STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

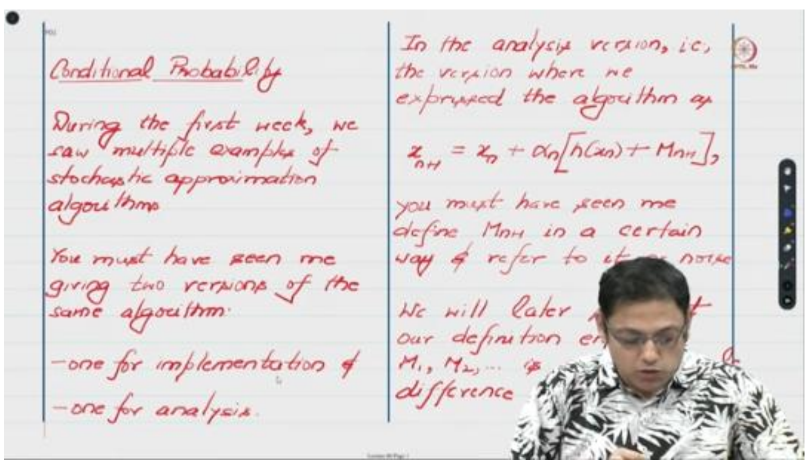
Department of Computer Science and Automation

Indian Institute of Science

Lecture 6

Probability Spaces- A Measure Theoretic Perspective

Hello and Namaste everyone. Welcome to the second week of this NPTEL course on Stochastic Approximation. So, in the first week of this course, you must have seen me describe several examples of Stochastic Approximation algorithms, right? And during this discussion, you must have seen me giving two versions of the same algorithm. I would have stated one algorithm which was, in some sense, useful for implementation, and you would have seen me give another version which was more suitable for analysis.



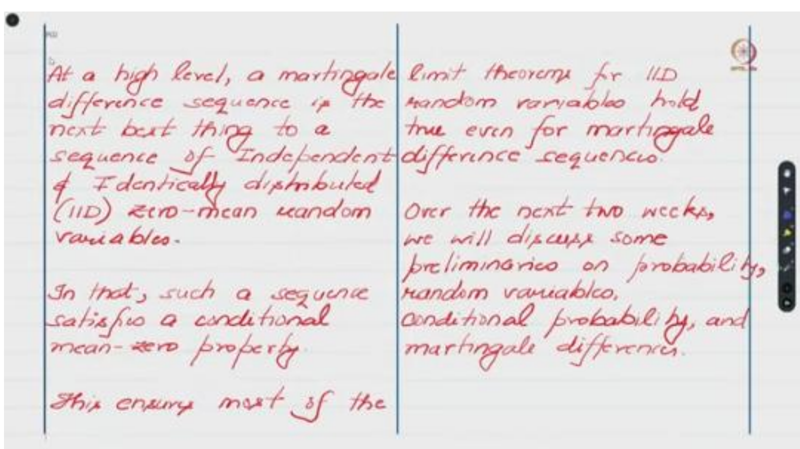
In the analysis version, You would have seen me express the algorithm in the form xn plus 1 equals xn plus alpha n times h of xn plus mn plus 1.

$$x_{n+1} = x_n + \alpha_n[h(x_n) + M_{n+1}]$$

And you must have seen me define mn plus 1 in a certain way and refer to it as noise. Right? Now, when you must have seen me do that, I hope a question arose in your mind: why did I define it in a certain way?

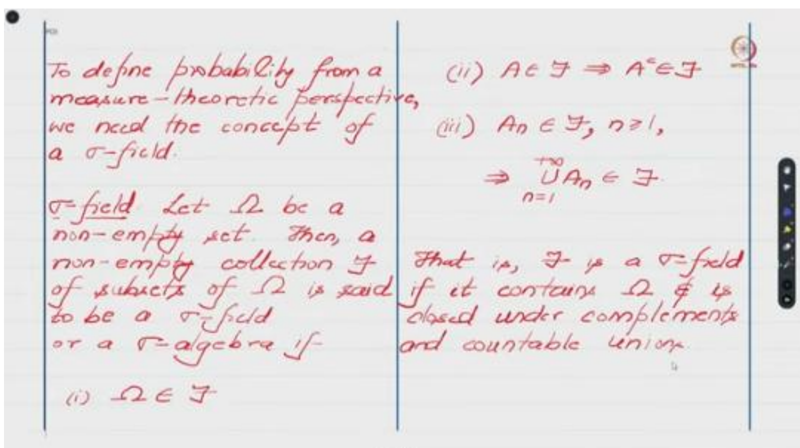
What was the, you know, reason for defining it in that way? Could there have been a different way to define this noise, and so on and so forth, right? So, these are the kind of questions that must have hopefully arisen in your mind, right? And we will later see, hopefully in a couple of lectures, that the definition that I provided ensures that the sequence M1, M2, M3, which is obtained by substituting different values of n over here, actually turns out to be what is called a martingale difference sequence, right? So, in order to understand what a martingale difference sequence is and so on and so forth, we will need to understand some basics of probability.

In particular, we will need to understand some foundations of conditional probability, right. So, at a very high level, what is a martingale difference sequence? Well, a martingale difference sequence is the next best thing to a sequence of independent and identically distributed zero-mean random variables. You will see that this martingale difference sequence actually satisfies some conditional zero-mean property. And because of this, most of the limit theorems that you must have studied—you know, during your undergrad or previous probability courses—such as the law of large numbers and so on, they carry over to the martingale difference sequences as well, right?



So, in short, you can view this martingale difference sequence as the next best thing to independent and identically distributed zero-mean random variables. And that is very helpful because, you know, we can work in a very generalized setup then without restricting ourselves to these noises being independent and so on. At the same time, we can also, you know, rely on standard limit theorems that apply for independent and identically distributed random variables to hold for these martingale difference sequences. So, as I said, over the next two weeks, we will go over some basics or preliminaries of probability. So, we will go over some basics or preliminaries of probability so that we can understand what this martingale difference sequence is, what I mean by, you know, the limit theorems applying and so on.

And in particular, over the next week, you will see me discussing some basics of probability, random variables, conditional probability, and in the following week, we will be talking about martingale differences. So, with this in place, let us begin our discussion on probability from a measure-theoretic perspective, and the first thing we need to understand is the concept of a sigma-field, right? So, what is a sigma-field? Well, suppose we have been given some set omega. Now, often this omega will correspond to the sample space of some experiment that we are interested in.



