**Continuous Random Variables**

Table of Contents

[Probability Models for Continuous Random Variables 3](#_Toc63620810)

[PMF 3](#_Toc63620811)

[CDF 5](#_Toc63620812)

[PDF 6](#_Toc63620813)

[Properties of PDFs 11](#_Toc63620814)

[Properties of CDFs 11](#_Toc63620815)

[Expectations and Variances 15](#_Toc63620816)

[Well-Known Continuous Random Variables 16](#_Toc63620817)

[Uniform Continuous Random Variables 16](#_Toc63620818)

[Exponential Random Variable 20](#_Toc63620819)

[Gamma Random Variables 24](#_Toc63620820)

[Gaussian Random Variables 27](#_Toc63620821)

[Beta Random Variables 35](#_Toc63620822)

For a continuous random variable, the number of possible values is uncountable, and thus is infinite. For example, if we were to select a random point on a line between to , it could have infinite possible values, or when a person arrives, we can state the time to hours, to minutes, to seconds, to milliseconds and so on.

If is a continuous random variable, (exclusive) or (inclusive). Notice that unlike with discrete random variables, we are not giving a set of specific values, since it is literally not possible. Essentially, we can be infinitely more specific.

## Probability Models for Continuous Random Variables

### PMF

For discrete random variables, we used PMF and CDF as probability models. However, PMF is not applicable for continuous random variables. This is because , always, for continuous random variables.

We can prove this. Consider, for simplicity, that all values for a continuous random variable are equally likely. We know that in this case. Since there are infinite possible values, , meaning is always .

More formally, PMF is nothing more than distributing unit of mass on a number line. If we have an infinite number of points on the number line, then the allocation per point is .

Consider a situation where we have a line of length . We will pick a random point on that line. Here, , so is a continuous random variable and . We know that .

Now consider the same situation, except that we have discretised the line, dividing it into equal parts. Thus, we can define a random variable , and . Due to this difference, is a discrete random variable, and , given all the parts are equally likely to be chosen.

There is a relationship between and here. Say and we pick the point . . In fact, for all values of between and (notice that this is recurring), . The question is, how well does approximate ? That depends on how large the value of is. The larger the value of , the more accurately can be approximated by .

We can see that, for (the PMF of can also be represented like this apparently) and }, many values of are represented by a single value of . Thus, . As such,

Again, as we increase the value of ,

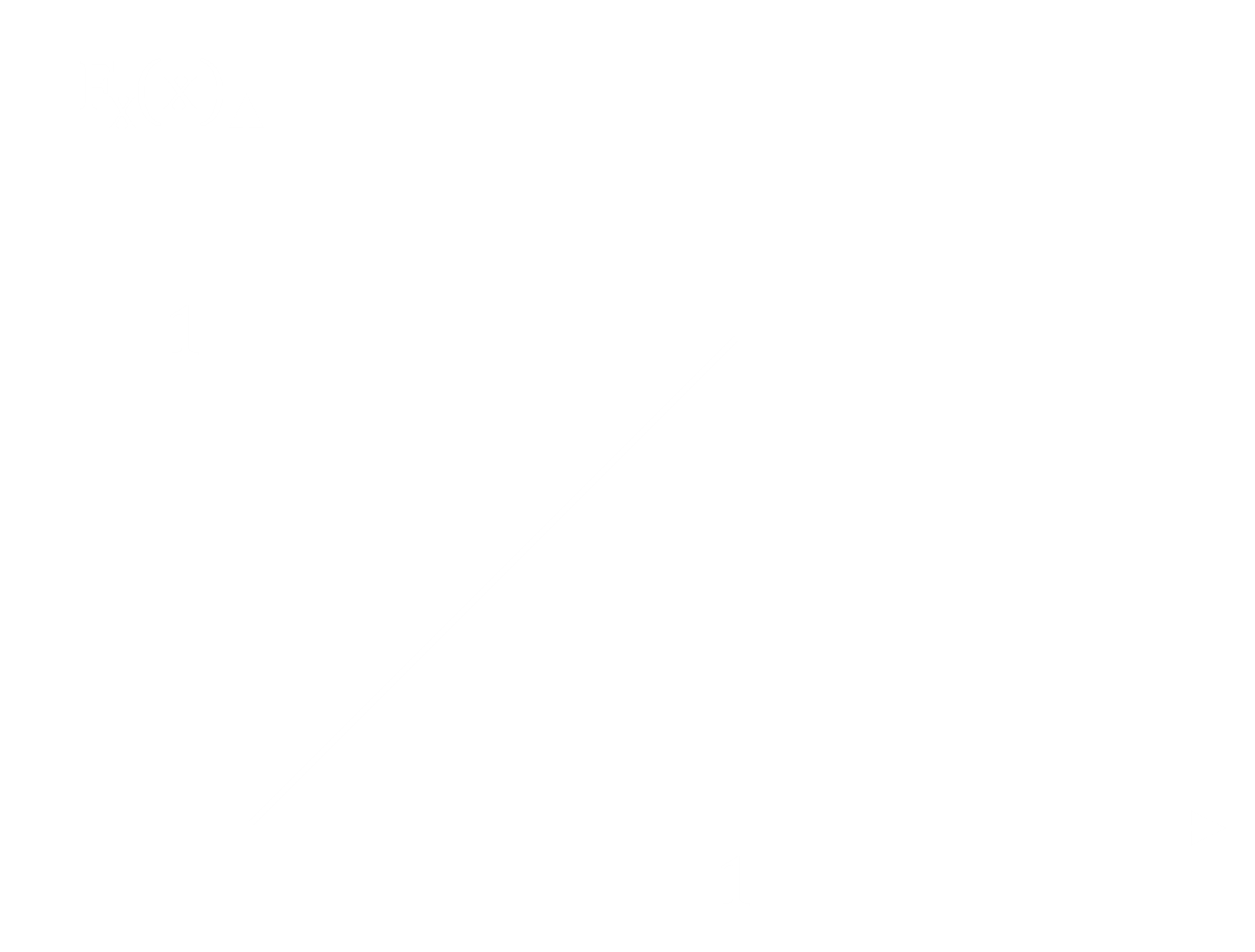
The very definition of the PMF probability model states that . Thus, the only takeaway we have from this is that for continuous random variables, . This was the mathematical proof of the same thing we discovered one whole page and a headache ago.

I don’t know if you’ve noticed but, we’ve kind of been using PMF a lot. It going poof here is a bit a problem for us. We stated that one of the goals of using random variables was to facilitate further processing, and PMFs have been helping us a lot with that. Now that they’re gone, we need to find something to replace them, and we will do exactly that. Wait a bit though.

### CDF

Say we pick the point on the line of meter. Here, . This is because, since the probability of picking any of the points is the same, we will be picking one of the points to the left of in half of the cases and one of the points to the right in half of the cases. Similarly, , and .

