**Discrete Random Variables**

Table of Contents

[Random Variables 3](#_Toc63619066)

[Limitations of Simple Probability Models 3](#_Toc63619067)

[Discrete and Continuous Sets 4](#_Toc63619068)

[Definition of Random Variables 5](#_Toc63619069)

[Distribution Functions 8](#_Toc63619070)

[Events Defined by Random Variables 8](#_Toc63619071)

[Probabilities Defined by Random Variables 9](#_Toc63619072)

[Probability Mass Function (PMF) 9](#_Toc63619073)

[Cumulative Distributive Function (CDF) 11](#_Toc63619074)

[Derived Random Variables 14](#_Toc63619075)

[Expectations 15](#_Toc63619076)

[Variances 17](#_Toc63619077)

[Well-Known Random Variables 18](#_Toc63619078)

[Bernoulli Random Variables 19](#_Toc63619079)

[Bernoulli Trials 21](#_Toc63619080)

[Geometric Distributions 24](#_Toc63619081)

[Binomial Distributions 26](#_Toc63619082)

[Negative Binomial Distributions 27](#_Toc63619083)

[Uniform Distributions 27](#_Toc63619084)

[Hypergeometric Distributions 28](#_Toc63619085)

[Poisson Distributions 29](#_Toc63619086)

[Truncated Geometric Distribution 31](#_Toc63619087)

[Expected Values of Well-Known Discrete Random Variables 32](#_Toc63619088)

## Random Variables

Random variables are key to this course, and understanding them is a core part of this course.

### Limitations of Simple Probability Models

We use random variables to define the probability models associated with a random experiment in a more sophisticated manner. We know how to define simple probability models, but if we look at these probability models a little closely, we will notice some limitations. The limitations are not severe, but they still exist.

The limitations include:

* Regarding the elements of the sample space
  + The element could be anything at all, including non-numeric objects
  + Further processing is not possible due to the above point
* Regarding the probability model as a whole
  + It is not compact
  + The sample space and probabilities are defined separately

Random variables help us overcome these limitations.

### Discrete and Continuous Sets

Random variables can be discrete, or continuous. It can also be mixed, but we will not be studying mixed random variables here. Before we can define the two types of random variables, we need to see what discrete sets and continuous sets are.

In a discrete set, the number of elements is finite, or at least countable for an infinite number of elements. For example, a random experiment that tests if a data packet has been sent successfully or not has a finite number of elements in its sample space. Another example could be the number of attempts needed to send a data packet successfully. There are an infinite number of elements in the sample space for this random experiment, but the elements are countable.

In a continuous set, the number of elements is infinite and uncountable. For example, consider a line from to , on which we have to choose a point. There are an infinite number of points we can choose, and the values can get infinitely more precise, making the set uncountable. A continuous set like this one is represented as , which signifies all the values between and . There are a few other representations, but we will look into those in the next chapter.

Obviously, the random variables that apply to each of these types of sets are called discrete random variables and continuous random variables respectively.

### Definition of Random Variables

The first limitation with simple probability models was the fact that the sample space did not necessarily consist of numbers, which made further processing difficult. For example, consider that our observation is the sum of two consecutive coin tosses. This would force us to deal with ordered pairs, which is problematic. Thus, we need a process to convert these non-numeric elements into numbers.

A random variable is a real valued function that converts each element of a sample space into a real number. If is an event that is in the sample space , then is a function that, when given the parameter , produces a numeric result , i.e. . This function is called a random variable. For simplicity, the function is simply written as .

The function can be one-to-one, mapping each element in the sample space to one value, or many-to-one, mapping several elements, or an event, to a single value.

Thus, gives us a compact way of mentioning an event while simultaneously capturing the uncertain variability (the fact that the value can change in an uncertain manner) and allowing us to use numbers to represent events.