So, let us say we have been given such a set, and this set is a non-empty set, right? Then, a non-empty collection F of subsets of omega is said to be a sigma field or, equivalently, a sigma algebra if this collection F satisfies three properties. The first property is that the sample space omega should be in F. The second property is that if A belongs to this collection, A complement should also belong to this collection. And the third property is that if you have a collection AN of sets in F, then we require that their countable union also be in F.

In other words, we will say this collection of subsets of omega is a sigma field if it contains omega and is closed under complements and countable unions.

$$\Omega \in \mathcal{F}$$

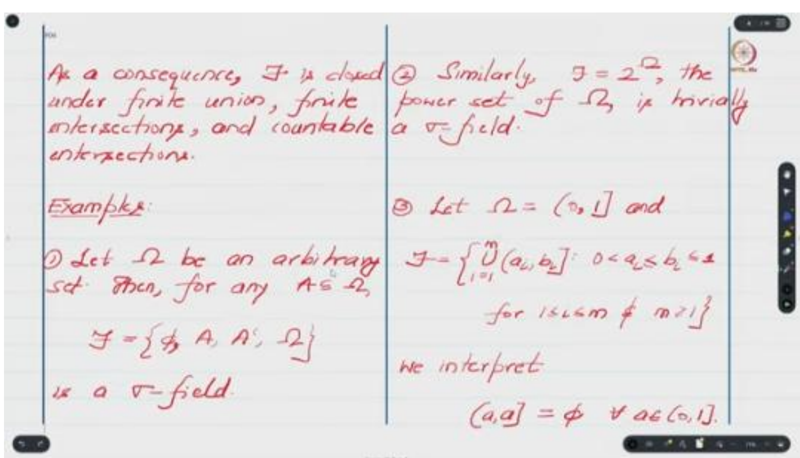
$$A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$$

$$A_n \in \mathcal{F}, n \geq 1,$$

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$$

So, let me quickly summarize. So, you have been given a sample space omega. Now, we are looking at a non-empty collection of subsets of omega, and we refer to this non-empty collection as calligraphic F. We will say this collection calligraphic F is a sigma field or a sigma algebra if it satisfies three properties. That is, it contains omega, and it is closed under complements and countable unions.

You will soon see what the advantage of working with this sigma field is. Now, a question that may arise in your mind is: What is so special about countable unions and complements? Why do we not talk about finite unions, finite intersections, countable intersections, and so on and so forth? Well, the short answer is we do not need to explicitly include them in the definition of a sigma field. The current definition of a sigma field already implies that this collection F is closed under finite unions, finite intersections, and countable intersections.



In other words, by ensuring that your calligraphic F satisfies those three conditions, we get for free that this calligraphic F satisfies these additional conditions as well. So, let us look at a quick example of a sigma field. So, let us say omega is some arbitrary set. So, this is some sample space that is of interest to you, and let us say A is some subset of omega. And let us look at this collection F consisting of the empty set A, its complement A, which we denote by A to the power C.

So, you have to interpret this as A complement, right? And then you have the sample space omega itself, right? Now, the claim is that this is a sigma field, right?

$$\mathcal{F} = \{\phi, A, A^c, \Omega\}$$

And the way to verify this is you check the three properties that I mentioned before. The first is that the sample space should belong to this collection.

So, you can see that that is trivially true. And then you can check if for every element in this collection, does its complement also belong to this collection. Like, for example, if you take the empty set, then its complement is the sample space, and you can see that that belongs over here. If you take A, its complement is A complement, and that belongs to this collection. If you take A complement, its complement is A itself, and that also belongs to the collection; and omega's complement is the empty set, and that also belongs to this collection.

So, in this way, you can see that F is actually closed under the complement operation, and similarly, one can check that F is also closed under the countable union operation. So, now let us move on to another example of a sigma field. So, let us say you have been given a set Omega, and you look at the power set associated with this set. The power set is basically the set containing all subsets of Omega. That is what the power set is, and one can see that this collection is trivially a sigma field.

So, it contains all subsets. So, it is obvious that it will be closed under complements and countable unions, and so on and so forth. So, in the first two of these examples, we have seen a collection which indeed is a sigma field, and whenever we study new things, it is always important that we also have access to an example which is not a sigma field. So, in this third example that I will now be talking about, we will look at an example where the collection is not a sigma field, right? So, in this example, let Omega be this set {0, 1}. So, notice that it is open at 0 and closed at 1.

So, let's look at this set, and let's say this is your Omega, and let us define calligraphic F to be made up of sets of this form. So, let's go over it carefully. So, what are the elements of F? Well, they are subsets of Omega which have the following form. So, what is this form? It is a finite union, right? And finite union of sets of this form.

So, each interval over here is open from the left and closed from the right, right? And you take a finite union of them, and we require that for each i, your a_i be strictly bigger than 0 and b_i be less than or equal to 1, and a_i be less than or equal to b_i. Is that okay? All right. In fact, we can actually—sorry, there is a typo here. We can actually allow a_i to be equal to 0 as well, right?

And this m that we have over here, right? This can be any finite number bigger than or equal to 1, right? So, you take some m. Whatever number you want—5, 600, 1 million—you take finite unions of intervals of this form, right? And put them together in this collection, right? And we will, you know, interpret an interval of this form—that is, you know, where you set a_i equals b_i and m equals 1, right. So, in that case, we will end up with an interval of this form.

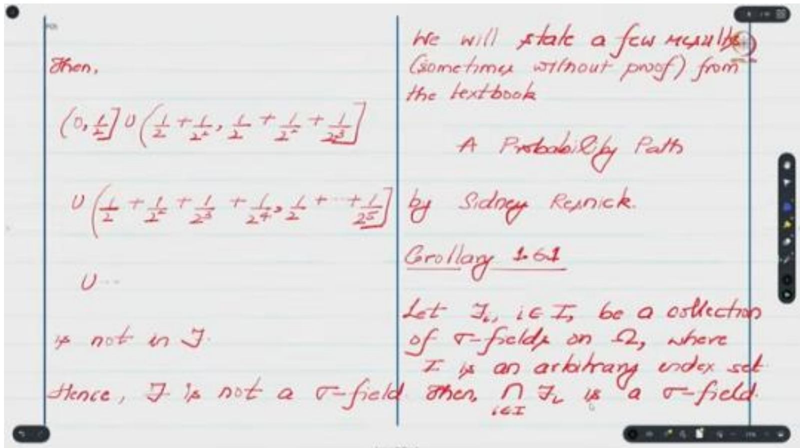
So, because it is open at this end and closed at this end, we will interpret an interval of this form to be the empty set for any a. Is that okay?