Even if we cannot find the probability for specific values of , we can find the probability for an interval. . As such, CDF can be used with continuous random variables.



A few points to notice:

This last point is correct because and we know that . Following the same logic:

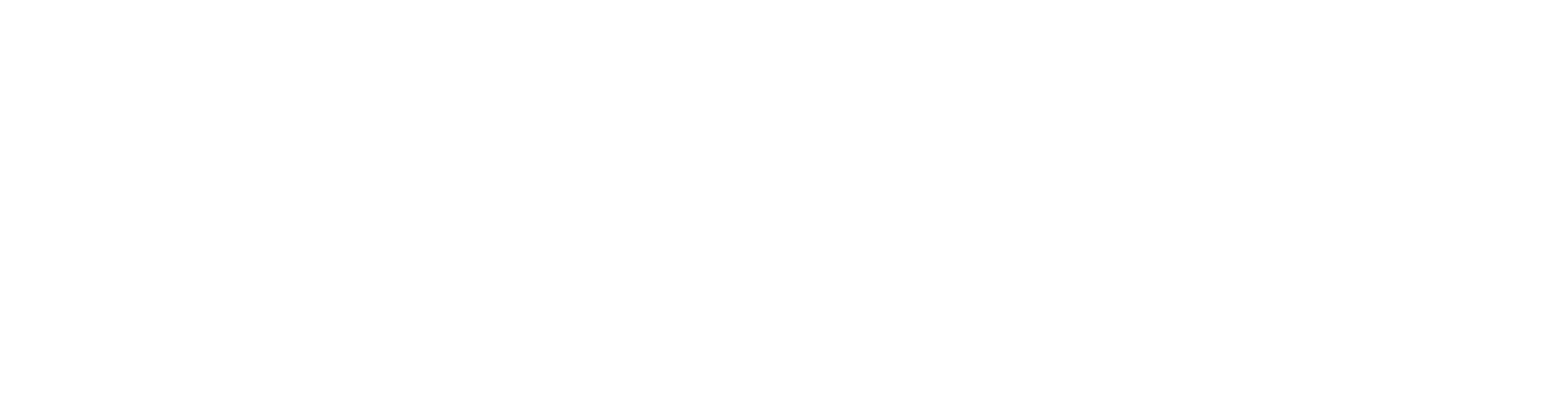
### PDF

Density is the mass per unit volume, but it is also possible to define it as a unit per unit area, for example the density of people per unit area in a field. We can even find examples for valid definitions of density per unit length.

The probability density function (PDF) defines the average probability per unit length.

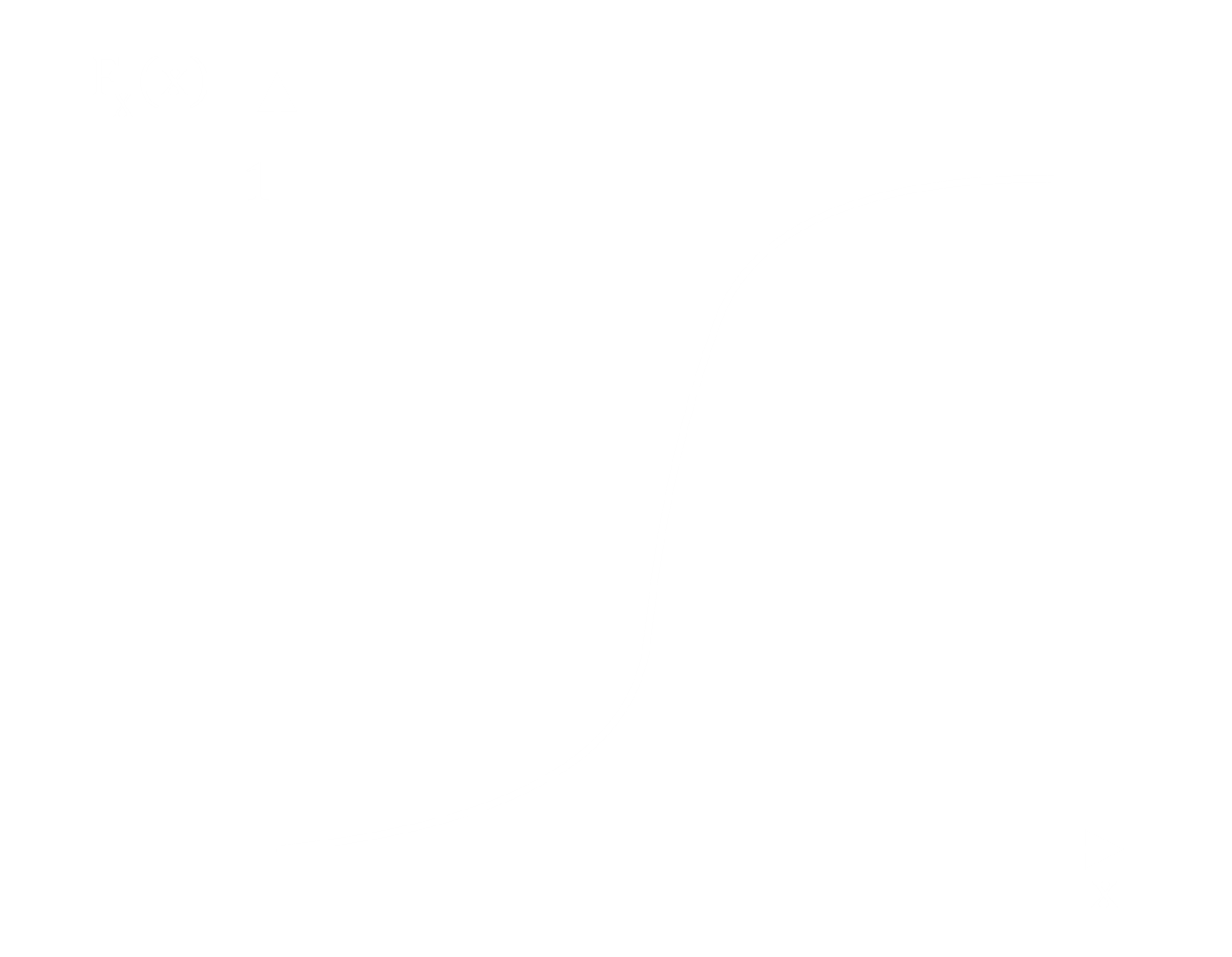
When we first started discussing random variables, we stated that one of its major purposes was to allow further processing of the probabilities of the different outcomes. One of the key ways we achieved this with discrete random variables was by using PMFs. However, as we have seen, PMFs cannot be used with continuous random variables. As such, we need a probability model that can work with continuous random variables in the way that PMFs worked with discrete random variables. This is where Probability Density Functions (PDFs) come in.

