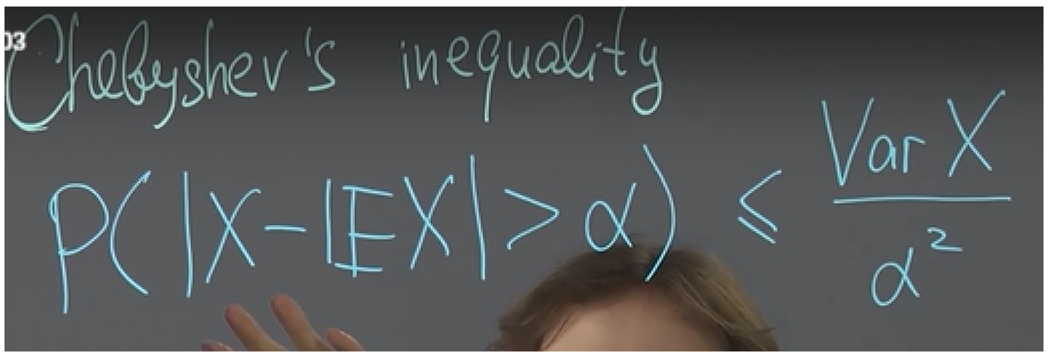


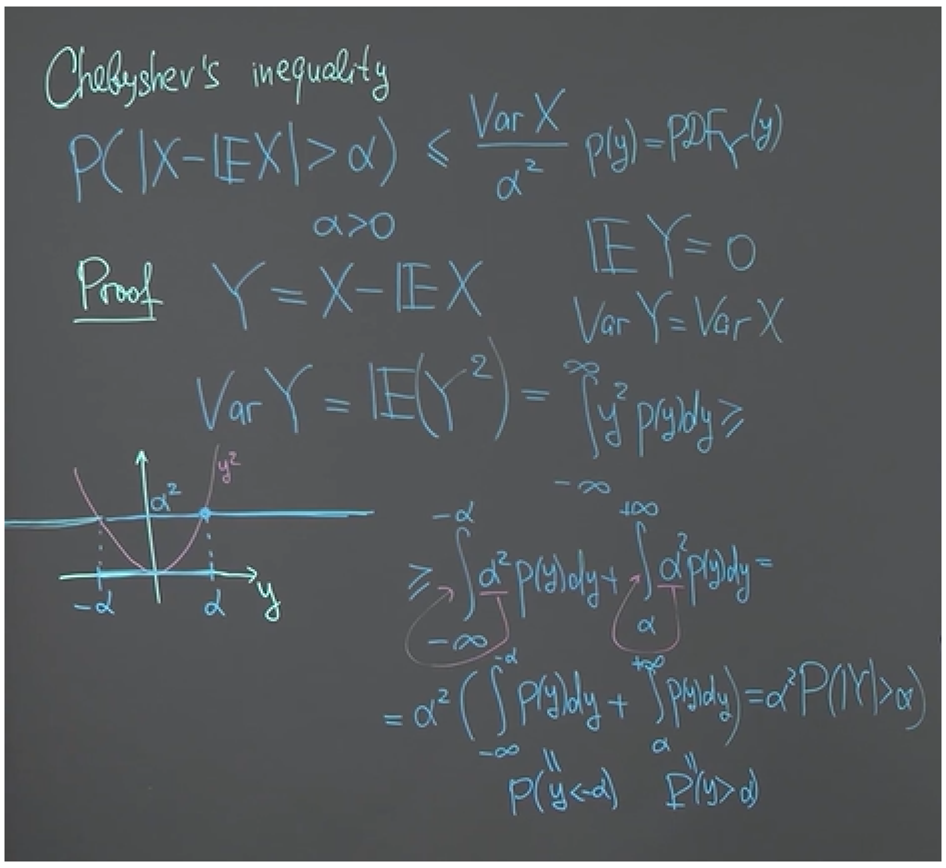
If these variables have pdf, a similar relation with facorization should hold for joint pdf for them to be mutual independent. Later, we will consider sequences of events which are mutually independent.

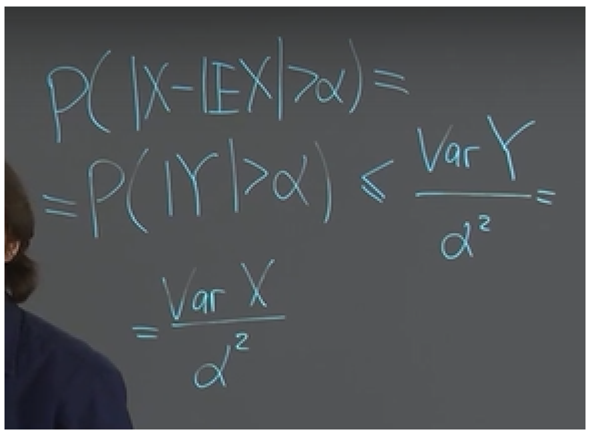
Chebyshev's inequality

Our nearest goal is to state and prove law of large numbers. But in the prof, we will need a Lemma which is called Chebyshev's inequality. Let's discuss this Lemma now:

Chebyshev's inequality ask: What is the probability that the value of random variable is too far from its expected value? Of course, this should be related to the variance, but Chebyshev's inequality gives us some explicit estimate.







This inequality has real consequences. For example, if we pick alpha^2 to be equal to twice VarianceX, then we see that the probability for r.v. to deviate from its expected value for square root of its double Variance is less than 1/2. We will use this arguments in the prof of law of large numbers.

Chebyshev's inequality practice

Question 1

Let X be a random variable with $Var X = 1.2$. We take random samples from X and average them. What's the minimum number of samples that should be taken to assure that the average deviates from $\mathbb{E}X$ by not more than 1.8 with probability at least 0.95?

Solution:

$$P(|X - \mathbb{E}X| > d) \leq \frac{Var X}{n \cdot d^2}$$

$$0.95 \leq \frac{1.2}{n \cdot d^2} \Rightarrow n \leq \frac{1.2}{0.95 \cdot 1.8^2} \leq 7.4 \qquad n \leq 7.4$$

$$n = 7$$

The correct answer is: 8

Chebyshev's inequality skill test

Question 1

Consider a sequence of 1000 independent Bernoulli trials with probability of success being equal to 0.3. Use Chebyshev's inequality to estimate probability of the event "deviation of the number of successful trials from its average value is less than 30". Choose the sharpest estimate that follows from Chebyshev's inequality

- more than 0.75
- more than 0.77
- more than 0.74
- more than 0.76 ✔
- more than 0.78

Solution:

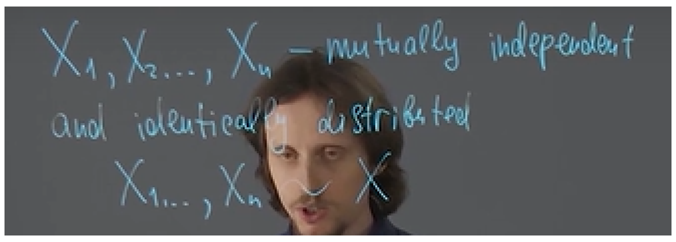
- **n:** Number of trials = 1000
- p: Probability of success = 0.3
- X: Random variable representing the number of successes
- μ: Expected value of X = n*p* = *1000* 0.3 = 300
- σ: Standard deviation of X = sqrt(n*p*(1-p)) = sqrt(1000 *0.3* 0.7) ≈ 14.49
  Chebyshev's inequality states:
  $P(|X - \mu| \geq k\sigma) \leq 1/k^2$

We want to find $P(|X - \mu| < 30)$. To use Chebyshev's inequality, we need to find 'k':
$k = 30/\sigma \approx 30/14.49 \approx 2.07$
Now we can apply Chebyshev's inequality:
$P(|X - \mu| < 30) = 1 - P(|X - \mu| \geq 30) \geq 1 - 1/k^2 \approx 1 - 1/(2.07)^2 \approx 0.76$

The correct answer is: more than 0.76

Law of large numbers (llm)

Llm allow us to estimate some quantities, for example, expected calue of a random variable by looking at a sample from this random variable.
Let us consider a sequence of independent and identically distributed random variables.

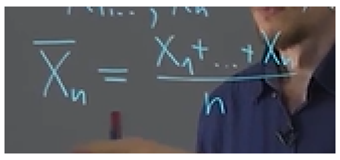


Identically distributed means that they all have the same distribution.
X - distribution
X1, ..., Xn - copies from X. They don't have the same value but they have the same distribution, the same pdf and so on. We consider these random variables(X1, ..., Xn) as a sample from random variable(X). I mean, for example, that we asked some question to ten randomly selected persons and their answers are encoded into this random variables. We can think about this selection process as sampling from some random variable. After that, we can think that we have not one random variable, but we ten random variables. And our experiment is sampling of all ten random variables together.

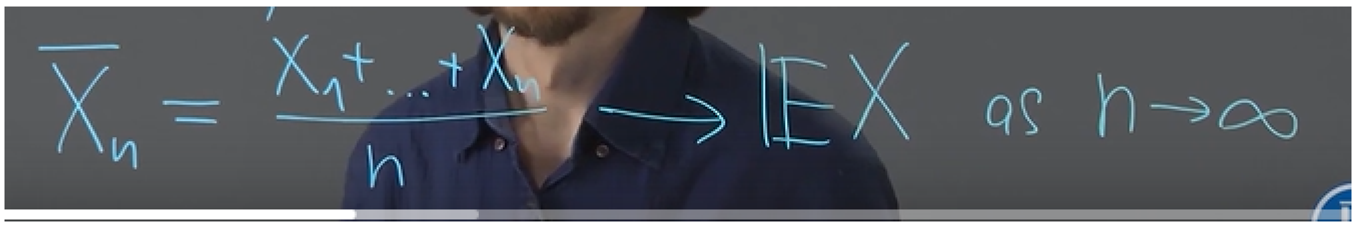
Additional material(wikipedia):
In probability theory and statistics, a collection of random variables is **independent and identically distributed** if each random variable has the same [probability distribution] as the others and all are mutually [independent]. This property is usually abbreviated as ***i.i.d.***, ***iid***, or ***IID***. In other words, the terms *random sample* and ***IID*** are synonymous. In statistics, "*random sample*" is the typical terminology, but in probability, it is more common to say "**IID**."

- **Identically distributed** means that there are no overall trends — the distribution does not fluctuate and all items in the sample are taken from the same probability distribution.
- **Independent** means that the sample items are all independent events. In other words, they are not connected to each other in any way; knowledge of the value of one variable gives no information about the value of the other and vice versa.

Then we're interested in the possibility of averaging of this result. What we will get if we consider average of these numbers(X1, ..., Xn)?

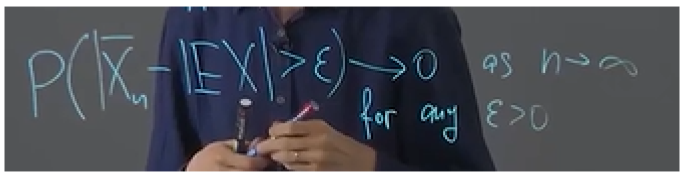


This is average. This thing is also random variable because it's a combination of random variables. Informally speaking, law of large number says that (Xn with -)

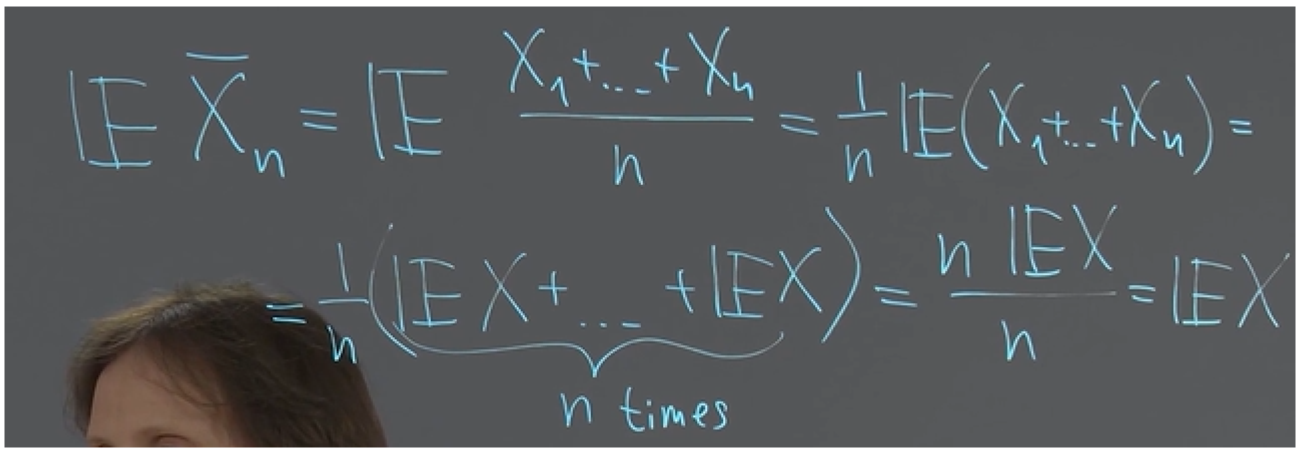


$$\overline{X}_n \to \mu \quad \text{as } n \to \infty.$$

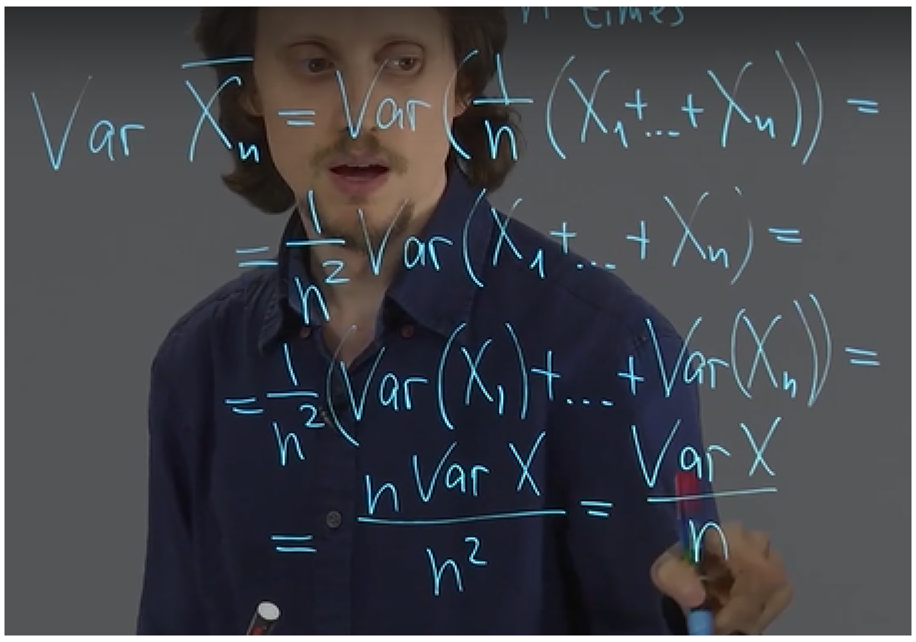
However, (Xn with -) is random varible and EX is just a number. We have to define more precisely what this limit means. And we can do it in several possible ways.



