Normal and derived distributions

Edoardo Ottavianelli, 1756005

Statistics course, Cybersecurity Master Degree
Sapienza University of Rome, December 2021

Contents

1 Abstract **3**

2 Introduction **4**

3 History **10**

4 Normal Distribution **12**

5 Derived distributions **18**

6 Implementation **24**

7 Conclusions **27**

1 Abstract

The goal of this document is to introduce novices to the Normal distribution, so here you won't find any cutting-edge research, just the basics. We will see an introduction that will try to explain some important concepts of Statistics and Probability doing real world examples to better understand the theorems explained. I tried also to include images to make the reading more comfortable. There is also an historical background to understand better the ideas that gave birth to these distributions. There isn't only the normal distribution, but also as the title suggests some 'derived' ones, like the Standard, the F-distribution and the Student's T distribution. The last section instead presents some implementations (using C, C# and Python) of the normal distribution and some efficient pseudo-random number generators.

2 Introduction

In order to understand the following sections we need to talk about some basic concepts of Statistics and Probability. To achieve this, we can make a real-world and simple example that will help us presenting these concepts. Let's take a non-loaded dice with six sides, so the most famous one. On each side there will be a number, specifically from one to six. We call the action of throwing the dice on a flat surface an 'experiment' and we will observe the result of experiment looking at the number on the side of the dice that will be on top. We call the 'sample space' the set of all possible results. In simple terms, you have to figure out every possibility that might happen. With dice rolling, your sample space is going to be every possible dice roll. Question: What is the sample space of a single experiment? So the action of throwing the dice just one time? The sample space will be: { 1, 2, 3, 4, 5, 6 }. This is the set of all the possible results we can have.
What about the sample space of a trial composed by two experiments? So the action of throwing the same dice two times? The sample space will be:

{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) }.

This is the set of all the possible results we can have. Of course there could be repetitions, because even if we are throwing the same dice two times these experiments are independent from each other. Look at your sample space and find how many add up to 4 or 7 (because we're looking for the probability of rolling one of those numbers). The rolls that add up to 4 or 7 are:

{ (1,3), (1,6), (2,2), (2,5), (3,1), (3,4), (4,3), (5,2), (6,1) }.

There are 9 possible combinations. Take the answer from step 2, and divide it by the size of your total sample space from step 1. What I mean by the "size of your sample space" is just all of the possible combinations you listed. In this case, Step 1 had 36 possibilities, so: $9/36 = 0.25$. But what does it mean this operation? And its result? The following table shows the probabilities for rolling a certain number with a two-dice roll. If you want the probabilities of rolling a set of numbers (e.g. a 4 and 7, or 5 and 6), add the probabilities from the table together. For example, if you wanted to know the probability of rolling a 4, or a 7: $3/36 + 6/36 = 9/36$ [1] [2].

Roll a ...	Probability
2	1/36 (2.778%)
3	2/36 (5.556%)
4	3/36 (8.333%)
5	4/36 (11.111%)
6	5/36 (13.889%)
7	6/36 (16.667%)
8	5/36 (13.889%)
9	4/36 (11.111%)
10	3/36 (8.333%)
11	2/36 (5.556%)
12	1/36 (2.778%)

From the Wikipedia page of "Probability": 'Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur, or how likely it is that a proposition is true. The probability of an event is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility of the event and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur. A simple example is the tossing of a fair (unbiased) coin. Since the coin is fair, the two outcomes ("heads" and "tails") are both equally probable; the probability of "heads" equals the probability of "tails"; and since no other outcomes are possible, the probability of either "heads" or "tails" is 1/2 (which could also be written as 0.5 or 50%). [3]
In our case this means that the most probable outcome of a two-dice rolling trial is the result having 7 as sum of experiments (having 6/36 or 1/6 or 16.6% as probability of success).
Then if we look at the records above and below (still looking at the table above), we can observe two results that are less likely to happen than the previous one (7), but they are still the most probable out of the previous one. It's the same looking at the records above and below the two discussed (6 and 8); we can continue until we reach the end of the table.
The next image shows this in a very clear and obvious manner. Take a look at the figure, specifically at its shape since we will see that shape many times in this document. What we are seeing is called a 'Distribution'. Here in particular it's a 'Probability Distribution', that is a mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment [4]. It's a visual representation of what we can write in mathematical language, firstly we wrote some considerations about the events and the related probabilities, here we are just showing in a visual manner what we wrote. In this case the mathematical function used is the probability mass function (pmf) $p(S)$ that specifies the probability distribution for the sum S of counts from two dice. For example, the figure shows that $p(11) = 2/36 = 1/18$. The pmf allows the computation of probabilities of events such as $P(S > 9) = 1/12+1/18+1/36 = 1/6$, and all other probabilities in the distribution.

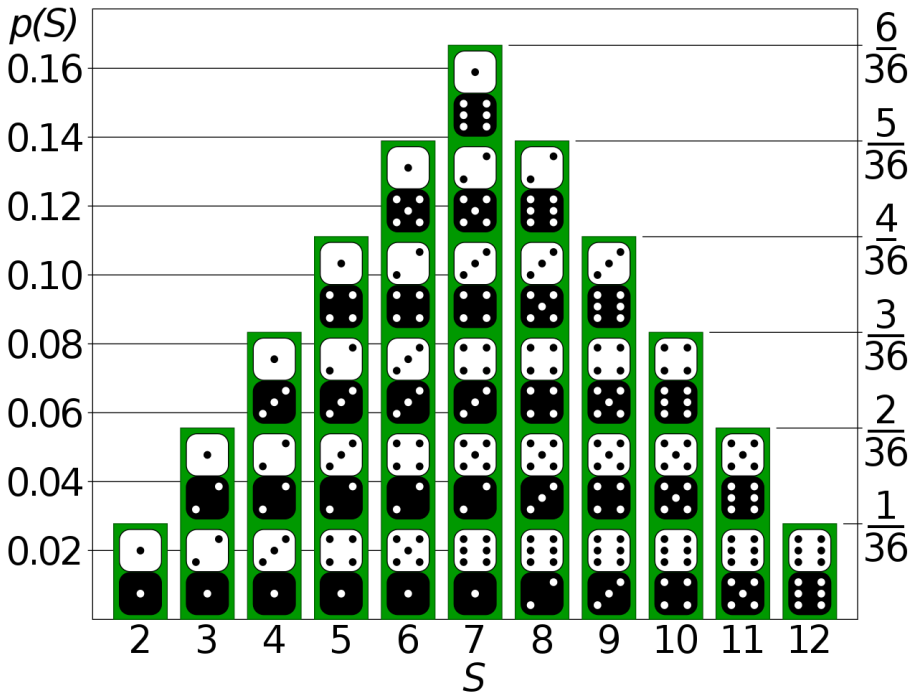

Figure 1: Distribution of two-dice rolls experiment
Source: https://en.wikipedia.org/wiki/Probability#/media/File:Dice_Distribution_(bar).svg