$$\Omega = (0,1]$$

$$\mathcal{F} = \left\{ \begin{matrix} \bigcup_{i=1}^{m} (a_i, b_i] : 0 \leq a_i \leq b_i \leq 1 \\ for\ 1 \leq i \leq m \notin m \geq 1 \end{matrix} \right\}$$

$$(a, a] = \phi \ \forall \ a \in (0,1]$$

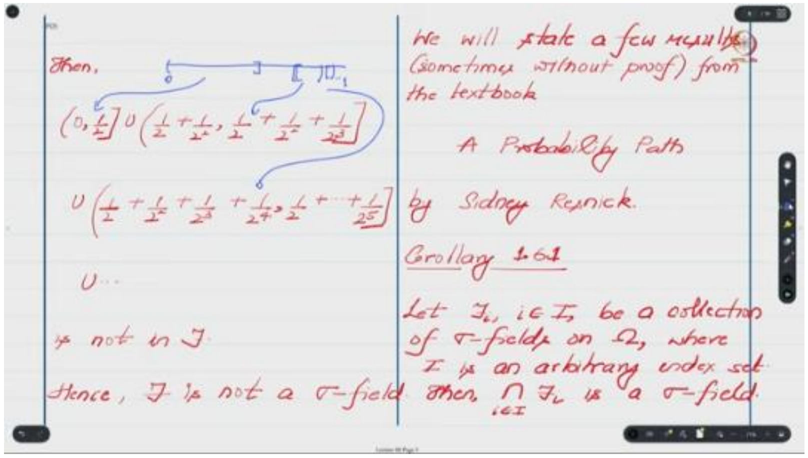
Alright, and now my claim is that this collection is actually not a sigma field, right? And in order to show that it is not a sigma field, we need to—it suffices to show that there is one property that is violated, and the property that is violated is the closeness under the countable union. So, the claim is that if you look at this countable union, so let us look at this—this is the, you know, interval going from 0 to half, so maybe if I, you know, draw this, you know—



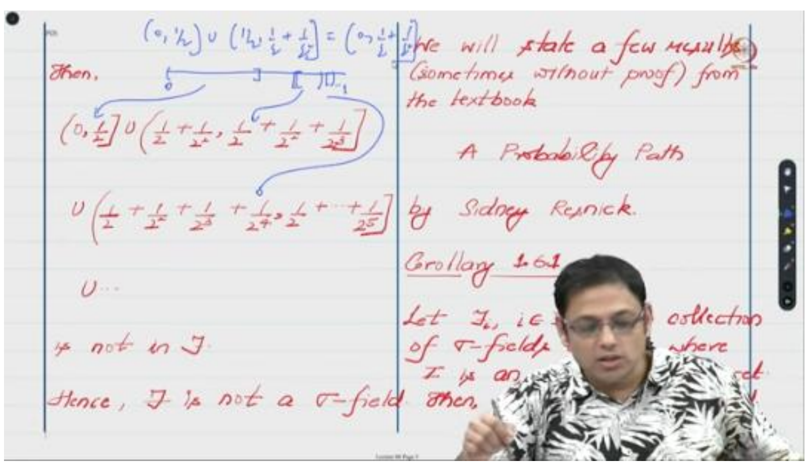
So, let us say this is 0, this is 1. So, I am going to take hold of this interval from 0 to half. This is the interval over here. Then, what I am going to do is give a gap. So, I am going to go half away.

And then I am going to construct an interval something like this. And then I am again going to give a gap and construct an interval something like this. And I will keep doing these intervals, right? So, this interval corresponds to this, this interval corresponds to this, and so on and so forth. And you can see that there is some gap present over here.

So, there is some gap present over here. And hence, if you take the countable union of them, it cannot be represented in some, you know, simplified fashion, right? Like, what do I mean by that? If you had 0, sorry, let me do it again. If you had



The interval (0, half) union (half, half plus 1 by 2 square) right. So, if you have this, you can compactly write it as (0, half plus half square). So, you can write it in this fashion, right? You are able to write it in this fashion because there is no gap between these two intervals. However, in this choice, we have carefully ensured that there is some gap present over here, and hence if you take this union, you cannot write it in a simplified form, right?



And you can see that each of these intervals actually belongs to F. However, they are—I mean, I am going to take their countable unions, and because it is a countable union, it will not belong to this F, because the elements in this F have to be represented only as finite unions. This N can be very large, but it has to be finite. On the other hand, we have now cooked up an example of a countable union wherein each element belongs to F, but their countable union does not belong to F, and hence this F that we are looking at is not a sigma field.

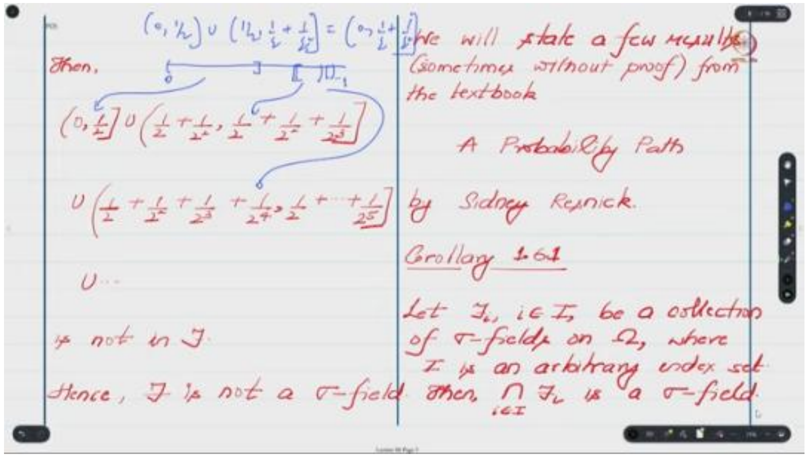
So, I hope you now have a good understanding of what a sigma field is and what is not a sigma field. Now, in order to talk about conditional probability, we first need to discuss expectations and probabilities, and in order to talk about probabilities, we need to discuss probability spaces, and in order to talk about probability spaces, we need this concept of the sigma field. So now we are going to see some properties of the sigma field, and for this, we will build upon some results from this textbook called Probability Path by Sidney Resnick. Is this okay? And these numbers I have ensured are from that textbook itself.

So, in case I have not provided you with the proof and you are interested in looking at the proof, I encourage you to find this textbook and go over the corresponding results. All right. So, the first of these results that would be of use to us is this corollary 1.6.1. So, what does this result say? It says that if you have a collection of sigma fields, then their intersection will also be a sigma field.