What does it mean? It means that we can use (Xn with -) as an approximation for expectes value for X with any given margin epsilon, if only we can make n large enough.

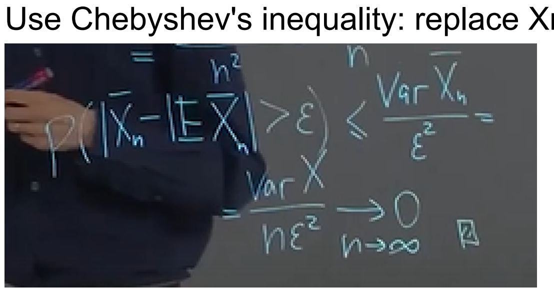


X1,..., Xn - are Identically distributed(similar)



We see that Variance of Xn decreases as n increases.

Use Chebyshev's inequality: replace Xn to Vn with -.



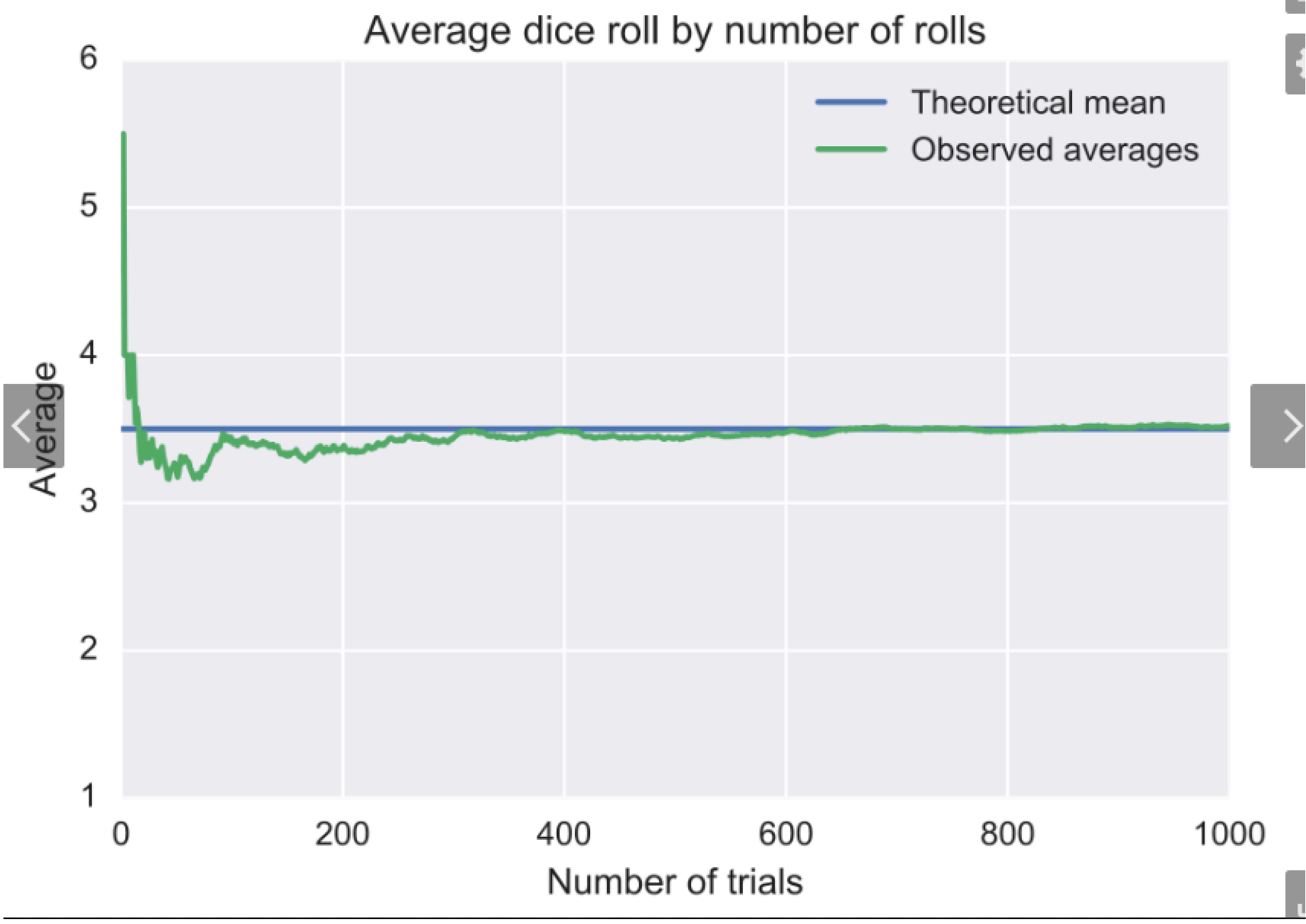
Law of large numbers says that we can use average over identically distributed random variables to estimate the expected value. However, to make our estimates exact, we have to understand how this value, this average is distributed. To answer this quiestion we ahve to use a new theorem which is called central limit theorem.

Additional material(wikipedia):

For example, a single roll of a fair, six-sided die produces one of the numbers 1, 2, 3, 4, 5, or 6, each with equal probability. Therefore, the expected value of the average of the rolls is:

(1 + 2 + 3 + 4 + 5 + 6) / 6 = 3.5

According to the law of large numbers, if a large number of six-sided dice are rolled, the average of their values (sometimes called the sample mean) will approach 3.5, with the precision increasing as more dice are rolled.



An [illustration](#) of the law of large numbers using a particular run of rolls of a single [die](#). As the number of rolls in this run increases, the average of the values of all the results approaches 3.5.

Limit of binomial distribution

Question 1

Let X_n be binomially distributed random variable with number of trials n and probability of success p. Consider random variable $Y_n = X_n/n$. What can you say about limit of Y_n as n tends to infinity?

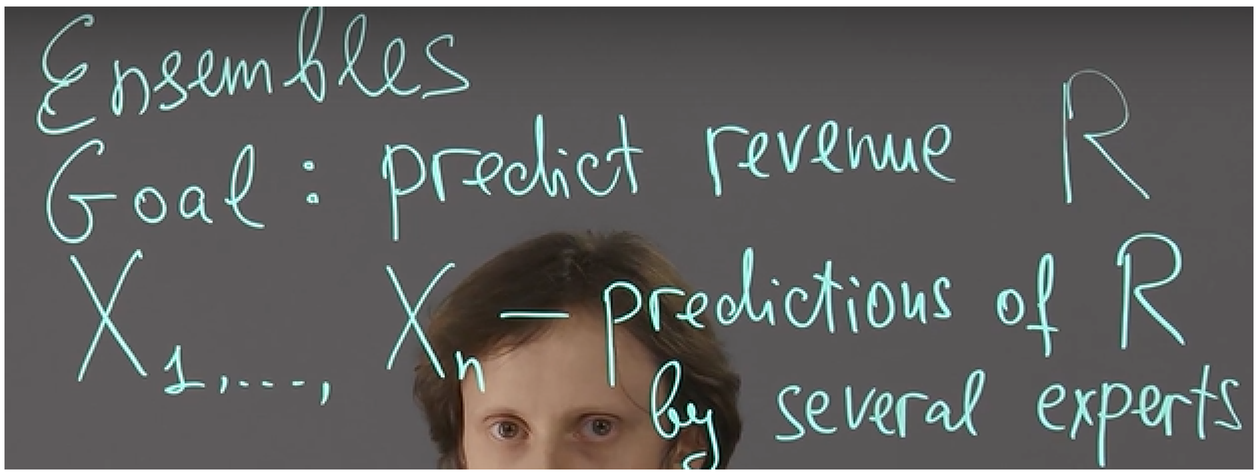
- Y_n does not have a limit
- Y_n tends to p ✔
- Y_n tends to infinity
- Y_n tends to 0
- It does not make sense to ask about limit of Y_n

Ваш ответ верный.

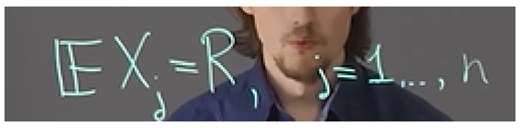
The correct answer is:
Y_n tends to p

Law of large numbers and ensemble models

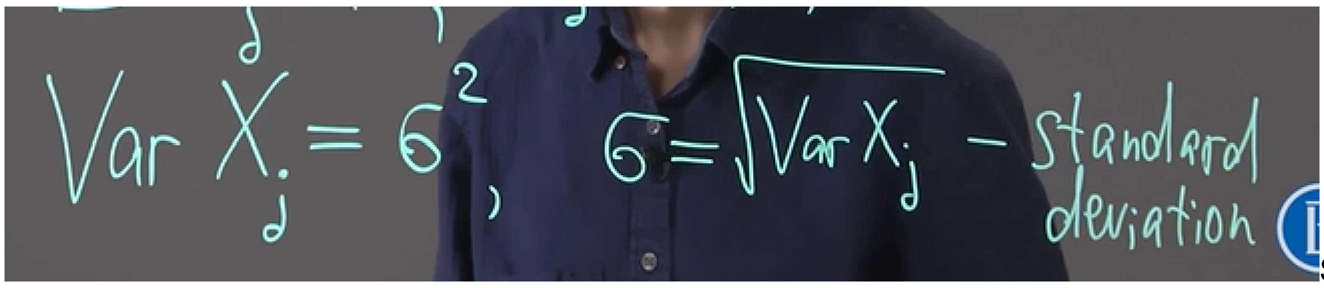
Let us discuss how we can apply ideas behinds of law od large numbers for measuring problems. For example, let us assume that we have a company that owns some shops and wants to predict the revenue of a particular shop. To do so, this company can hire an expert and ask this expert: "What will be the revenue of this shop?". However, this company can also use maching learning tools and to make a prediction model that will predict the value of revenue. In both cases, it's possible that our predictions can have some errors due to probabilistic nature of the problem. However, we can use some techniques to reduxe this error.This techniques can be applied both for human experts and to measure learning modals in more or less the same way.
Let us discuss first the human experts because it's easier to think about them. Let us assume that we have not one, but several experts. As I said, the predictions of these experts can have some errors. So we will model this prediction as random numbers.



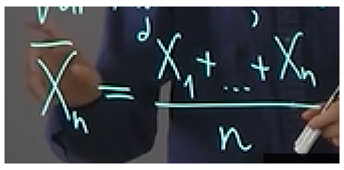
We assume that experts more or less similar to each other and on average they're good. So we assume that Expected value of each X equals to the true value that we want to predict.



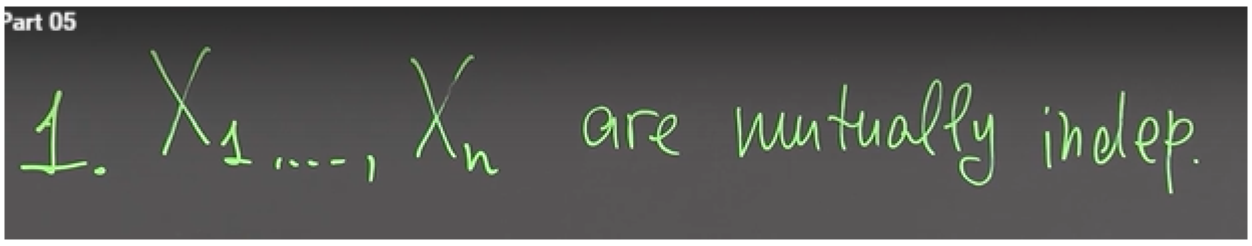
Now, the quality of this prediction is given by its variance. The larger possible deviations of prediction to the actual value => the worse prediction. So learge variance are bad and low variance are good.
Let us assume that our experts have the same variance of the predictions.