Consider two cases of CDFs. In the first case, and in the second case, . If we draw the CDF graphs for both these cases, we will find that they both increase linearly, but the graph for the first case has a much steeper slope.



We can understand this by considering the probability distributions as mass distributions. Thus, in the first case, where has values between and , a total probability of is distributed over a smaller range, thus giving a steeper slope. In the second case, the same total probability of is distributed over a larger range, between and , thus giving a less steep slope.

Now consider a more complex example:



Since we cannot find the probability for a specific value, we will consider the probabilities for two intervals, from to and from to . Notice that we are taking intervals of the same length, just starting at different points, and that is a very small value.

Even though we are taking intervals of the same length, the probability values we find for those intervals are greatly different. This is because the density of probabilities in the latter interval is higher. Thus, the probability of an interval is proportional to the probability density in that interval. This proportionality becomes an equality when we multiply the probability by the length of the interval.

Another indication of the density is the steepness of the CDF curve. A flat area indicates a lower density, while a steep area indicates a higher density and vice versa.

The slope of the CDF curve is of course given by its first derivative, the rate of change. For a specific value of ,

CDF values simply indicate a probability. Unlike PMFs, they indicate the probability over a range, but the value is still a probability. Thus, the first term in the equation above can be thought of as the probability in an interval divided by the length of the interval, the average probability density. From this, we can go back to finding the total probability by multiplying it with the length of the interval.

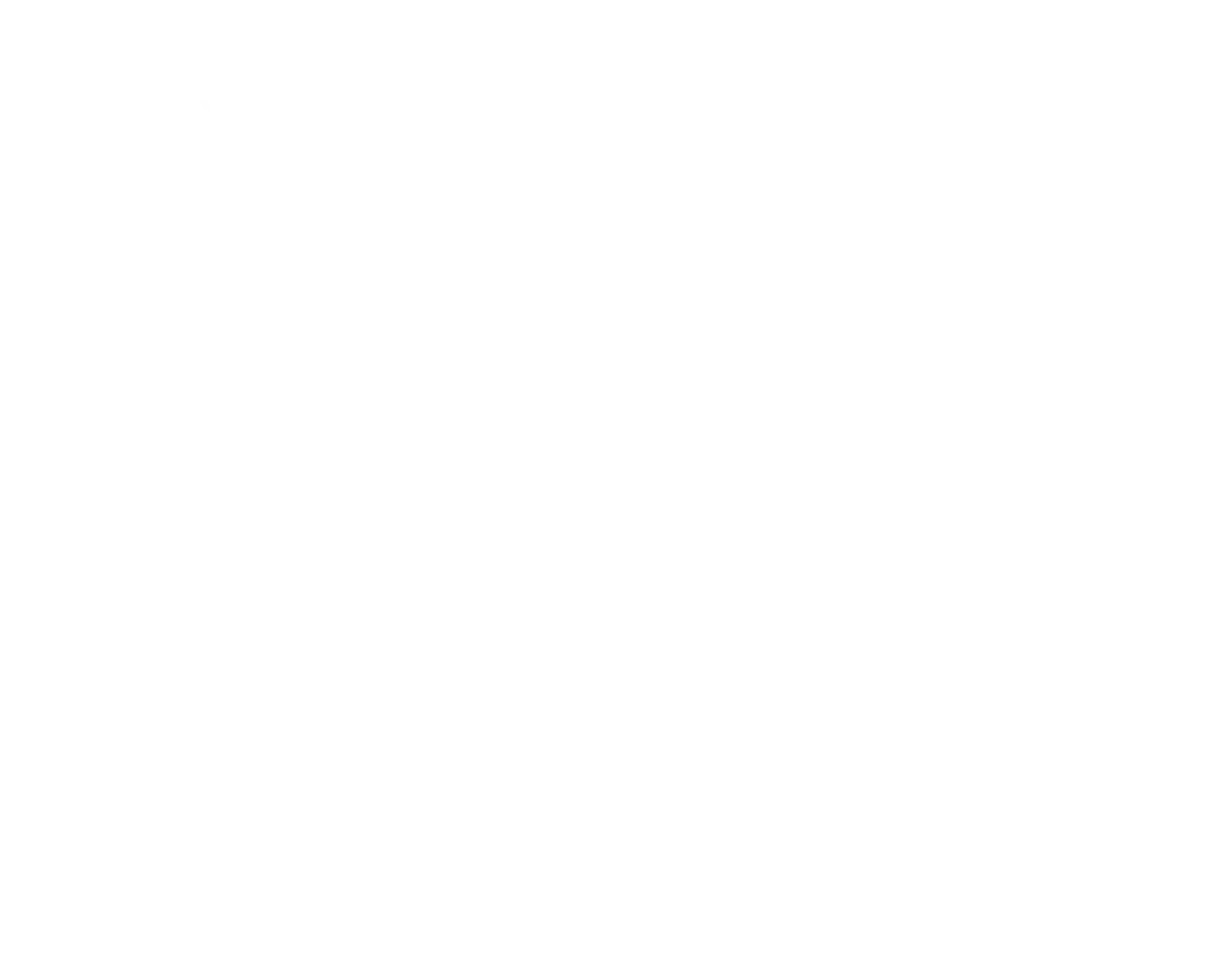
If we use very small values of , we could even go as far as to say the probability at the single point is approximately equal to the probability density for the interval from to . Of course, this approximation will be extremely inaccurate if we use large values of . If , represents the probability density near . This probability density near is denoted by . Thus,

This is the PDF of . The PDF is the probability density per unit length for a random variable near the point .

Always remember that the PDF is not a probability, but rather a probability density. By the axioms of probability, the probability at a point must be between and . However, the probability density at a point can be more than , if there is a high probability within a small interval.