This figure shows the 'Probability Distribution', but there is also a Distribution for Statistics, or better a 'Frequency Distribution'. What is the difference? While in Probability we are talking about events 'in theory', in Statistics we are observing real observations; the name involves in fact Frequency because we are visualizing frequencies of results and not the probabilities of them. The figure 2 shows a Frequency Distribution of ten trials of a dice roll experiment.
We can observe that throwing the same dice ten times, we got three times the result 1, one time the result 2, four times the result 3, one time the result 4, one time the result 5 and never the result 6. This is an example of Frequency Distribution. We can also show the same observations in a table having as first row the headers, like 'Result' and 'Frequencies', and then each row will have as first item the number from 1 to 6 and as second item its frequency.
We can also compute the 'mean' (or average) of these observations. From the Mean Wikipedia page: "For a data set, the arithmetic mean, also known as arithmetic average, is a central value of a finite set of numbers: specifically, the sum of the values divided by the number of values. The arithmetic mean of a set of numbers $x1, x2, ..., xn$ is typically denoted by $\bar{x}$. If the data set were based on a series of observations obtained by sampling from a statistical population, the arithmetic mean is the sample mean (denoted $\bar{x}$) to distinguish it from the

mean, or expected value, of the underlying distribution, the population mean" [5]. We can also define the mean as the value that if substituted to a value in

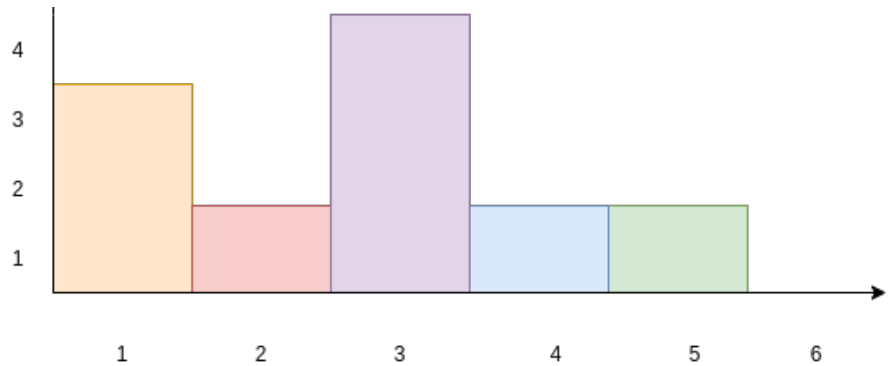

Figure 2: Distribution of 10 dice rolls experiment

the sample, will not modify the sum of the observations. In this example the mean is

$(1+1+1+2+3+3+3+3+4+5)/10 =$
$((3*1)+(1*2)+(4*3)+(1*4)+(1*5))/10 = 2,6.$

In fact, the sum of observations is $1+1+1+2+3+3+3+3+4+5 = 26$, that is equal to $2,6*10 = 26$. Another important concept to talk about is the 'Median'. From the Median Wikipedia Page: "In statistics and probability theory, the median is the value separating the higher half from the lower half of a data sample, a population, or a probability distribution. For a data set, it may be thought of as 'the middle' value. The basic feature of the median in describing data compared to the mean (often simply described as the "average") is that it is not skewed by a small proportion of extremely large or small values, and therefore provides a better representation of a 'typical' value." [6]. In this example, we could easily say 'the median is 3, because it sits in the middle of all possible values.', it's not wrong but it's better to understand how. As we did for the mean, we have to take into account also the repetitions (!!) of the observations (three times the value 1 and four times the value 3). So these are our observations: { 1, 1, 1, 2, 3, 3, 3, 3, 4, 5 }, and the value in the middle is: .... wait! we don't have a value in the middle, since the number of elements is even (10), but this is not a problem since we can take the two values in the middle. So in this case we should pick 3 and 3, but it's okay to take only one of the two and just saying 3 because they are equal. Another important concept is the concept of 'Mode'. From the Mode Wikipedia page: "The mode is the value that appears most often in a set of data values. If X is a discrete random variable, the mode is the value x (i.e, X = x) at which the probability mass function takes its maximum value. In other words, it is the value that is most likely to be sampled.". It's straightforward to see that the mode of our example is 3, since it has the highest number of observations. Unlike mean and median, the concept

of mode also makes sense for "nominal data" (i.e., not consisting of numerical values in the case of mean, or even of ordered values in the case of median). For example, taking a sample of Korean family names, one might find that "Kim" occurs more often than any other name. Then "Kim" would be the mode of the sample. [7] These three measures are called 'Measures of central tendency'. The goal of these measures is to represent a large set of data in a single value. Of course only these measures can't describe in a good manner the set, they are vague, but still we can get some meaning from them. The last concept I want to introduce is the 'Standard Deviation'. From Wikipedia: "In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the values are spread out over a wider range. Standard deviation may be abbreviated SD, and is most commonly represented in mathematical texts and equations by the lower case Greek letter sigma σ, for the population standard deviation, or the Latin letter s, for the sample standard deviation." [8]. For a finite set of numbers, the population standard deviation is found by taking the square root of the average of the squared deviations of the values subtracted from their average value. These are the observations: { 1, 1, 1, 2, 3, 3, 3, 3, 4, 5 }; we already know the mean: 2,6. First, calculate the deviations of each data point from the mean, and square the result of each:
$(1-2,6)^2 = 2,56$ (3 observations)
$(2-2,6)^2 = 0,36$
$(3-2,6)^2 = 0,16$ (4 observations)
$(4-2,6)^2 = 1,96$
$(5-2,6)^2 = 5,76$

The variance is the mean of these values:
$((2,56 * 3) + 0,36 + (0,16 * 4) + 1,96 + 5,76)/10 = 1,64.$

And the population standard deviation is equal to the square root of the variance:
$\sigma = \sqrt{1,64} = 1.28.$

Frequency table :

element	frequency	cumulative frequency	relative frequency	cumulative relative frequency
1	3	3	0.3	0.3
2	1	4	0.1	0.4
3	4	8	0.4	0.8
4	1	9	0.1	0.9
5	1	10	0.1	1

Figure 3: Frequency table

The figure above is called the ‘Frequency table‘. This is another way to represent our experiments. The first column represents the unique values we have (6 is not shown just because we don't have data for that entry) for our observations. The second column describes the frequencies. The third column is the ‘cumulative frequency‘, that is another way to represent the frequencies, while in the precedent column we wrote the frequencies not caring about precedent values, here we are summing the frequencies, so at the n-th line we have the sum of the n-th value plus all the precedents. Obviously at the last line we will have the total number of experiments we made (10 in this case). The relative frequency instead is the frequency of the element divided by the total number of experiments (for the first line for example we have 3 experiments and so 3/10 = 0.3). Finally the cumulative relative frequency (as the name suggests) is the sum of the relative frequencies (same as the cumulative frequency, so at the n-th line we have the sum of the n-th value plus all the precedents).