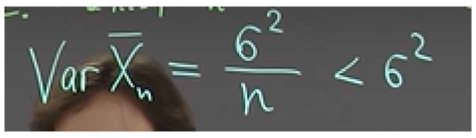
So let's discuss **how can we reduce the variance of the prediction.** Let us make a committee out of this experts and use the decision of this committee(комиссия), the prediction of this committee as the prediction that we will use later. It means that we can introduce:



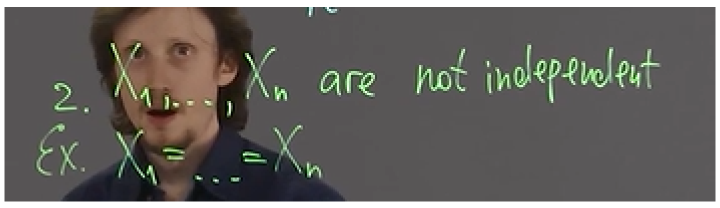
Is it true that this prediction of committee is better than prediction of individual experts? To answer this question, we have to analyze variance of this random variable. And here we have two cases:



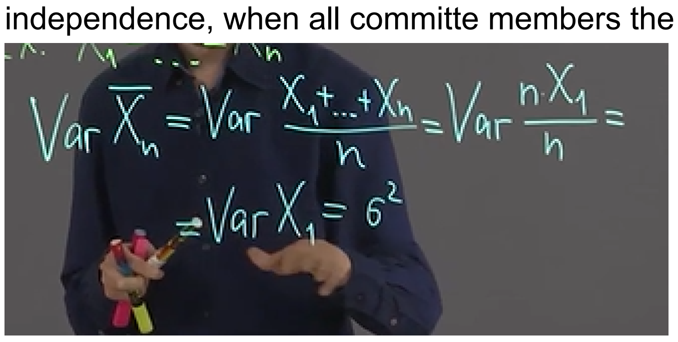
It means that oer comitee memebrs work independently to each other and their predictions don't affect on each other's predictions. We use theorem of large numbers:

1. 

   As we discussed in this case, Variance of average be found as variance of individual expert predictions devided by n. For any positive n greater that 1, this number is less than variance. It means that the prediction of the committee, which is calculated as an average of predictions of each members, has a lower variance than predictions of individual member.

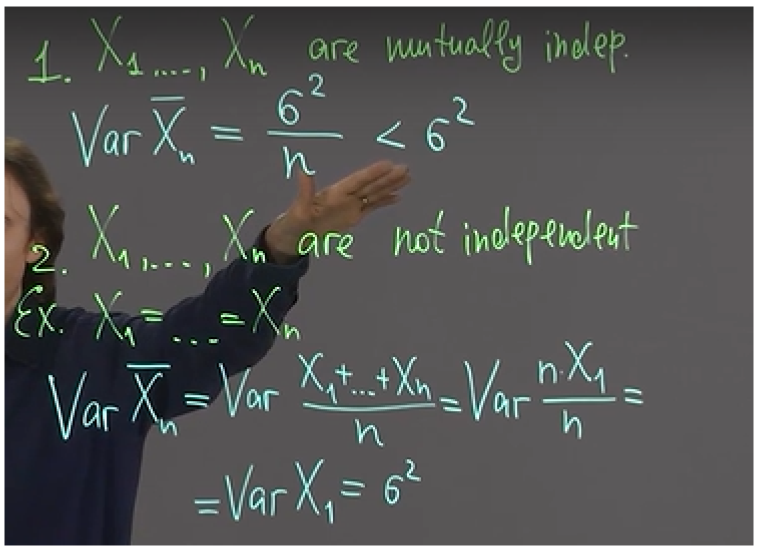
2. 

   On the other hand, it's possible that committee members don't work independently with each other. In this case, this condition(X1, .., Xn are mutually independent) is violated and the reasoning that allowed us to prove this formula doesn't work.
   So let's us assume that all committe members give the same prediction. For example, some famous member of the committe and all other members of the committe just do the same prediction as this famous member. This is in a sense an extreme case of non-independence, when all committe members the sam epredictions.

   Int his extreme case we see that variance of the predictions of the committee the is the same as variance of the prediction of its member. It means that it's not very useful to make a committe here. We just get ome member of the committe, one expert and ask this expert.

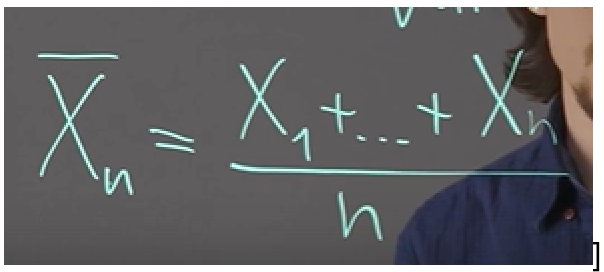
A very similar situation occurs in ML problems. We can replace these human experts by different ML models. If these model work in a sense, independently of each other. We can be close to the 1st case when considering a committe or as it's called in ML ensamble of models. We can reduce variance of the prediction and make the prediction better as 1st case.



But if the predictions of several model are not independent and if there are correlated to each other, then this variance will be not much or lrss than the variance of individual model. In this case, it doesn't make sense to consider ensemble of models, we can just use one model. When we discuss ensemble models like random forest, we will discuss how can we do behave different models behave independently and give independent prediction.

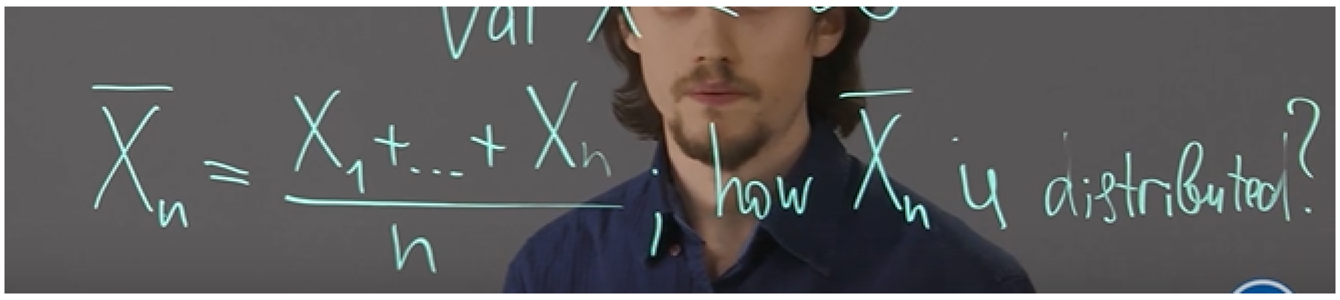
Approaching central limit theorem

This theorem is a crucial fact of probability theory, it gives us information about the distribution of average of several random variables that are independent of each other. And this theorem allows us to make precise estimates that we can use in our statisticak investigations.
i.d.d - independent and identically distributed
Xj ~ X - have the common distribution

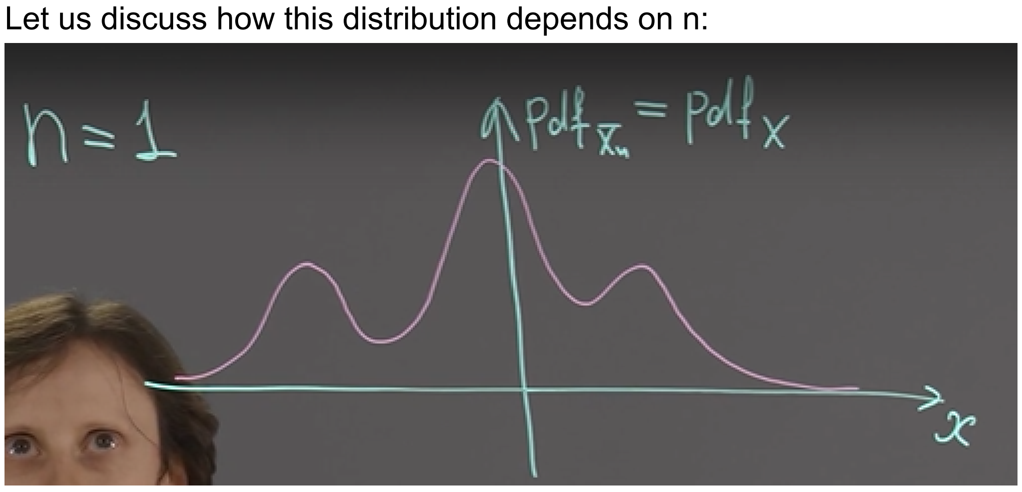


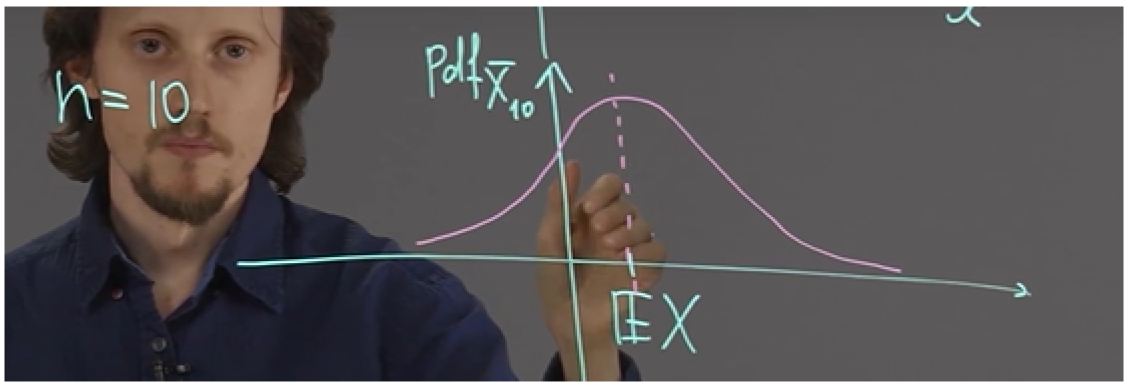
]

And the question that we are interested in: "What is the distribution of Xn with -"

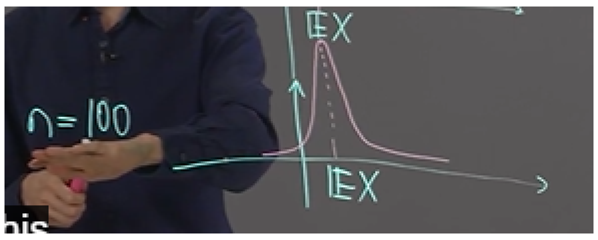


Let us discuss how this distribution depends on n:





The distribution of average have most of their probability somewhere near the expected value of X due to law of large numbers. And the actual shape of the graph of the pdf becomes closer to this bell curve. Now if we increse n further, this bell curve becomes more narrow because the variances will be decreased as we discussed in prof of law of large numbers. For example: n*10 = 100 => standard deviation will be decreased, so pdf will be* 1/sqrt(100).

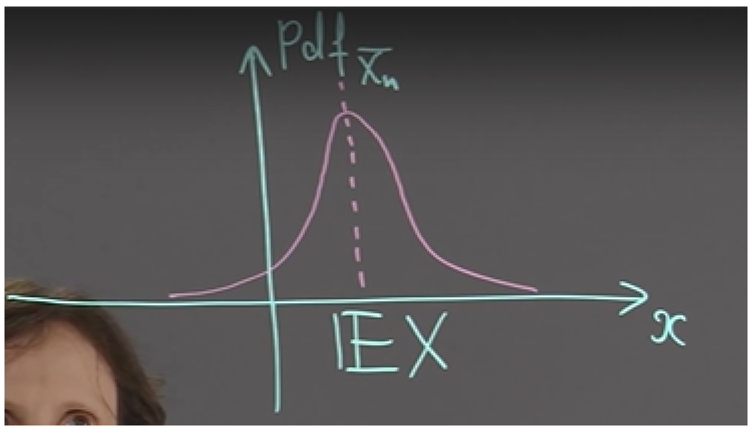


The curve becomes more narrow. If we increase n further this shape will be more an more narrow and it will be more and more concentrated near the EX. It works like we duscussed work transformations with pdf.

So, **what central limit theorem says that there is, in a sense, limit shape which is very close to the pic with n=10 for this distribution.** But to uncover this limit shape we have to deal with the fact the actual limit tends to smth like **very narrow bell curve near EX.** To overcome this issue we have to introduce a new ranfom variable which is called z-value or z-score.

Z-score and statement of central limit theorem

To state CLT we have to introduce the so-calles z-score. Z-score - transform version of the average (Xn with -). First, let us recall how this average can be distributed.