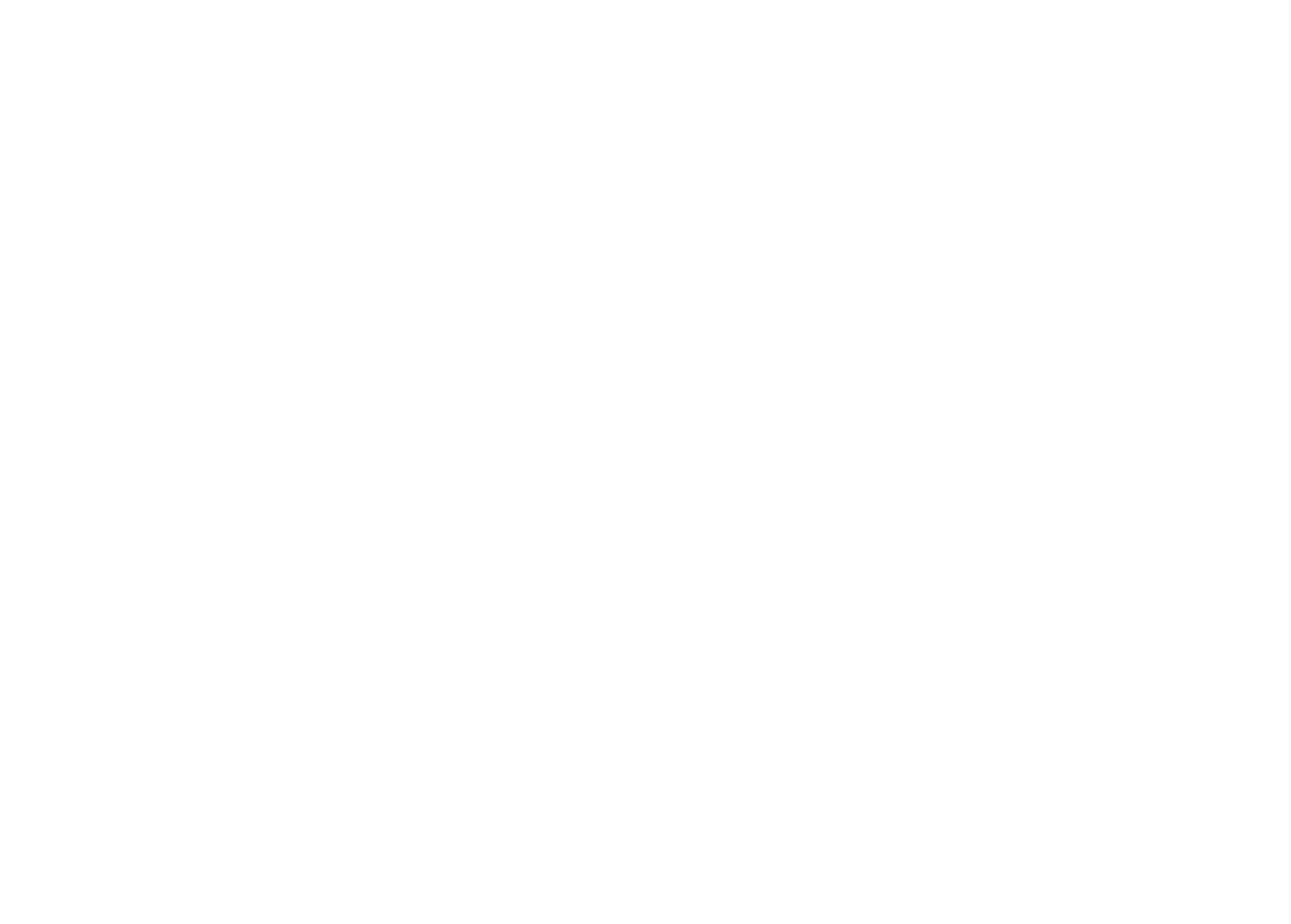
#### PDF Curves

The PDF curve understandably depends on the CDF values. For the case where probabilities are uniformly distributed over an interval, the PDF curve will be flat.



Of course, the area of the PDF curve for the interval will represent the probability in that interval. As such, the total area will be . This means the height of the line is .

For non-uniform distributions, the PDF curve will be curved.



Thus, is given by the area of the curve, which, in this case, can be calculated by . If , this just gives us .

We previously saw an example where . Here, . Inversely, .

Another thing to note is that .

Of course, the PDF formulas are not always simple, and since integration and differentiation have become involved, we will soon start facing problems like having to integrate by parts.

### Properties of PDFs

### Properties of CDFs

* Continuous function, and thus differentiable

Example

We need to find

1. Find the value of ,
2. Derive ,
3. Retrieve the PDF from the CDF, although it is already given,
4. Find using the PDF and the CDF separately.

This problem involves integration by parts ().

i)

ii)

iii)

iv)

Using CDF,

Using PDF,

## Expectations and Variances

For discrete random variables, we defined the expectation as

For continuous random variables, it is defined as

For discrete random variables, we defined variance as

For continuous random variables, it is defined as

Note that the limits for integration are replaced by the given limits for individual problems. has been used for show here.

## Well-Known Continuous Random Variables

### Uniform Continuous Random Variables

We have already seen a uniform random variable for a discrete set of values. In the continuous version of this, the values are given by a set . This set defines the possible values of .

Note that the brackets do not matter for continuous random variable and can be used interchangeably.

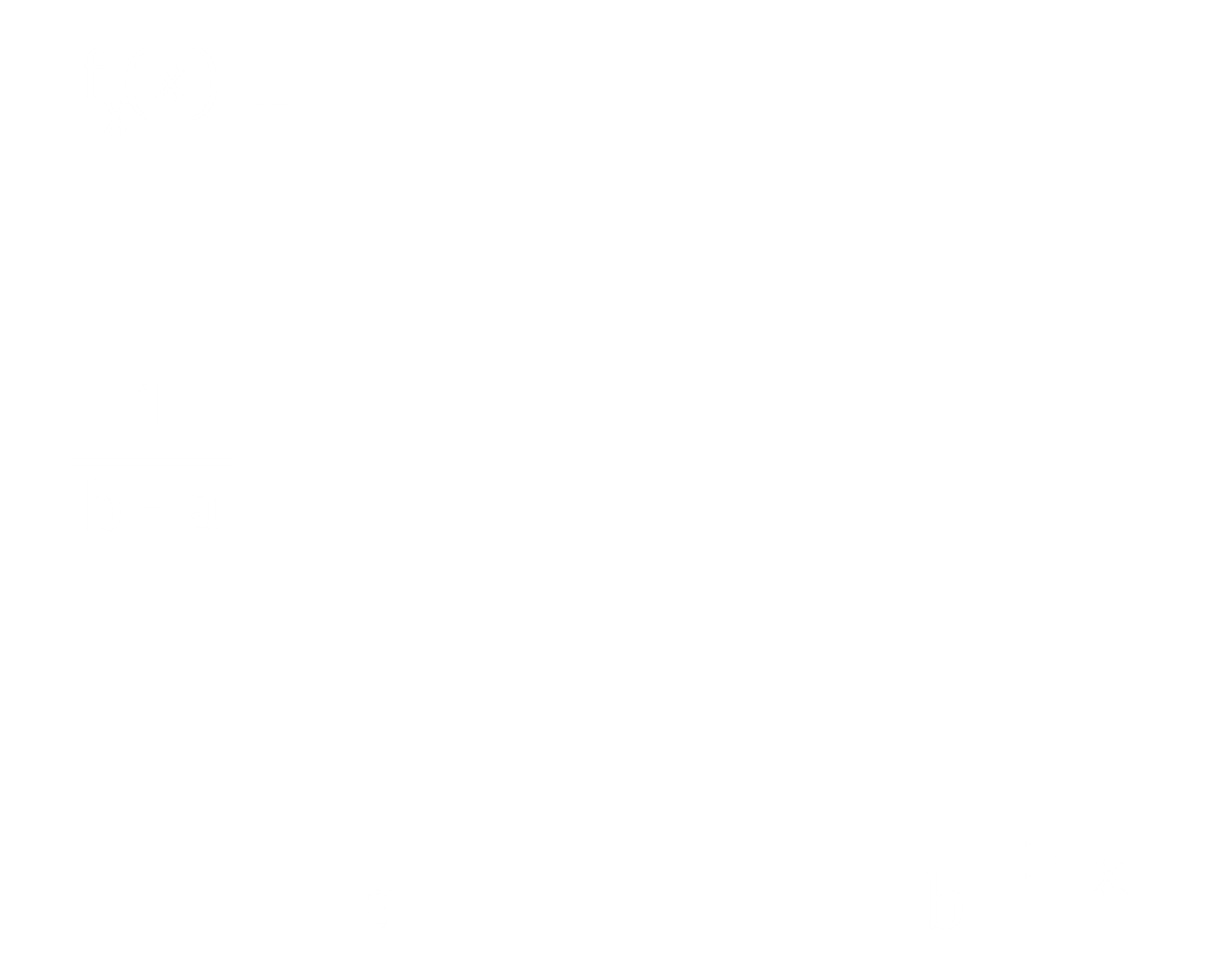
Unlike discrete random variables, we cannot simply define the uniform continuous random variable as one that has an equal probability for every possible value. This is because all values for continuous random variables always have the value , as we have previously discussed. As such, there are two possible definitions:

1. is a uniform continuous random variable if the probability density is constant
2. is a uniform continuous random variable if it is equally likely that a point from any interval within the specified range will be picked. This definition is a little more confusing and is less preferred.

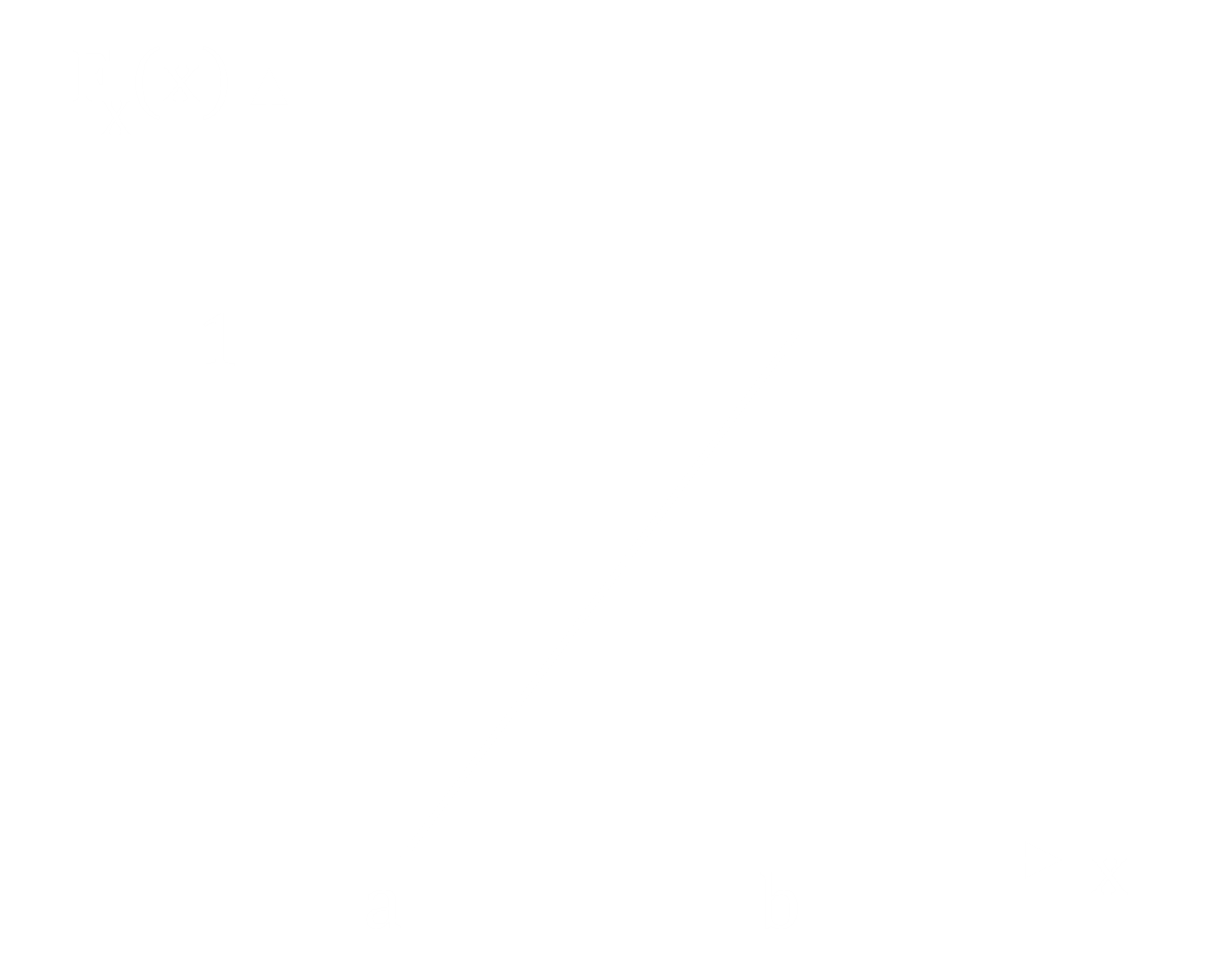
By this definition, the PDF can be defined as:

We know,

Thus, the PDF is given by



Using this, we can calculate the CDF as



For uniform random variables, the expectation is given by

and the variance is given by

Example

Say a student requires between to minutes to travel from home to school. The probability density between this interval is constant. Say the class starts at and the student starts from home at . What is the probability that they reach in time?

Let be a uniform continuous random variable, where . The best possible case is that minutes are needed, which puts the arrival time at , meaning . The worst possible case is minutes are needed, which puts the arrival time at , meaning . Thus, is uniformly distributed between .

An easier way we could have calculated this probability is if we defined the ‘length’ of valid values as , from to , and the ‘length’ of the total range as , from to . Thus,

### Exponential Random Variable

Exponential random variables are related to Poisson distributions. Consider a scenario where events are occurring at a particular Poisson rate, . Say events occur in a time period .