Procedure: Send 3 data packets

Observation: Number of successful deliveries

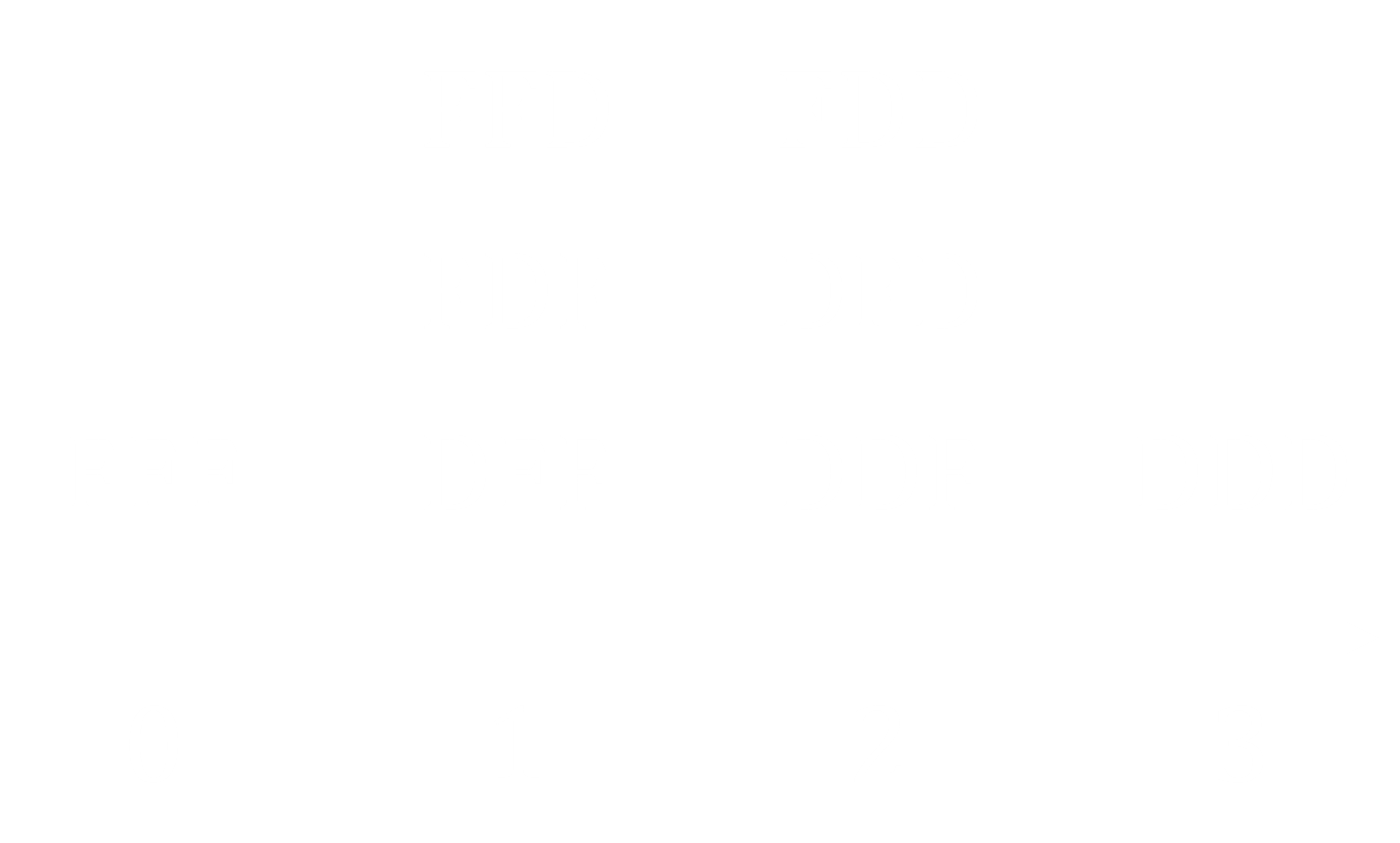
Sample Space:

A random variable associated with an experiment depends on how we define the function. Thus, we can have multiple random variables for a single experiment if we define them in multiple different ways.

We need to define the function in a way such that each of the events is converted to a real number. In this case, the events can be represented as

Thus,

We can even represent using a diagram.



Thus, represents the probability of an event occurring, where the elements of the event are and , or , , and . Essentially, it means that is an element of such that , i.e. .

If , .

To complete the probability model, we still need to identify the set of values associated with random variables. A discrete random variable is usually associated with a finite number of values, the set of which is represented as .

.

Procedure: Rolling 2 dice

Observation: Sum of rolls

} where

## Distribution Functions

### Events Defined by Random Variables

We have previously seen how taking any set of elements from the sample space creates an event. Now let’s look at how random variables can be used to define events.

Say is a random variable. We know that maps elements of into an element of . The result of the function is a point on a number line. Thus, since is an event, we need to define in such a way so that it defines an event space, meaning each event is mapped to a number. If is a result of the function , then when , it represents an event. More specifically, if , then this is an event. is defined by a number of elements from the sample space that map to the number .

For example, if , then . Thus, defines that event space.

Given some value , we are interested in the probabilities , and . These three are said to define the distribution function.

### Probabilities Defined by Random Variables

The probability that the random variable has a value is given by . This means that when we ran the experiment, we got an outcome from that is mapped by the function to the number . Thus, not only is each value of mapped from an element or event from , the probabilities are mapped as well.

For the previous example where we send three data packets, the sample space is , creating the set . Here, , , and .

gives us the amount of probability assigned to the number . Using this, we have assigned a probability to every possible number. For example, .