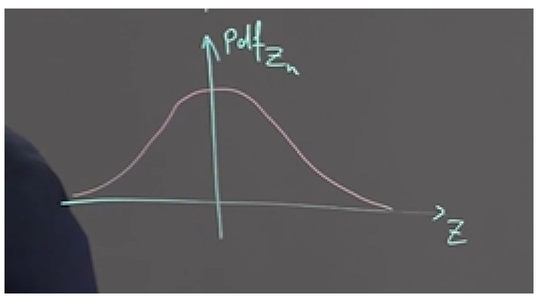
This distribution is concentrated near the EX and its variance becomes lower than when n increases (n -> + infinity). What do we want to do is to introduce a transformation of pdf in such way that this transformed pdf have some limit as n -> 0. To do this transformation, we ahve to do next 2 things:

1. We have to shift the graph on the pic above in such way that it will be cocentrated not near EX, but near zero.
2. We have to stretch it in a way that gives us some limiting variance of our distribution.
   So, when we increase n, variance of distribution tends to 0. But we want a new distribution that is related to this one . This relation allow us to make the variance of this new distribution on this new random variable not to tend to zero, but to have some limit.

To do so, we introduce a new random variable:

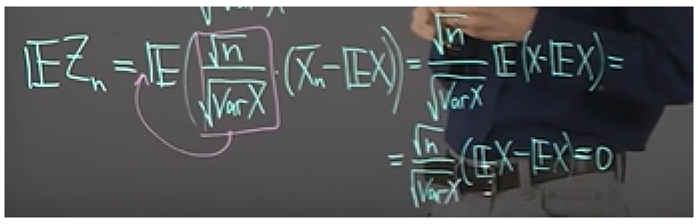
$$Z_n = \frac{\bar{X}_n - \mathbb{E}X}{\sqrt{\operatorname{Var} X}} \cdot \sqrt{n}$$

If we draw distribution of this new variable, we get a distribution which is closed to the standard normal distribution.

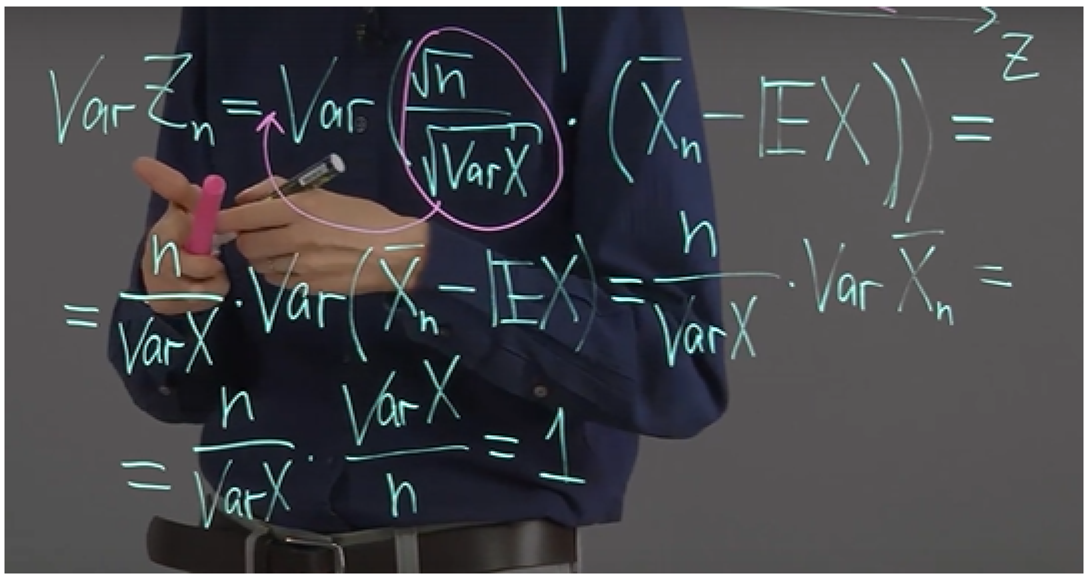


So how Z-score works?

1. Numerator: shift graph (EX -> 0)
2. Multiply by constant which depends on varianceX and on n. This allows us to stendardize VarianceZ in such a way that this varianceZ = 1

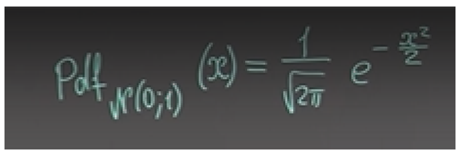


X - EX = 0 as we shifted EX to 0.

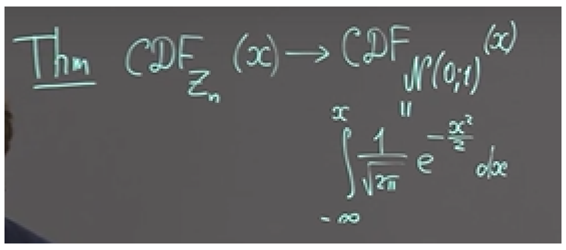


X - EX doesn't change Variance it simple move to 0.

To state the central limit theorem, let us recall on **pdf of standard normal distribution:**

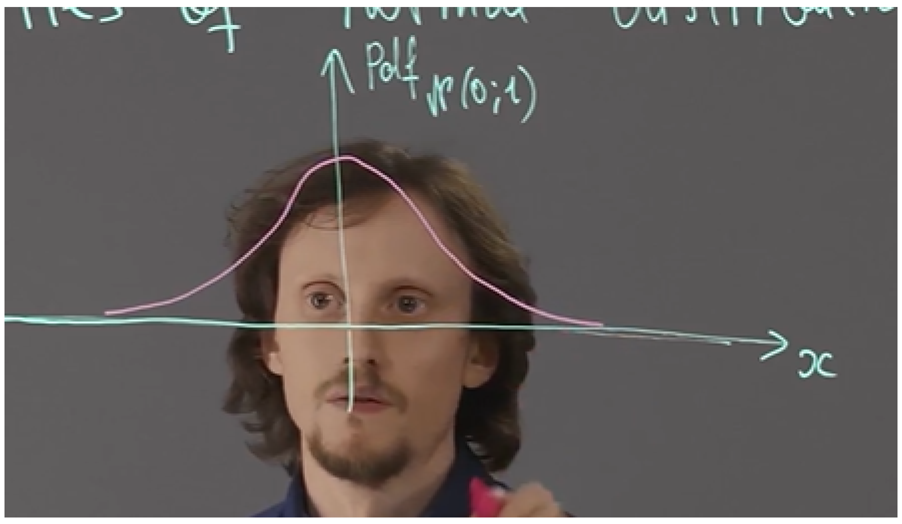


The sequence of our z-scores tends to standard normal distribution.
When we discussed law of large numbers, we said that some sequences of random variables tends to some one non-random number (EX). But now we want say that sequence of random variables tends to some other random variable with standard normal distribution.
CLT says that distribution of Z-score tends to standard normal distribution:

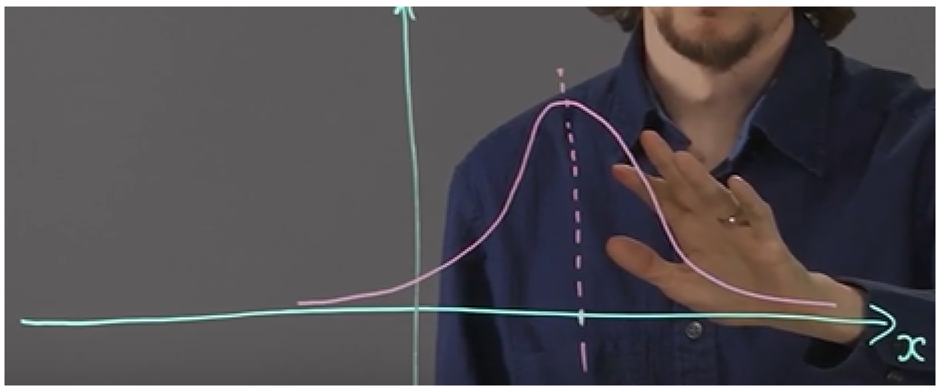


Introducing normal distribution

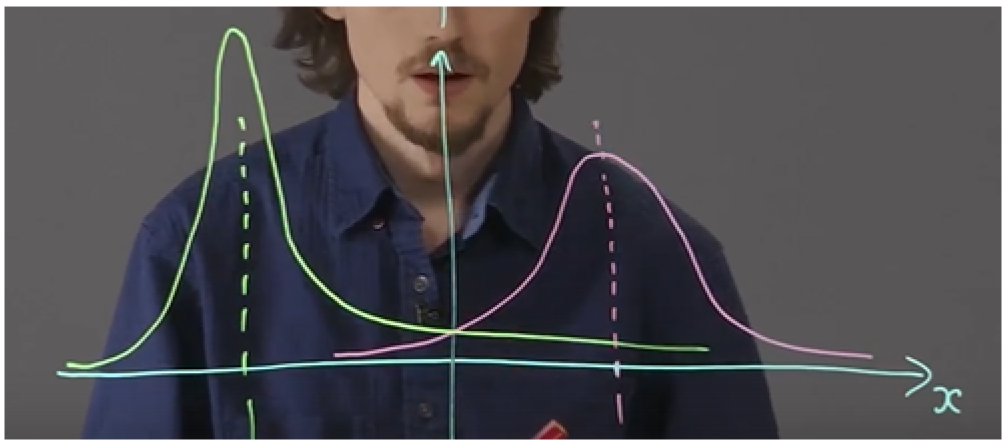
CLT says that distribution of Z-score tends to standard normal distribution. At the same time Z-score is just a linear transformation of the average (X with -). This allows us to find distribution, which approximates the distribution of (X with -) which we're interested in. To do this let us discuss properties of normal distribution beyond standard normal distribution.



Now, if we shift this pdf to the left or to the right or if we rescale the corresponding r.v. and so expand or contract the pdf we'll get different distributions which also called normal distributions, but not standard normal distribution. Some examples of normal distributions:

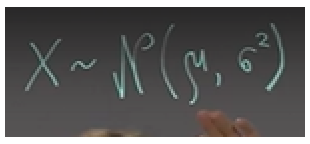


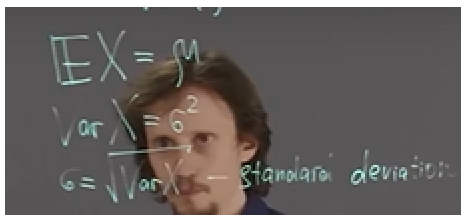
And we can do another linear transformation and get another variant of normal distribution, which is not normal.



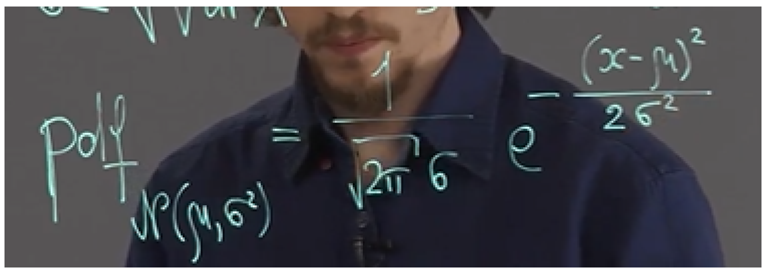
These two distributions are different: different variance and different EX. But they both can be considered as graphs of normal distributions. So, to specify a normal distribution, we have to specify two parameters:

1. EX = mu
2. Variance = standard deviation^2

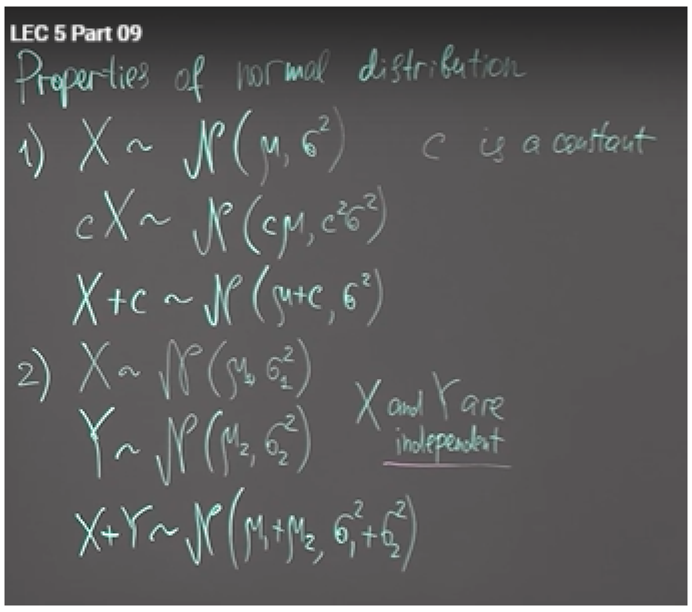
   

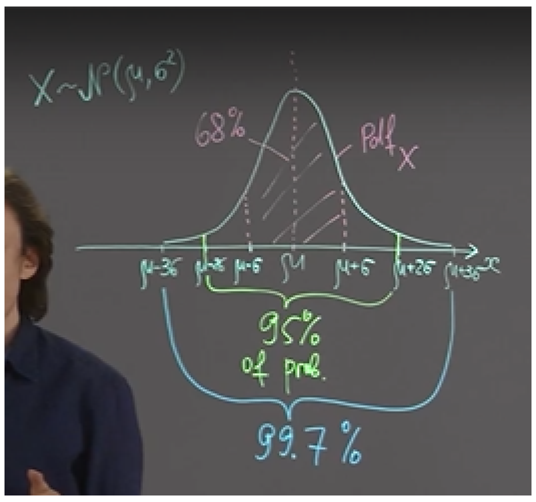
   There exists only one normal distribution with the defined EX and VarX