Let , and so on. This can be generalized to

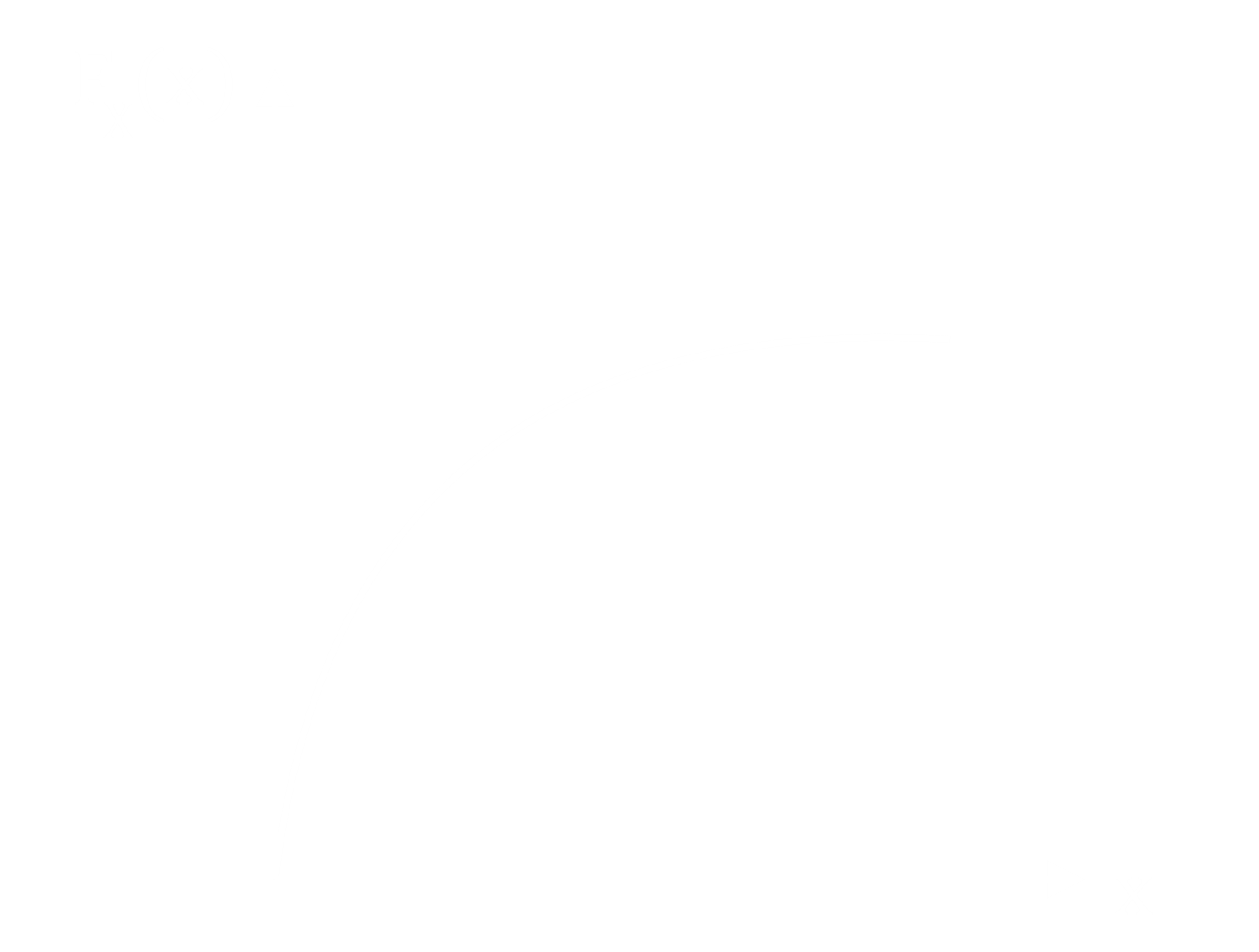
where

Since the events could occur at absolutely any moment, all the we just defined are continuous random variables. For exponential distributions, we want to find the distribution, i.e. the CDF and PDF, of .

We know that . Instead of finding this, we will instead find , the complimentary CDF. The complimentary CDF tells us the probability that the first event occurs after the time . This is the same as the probability that there is no event between time .

counts the number of arrivals in an interval. We know that this is defined by the Poisson distribution. Thus

From this,



We can also find the PDF from this.



The expected value of an exponential random variable is given by

and the variance is given by

The expected value actually tells us something very obvious in this case. If there are events per hour, . Thus, the expected interval time will be hours, or minutes, meaning on average, there will be an arrival every minutes. Note that this is on average. In reality, the arrivals are random.

Exponential random variables can be related to geometric random variables, since in geometric random variables we were interested in the number of Bernoulli experiments before the first success and here we are interested in the time before the first event.

Example

On an average, there is an earthquake every months. Thus, . We need to find the probability that the next earthquake will occur after months, but before months.

Say , meaning . Whenever we are given problems related to some rate, if we have to work with the rate of events occurring then we want a Poisson random variable, and if we have to work with the time interval between events, then we want an exponential random variable.

### Gamma Random Variables

Gamma distributions are a generalization of the exponential distribution, and as such, are also related to Poisson distributions.

Say .

In this case, the PDF is given by

Clearly, this becomes the same as the exponential distribution for .

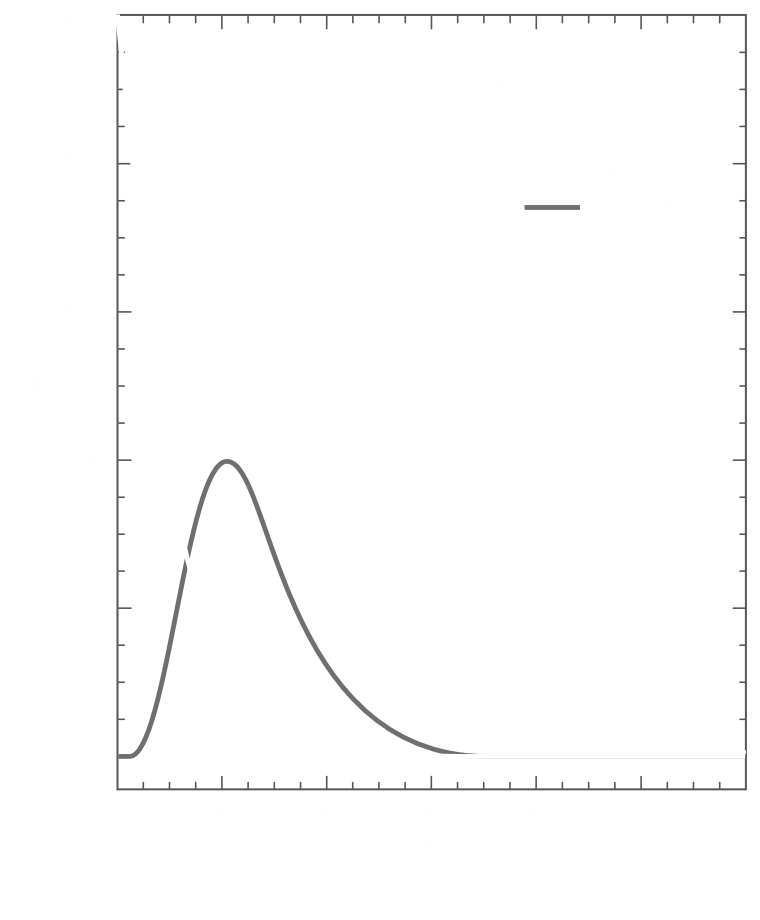
Also, since we have a factorial in the denominator, has to be an integer.