3 History

The first person who discovered the Normal Distribution formula was Abraham de Moivre (1667 – 1754). He was a French mathematician known for de Moivre's formula, a formula that links complex numbers and trigonometry, and for his work on the normal distribution and probability theory.

He moved to England at a young age due to the religious persecution of Huguenots in France which began in 1685. He was a friend of Isaac Newton, Edmond Halley, and James Stirling. Among his fellow Huguenot exiles in England, he was a colleague of the editor and translator Pierre des Maizeaux. De Moivre wrote a book on probability theory, "The Doctrine of Chances", said to have been prized by gamblers [9]. In the later editions of his book, de Moivre included his unpublished result of 1733, which is the first statement of an approximation to the binomial distribution in terms of what we now call the normal or Gaussian function. The core problem for DeMoivre was to find the sum of several terms in a binomial expansion.

n	Bin. expansion	Coefficients
1	a+b	1 1
2	$a^2 + 2ab + b^2$	1 2 1
3	$a^3 + 3a^2b + 3ab^2 + b^3$	1 3 3 1
4	...	1 4 6 4 1
5	...	1 5 10 10 5 1

Although this can be interpreted as the first obscure expression for the normal probability law, Stigler points out that de Moivre himself did not interpret his results as anything more than the approximate rule for the binomial coefficients, and in particular de Moivre lacked the concept of the probability density function. In the table above, but better in the figure below, it's easy to see the pattern of coefficients that are high in the center and when we move to the edges we can see lower values.

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1

Figure 4: Coefficients in binomial expansion

In 1823 Gauss (1777 – 1855) published his monograph "Theoria combinationis observationum erroribus minimis obnoxiae" where among other things he introduces several important statistical concepts, such as the method of least

squares, the method of maximum likelihood, and the normal distribution. Although Gauss was the first to suggest the normal distribution law, Laplace (1749 – 1827) made significant contributions. It was Laplace who first posed the problem of aggregating several observations in 1774, although his own solution led to the Laplacian distribution. Finally, it was Laplace who in 1810 proved and presented to the Academy the fundamental central limit theorem, which emphasized the theoretical importance of the normal distribution. In the middle of the 19th century Maxwell (1831 – 1879) demonstrated that the normal distribution is not just a convenient mathematical tool, but may also occur in natural phenomena.

Since its introduction, the normal distribution has been known by many different names: the law of error, the law of facility of errors, Laplace's second law, Gaussian law, etc. Gauss himself apparently coined the term with reference to the "normal equations" involved in its applications, with normal having its technical meaning of orthogonal rather than "usual". However, by the end of the 19th century some authors had started using the name normal distribution, where the word "normal" was used as an adjective – the term now being seen as a reflection of the fact that this distribution was seen as typical, common – and thus "normal".

" Many years ago I called the Laplace–Gaussian curve the normal curve, which name, while it avoids an international question of priority, has the disadvantage of leading people to believe that all other distributions of frequency are in one sense or another 'abnormal'. " - Pearson (1920)

4 Normal Distribution

A normal (or Gaussian or Gauss or Laplace–Gauss) distribution is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The parameter μ is the mean or expectation of the distribution (and also its median and mode), while the parameter σ is its standard deviation. The variance of the distribution is σ^2. A random variable with a Gaussian distribution is said to be normally distributed, and is called a normal deviate [10]. We define

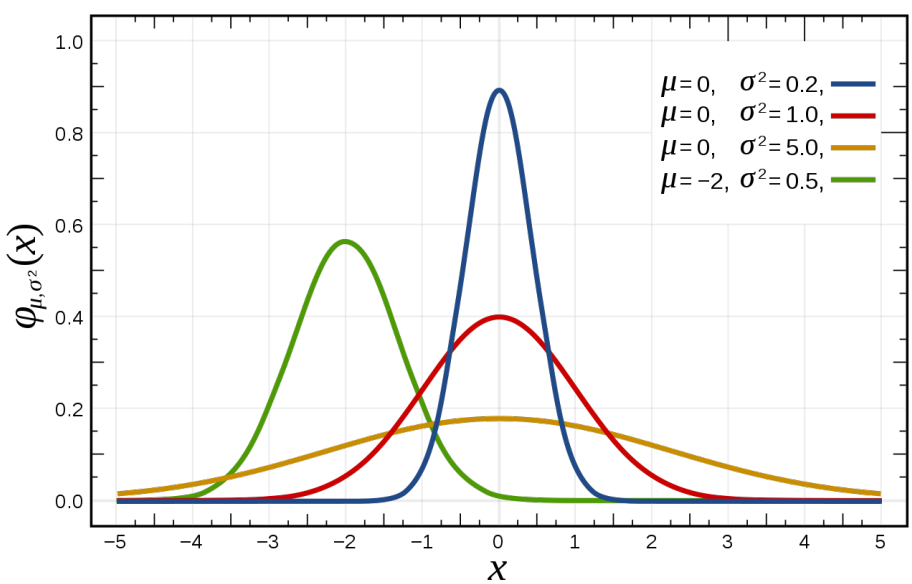

Figure 5: Normal distribution
Source: https://en.wikipedia.org/wiki/Normal_distribution
#/media/File Normal_Distribution_PDF.svg

a Normal Distribution using a capital letter N followed by the mean μ and variance σ^2 and we read the following notation as "X follows a Normal Distribution with mean μ and variance σ^2": $X \sim N(\mu, \sigma^2)$. The importance of this type of distribution is the fact that we can see it in many different places, examples, facts in real world (economics, finance, social sciences, physics...), we will use some real-world examples to understand it better. Behind the concept of Normal Distribution there is another very important theorem, that is the Central Limit Theorem. The Central Limit Theorem (CLT) establishes that, in many situations, when independent random variables are summed up, their properly

normalized sum tends toward a normal distribution (informally a bell curve) even if the original variables themselves are not normally distributed. The theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions [11].
The simplest case of a normal distribution is known as the standard normal distribution or unit normal distribution (also known as Z-distribution). This is a special case when $\mu = 0$ and $\sigma = 1$, and it is described by this probability density function:

$$\varphi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

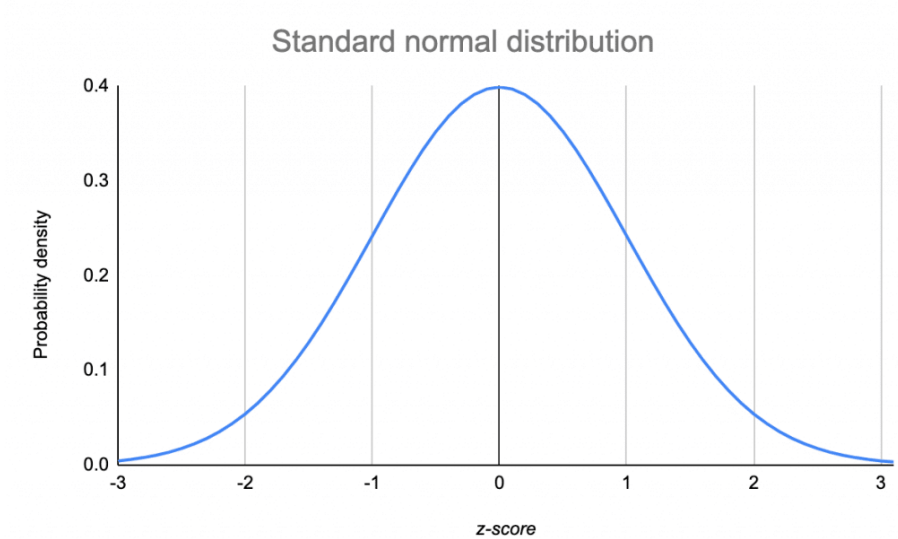

Figure 6: Standard Normal distribution
Source: https://www.scribbr.com/statistics/standard-normal-distribution/