Properties of normal distribution

Due to CLT many random variables can be approximated by normal distributions. Of course, not every random variable is normally distributed, but there are lot of situations wher ewe can assume that our random variable can be considered as a kind of average of some other independent random variables. Like, for example, it's possible that the variable that we are interested in is a kind of sum of some random factors that work independetly of each other. In this case, it's natural to assume that our random variable approximately is normally distributed. Fortunately, normal distribution have very good mathematical properties:



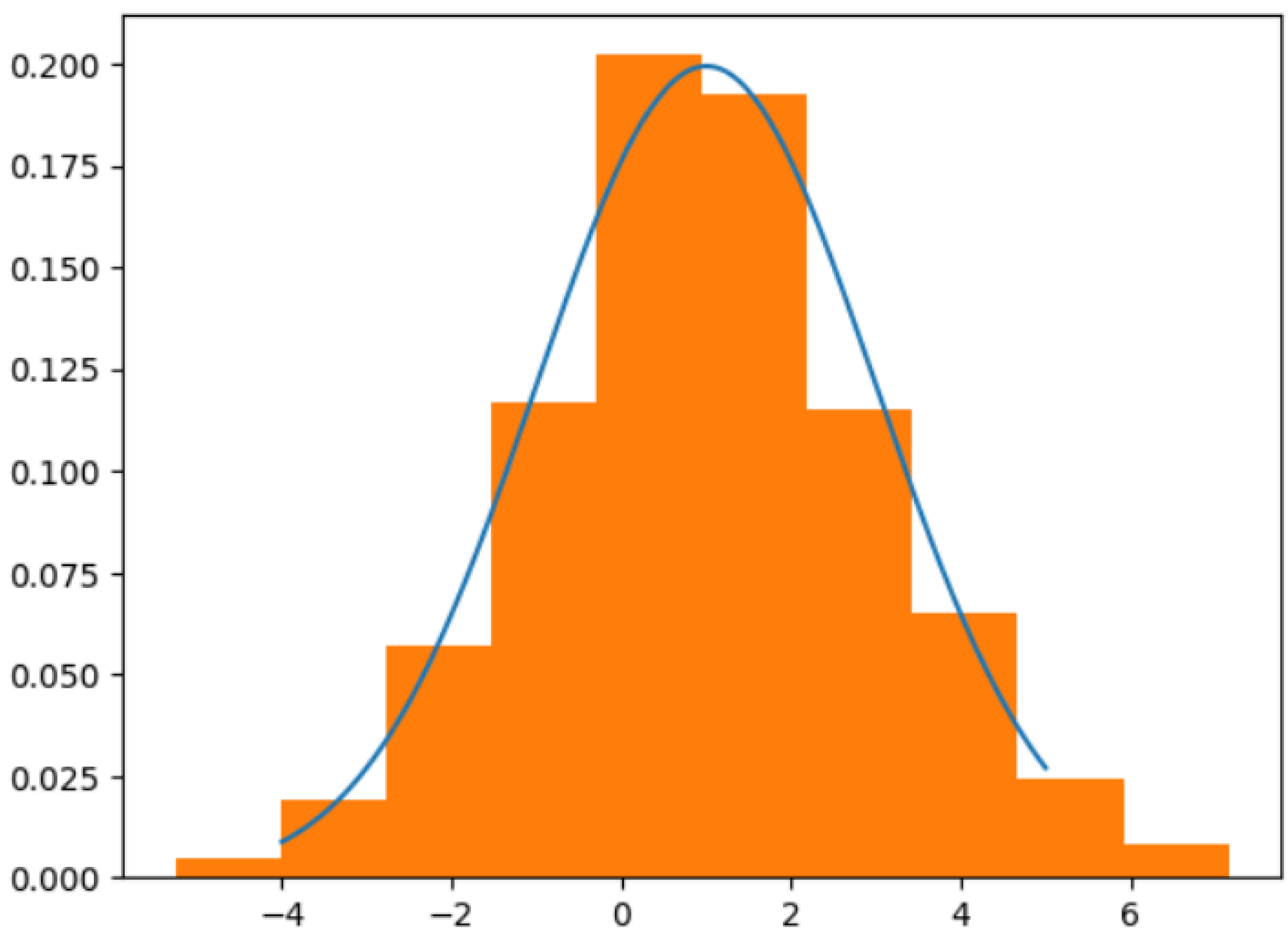


So if know that X is normally distributed with parameters, at least approximately, for example, we can get this results from CLT. Then we can say, for example, that the value of X will be most likely in the segment from mu-3sigma to mu+3sigma.So the distance between X and its EX is most likely not larger than 3Sigma and with rather good probability, the distance is not larger than 2Sigmas. This pic is called 3-Sigma law because the most from normal distribution within 3Sigma from its EX. This is very useful because sometimes we know the value of r.v.X and we to say smth about its EX. This is called a problem of estimation value in statistics.

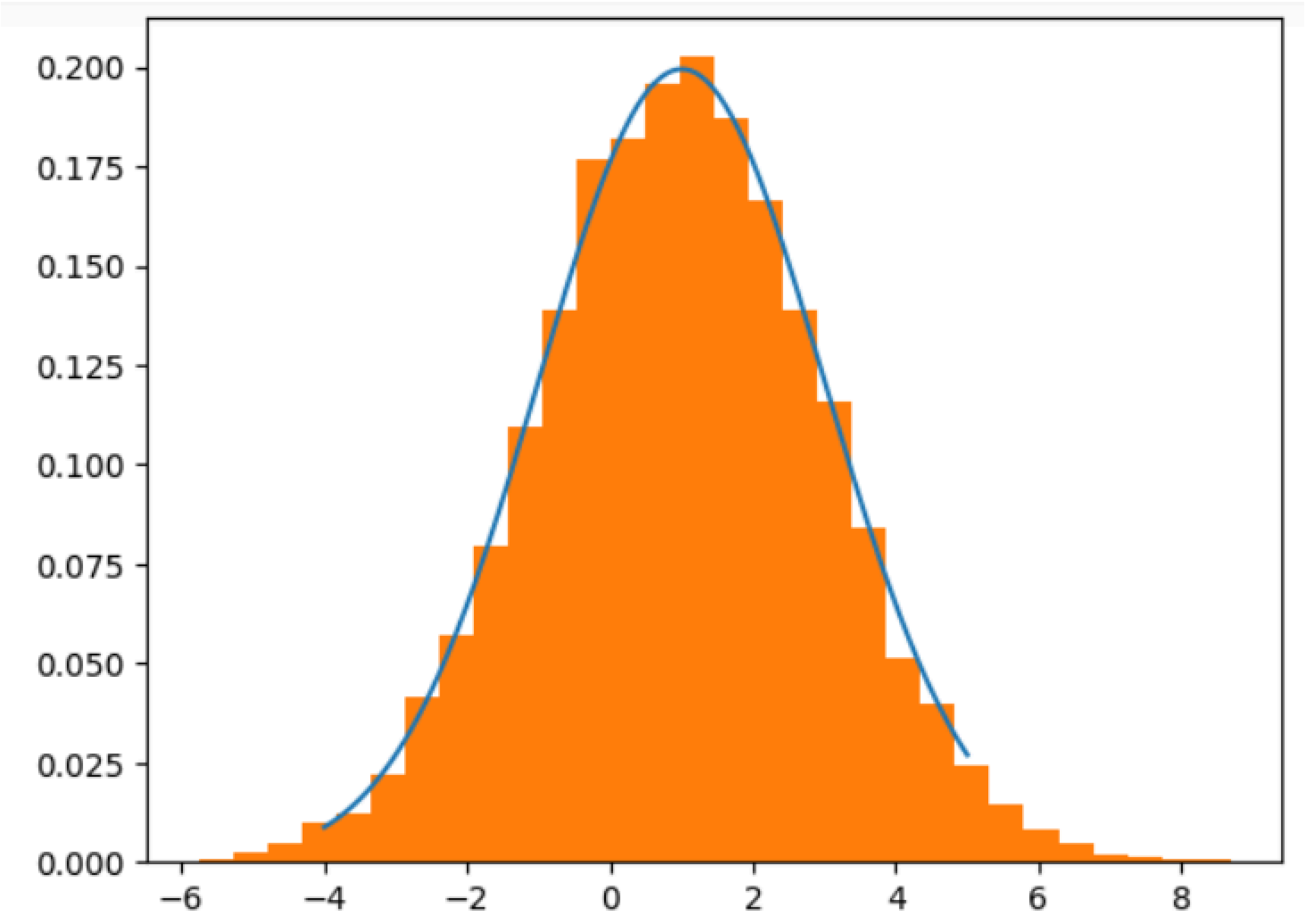
Generating and visualizing normal distribution with Python

Let us study normal distribution.

```
from scipy.stats import norm
import matplotlib.pyplot as plt
%matplotlib inline
import numpy as np

X = norm(loc=1, scale=2)#loc = EX, scale = SD
X.rvs(10) # generate 10 values of r.v.
#array([-0.8576428 , -0.69187098,  0.36283213,  0.2916131 ,  4.52017368,
        1.36763074,  0.3707561 ,  4.52793754, -2.115475  , -2.00506348])

X.pdf(1.2) #0.1984762737385059

x = np.linspace(-4, 5, 200)
sample = X.rvs(1000)
plt.plot(x, X.pdf(x))
plt.hist(sample, density=True)
```



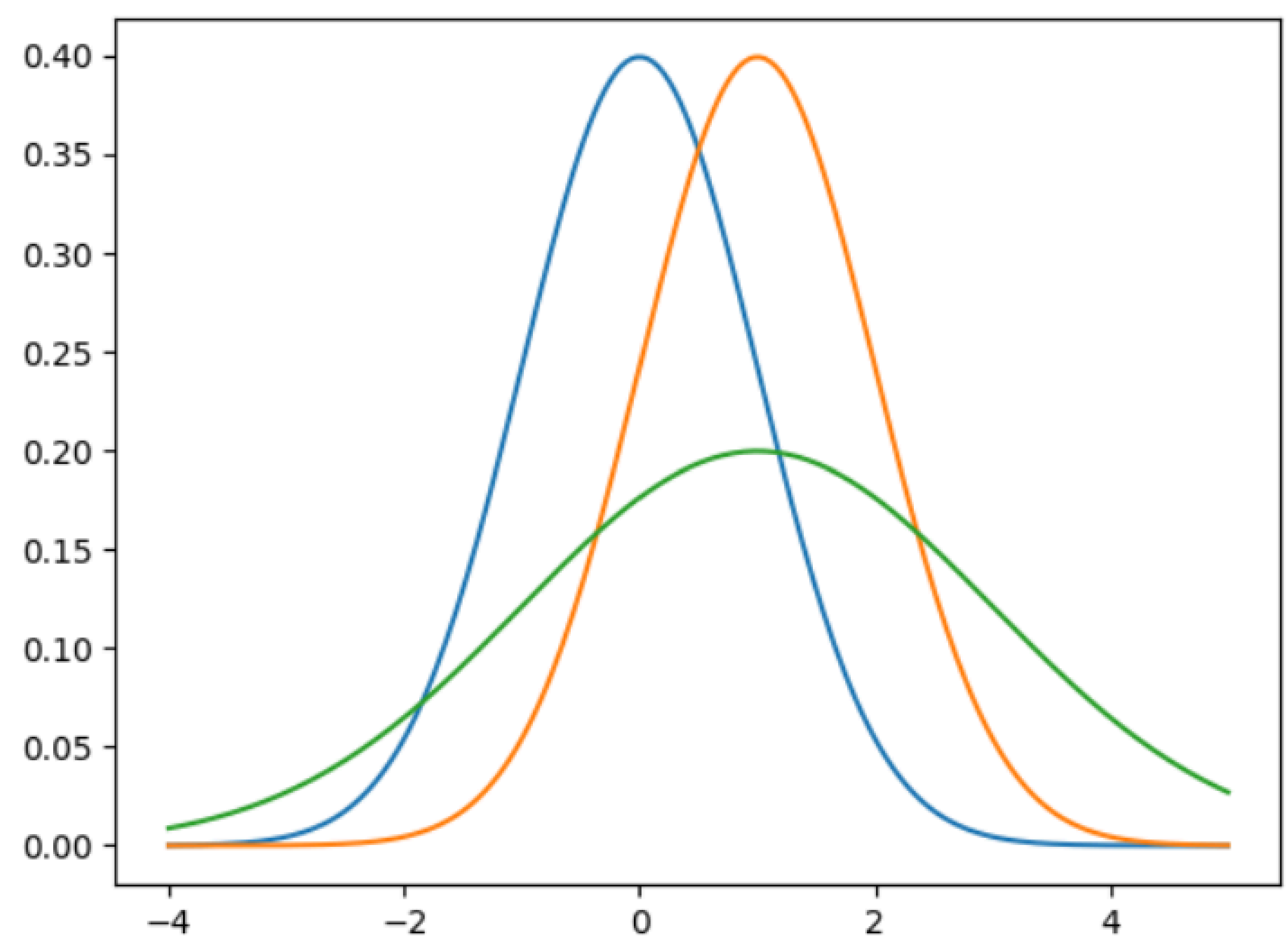
```
x = np.linspace(-4, 5, 200)
sample = X.rvs(10000)
plt.plot(x, X.pdf(x))
plt.hist(sample, density=True,bins = 30)
```



increase n and increase number of beans

So, let ud investigate how the parameters of the distribution affect the graph of pdf.

```
plt.plot(x, norm.pdf(x, loc=0, scale=1)) #standard normal distribution
plt.plot(x, norm.pdf(x, loc=1, scale=1)) #shifted to the right
plt.plot(x, norm.pdf(x, loc=1, scale=2)) #shifted to the right and rescaled
```



Python skill test

Question 1

Let X be normally distributed random variable with expected value 2 and variance 9. Find $P(X \leq 4)$. Use **scipy.stats.norm** function to create appropriate distribution, then use **.cdf** method for this distribution. Give an answer with 5 digits after decimal point.