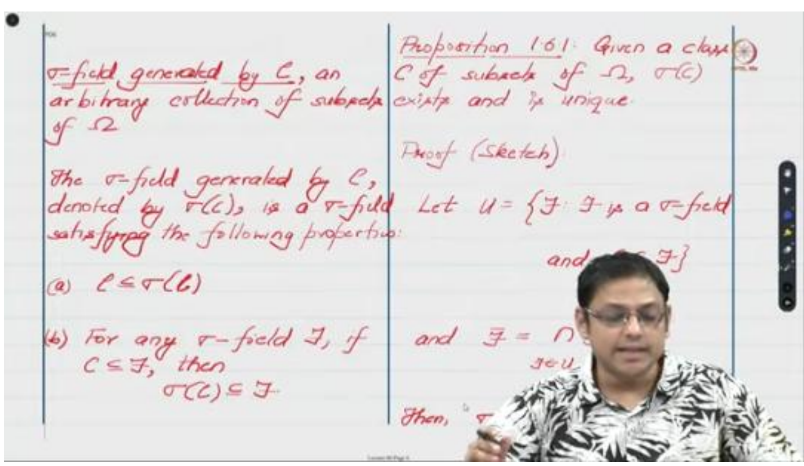
So, formally, what is the result?

Let $\mathcal{F}_i, i \in I$, be a collection of sigma fields on omega, right? So, so far we were talking of only one sigma field. Now, what we are going to do is we are going to talk about multiple sigma fields, okay?

And in this result, we are talking about, you know, a sigma field for every i in the index set capital I. So, we have this collection of sigma fields, and this index set I can be an arbitrary index set. By that, I mean it could be a finite index set, it could be a countable index set, it could be an uncountable index set—whatever you want. Then the claim is that if you take the intersection of all these sigma fields, then that intersection itself will be a sigma field. Actually, this is a very easy statement to verify.



You just have to verify that all three properties required in the definition of the sigma field are, I mean, all those three properties are actually satisfied by this intersection of Fi. So, I mean, you just do that exercise, and you can check that. So, you know, your intersection of FIs is indeed a sigma field. So, I am not going to give you the proof. I suggest you do it as an exercise yourself. You just have to verify the definition, and if you are not able to do it, I request you to see if you can find or get hold of this probability textbook and get the proof from there. The next concept that we will require in this discussion on conditional probability and expectations is this concept called a sigma field generated by C, where C is an arbitrary collection of subsets of omega.



So, we have been given a set omega and a collection of subsets. Now, this collection of subsets need not by itself be a sigma field. So, recall that a sigma field itself is a collection of subsets that satisfies some special properties. Now, let us say we have been given a C, which is a collection of subsets that does not have these special properties already, and the question we are interested in is: can we come up with a sigma field that includes this collection C that is given to us? That is what we mean by a sigma field generated by C.

So, the sigma field generated by C, which we will denote as sigma of C, is a sigma field that has the following properties. So, the first property is that the given collection C should be a subset of this sigma field. That is the first property. And the second property is that if you are given any other sigma field F and if this collection C is a subset of F, then we require that sigma of C, which is the sigma field generated by C, should be a subset of F.

$$\mathcal{C} \subseteq \sigma(\mathcal{C})$$

$$\mathcal{C} \subseteq \mathcal{F}$$

$$\sigma(\mathcal{C}) \subseteq \mathcal{F}$$

So, what this implies is that, or these two properties together imply that sigma of C is the smallest or the minimal sigma field containing C. Is this okay?

So, if you give me C, you construct a sigma field containing C, and that sigma field would be sigma of C if it is the smallest sigma field that you can construct. And one can ask: does such a sigma C exist? Is it unique? And so on and so forth. Proposition 1.6.1, again from this probability textbook, actually gives an answer to this question. It says that, suppose you give me a collection C of subsets of omega, then Sigma of C—that is, the sigma field generated by C—exists and is unique. So, recall that sigma of C is the smallest sigma field containing C. You can go over the proof, but here is a quick sketch of the proof.

So, the proof proceeds as follows. You first define calligraphic U to be the collection of all sigma fields containing C, right? So, what is calligraphic U? It is the collection of all sigma fields containing C. Is this okay? And then you define F bar to be the intersection of all such Fs.

So, what is calligraphic Q? It is the collection of all sigma fields containing C, and what is F bar? It is the intersection of all such Fs. Now, in the previous proposition, we saw that—sorry—F bar would be a sigma field by itself, right? And, you know, the proof then goes on to show that F bar is actually the smallest sigma field containing C, and hence it is sigma of C, okay?

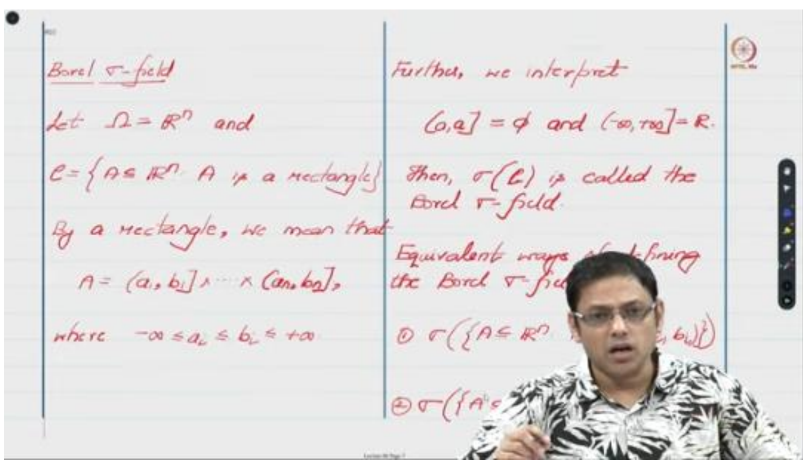
So, you can verify these things on your own.

$$\mathcal{U} = \{\mathcal{F} : \mathcal{F} \text{ } is \text{ } a \text{ } \sigma - field \text{ } and \text{ } \mathcal{C} \subseteq \mathcal{F}\}$$

$$\overline{\mathcal{F}} = \cap \text{ } \mathcal{F}$$

$$\sigma(C) = \text{ } \overline{\mathcal{F}}$$

It is not that hard. So, the summary of what we have discussed so far is: we have seen what a sigma field is, and we have also seen that if you are given an arbitrary collection of subsets of omega, you can actually come up with the smallest sigma field generated by this collection. This smallest sigma field we refer to as sigma of C. So, now let us look at some special sigma fields, and at the top of the list is this Borel sigma field. This Borel sigma field is something that we will often encounter throughout this course.