### Probability Mass Function (PMF)

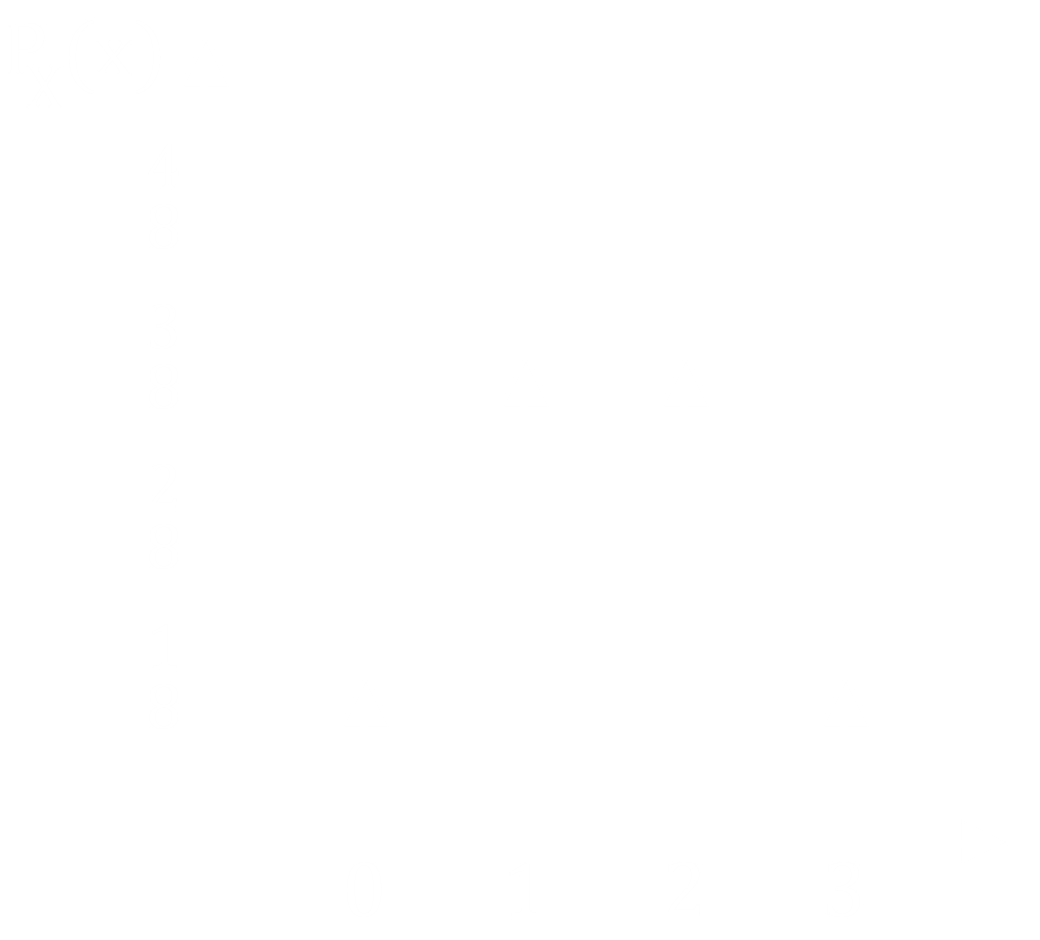
is called the probability mass function (PMF), and is denoted by . It is one of the distribution functions associated with random variables. It represents the probability of the event occurring, where is the value for some event for the random variable . It is called a distribution function, since it describes the distribution of probability over the number line. It is presented as a list of all its possible values.

Notice that using this representation, we are able to represent the whole probability model in a compact manner, and have the sample space and the probabilities together. Thus, we have solved the second limitation of simple probability models.

There are three conditions associated with the PMF.

For ,

For ,



Finally, the reason this is called a probability **mass** function, is because the probabilities can be considered to be masses.

Example

Consider a gambling game. Say we need to pay to play. The game is that three coins are tossed one after another and the outcomes are seen. There are four possible outcomes to this game.

* If the number of heads is , we get nothing back, making the net gain .
* If the number of heads is , we get , making the net gain .
* If the number of heads is , we get , making the net gain .
* If the number of heads is , we get , making the net gain .

Given that the coins are biased and , we need to develop the probability model.

Let’s assume the random variable . Thus, and .

### Cumulative Distributive Function (CDF)

The word ‘cumulative’ means starting from the very first possible instance until the given instance, unless otherwise mentioned. Thus, the cumulative distributive function gives us the probability of the event occurring. The CDF is denoted by , where .

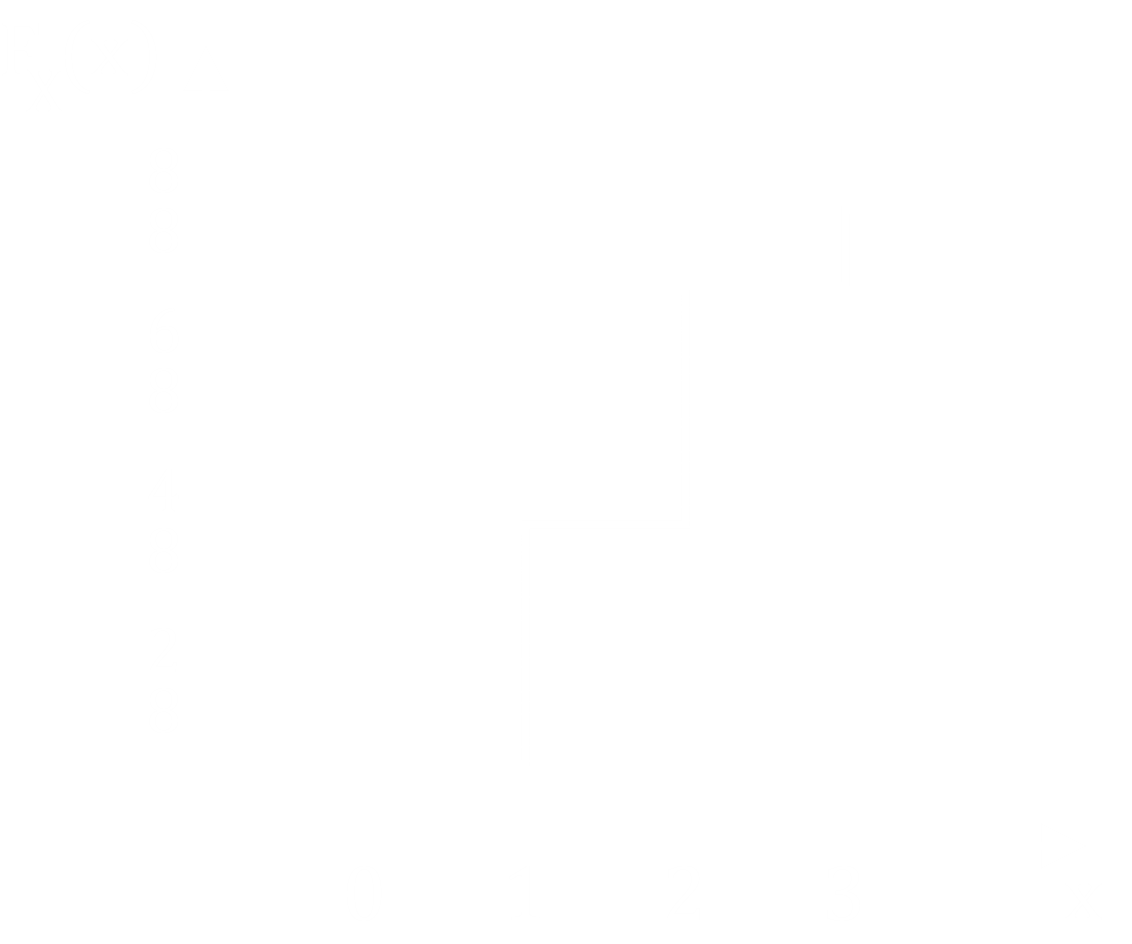
The only difference between the CDF and the PMF is that the PMF deals with explicit values of , while the CDF deals with all possible values of . Thus, there is a possibility that a value of such that will have a probability under CDF.