The factor 1/2 in the exponent ensures that the distribution has unit variance (i.e., variance being equal to one), and therefore also unit standard deviation. This function is symmetric around $x = 0$, where it attains its maximum value $1/\sqrt{2\pi}$ and has inflection points at $x = +1$ and $x = -1$. Authors differ on which normal distribution should be called the "standard" one. Gauss defined the standard normal as having a variance of $\sigma^2 = 1/2$, while Stigler defined the standard normal as having a variance of $\sigma^2 = 1/(2\pi)$.
We called that type of curve a 'bell shape curve'... because it looks like a bell. We can make an example of a normal distribution that everyone can understand in a clear way. Think about a group of people, and let's measure

their height. The chart we will plot has the frequencies on the Y axis and the height (measured in centimeters, inches...as you want) on the X axis. The shape of the distribution will be a bell curve; people can be short, but it is less likely to have really short people; people can have an average height, and this is the most likely case; and finally people can be tall, and still (like short people) this is less likely. Of course, we can have outliers (so for example people really tall or really short), but this is less likely to happen. Being a bell shaped curve means that the majority of the data are centered around the mean, while values further away from the mean are less likely to occur. Furthermore, the graph is symmetric with regards to the mean, that suggests values equally far away in opposing directions would be still equally likely.
To better understand this concept, we can extend this example. Let's consider the height of people, but now we differentiate humans when born and as adults. The green distribution (in Figure 7) represents the new born heights, while the

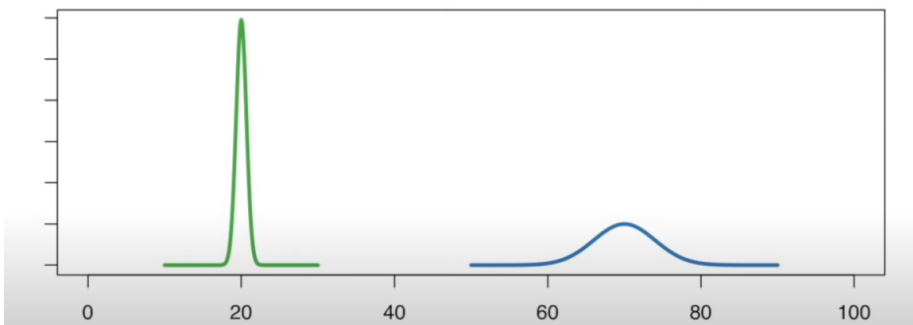

Figure 7: Two normal distributions of the height of male humans when born and as adults
Source: https://www.youtube.com/watch?v=rzFX5NWojp0

blue one represents the adults' heights. In both cases we can observe a bell shape curve, but it's also clear that there is a huge difference in the two representations. The average baby height is 20 inches (50.8 centimeters), while the average for the adults is 70 inches (177.8 centimeters), and those distributions, being normal, are centered in the average. If we look at the green graph, we can say that there is a high probability that a newborn baby will be between 19 and 21 inches tall. In contrast, adults are between 60 and 80 inches tall. We can easily notice that the curve for the babies is relatively tall compared to the adults one, this is because there are many more possibilities for adults' height than for babies. The more options there are for height, the less likely any specific measurement will be one of them. This is an important concept: the width of the curve is defined by its standard deviation. The standard deviation of the babies is relatively small compared to the adults. The standard deviation for babies is 0.6 inches, while for the adults is 4. Knowing the standard deviation is important because about 68% of values drawn from a normal distribution are within one

standard deviation σ away from the mean; about 95% of the values lie within two standard deviations; and about 99.7% are within three standard deviations. This fact is known as the "68-95-99.7" (empirical) rule, or the "3-sigma" rule. This means that 95% of the baby measurements fall between 20 +/- 1.2 inches; and 95% of the adults measurements fall between 70 +/- 8 inches [12] [13]. The value of the normal distribution is practically zero when the value x lies more than a few standard deviations away from the mean (e.g., a spread of three standard deviations covers all but 0.27% of the total distribution).

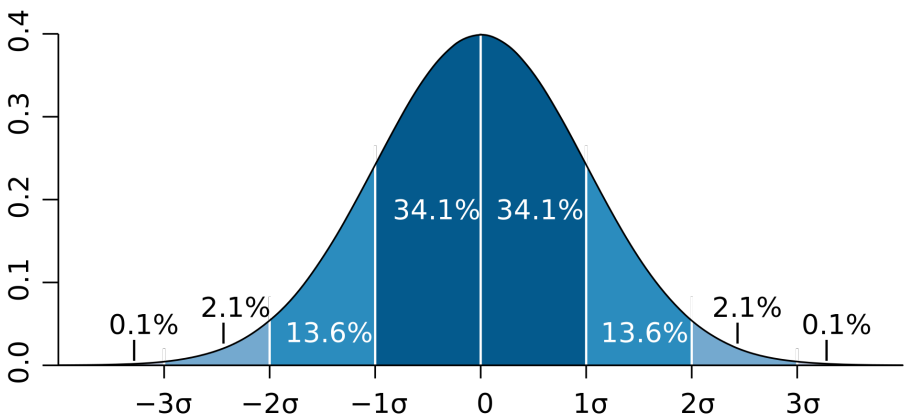

Figure 8: 3-Sigma rule
Source: https://en.wikipedia.org/wiki/Normal_distribution#/media/File: Standard_deviation_diagram.svg

Therefore, it may not be an appropriate model when one expects a significant fraction of outliers (values that lie many standard deviations away from the mean) and least squares and other statistical inference methods that are optimal for normally distributed variables often become highly unreliable when applied to such data. In those cases, a more heavy-tailed distribution should be assumed and the appropriate robust statistical inference methods applied.

A lot of things are normally distributed, we made an example mentioning height, but this can be replicated with many observations from the real world. I will present other two examples that clearly show this concept: Stock market and shoes size.