So, the Borel sigma field arises in the context where the sample space omega is actually your Euclidean space, right? So, what is the Borel sigma field? Well, the Borel sigma field is—let us say the sample space omega is Rn—and you start with some collection C, which is defined in the following way. The collection C consists of all those subsets of Rn such that A is a rectangle, right? So, by a rectangle, we mean that the set A has a Cartesian product of the following form, right?

So, you have an interval which is open at one end and closed at the other end, and you have the Cartesian product of sets of intervals of this form. Again, we allow your a_i to be any number, which is either infinity, minus infinity, or strictly bigger than it, and b_i can be any number, which is either plus infinity or strictly less than it, and a_i and b_i should satisfy this relation, ok. So, any set which can be written in this way we will refer to as a rectangle, and you collect all these rectangles together and put them in this collection C, right? So, this is a C, and one can again see that this C is not going to be a sigma field, right? But you can see that these are, in some sense, the collection that is of interest to us when we talk about probability and so on and so forth.

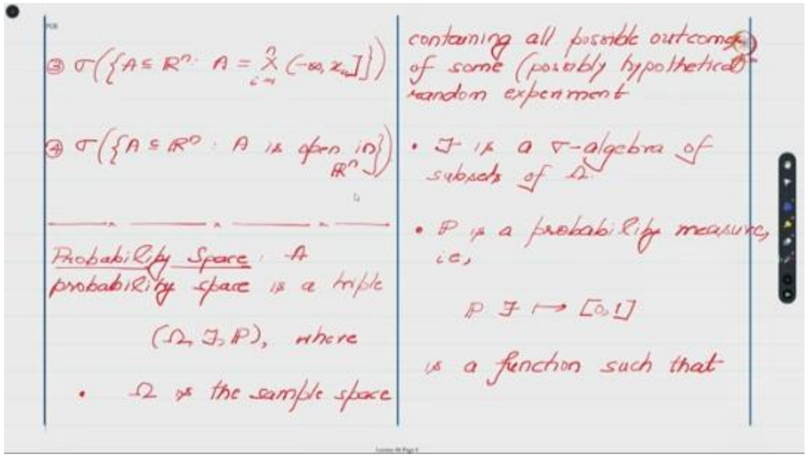
You will soon see that if you have not already noticed, right? And hence, we would like to come up with the sigma field containing C, right? And the smallest sigma field that is generated by such a collection is referred to as the Borel sigma field. There are several equivalent ways of defining the Borel sigma field. For example, you could look at this collection where your collection includes sets of this form, where A is a Cartesian product of intervals of this form, where the interval is open at both ends.

Notice that here it was open at one end and closed at the other end, whereas here it is open at both ends. One can show that if you look at the smallest sigma field containing this and the smallest sigma field containing this, they will both turn out to be the same, meaning every element in here will be in here and vice versa.

$$\sigma(\{A \subseteq \mathbb{R}^n : A = \times_{i=1}^{n} (a_i, b_i)\})$$

$$\sigma(\{A \subseteq \mathbb{R}^n : A = \times_{i=1}^{n} [a_i, b_i]\})$$

So, this is very nice, and similarly, one can come up with this alternative definition of the Borel sigma field by looking at a different collection. So, here we look at all subsets of Rn which are the Cartesian product of intervals of this form. Here you see that you are closed at both ends. So, again, one can show that the sigma field generated by this collection, the sigma field generated by this collection, and the sigma field generated by this collection are all one and the same, right? And there are some other alternative definitions as well.



The third of these is the sigma field generated by the collection where the set has the form given over here. It is the Cartesian product of intervals of this form, where the first or the starting place is minus infinity and the interval ends at some xi, right? And you take a finite Cartesian product of intervals of this form. Right, and then look at the sigma field generated by this. Again, one can show that this sigma field and the previous sigma fields turn out to be one and the same. So, there are several ways—I am only listing a few of them—and here is another common way to define the Borel sigma field: it is the sigma field containing open subsets of Rn.

$$\sigma(\{A \subseteq \mathbb{R}^n : A = \times_{i=1}^{n} (-\infty, x_i]\})$$

$$\sigma(\{A \subseteq \mathbb{R}^n : A \text{ is open in } \mathbb{R}^n\})$$

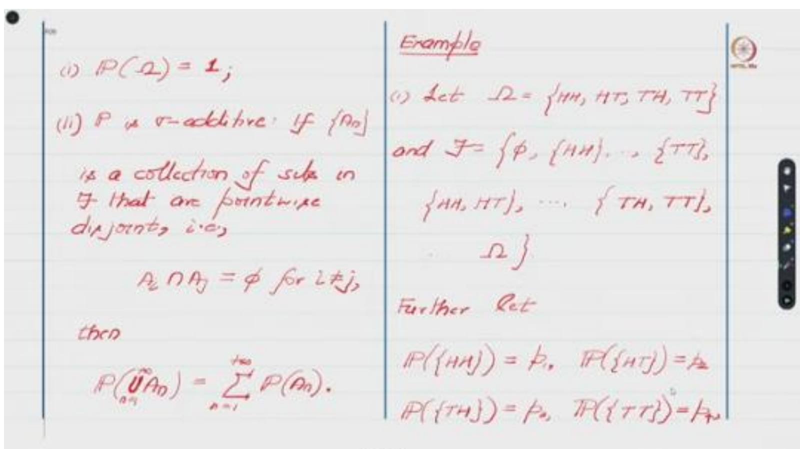
So, these are all equivalent ways, and if you are interested in why they are equivalent, I again encourage you to look at this probability textbook to figure out why they are equivalent. So, now we have some basics of what a sigma field is.

We will now move on to looking at this topic called probability space. So, what is a probability space? Well, a probability space is a triple, meaning that whenever I have to define a probability space, I need to give you three things. So, what are these three things? The first is that I have to give you a sample space.

The next is that I have to give you a sigma field F, and the third is that I have to give you a probability measure. So, what is omega? Well, it is a set. And we will refer to this as the sample space, and typically this will consist of all possible outcomes of some random experiment, and often this experiment could be hypothetical. And this F over here will be some sigma field or sigma algebra of subsets of omega.