* If is less than the least possible outcome in the given situation, meaning , the probability will be .
* However, if the value of is within the valid range of outcomes, meaning , even if itself is not one of the outcomes, it will have some probability depending on its value.

The exact value depends on where in the range is. If such that and , meaning is between two valid members of , then , the CDF of the lower value.

* If is larger than the greatest possible outcome, i.e. , then since the given is always greater than any possible value of .

For the example of sending three data packets,



Comparing this diagram to the one for PMF, we notice that the value of for each value of corresponds to the change in .

Thus, .

Between any two numbers and , the curve is flat. If , the curve is flat for . If , the curve is a straight vertical line at . This vertical line corresponds to a jump at , the amount of which is equal to .

We want to find the probability , where .

The event can be broken down into two events, . Essentially, we are taking the two parts that make up the event separately. Thus,

which makes sense.

## Derived Random Variables

A derived random variable is defined from another random variable. Thus, the derived random variable converts every element into an element . Here, we say , or .

The derived random variable can be a one-to-one or a many-to-one function. The function .

Consider a scenario where and . Thus, the function is such that . This is a many-to-one function. Here, to find , we need to sum for all such where .

The same formula could be used for a one-to-one function, but of course in that case we would just be summing a single value.

Thus, for , we have .

From this, we can easily calculate .

## Expectations

One of the criticisms we made of the simple probability model is that it does not allow for further processing of the probability model. We mentioned that once the sample space had been converted to real numbers and a more sophisticated probability model, like the PMF, was being used, further processing would be possible. One such further processing is expected values or expectations.

Say someone gambles in a game times. One scenario is where is very small, like or or . The net gain in this case depends entirely on luck. A lucky person may play the game once and win a large amount, while an unlucky person could play a few times and win nothing.

However, if is very large, to the extent that extends to , the net gain no longer depends on luck, but rather on the laws of probability. There will be a certain number of outcomes of each possibility, and the net gain can be found by summing all of those outcomes.

The number of wins is related to the PMF, which is the probability. In this case, the probability depends on the frequencies of the different outcomes occurring. If a certain outcome has a frequency of , that outcome will appear times. This will not be an exact result, but any discrepancies will be ignorable.

From all of this, we will be able to create a table of sorts.

|  |  |  |
| --- | --- | --- |
| Net Gain | Probability | Total Gain |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Grand Total | |  |

The value of the grand total that we just found is an important quantity, which is called the expected value. Regardless of our luck, if we play the game a large number of times, on average, this will be the net gain, or rather net loss, of per game. If this number was , this would be a fair game, since on average a player would neither lose nor win.

This is the strategy used by casinos to ensure that they do not lose money. They do not bother about winning every single game, but the games are rigged so that, on average, they profit.

The expected value is dependent on two things, the values of the random variable and the probabilities with which those values appear.

If conflicts occur, the expected value may also be represented as or .

If we use the analogy of mass that we used with PMF, the expected value represents the centre of mass.

## Variances

Although expectations are generally a good representation of random variables, there are still situations where expectations might not work well. Variance gives us the average deflection of the values of random variables from the expected value. This deflection could be positive or negative. Simply put, it is the difference of the actual value of from the expected value.

The variance of a random variables is given by . However, programmatically, this would require us to pass over all the values in two times. Thus, a simpler alternative formula, , is used.

Using the example from the previous table,

## Well-Known Random Variables

Finding the probability model for a random experiment in real life can be tedious in some cases, even involving multiple random variables sometimes. However, most random experiments actually follow the same general model. For example, tossing a coin repeatedly until we get a head is the same as sending a data packet repeatedly until it is successfully delivered. There are a few well known random variables, and we should always try to fit a given random experiment into one of these random variables first, before going through the manual process of defining the probability model. We will generally find most random experiments can be fit into one of them.

Well-Known random variables can be divided into two categories. Uniform random variables, Poisson random variables and Hypergeometric random variables fall under one category, while Bernoulli random variables, Geometric random variables, Binomial random variables and Negative Binomial random variables, also known as Pascal random variables, fall into another category.