There are situations where is a real number, and in those cases, we need to convert this into a gamma function.

Thus, in general we can say

and

The gamma distribution curves are different from other distributions, due to the existence of two parameters, . Thus, the graph depends on which parameter we keep fixed.



We do not need to remember these graphs.

The expected value of a gamma distribution is given by

and the variance is given by

Gamma distributions are related to negative binomial distributions, since in negative binomial distributions, we are interested in the number of attempts required for successes, and here we are interested in the time until the -th event.

Example

Say customers are arriving at a restaurant at a rate of customers per hour and that the restaurant will start having profits after customers have arrived. We want to find the expected time until this happens.

Let , where .

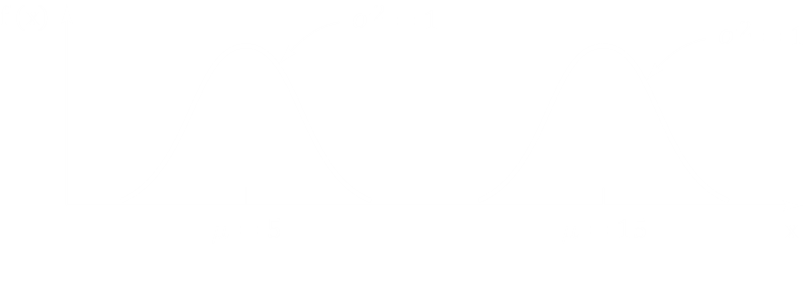
### Gaussian Random Variables

Also called the normal random variable, the gaussian random variable is possibly the most important random variable due its wide usage in statistical inferences. Whenever we are unsure about what the random variable should be in a situation, it is best to use the gaussian random variable.

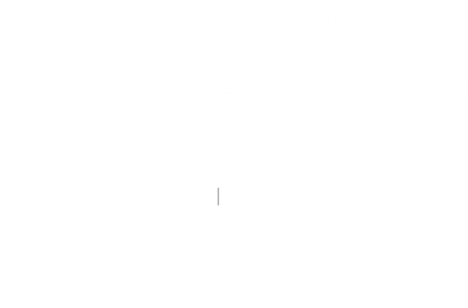
The PDF is given by

, where .

The two parameters are , which is the variance of the random variable and , which is the expectation of the random variable.



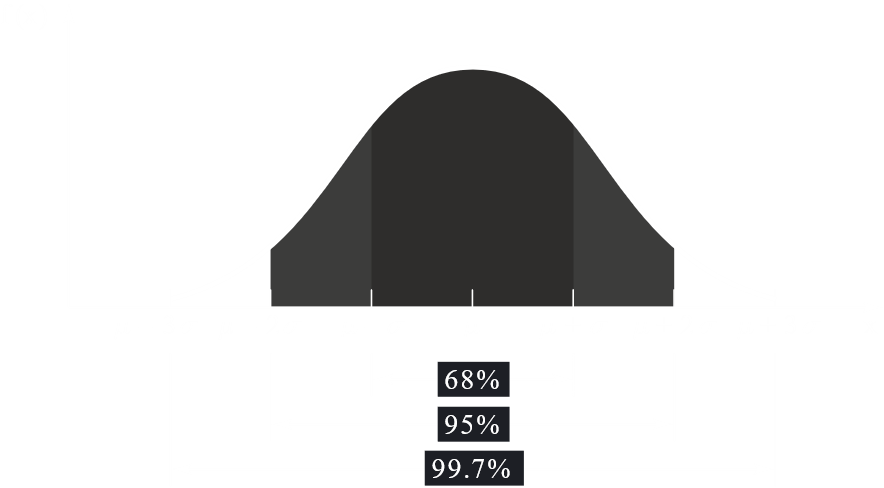
Notice that the curve is symmetric, and if we change the value of , the only thing that happens is that the curve shifts. The shape of the curve is unaffected. However, increasing the value of will flatten the curve.



Note that the width of the curve has to increase if the height decreases, so as to maintain the area under the curve at .

The problem with gaussian random variables is that, if we integrate the PDF, we do not get a closed-form solution. As such, we cannot mathematically define a value for the CDF. This means we cannot find the probability associated with a particular interval using gaussian random variables.

The probabilities in these scenarios are calculated using numerical techniques that we will not cover in this course. Another option is to use tables and the PDF curve.



For any given curve,

We can use tables or computer applications or some specific calculators to calculate the CDF values for Gaussian random variables. Here, we shall be looking into how to use those tables.