For stock returns, the standard deviation is often called volatility. If returns are normally distributed, more than 99 percent of the returns are expected to fall within the deviations of the mean value. Such characteristics of the bell-shaped normal distribution allow analysts and investors to make statistical inferences about the expected return and risk of stocks.

It's easy to see that the vast majority of the shoe sizes are between 7 and 9 (about 75%) and then if we move to the edges the percentage tends to lower

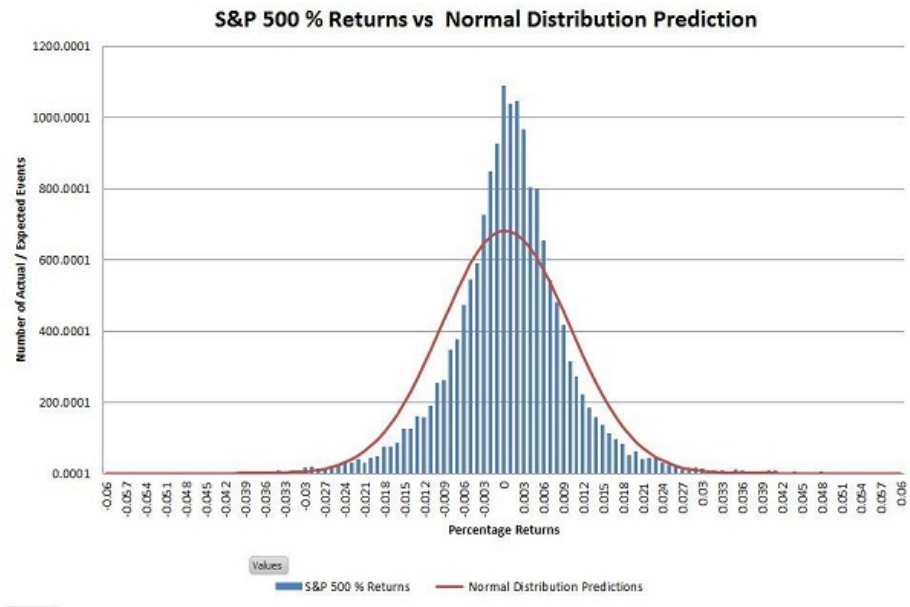

Figure 9: Stock market
Source: https://studiousguy.com/wp-content/uploads/2019/04/Stock-Market-Return-normal-distribution.jpg

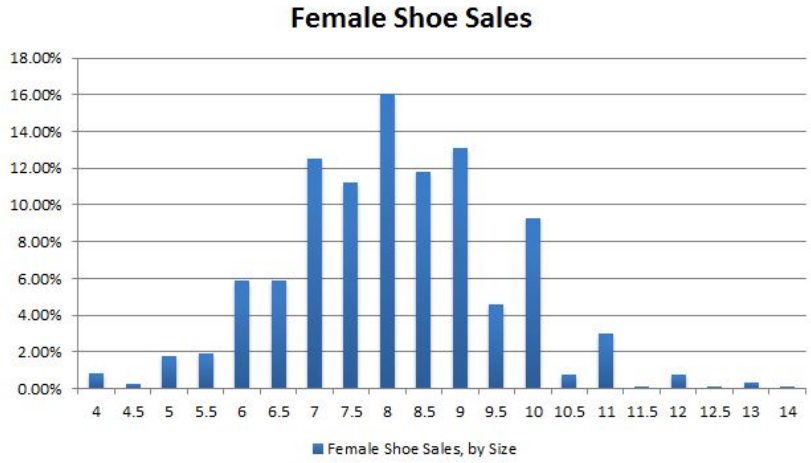

Figure 10: Shoe sales by size
Source: https://studiousguy.com/wp-content/uploads/2019/04/Shoe-chart-normal-distribution.jpg

values till we encounter outliers, but be aware that the theoretical distribution is different from the empirical one! Look at the size 4 (36 in Europe), 1% of the population bought it; instead for the size 4.5 the percentage is near zero (The same works for size 12).

5 Derived distributions

From the Normal Distribution we can derive other important distributions. One of these is the "Chi Square" Distribution. In Probability theory and Statistics, the chi-squared distribution (also chi-square) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. The chi-squared distribution is a special case of the gamma distribution and is one of the most widely used probability distributions in inferential statistics, notably in hypothesis testing and in construction of confidence intervals. This distribution is sometimes called the central chi-squared distribution, a special case of the more general noncentral chi-squared distribution. The chi-squared distribution is used in the common chi-squared tests for goodness of fit of an observed distribution to a theoretical one, the independence of two criteria of classification of qualitative data, and in confidence interval estimation for a population standard deviation of a normal distribution from a sample standard deviation [14]. If $Z1, ..., Zk$ are independent, standard normal random vari-

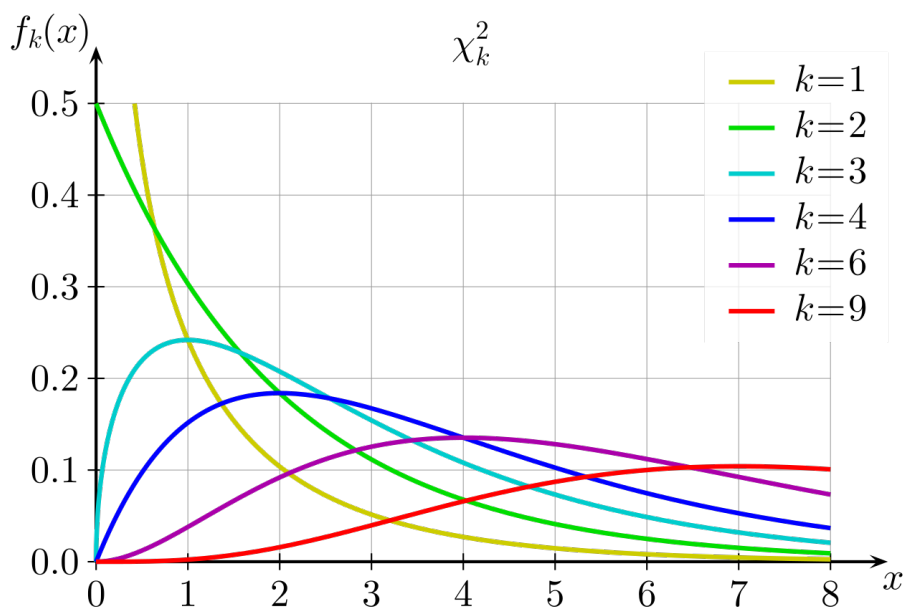

Figure 11: Chi-square distribution
Source: https://en.wikipedia.org/wiki/Chi-squared_distribution#/media/File Chi-square_pdf.svg

ables, then the sum of their squares, $Q = \sum_{i=1}^{k} Z_i^2$, is distributed according to the chi-squared distribution with k degrees of freedom. This is usually denoted as $Q \sim \chi^2(k)$ or $Q \sim \chi_k^2$. The chi-squared distribution has one parameter: a positive integer k that specifies the number of degrees of freedom (the number of random variables being summed, $Z_i s$). The chi-squared distribution is used