Solution:

```
from scipy.stats import norm

mean = 2
std_dev = 3

# Create a normal distribution object
distribution = norm(loc=mean, scale=std_dev)

# Calculate the probability P(X <= 4)
probability = distribution.cdf(4)

print(f"P(X <= 4) = {probability}")
```

The correct answer is: 0.7475

Question 2

Let X be normally distributed random variable with expected value 2 and variance 9. Find $P(X \in [1, 5])$. Give an answer with 5 digits after decimal point.

Solution:

```
from scipy.stats import norm

mean = 2
std_dev = 3

# Create a normal distribution object
distribution = norm(loc=mean, scale=std_dev)

# Calculate the probability P(1 <= X <= 5)
probability = distribution.cdf(5) - distribution.cdf(1)

print(f"P(1 <= X <= 5) = {probability:.5f}")
```

The correct answer is: 0.4719

Question 3

Let X be standard normal random variable. Find value x such that $P(X \leq x) = 0.3$. Give an answer with 5 digits after decimal point.

Hint: Use **.ppf** method.

Solution:

```
from scipy.stats import norm

# Find the value x such that P(X <= x) = 0.3
x = norm.ppf(0.3)

print(f"x = {x:.5f}")
```

The correct answer is: -0.52441

Question 4

Let X be standard normal random variable. Find value x such that $P(|X| > x) = 0.2$. Give an answer with 5 digits after decimal point.

Hint: Use **.ppf** method and symmetry of normal distribution.

Our solution:

from scipy.stats **import** norm
print(-norm.ppf(0.2 / 2)))

This gives a value x such that $P(X < -x) = 0.1$. Due to symmetry of normal distribution, probability that $|X| > x$ is twice as large.

Solution:

```
from scipy.stats import norm

# We want P(|X| > x) = 0.2, which means: # P(X < -x) + P(X > x) = 0.2
# Since the normal distribution is symmetric,
# P(X < -x) = P(X > x) = 0.1
# Find the x value corresponding to P(X > x) = 0.1

x = norm.ppf(1 - 0.1)
```

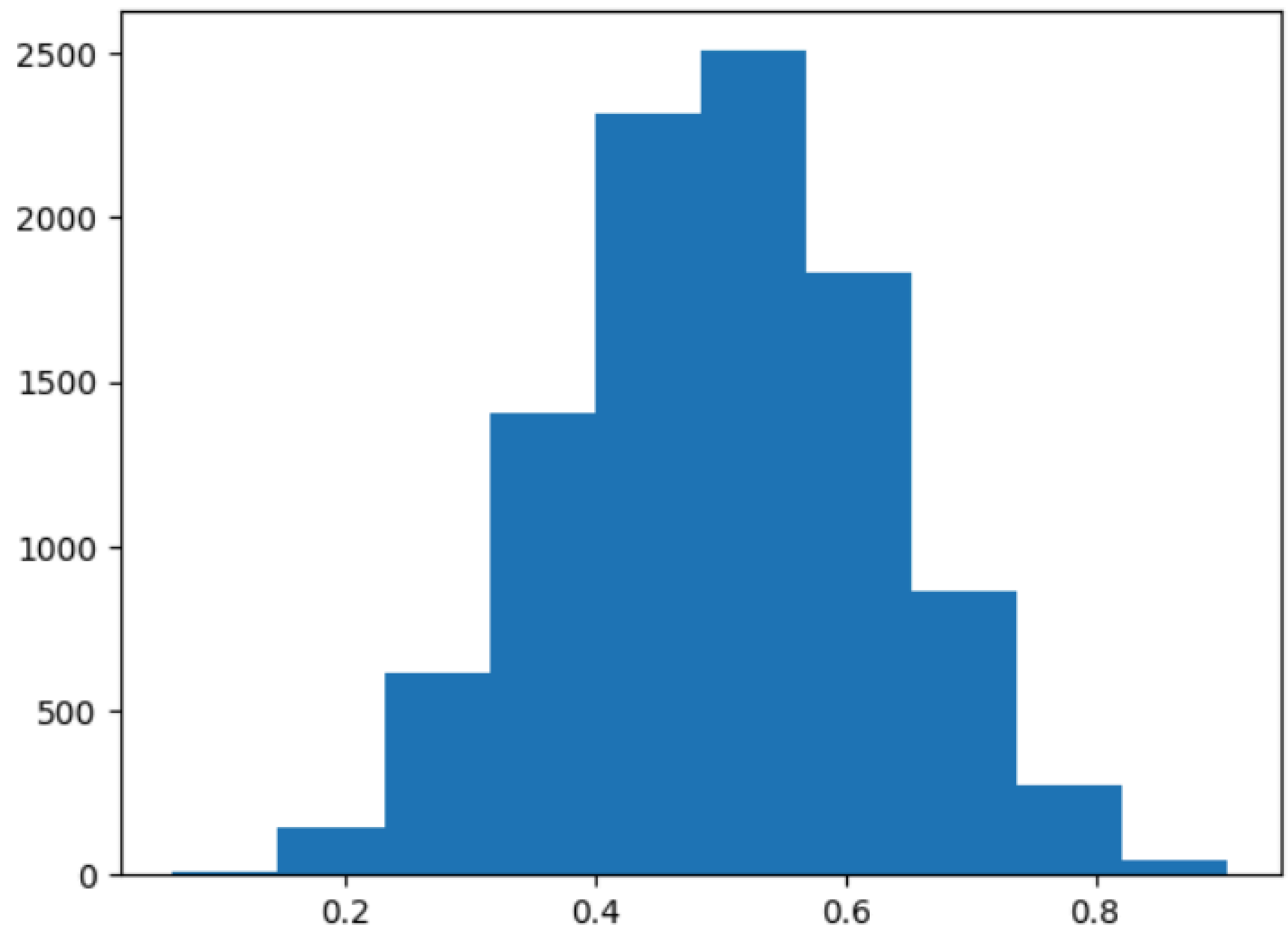
The correct answer is: 1.28154

Modeling central limit theorem with Python

We will simulation to check CLT: the distribution of sample mean resembles the bell curve that is given by the normal distribution. We use uniform dostribution as initial.

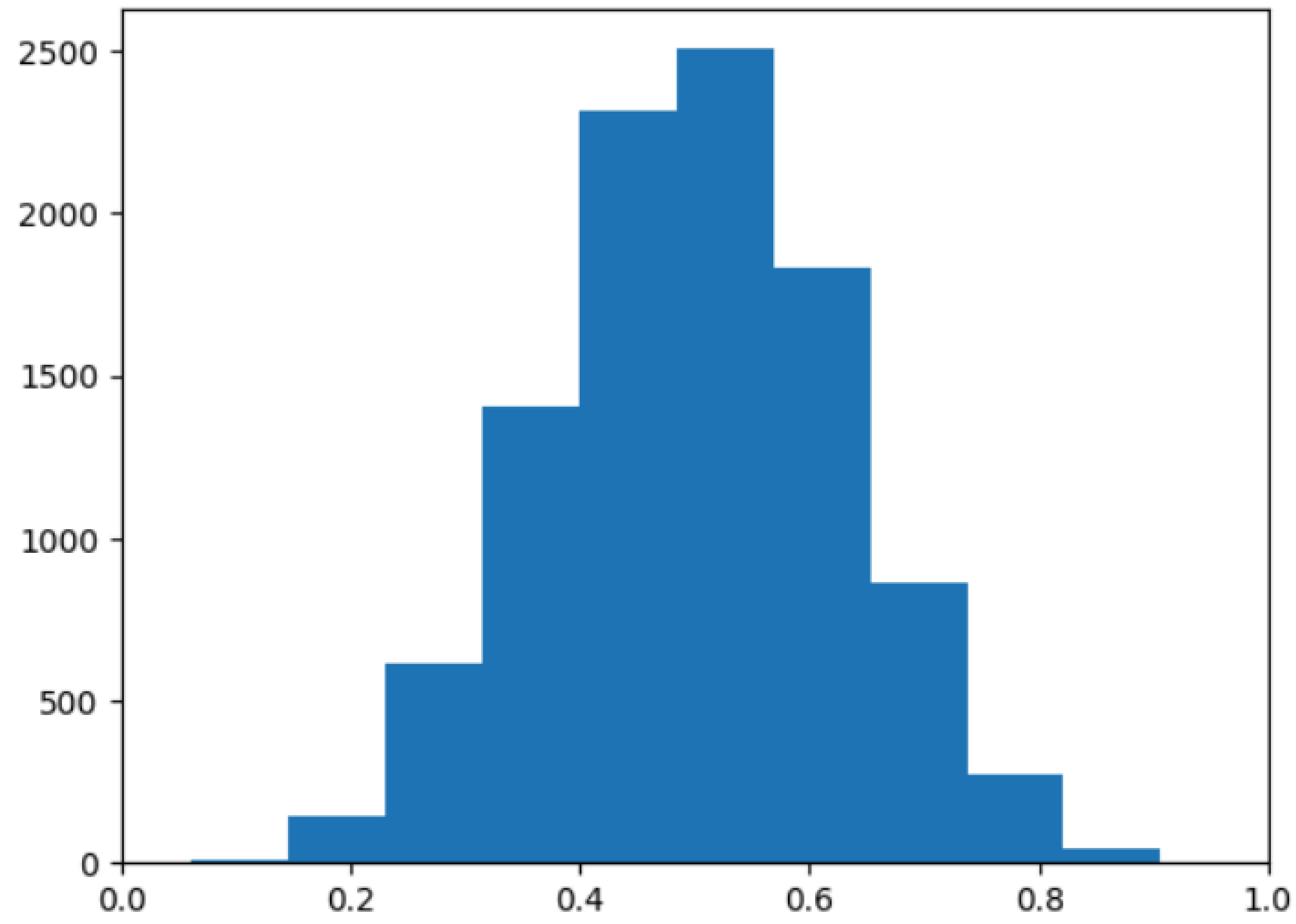
We can try the same experiment with another types initial distribution or with discrete r.v.

```
import matplotlib.pyplot as plt
%matplotlib inline #to get picture inside notebook
import numpy as np
from scipy.stats import norm

sample_size = 5
number_of_samples = 10000
means = []
for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    means.append(mean)

plt.hist(means); #; - avoid numbers above
```

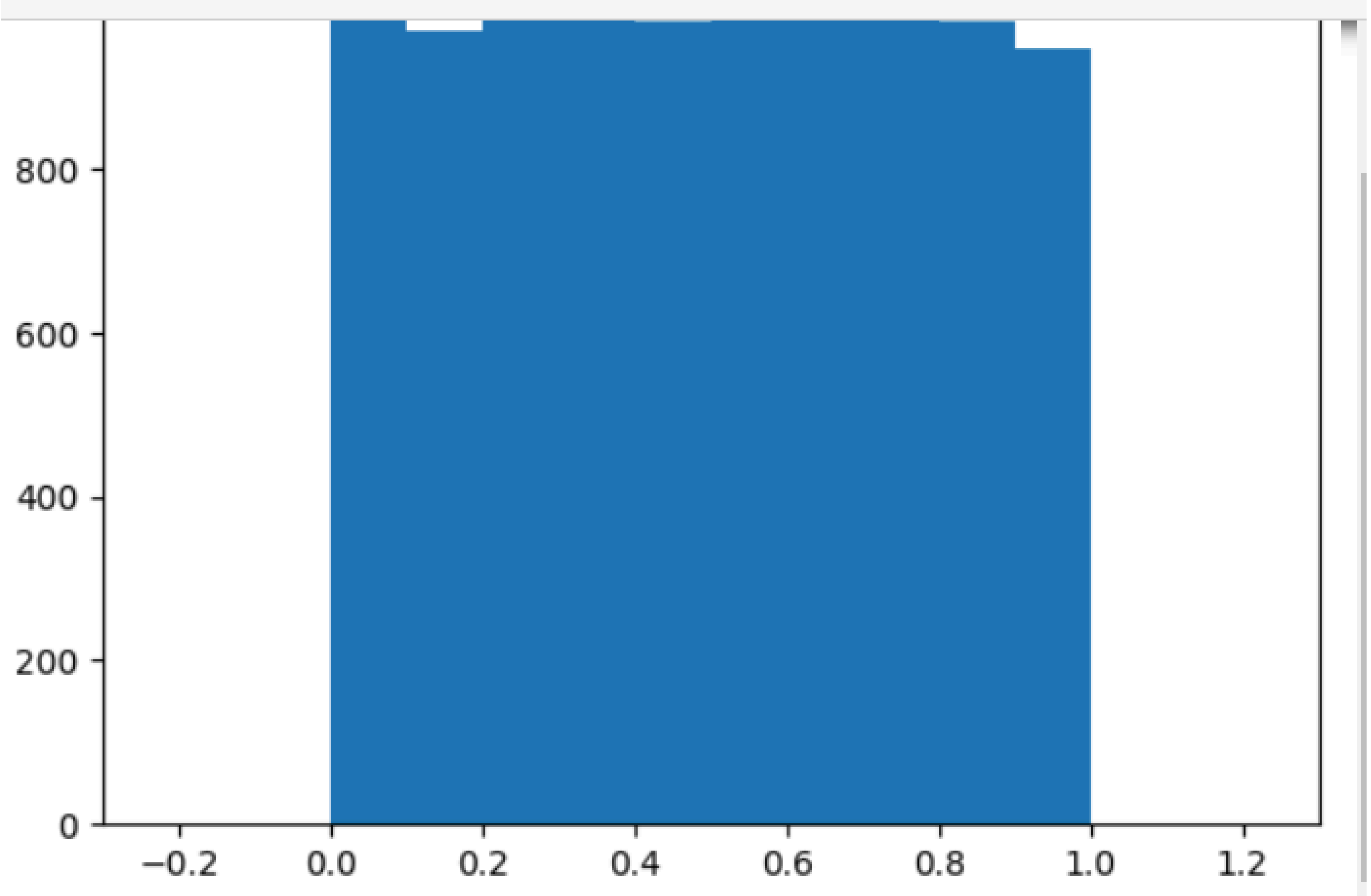


We see a bell curved which we discussed in CLT. We can change the size of sample and look at the result.

```
plt.xlim(0, 1)
plt.hist(means); #; - avoid numbers above
```