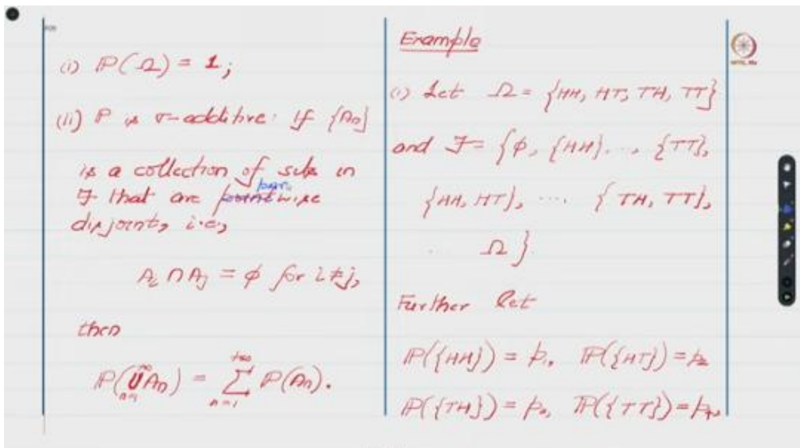
This will depend on what are the kind of sets on which we want to define probability, and so on and so forth. And the third element in this probability space is the probability function. So, probability is actually a function, or probability measure is a function. So, whenever I talk of a function, the question that should arise in your mind is: what is the domain of this function, and what is the range of this function? For the probability measure, the domain is actually the sigma field of interest.

And the range is 0 to 1, the interval 0 to 1. In other words, you give me an element of the sigma field, and P will tell you what number between 0 and 1 is assigned to that element in F. Now, this P cannot be any arbitrary function. It has to be a function that satisfies certain properties, which are listed over here. In particular, there are two properties.



The first property is that for the whole sample space, P should necessarily assign the value 1. And P should be sigma-additive. What does sigma-additive mean? Suppose you have been given a collection of sets in F, right? A countable collection of sets in F, which are—sorry, there is a typo here.

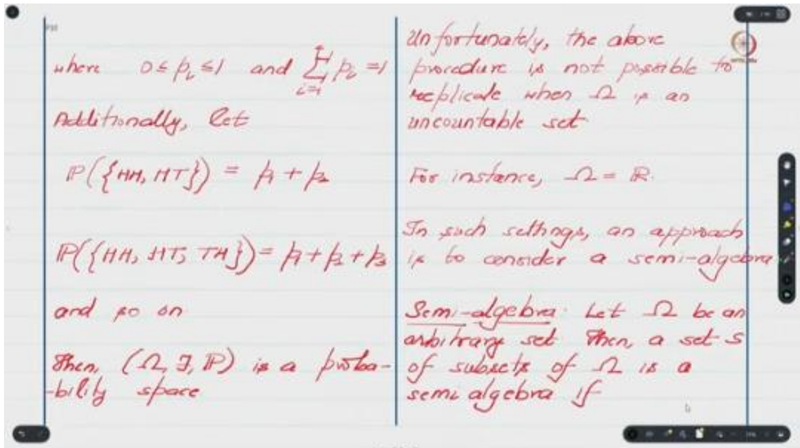
This should be pairwise disjoint, okay? So, suppose you have a collection of sets in F that are pairwise disjoint, which means if you look at A_i and A_j where i is not equal to j, then their intersection should be empty—that is what pairwise disjoint implies. Whenever you have such a collection, then we require that P satisfies this additive property—that is, the probability of this countable union should be the sum of these individual probabilities. So, whenever we have a function P that satisfies this property, we will say that P is sigma-additive. So, let us quickly summarize. What is the probability space? To define it, we need three things, okay. First, we should have a sample space. Then, we should have a sigma field. And then, we should have a function that is defined on top of this sigma field, right? And that function P should satisfy these two properties: on one hand, the probability of the full sample space should be one, and on the other hand, this function P should be sigma-additive, which basically means that it satisfies this condition over here.



So, here is an example of this probability space. So, you must have, you know, often looked at this coin-tossing experiment. So, in this coin-tossing experiment, the sample space is made up of HH, HT, TH, and TT. And we can, you know, for convenience, look at the sigma field, which is the power set of omega. So, the power set of omega means your calligraphic F consists of all possible subsets of omega.

So, the first subset is the empty set, then you have the singleton subset, that is, the subset consisting of only one element. So, you have the set consisting of HH, HT, TH, and TT. Then you have sets consisting of two elements: HH, HT, and so on. And you have sets containing three elements, and the set containing four elements is basically omega.

So, in that sense, your calligraphic F is the power set of omega, and this F is really a sigma field. And now we want to define a probability measure on F, right, on calligraphic F. So, which means I have to assign some value to every element in this collection. So, what do I do? I assign, you know, the singleton set some value P1. The singleton set HT, I assign some value P2; the singleton set TH, I assign the value P3; and the singleton set TT, I assign the value P4.



$$\Omega = \{HH, HT, TH, TT\}$$

$$\mathcal{F} = \{\phi, \{HH\}, \ldots \{TT\}, \{HH, HT\}, \ldots \{TH, TT\}, \ldots \Omega\}$$

$$\mathbb{P}(\{HH\}) = p_1$$

$$\mathbb{P}(\{HT\}) = p_2$$

$$\mathbb{P}(\{TH\}) = p_3$$

$$\mathbb{P}(\{TT\}) = p_4$$

And I ensure that all these P's are between 0 and 1, and they add up to 1, right? In addition, I ensure that, you know, whenever I have a set consisting of two elements—let us say HH and HT—then I assign the value P1 plus P2. If it consists of three elements, I assign the value P1, P2, P3, and so on.

$$\mathbb{P}(\{HH, HT\}) = p_1 + p_2$$

$$\mathbb{P}(\{HH, HT, TH\}) = p_1 + p_2 + p_3$$

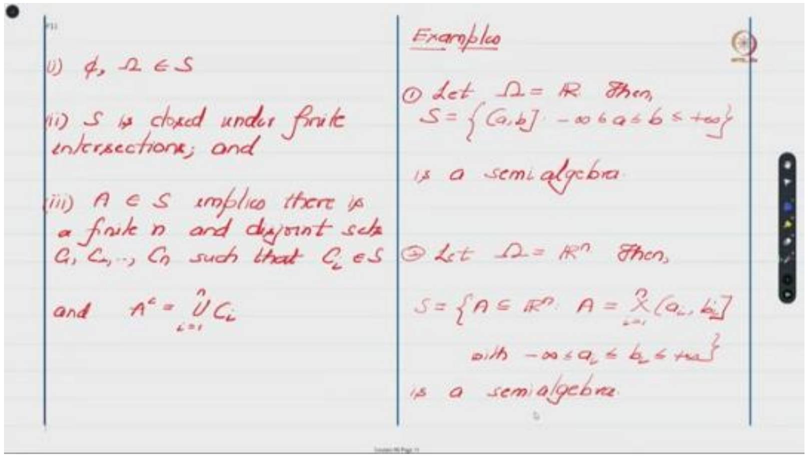
So, one can ask, okay, why am I, you know, trying to define it in this way? I first assign these P1, P2, P3, P4, and then if I assign it as P1 plus P2 and so on and so forth, one can check that these properties of P will actually be satisfied.