primarily in hypothesis testing, and to a lesser extent for confidence intervals for population variance when the underlying distribution is normal. Unlike more widely known distributions such as the normal distribution and the exponential distribution, the chi-squared distribution is not as often applied in the direct modeling of natural phenomena.
To understand better how we can use the chi-square distribution we can make an example, as always. Let's take 2 dices, both with six sides, but one of them is fair, while the other is (1, 5, 6)-loaded, this means that due to the weight of some sides it is more likely that the result of a roll will be 1, 5 or 6. We take one of them, but we don't know if it's the loaded, or is the fair one. We can play with that dice for six days, then we have to figure out if that dice is loaded or not. We decide to perform 100 rolls per day, for six days, so at the end of the week (sunday is free) we will perform 600 rolls. At the end of the week we have to give our choice being sure at 95%; of course we have two hypothesis: the dice is fair or the dice is not fair. What we have to understand, basically, is if the variation in our observed data simply is due to chance, or if the variation is beyond what random chance should allow. How far can our data vary before we have to reject our null hypothesis and conclude the dice is not fair? We know that the confidence level is 95%, the error probability is less than 0.05. The degrees of freedom in this case is 5 (We have 6 random variables: 6 - 1). Due to these values, the chi-square critical value is $CHIINV(error, df) = CHIINV(0.05, 5) = 11.07$. The critical value of a statistical test is the value at which, for any per-determined probability (p), the test indicates a result that is less probable than p. How to know which is the critical value we need for a specific test? There exists a Chi-Square distribution table in which we can find this exact value, be aware that that value differs depending on the degrees of freedom and the error you want to consider. If you want to find this table just google it (I've used this one: https://programmathically.com/wp-content/uploads/2021/04/Chi-Square-table-1024x964.png) .Such a result is said to be statistically significant at that probability. This means that if our result is greater than 11.07, then we must reject the first hypothesis and claim that the dice is not fair.

Number	Expected frequency	Observed
1	100	111
2	100	90
3	100	81
4	100	102
5	100	124
6	100	92

In this table we can observe the expected frequency, so the values we expect to see (but of course the reality is different from a simulation), and the observed values, so the actual frequencies.
Our first step is to subtract the expected values from the observed ones (so as example in the first case we have $(111 - 100) = 11$). The next step is take the

results of the previous operation and square them. The third step is take the previous results and divide each of them by the expected value (to normalize the values, the expected value is always 100, so it's quite easy). The next step is just sum up all the values we got: $1.21 + 1 + 3.61 + 0.04 + 5.76 + 0.64 = 12.26$. 12.26 is greater than 11.07, so we have to reject the first hypothesis and claim that the dice is not fair. The observed frequency of each number on the dice differed significantly from what would be expected on a theoretically fair dice, $CHI - SQ(5, N = 600) = 12.26, p <= 0.05$. Our problem was 12.26. Our critical value was 11.07. Therefore our variation was too great to be explained by chance alone. Therefore we must reject first Hyp. and accept the second one. In conclusion: we are 95% confident that we have the loaded dice [16].
Another derived distribution important to be mentioned is the F-distribution, also known as Snedecor's F distribution. It's a continuous probability distribution that arises frequently as the null distribution of a test statistic, most notably in the analysis of variance (ANOVA) and other F-tests [15]. Suppose U_1 has a X^2 distribution with v_1 degrees of freedom and U_2 has a X^2 distribution with v_2 degrees of freedom. We say that $F = \frac{X_1/v_1}{X_2/v_2}$ follows a F-distribution with a v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator. The F-distribution often arises when we are working with ratios of variances. Notice in the Figure 9 that numerator and denominator are not commutative, the Pdf (Probability density function) with 1 as numerator and 2 as denominator is different with respect to the one having 2 as numerator and 1 as denominator.
The last distribution we see is the "Student's t-distribution". In probability and statistics, Student's t-distribution (or simply the t-distribution) is any member of a family of continuous probability distributions that arise when estimating the mean of a normally distributed population in situations where the sample size is small and the population's standard deviation is unknown. It was developed by English statistician William Sealy Gosset under the pseudonym "Student". The t-distribution plays a role in a number of widely used statistical analyses, including Student's t-test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervals for the difference between two population means, and in linear regression analysis. Student's t-distribution also arises in the Bayesian analysis of data from a normal family. If we take a sample of n observations from a normal distribution, then the t-distribution with $\nu = n - 1$ degrees of freedom can be defined as the distribution of the location of the sample mean relative to the true mean, divided by the sample standard deviation, after multiplying by the standardizing term $\sqrt{n}$. In this way, the t-distribution can be used to construct a confidence interval for the true mean. The t-distribution is symmetric and bell-shaped, like the normal distribution. However, the t-distribution has heavier tails, meaning that it is more prone to producing values that fall far from its mean. This makes it useful for understanding the statistical behavior of certain types of ratios of random quantities, in which variation in the denominator is amplified and may produce outlying values when the denominator of the ratio

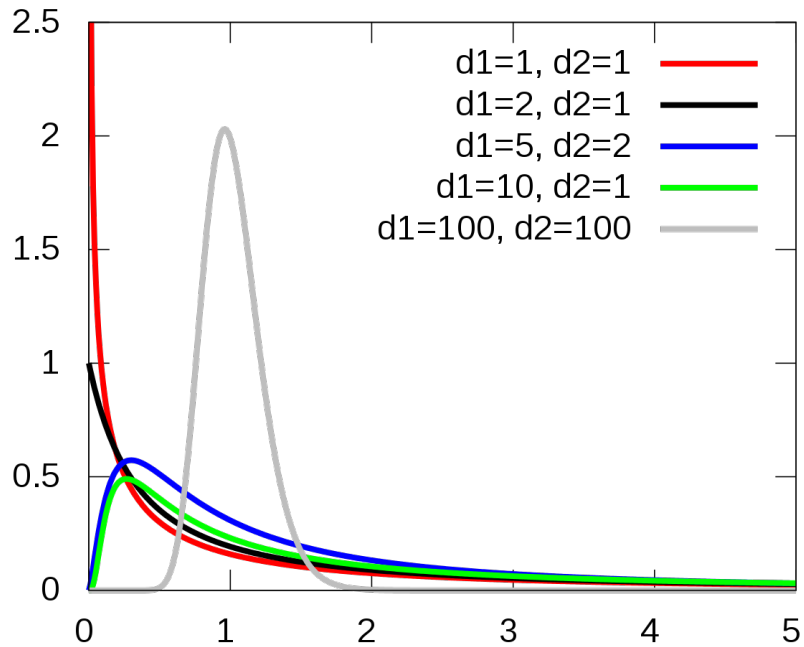

Figure 12: F-distribution
Source: https://en.wikipedia.org/wiki/F-distribution#/media/File:F-distribution_pdf.svg

falls close to zero. The Student's t-distribution is a special case of the generalised hyperbolic distribution [19]. The t-test was developed by William Sealy Gosset, he worked at the Guinness brewery over a hundred years ago and he developed this test to determine things like the difference between barley yield. Let's imagine we have two fields of barley: F_1 and F_2, and I want to compare them but I don't want to cut down the whole field, let's take just some samples. The two distributions will follow a normal distribution, each of them having a specific mean and standard deviation. We could say that the average of the second sample is greater than the other one, but that's only part of the picture; depending on the variance we could have a significant difference or not. That's when the t-value comes into play, it's literally a signal to noise ratio. The signal will be calculated as the difference between the two means, and the noise as the variability of the groups. This is the formula we will use:

$$t - value = (\bar{X}_1 - \bar{X}_2)/\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}$$