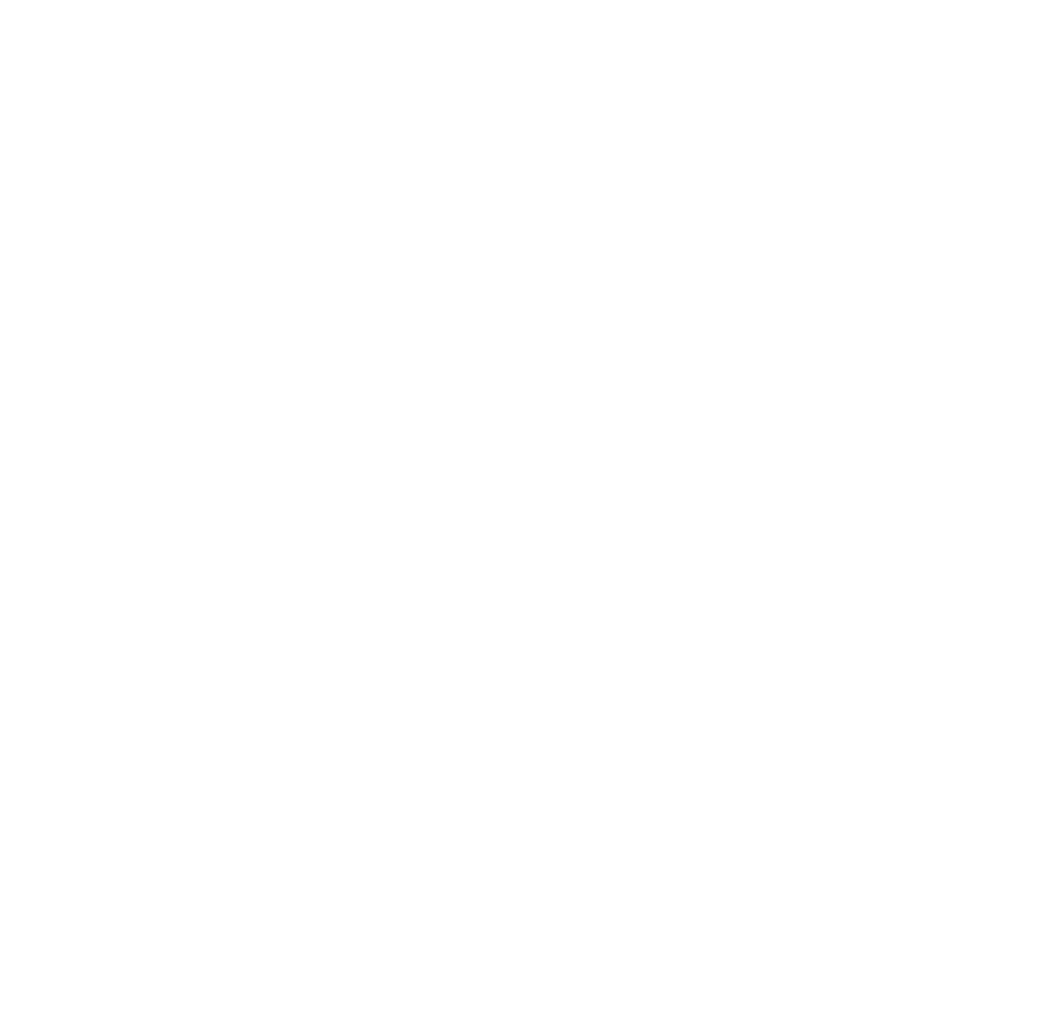
### Bernoulli Random Variables

Bernoulli random variables are the simplest random variables, but they are also one of the most important. They are associated with Bernoulli experiments. Bernoulli experiments have two outcomes and the outcomes are classified as success/failure or on/off. Essentially, the outcomes are opposites of each other. Tossing a coin falls into this category, as does sending a packet of data. If there are more than two outcomes for a Bernoulli experiment, they will still be dividable into two groups, one related to success and another to failure. For example, rolling a die and trying to get an even number is a Bernoulli experiment since the outcomes can be divided into two groups.