Hence, this way of defining P would ensure that P is indeed a probability measure. Is this okay? And because it is a probability measure, right, you know, this collection of omega, f, and P will turn out to be a probability space. Is this okay? Now, one can ask, okay, you know, this looks very simple.

Can we follow this procedure in general? Well, the answer is, you know, replicating this procedure often is not possible, especially when your sample space is an uncountable set. For instance, if omega is R, we cannot, you know, repeat this procedure. So, what is the summary of the previous procedure? You first assign or define the probability value for singleton sets and, you know, use that probability definition to extend to all other elements of your sigma field, right.

So, that idea is not always going to work. So, for such sets, there is an alternative approach, and the alternative approach is to first begin with what is called a semi-algebra. So, what is a semi-algebra? So, let us say omega is some arbitrary set, then a collection of subsets of omega. So, S is some collection of subsets of omega.

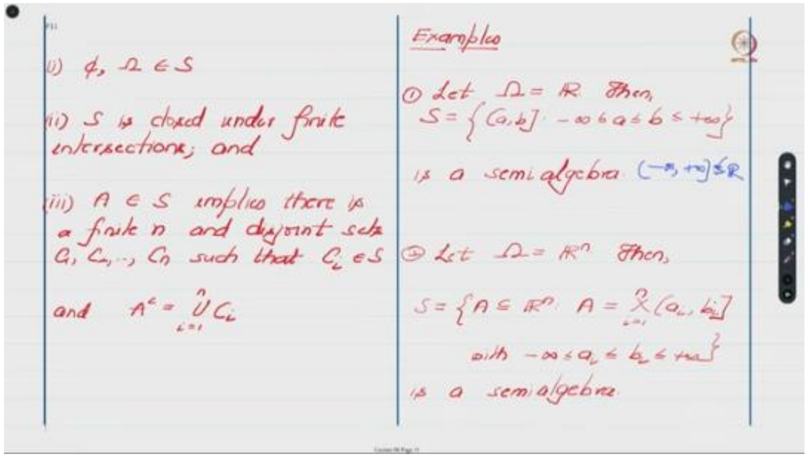
This is a semi-algebra. It satisfies the following three properties. The first is that your empty set and omega should belong to S, right? And this collection should be closed under finite intersections. And if any A belongs to S, it implies that the complement of A can be expressed as a finite union of disjoint sets in S, right?

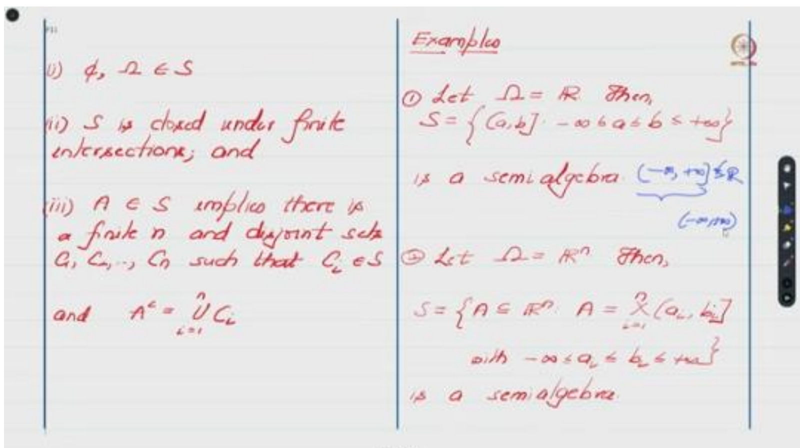


So, this is what this property means. So, if your S actually satisfies these three properties, then one can say that S is a semi-algebra. So, let us look at an example of a semi-algebra. Let us say omega is R, and let us say S consists of elements of the following form. It is basically made up of intervals which are open at one end and closed at the other end, and A and B satisfy this condition.

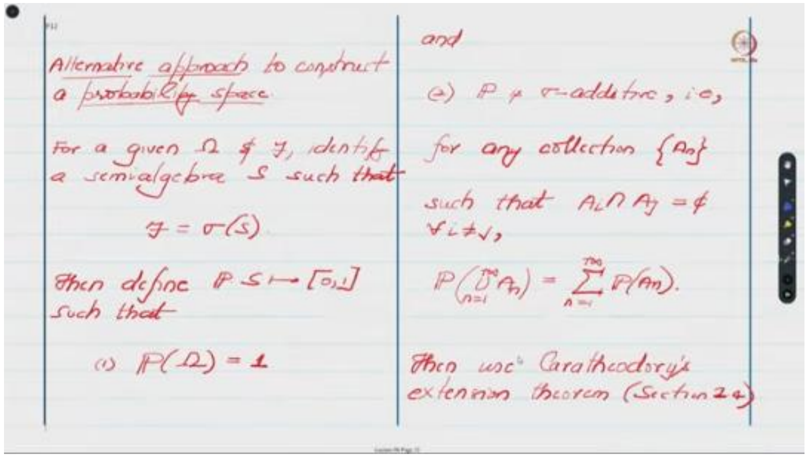
So, one can check that this collection is actually a semi-algebra. So, one can see that if I put A equals B, then this interval will be empty. So, this will be there, and if I put A equals minus infinity and B equals A equals minus infinity. And B equals plus infinity, right? So, we will interpret this result as this set as R, and from this interpretation, we can see that this collection S is actually, you know, containing your omega set, right?

This is a way we will interpret. So, when I put the square bracket at plus infinity, it is not to imply that there is some element called plus infinity that has to be included, like we do in the context of extended real numbers. No, we are not doing that. We will just interpret this set to imply this set. So, we are just saying that whenever I write, you know, open at minus infinity and closed at plus infinity, for that special case, we will interpret this set to imply this, and under this interpretation, one can see that omega actually belongs to this, and in the same way, you can check that, you know, these properties are also satisfied.