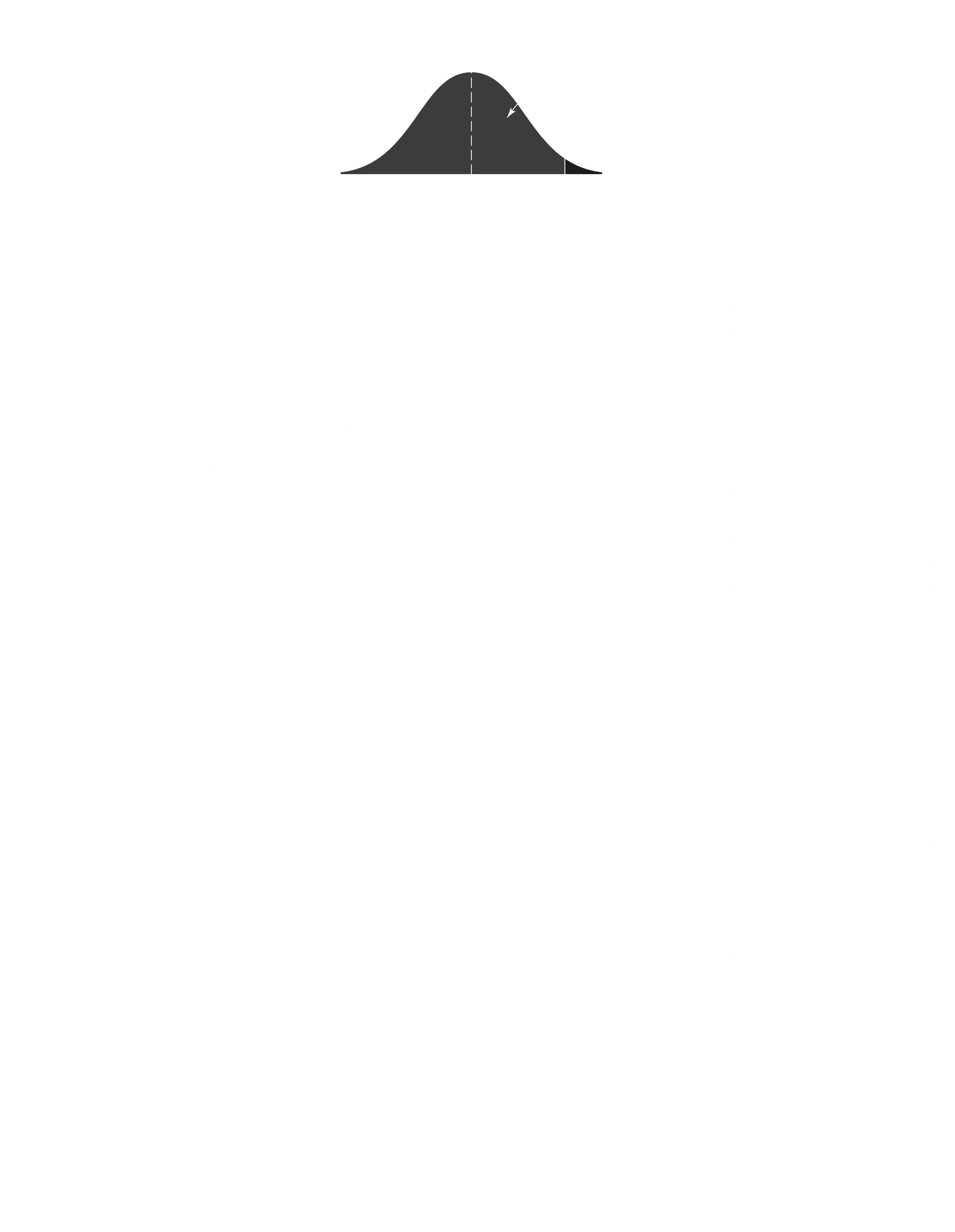
where is the mean and and is the variance and

The table should contain all the CDF values for each pair of and . There could be millions of values like this, and it would be unrealistic to expect a table to include all of the values.

Because of this limitation, we will be considering a special gaussian random variable, . Such a gaussian random variable, which has a mean and a variance, is called a standard normal random variable.

For ,

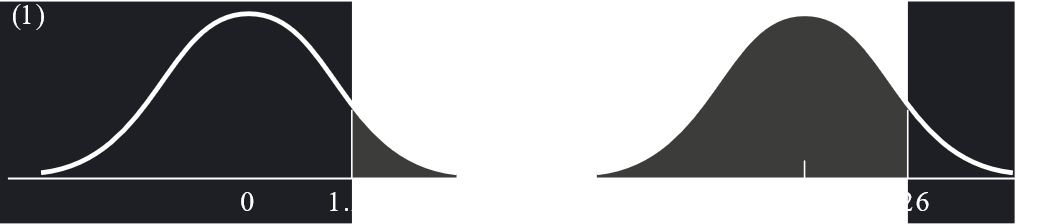
Normally, we would denote the CDF of a random variable as , but the CDF of uses a special character, .



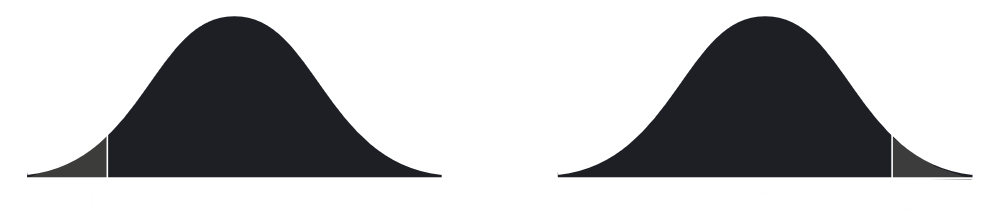
In most cases, the value of in tables is between to . The table is setup in a weird way to save space. Each for indicates the values for some values of and the column specifies the digit in the ths place of . Thus, to find the CDF for , we need to go to the row marked and the column marked , where we find the value .

However. there are no probability values available for negative values of in the table. For negative values, we use the symmetric property of the curve.

Each value from the table tells us the area under the curve from the leftmost point of the curve to the specified value of . Say we are asked to find . This is just . We took the area on the left of and subtracted it from to get the area on the right of .

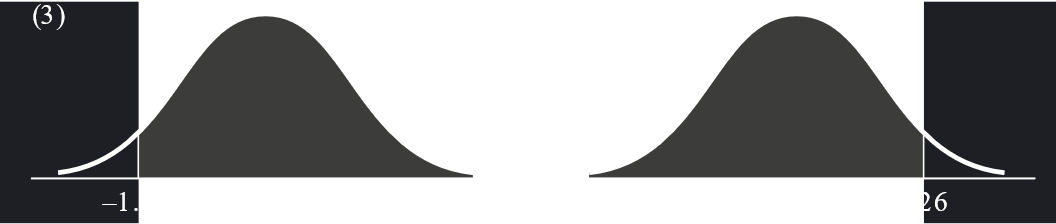


Now, using the symmetric property of the graph, we can tell that the area on the right of is the same as the area on the left of .



Thus, .

Following a similar pattern of thinking, .



A few more cases are:

We are approximating from the table since the exact value is not available in given table. There are two possible values ( and ) and one has been chosen that felt more accurate. If a more accurate table were available that had more accurate values of , the exact value could be found.

Now to get back to non-standard normal random variables, i.e. where and . For these cases, we can do two things. We could either have this huge table for all possible values of and , or we could have some formula that allows us to convert a normal random variable into a standard normal random variable. This formula is

Say is a random variable that is derived from . Any random variable that is derived from a gaussian random variable is also a gaussian random variable. Thus,

For a derived random variable where , two formulae we need to remember are:

Using these formulae,

Thus, we have standardized by using the formula .

Example

Say some data packets we are sending have an average delay that is a normal distribution with and . We want to find the probability that the delay for a specific packet is greater than .

Let be a random variable such that , i.e. .

### Beta Random Variables

For beta random variables, the PDF is defined as

Here,

Beta distributions are used for calculating probabilities that are fractions.

For example, when send data packets, what is the portion of packets that were successfully sent? (I have absolutely no clue what this means.)

The graphs of beta distributions are complex and change depending on the values of and .

Beta distributions will be covered in far more details in later chapters.