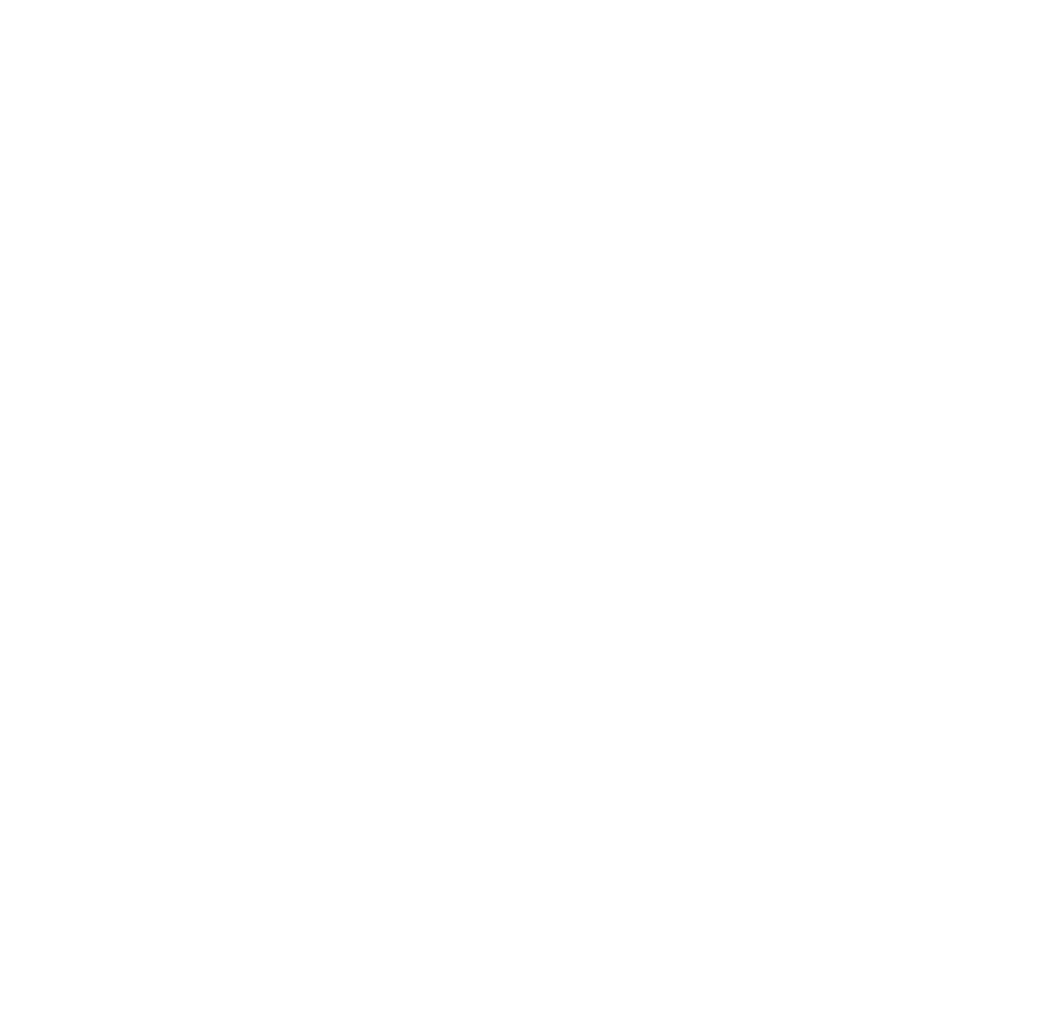
The Bernoulli random variables are defined such that any outcomes related to successes are converted to , while outcomes related to failures are converted to . Thus, .

If we consider the total probability of all successful outcomes to be , then

It is also possible to write this PMF in a single line as , .



Similarly, the CDF can be expressed as



In some advanced courses, all of this information is expressed simply as , where is used to express ‘distributed as’. In this case, instead of using the term ‘distribution function’, the PMF and CDF can be called a family of distribution. This is because the parameter, , is not a specific value in this case.

### Bernoulli Trials

Bernoulli experiments are actually building blocks for a wide range of random variables. It is possible to combine two or more Bernoulli experiments to create a complex experiment that represents another well-known random variable.

A Bernoulli trail is a sequence of Bernoulli experiments. A single Bernoulli experiment is repeated multiple times. For example, tossing a coin repeatedly is a Bernoulli trial.

However, not all sequences of repeated Bernoulli experiments count as Bernoulli trials. A Bernoulli trial must satisfy a few conditions:

* Individual Bernoulli experiments must be independent, meaning one experiment cannot depend on another one. Independence is ensured by:
  + Checking that the outcome of one experiment does not affect the outcome of another
  + The probability of success is a fixed number, i.e. is constant
* The number of repetitions must be either
  + Fixed, meaning the experiment is repeated times exactly
  + Dependant on a condition, such as a coin being tossed repeatedly until a head occurs

Example 1

Procedure: Keep sending data packets until a success occurs.

Sample Space:

Here, say a single attempt has a probability of success .

Let’s say we have a random variable . Thus, .

Now we need to find for all values of . Once we find this, the probability model is complete.

Notice how we do not have just two outcomes here. This is because this is not a single Bernoulli experiment, it is a sequence of Bernoulli experiments, a.k.a. a Bernoulli trial.

Example 2

Procedure: Send data packets and count the number of successes

Sample Space:

Here, let . Thus, . Here, we do not care about the order of the successes. Finally, we need to calculate .

Example 3

Procedure: Keep sending data packets until successes occur.

Observation: The number of packets sent.

Sample Space:

Here, and . Thus, we need to find for .

We can consider all three examples above together. We have a Bernoulli trial, meaning the probability is fixed and the number of repetitions is either fixed, or conditional.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Example Number | Number of Repetitions | Condition | Random Variable | Parameters | Distribution Type | Short-Form |
| 1 | Conditional | Until first success | Number of repetitions | Probability of success, | Geometric |  |
| 2 | Fixed | N/A | Number of successes | Number of successes, , Probability of success, | Binomial |  |
| 3 | Conditional | Until successes | Number of repetitions | Number of repetitions, , Probability of success, | Negative Binomial (Pascal) |  |

Thus,

* Geometric distribution deals with the number of trials required for a single success
* Binomial distribution deals with a specific number of successes
* Negative binomial distribution deals with the number of failures before a specified number of successes

Keep in mind that the three distributions given above are not single distributions, but rather a family of distributions.

### Geometric Distributions

Geometric distributions are connected to a family of random variables, called Geometric Random Variables. They are also related to binomial random variables, since they are special case of binomial random variables, where .