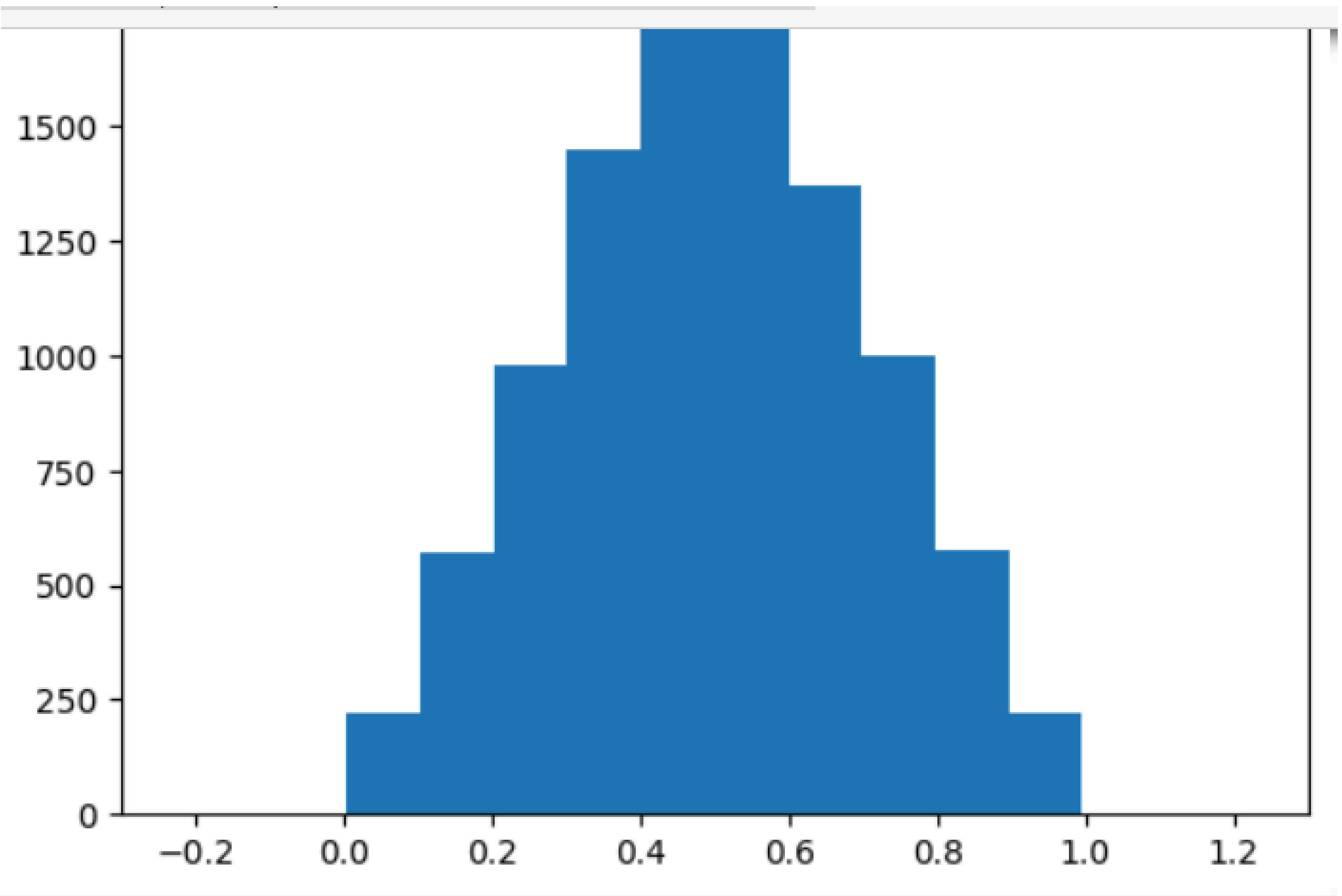

```
sample_size = 1
number_of_samples = 10000
means = []
for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    means.append(mean)


plt.xlim(-0.3, 1.3)
plt.hist(means); #; - avoid numbers above
```



We see that this graph looks like a graph pdf of uniform distribution. We make X a little bit wider to check that no points are outside of this segment from zero to one.

Increase sample size: 2

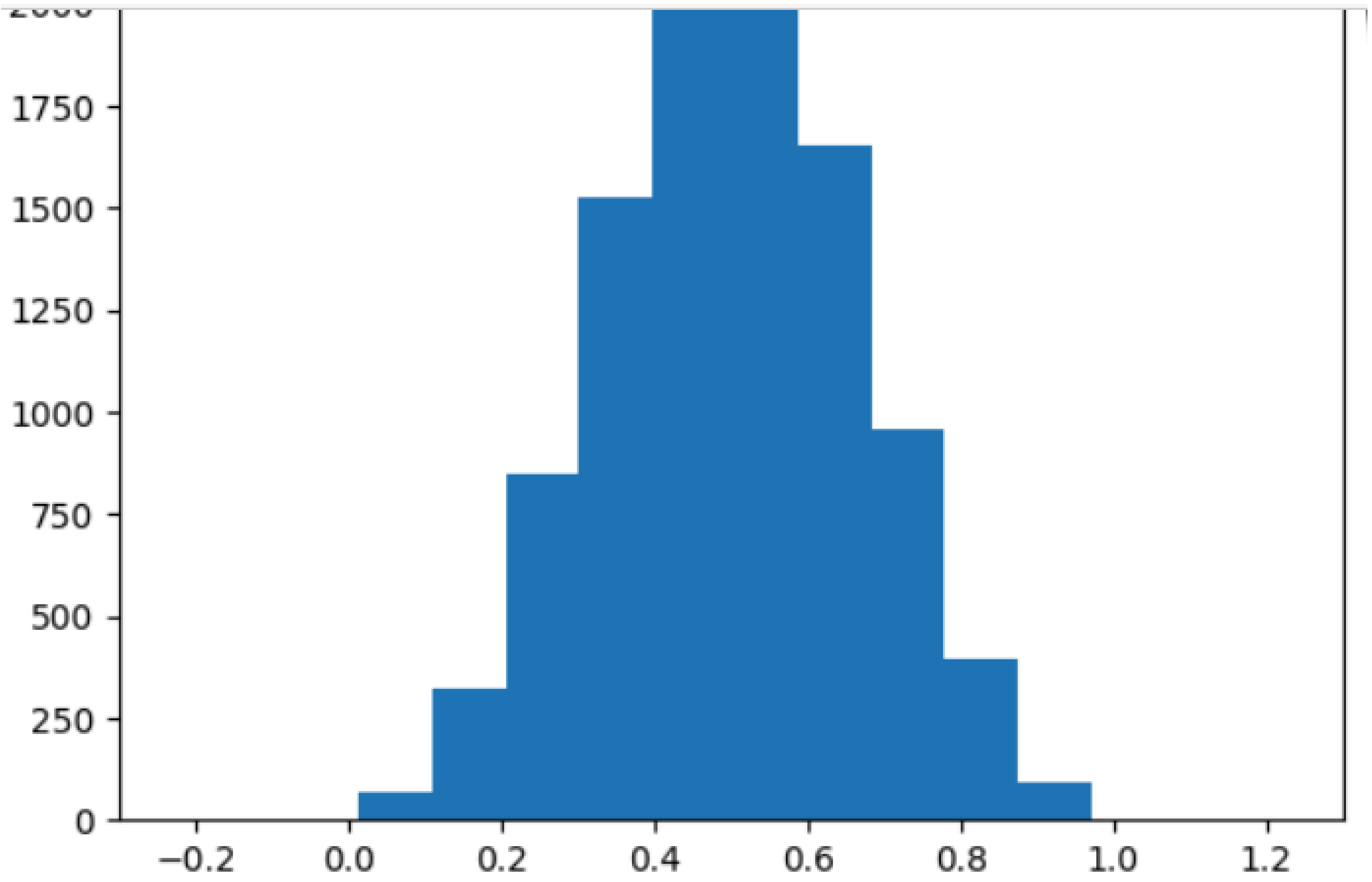
```
sample_size = 2
number_of_samples = 10000
means = []
for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    means.append(mean)

plt.xlim(-0.3, 1.3)
plt.hist(means); #; - avoid numbers above
```



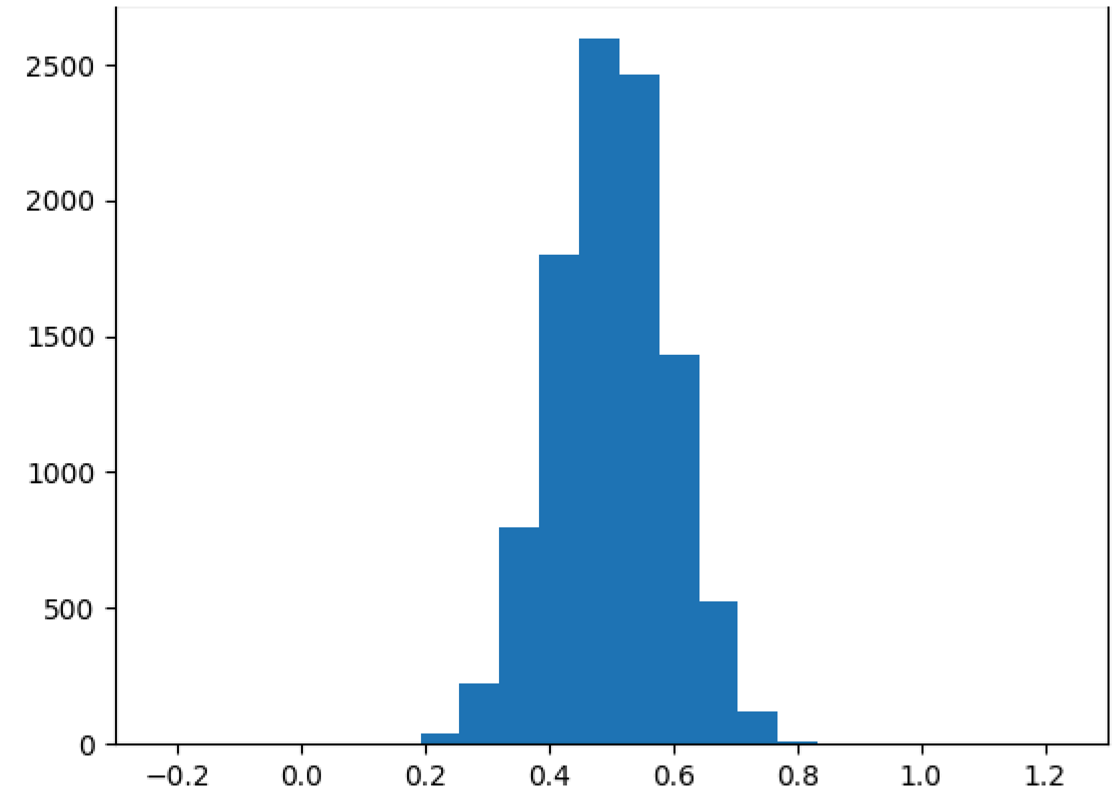
It's not very close to bell curve(normal distribution), but anyway, it's not uniform now.

```
sample_size = 3
number_of_samples = 10000
means = []
for i in range(number_of_samples):
```

```
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    means.append(mean)


plt.xlim(-0.3, 1.3)
plt.hist(means); #; - avoid numbers above
```