These are the samples:

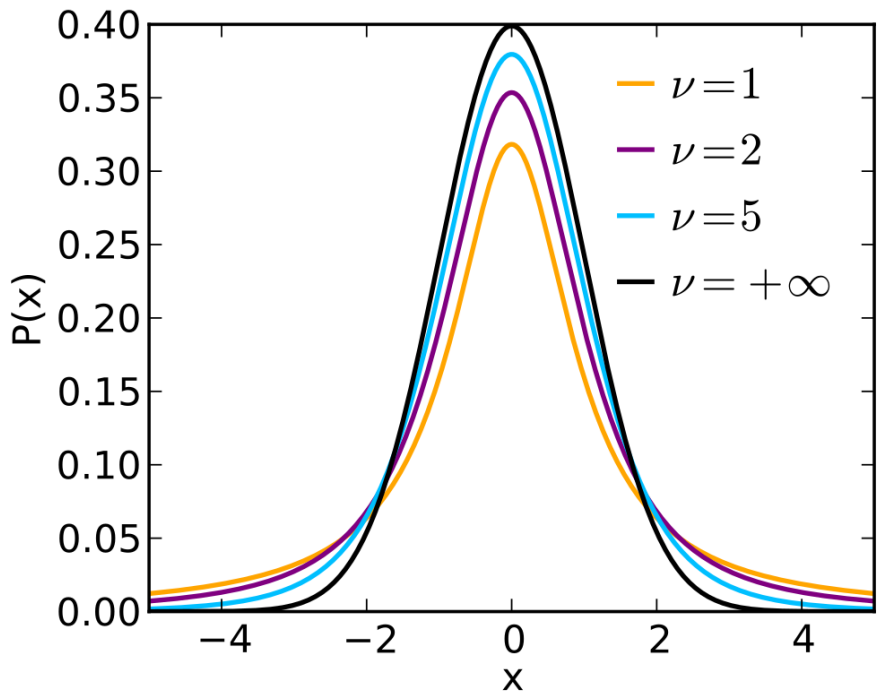

Figure 13: Student's t-distribution
Source: https://en.wikipedia.org/wiki/Student%27s_t-distribution#/media/File:Student_t_pdf.svg

Field 1	Field 2
15.2	15.19
15.3	15.9
16.0	15.2
15.8	16.6
15.6	15.2
14.9	15.8
15.0	15.8
15.4	16.2
15.6	15.6
15.7	15.6
15.5	15.8
15.2	15.5
15.5	15.5
15.1	15.5
15.3	14.9
15.0	15.9

The mean of the first sample is 15.38125, while for the second sample is 15.68125. The standard deviation for the first sample is 0.31245, while for the second sample is 0.40697. The variance for the first sample is 0.097625, while for the second sample is 0.165625. Using the formula we get $0.3/0.13 = 2.3$. Since this value is higher than 1, it means there is more signal than noise. Like in the chi-square test, we take an hypothesis; in this case our first hypothesis is that there is no significant difference between the two fields, on the contrary, the second hypothesis is that there is a significant difference between the two. To do so, we just need a t-test table (Google it to find a bunch of them, I've used this one https://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf) and figure out some values. We will use a p value of 0.05 ($p = 0.05$), since we are in the inferential statistics, this means that if we repeat the test one hundred times we will reject 95 times the null hypothesis. Which row we should choose? Since we used 16 samples per field, the degrees of freedom will be 30 $(16 + 16 - 2)$. Our critical value is 2.04, and since our t-value is 2.3, this means we will reject the null hypothesis, so there is a significant difference between the two fields [20].

6 Implementation

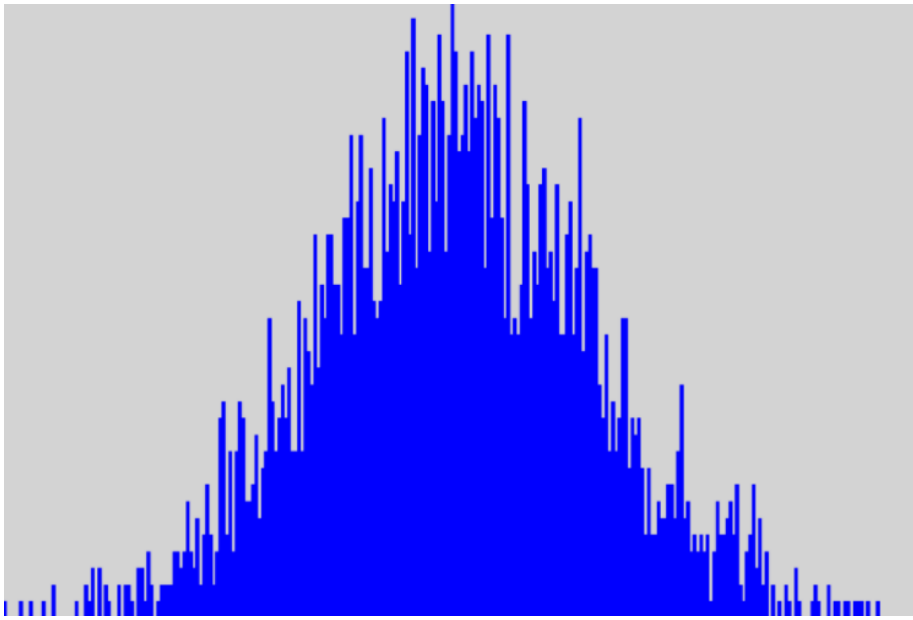

Figure 14: Normal Distribution;Source: My homeworks

```
// == PLOT EXP. NORMAL DISTRIBUTION ==
        private void button4_Click(object sender, EventArgs
            e)
        {
        /*
        Check the input before
        */
            if (checkInput4())
            {
                initGraphics();
                g.DrawRectangle(Pens.Black, viewport);
                g.FillRectangle(Brushes.LightGray,
                    viewport);
                observations = new List<double>();
                richTextBox1.Text = "";
                richTextBox1.Text += "N: " + n + "\n";
                for (int i = 0; i < n; i++)
                {
                /*
                Marsaglia polar method
                */
                    double u1 = 1.0 - rand.NextDouble();
```

```
                    //uniform(0,1] random doubles
                double u2 = 1.0 - rand.NextDouble();
                double randStdNormal = Math.Sqrt(-2.0 *
                    Math.Log(u1)) * Math.Sin(2.0 *
                    Math.PI * u2); //random normal(0,1)
                /*
                     random normal(mean,stdDev^2)
                */
                double trial = 0 + 1 * randStdNormal;
                trial = Math.Exp(trial);
                observations.Add(trial);
                richTextBox1.Text += trial + "\n";
            }
            intervals =
                continuousDistribution(observations,
                observations.Count);
            preparePrint(intervals);
            drawHistograms(intervals);
            // Update the pictureBox
            pictureBox1.Image = b;
        }
    }
```

Listing 1: C# Normal Distribution example

This is a snippet of code I have used during our homeworks in the Statistics course to draw a Normal Distribution in a pictureBox.