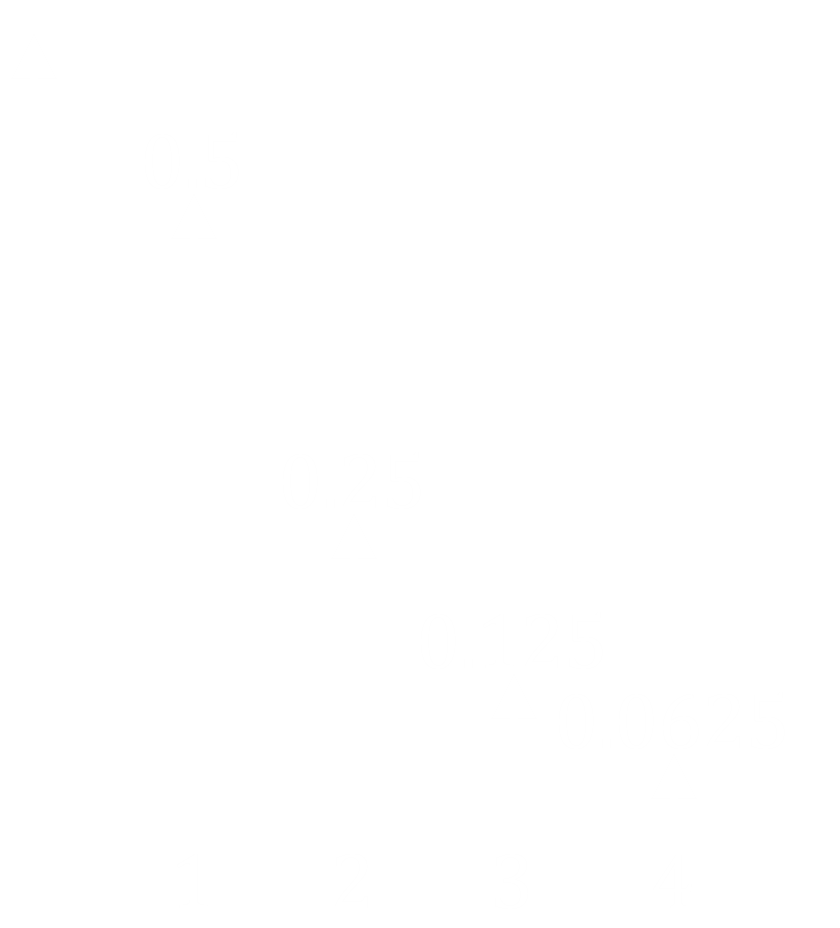
In a geometric distribution, we repeat a certain Bernoulli experiment independently until we get a success. The sample space is thus . We can define the random variable as

As such, . This is of course a discrete set, since there are an infinite number of elements, but the elements are countable.

Our goal is to find the PMF probability model for a geometric distribution. Whenever we are describing the probability model, we need to give a general solution. To do this, we need to start calculating the probability of every value in .

Since this is a sequence of Bernoulli experiments, there are only two possibilities. The probability of a success can be denoted as . Thus . Notice that this pattern follows the pattern for a geometric series. This is exactly why this is called a geometric distribution, and the random variables associated with it are called a family of geometric random variables.

If , the PMF graph would look like this:



The CDF will be given by when .

A simple variation of the geometric distribution is where we define the random variable as . This means and .

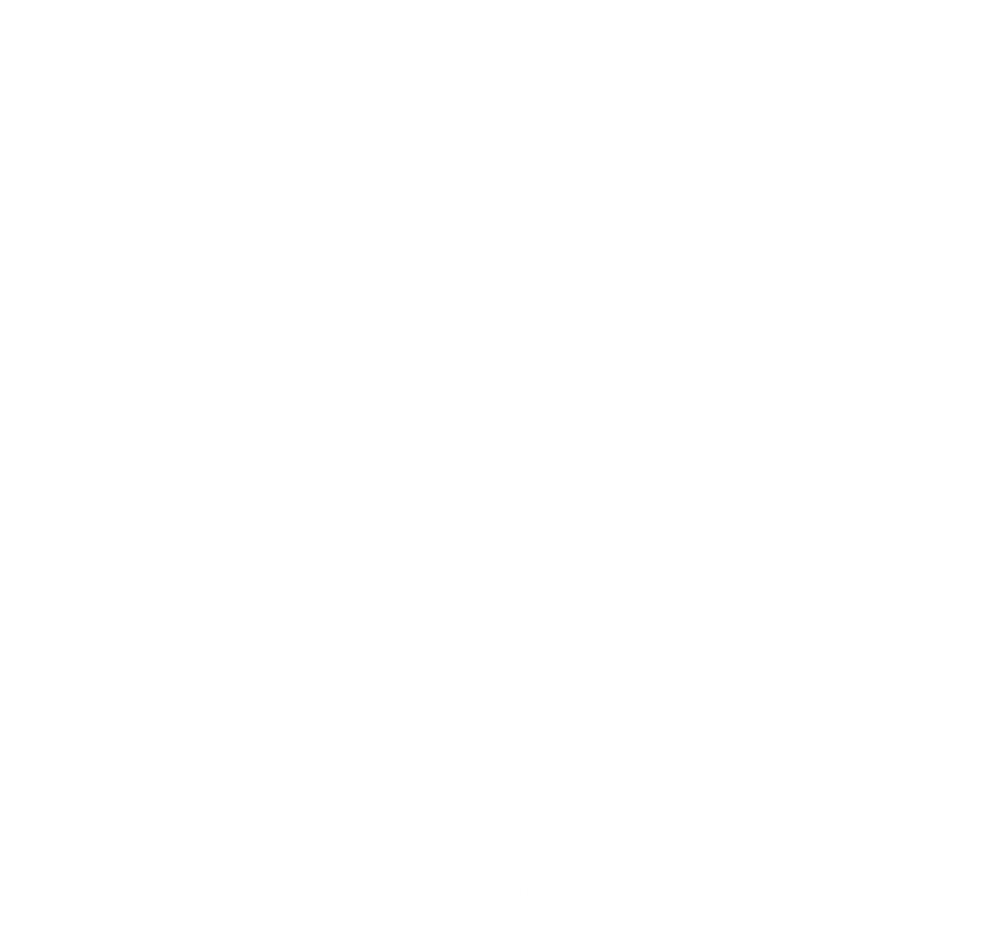
### Binomial Distributions

A binomial random variable defines the number of successes in attempts. For , . Thus, .