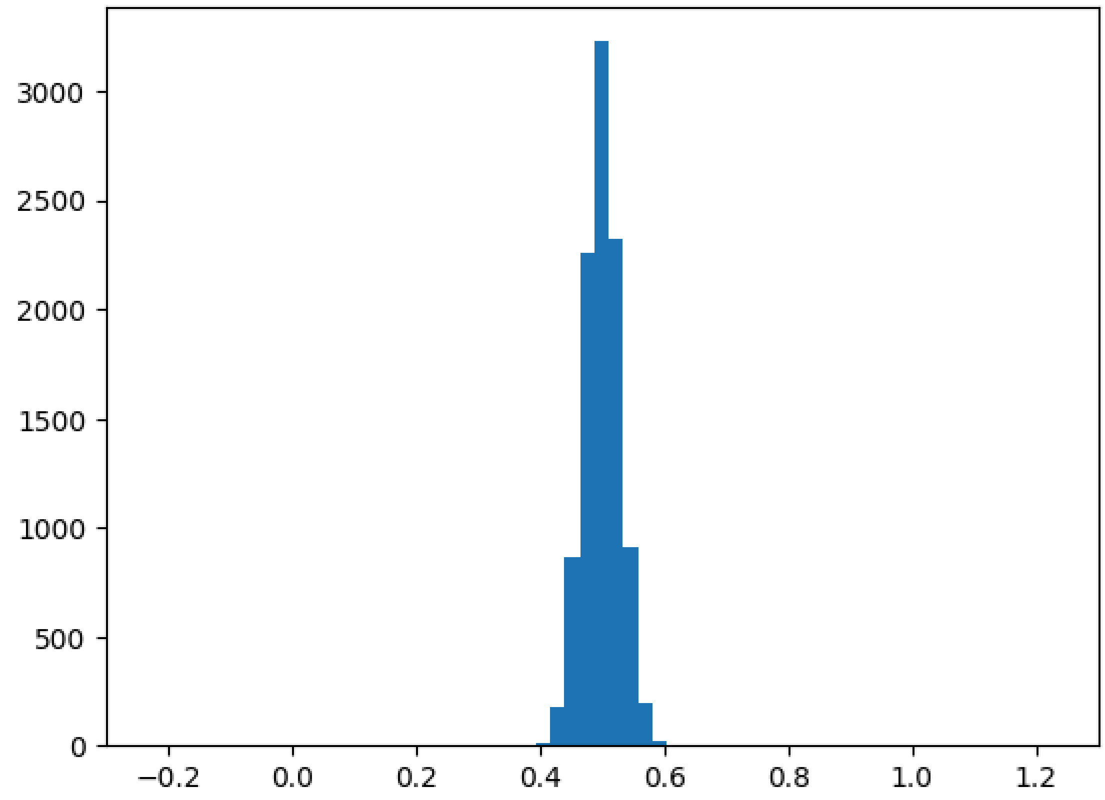



```
sample_size = 10
number_of_samples = 10000
means = []
for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    means.append(mean)


plt.xlim(-0.3, 1.3)
plt.hist(means); #; - avoid numbers above
```



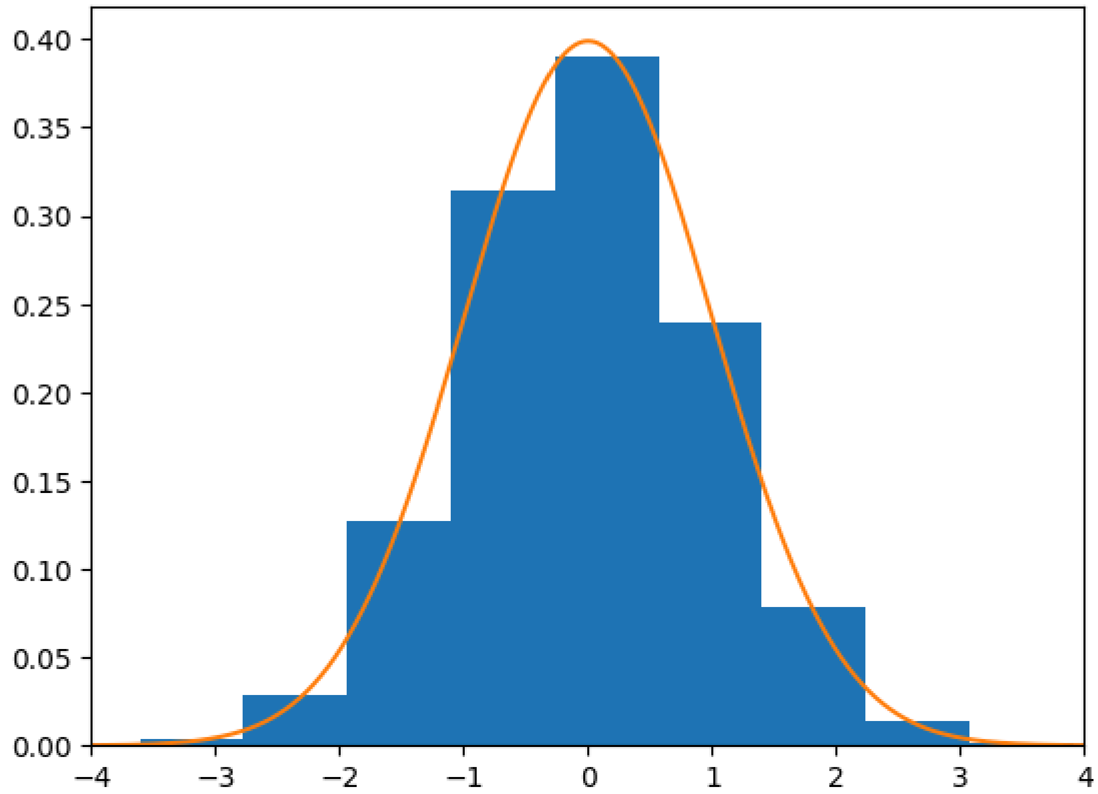

```
sample_size = 100
number_of_samples = 10000
means = []
```

```
for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    means.append(mean)


plt.xlim(-0.3, 1.3)
plt.hist(means); #; - avoid numbers above
```

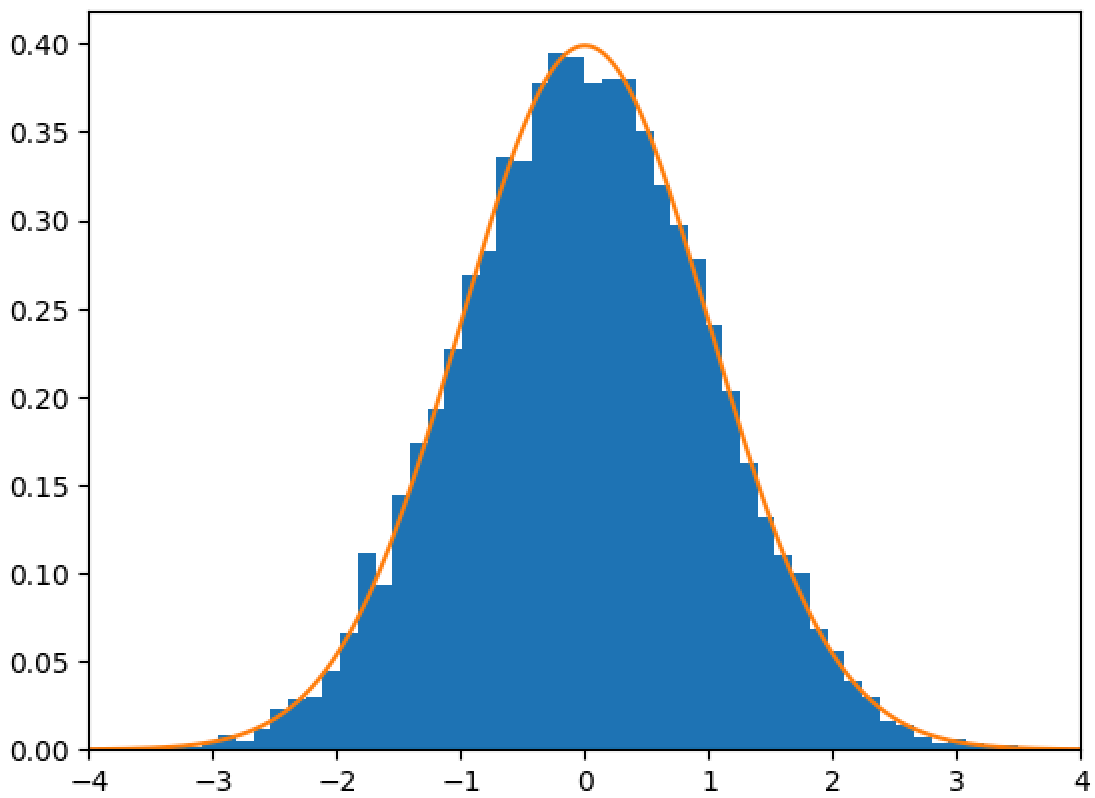


As we see, as we increase sample size, this graph becomes narrow and it's difficult to inderstand what is the shape of this graph. So we have to transorm to the corresponding z-score.

```
sample_size = 100
number_of_samples = 10000
means = []
zscores = []
EX = 0.5 #for uniform distribution
VarX = 1./12

for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    zscore = (mean - EX) * np.sqrt(sample_size)/np.sqrt(VarX)
    means.append(mean)
    zscores.append(zscore)


plt.xlim(-4, 4)
plt.hist(zscores, density= True); #; - avoid numbers above
x = np.linspace(-4, 4, 1000)
plt.plot(x, norm.pdf(x))
```



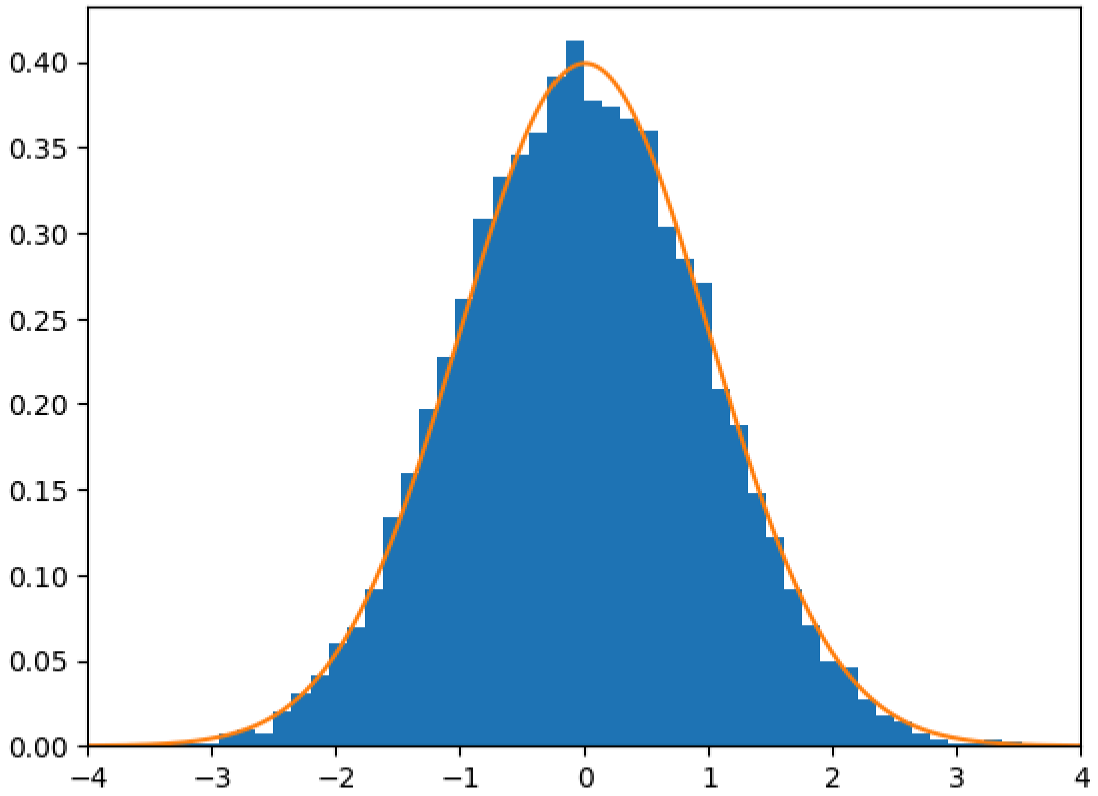

```
sample_size = 100
number_of_samples = 10000
means = []
zscores = []
EX = 0.5 #for uniform distribution
VarX = 1./12

for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    zscore = (mean - EX) * np.sqrt(sample_size)/np.sqrt(VarX)
    means.append(mean)
    zscores.append(zscore)


plt.xlim(-4, 4)
plt.hist(zscores, density= True, bins = 50); #; - avoid numbers above
x = np.linspace(-4, 4, 1000)
plt.plot(x, norm.pdf(x))
```



decrease sample size

```
sample_size = 10
number_of_samples = 10000
means = []
zscores = []
EX = 0.5 #for uniform distribution
```

```
VarX = 1./12

for i in range(number_of_samples):
    sample = np.random.uniform(low=0, high=1, size=sample_size)
    mean = sum(sample) /len(sample)
    zscore = (mean - EX) * np.sqrt(sample_size)/np.sqrt(VarX)
    means.append(mean)
    zscores.append(zscore)


plt.xlim(-4, 4)
plt.hist(zscores, density= True, bins = 50); #; - avoid numbers above
x = np.linspace(-4, 4, 1000)
plt.plot(x, norm.pdf(x))
```



You see that result is more or less the same. In fact, the CLT works rather fast. Even you have 30 or 50 observations, then you in most of cases, you can apply CLT and het good result.