We have many methods to generate a normal random variable, we will see some of them with their respective snippets of code. The Marsaglia polar method is a pseudo-random number sampling method for generating a pair of independent standard normal random variables:

```
public static float NextGaussian() {
    float v1, v2, s;
    do {
        v1 = 2.0f * Random.Range(0f,1f) - 1.0f;
        v2 = 2.0f * Random.Range(0f,1f) - 1.0f;
        s = v1 * v1 + v2 * v2;
    } while (s >= 1.0f || s == 0f);
    s = Mathf.Sqrt((-2.0f * Mathf.Log(s)) / s);

    return v1 * s;
}
```

Listing 2: Marsiglia Polar method

The Box–Muller transform, by George Edward Pelham Box and Mervin Edgar Muller, is a random number sampling method for generating pairs of independent, standard, normally distributed (zero expectation, unit variance) random numbers, given a source of uniformly distributed random numbers. The

method was first mentioned explicitly by Raymond E. A. C. Paley and Norbert Wiener in 1934. Be aware that the Marsiglia Polar method is more efficient that this one (Source: https://en.wikipedia.org/wiki/Marsaglia_polar_method).

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
np.random.seed(521)
U1 = np.random.uniform(size = 1000)
U2 = np.random.uniform(size = 1000)
R = np.sqrt(-2 * np.log(U1))
Theta = 2 * np.pi * U2
X = R * np.cos(Theta)
Y = R * np.sin(Theta)
fig,(ax1,ax2) = plt.subplots(1,2)
temp = ax1.hist(X)
ax1.set_title("X")
temp = ax2.hist(Y)
ax2.set_title("Y")
plt.show()
```

Listing 3: Box muller method

The fastest method is the Ziggurat algorithm, but I never used that in my homeworks. From Wikipedia: "The ziggurat algorithm is an algorithm for pseudo-random number sampling. Belonging to the class of rejection sampling algorithms, it relies on an underlying source of uniformly-distributed random numbers, typically from a pseudo-random number generator, as well as precomputed tables. The algorithm is used to generate values from a monotonically decreasing probability distribution. It can also be applied to symmetric unimodal distributions, such as the normal distribution, by choosing a value from one half of the distribution and then randomly choosing which half the value is considered to have been drawn from. It was developed by George Marsaglia and others in the 1960s.". A C implementation of this algorithm can be found here https://people.sc.fsu.edu/∼jburkardt/c_src/ziggurat/ziggurat.html.

7 Conclusions

We encountered a lot of new concepts in this document: Sample space, probability of an event, measures of central tendency (mean, median and mode), variance, standard deviations...
We also saw the concepts behind Normal and its derived distributions. I hope the examples made to explain the concepts were clear and helped while reading. I think these concepts are very useful in real life, just because as we said in the section 3 and 4, the normal distribution is everywhere and a lot of things naturally follow a normal distribution. To go deeper in the field I suggest to take a look at the references (next page), I think they are very valuable resources, both from a mathematical perspective and also for personal curiosity. There are many videos (free, on Youtube) that visually explain what is written here and also give real world examples that can aid in learning.

References

[1] StatisticsHowTo - Dice Roll Probability: 6 Sided Dice
https://www.statisticshowto.com/probability-and-statistics/probability-main-index/dice-roll-probability-6-sided-dice/

[2] Prof. Essa - Probability: Dice Rolling Examples
https://www.youtube.com/watch?v=FM-3yQtSbxQ

[3] Wikipedia - Probability
https://en.wikipedia.org/wiki/Probability

[4] Wikipedia - Probability Distribution
https://en.wikipedia.org/wiki/Probability_distribution

[5] Wikipedia - Mean
https://en.wikipedia.org/wiki/Mean

[6] Wikipedia - Median
https://en.wikipedia.org/wiki/Median

[7] Wikipedia - Mode
https://en.wikipedia.org/wiki/Mode_(statistics)

[8] Wikipedia - Standard Deviation
https://en.wikipedia.org/wiki/Standard_deviation

[9] Wikipedia - Abraham de Moivre
https://en.wikipedia.org/wiki/Abraham_de_Moivre

[10] Wikipedia - Normal distribution
https://en.wikipedia.org/wiki/Normal_distribution

[11] Wikipedia - Central limit theorem
https://en.wikipedia.org/wiki/Central_limit_theorem

[12] StatQuest with Josh Starmer
https://www.youtube.com/watch?v=rzFX5NWojp0

[13] 365DataScience - Understanding the Normal Distribution
https://www.youtube.com/watch?v=IhtmW28slDw

[14] Wikipedia - Chi-squared distribution
https://en.wikipedia.org/wiki/Chi-squared_distribution

[15] Khan Academy - Chi-square distribution introduction
https://www.youtube.com/watch?v=dXB3cUGnaxQ

[16] Brandon Foltz - Statistics 101: Introduction to the Chi-square Test
https://www.youtube.com/watch?v=SvKv375sacA

[17] Wikipedia - F-distribution
https://en.wikipedia.org/wiki/F-distribution

[18] jbstatistics - An Introduction to the F Distribution
https://www.youtube.com/watch?v=G_RDxAZJ-ug

[19] Wikipedia - Student's t-distribution
https://en.wikipedia.org/wiki/Student%27s_t-distribution

[20] Bozeman Science - Student's t-test
https://www.youtube.com/watch?v=pTmLQvMM-1M

[21] Alan Zucconi - How to generate Gaussian distributed numbers
https://www.alanzucconi.com/2015/09/16/how-to-sample-from-a-gaussian-distribution/

[22] Wikipedia - Marsiglia Polar Method
https://en.wikipedia.org/wiki/Marsaglia_polar_method

[23] Wikipedia - Box Muller transform
https://en.wikipedia.org/wiki/Box%E2%80%93Muller_transform

[24] Statistics by steps - Box-Muller Algorithm by Python
https://rh8liuqy.github.io/Box_Muller_Algorithm.html