Here, . Thus,

Notice how this formula follows the pattern of the binomial formula. Thus, this is called a binomial distribution and the random variables associated with it are called a family of binomial random variables.

For , the PMF graph would look like this:



The CDF for binomial variables must be found by summing up the individual PMFs. There is no general formula.

### Negative Binomial Distributions

In negative binomial distributions, is defined as

.

For , and .

Here, . Notice the condition that the last delivery has to be a success. Thus,

### Uniform Distributions

Uniform distributions are also a family of distributions, since they have parameters and depending on the parameter, we will get a different distribution.

Say is a discrete random variable and , where is the number of possible values of .

We can say that is a random variable for a uniform distribution if all of the possible values of are equally likely, meaning they have the same probability.

In the above case, we have considered that can have any value at all. There is a slightly different variant that assumes that the values of are successive integers. It can start from any integer number, but they should be within an interval, say , where . Thus, . Since there are values, the probability of each will be .

### Hypergeometric Distributions

Say we have a box with some ICs, of which are in a good condition and of which are defective. We now pick the ICs from the box in two conditions, firstly, with replacement, and secondly, without replacement. The probability of picking any individual IC is the same.

Say in the first scenario we pick ICs with replacement, putting each IC back after we pick them. Here, let . Thus, .

Here, the probability of picking a good IC is and the probability of picking a defective IC is . Since there are only two possible outcomes for each experiment, this means each experiment is a Bernoulli experiment, and the entire process of picking ICs with replacement is a Bernoulli trial. More specifically, we are technically just counting the number of successes, so . A such, .

Now consider the scenario where we pick ICs, but without replacement. In this scenario, every time we pick an item, the number of items remaining in the box changes. Say , and we want to find .

This case is a hypergeometric distribution, and we will solve this by counting the number of ways we can do the operations. First, consider how many ways we can pick ICs from a set of ICs. This is obviously . Next, consider how many ways we can pick exactly good and defective ICs. This is . Thus,

### Poisson Distributions

In Poisson distributions, some event always occurs. For example, a car arriving at its destination, a customer arriving at a shop, arrival of packets at a router. For such scenarios, the average arrival rate will be given to us, i.e. the number of arrivals per unit time, denoted by . This unit of time could be anything from a second to months or years, whatever is appropriate for the situation.

We need to define the time period, , and the random variable . For example, we might want to know the number of earthquakes that might occur in the next two years. As such, .

We are interested in calculating , where . However, we will not be deriving the formula for the PMF of Poisson distributions, due to the complicacy of the formula. We will also be provided the formula during any examinations, so it is not necessary to remember it.

Poisson distributions are actually approximations to binomial distributions. As we know, in binomial distributions, we need to calculate . For a value of like and a value of like , can give us huge numbers. In such a case, calculations can become very cumbersome. Consider the case of a shop owner who knows customers arrive per hour in their shop. If , we want to know .

If we divide the one hour into seconds, then the average number of customers arriving per second is . From this, we can assume that this value is the probability of a customer arriving in a particular second and the probability of a customer not arriving in a particular second is this value. Then, in each second, we have a Bernoulli experiment. Thus, the repetition of this experiment for seconds is a Bernoulli trial.

Say . Thus, . Thus, . This is possible to calculate, but cumbersome. Thus, we can make an approximation using a Poisson random variable.

Thus, Poisson random variables are only defined for positive values of . Here, . For the example above, .

Example

We frequently experience misdialled calls or cross connections. Say in an office, the average number of misdialled calls, . We want to find the probability of getting at least wrong calls by tomorrow.

### Truncated Geometric Distribution

When we calculated geometric random variables, we did not place an upper limit on . Thus, . For example, if we are sending data packets and we are not getting a message telling us a particular packet was delivered, we assume there was an error and keep sending it.

However, how many times do we keep doing this? What if the device we are trying to send the packet to is offline? If we just keep sending the packet, it will be a huge waste of network resources. Thus, we need to put a limit on it. This limit is denoted by . The geometric random variable is said to be truncated. Thus, .

For the first attempt, and . For the second attempt, and . For the third attempt, and . If we truncate the geometric distribution at , we cannot have any more attempts after this. Geometric distributions always end in a success, but here, we have a situation where we may have to end without any successes.

Remember that our goal originally was find the probability that attempts are needed, not the probability that the packet is sent successfully. Till , we have no problems, but at , we will not try again. As such, the probability that attempts are needed, regardless of whether it gives us a success or a failure, is the same as the probability that there are failed attempts.

### Expected Values of Well-Known Discrete Random Variables

For Bernoulli random variables,

Actually, . (A little unsure about why, but that’s what he said.)

Similarly, for geometric random variables:

For negative binomial random variables:

For Poisson random variables:

We just need to know these values, nothing more.