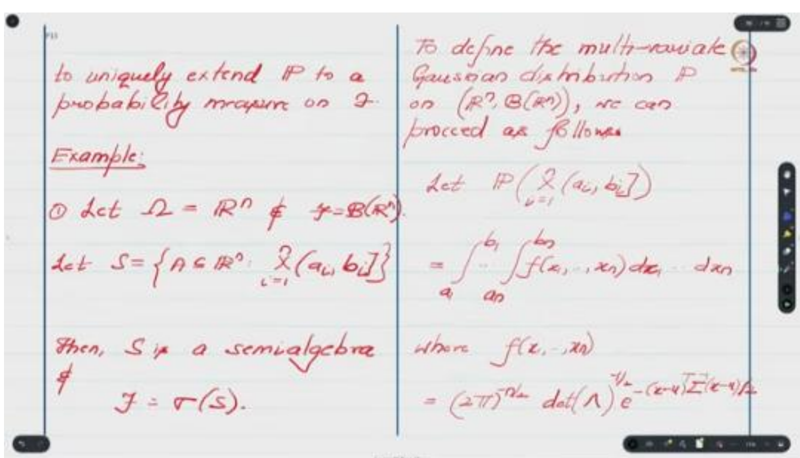


And because of this, this collection is a semi-algebra. And similarly, if your omega is Rn and let us say S is made up of subsets of Rn, which are finite Cartesian products of intervals of this form, one can again show that this is a semi-algebra. Now, once we have a semi-algebra, we can use it to define a probability measure on the sigma field that is generated by S. So, let us say you are given some omega and some sigma field F, and you want to assign a probability measure or define some probability measure on F, then the way to go about doing it would be to first identify a semi-algebra S that generates your sigma field F. And once you have this, you define P on this semi-algebra S so that it satisfies the following two properties.



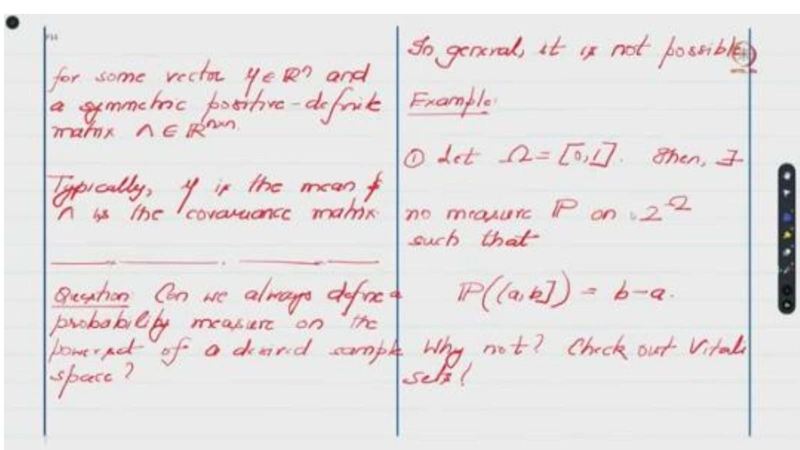
The first is that for the sample space omega, P should assign the value 1, and it should be sigma-additive. That is, for any collection A_n in S, such that they are pairwise disjoint, your P should satisfy this countable additivity property. And then, you know, you can look at Carathéodory's extension theorem from section 2.4 of this probability textbook to see that once you have defined P on a semi-algebra, this probability measure can actually be extended to the smallest sigma-field containing S. Is this okay? So, one can actually do that.

So, here is a quick example, right? So, let us say omega is R^n, and let us say your sigma-field is the Borel sigma-field of R^n, right? Now, you want to define, let us say, the standard multivariate Gaussian distribution on R^n, right? And you must have seen the way we define it. How do we define it?



Well, we first, you know, look at the collection of sets of the following form—that is, let us say we look at Cartesian products of the following form, right? And then, what we do is, in order to define the multivariate Gaussian distribution, we will say that if you have a set of this form, right, then the probability measure is given by this integral, where this little f that you have has this form that is given over here, right? So, if you look at any textbook that talks about multivariate Gaussian distribution, you can see that, okay, there is some little f which has this form, right? So, if it has this form, you can do this integration, where you know this a1 to b1 comes from the first interval, the a2 to b2 comes from the second interval, and so on, and finally, you end up with a_n to b_n. So, you look at this, you know, joint integral, right, and you use this joint integral to define the probability of this Cartesian product.

However, as I told you in the probability space, this probability measure should assign a value to every element in the sigma field. So, how do we ensure that? Well, the idea is that we will rely on Carathéodory's extension theorem to say that once we have defined a probability measure on this semi-algebra, we can extend that definition in a unique way to a probability measure on the sigma field. So, I mean, in this course, I will not have time to go over those details, but the idea is that when I have to define a probability measure on the sigma field—particularly in the case where your sample space is uncountable—the way we would do it is we would first identify a semi-algebra, define P on that space, and then use that to extend it to the whole sample space. One can now ask, can we always define a probability measure on any sigma field of our choice?



In particular, what is the purpose of working with sigma fields? Why can we not work with the power set itself, which we know is a trivial sigma field, and so on and so forth? So, the reason we have to work with this sigma field is that we cannot often work with the power set, and here is an example where we cannot work with the power set. So, let us say you have this set [0,1], and let us say you look at the power set and you want to define a probability measure on this power set such that it satisfies this property. So, you must have seen the uniform measure on this.

And for the uniform measure, you can see that the probability that the uniform random variable takes values between A and B is B minus A. So, in that sense, we want this probability measure to assign a value B minus A to the interval [A,B]. So, one can ask, can such a probability measure exist when your sigma field is actually the power set? So, one can show that one would not be able to construct a probability function P. So, what do you mean by 'one cannot construct'? You would not be able to assign the value of P to every element in the power set that satisfies those three properties—sorry, the two properties of P that we had mentioned in the definition, right? And one can ask, okay, why can we not do that?

Well, I would encourage you to look at what are called Vitaly sets and, you know, get the answer from there. So, now let me give a quick summary of what we have studied in this lecture. So, in this lecture, we went over the basics of probability. I rushed through it because, you know, this is going to be—I mean, a course that is focused on stochastic approximation—but to discuss stochastic approximation, we need this concept of martingale differences. In order to know about martingale differences, we need to know something called conditional probability. And that, in some sense, would be the core focus during this week.

And, you know, the first couple of lectures, the hope is that we can quickly cover the background. And, you know, you can read this textbook called Probability Path by Resnick and then use that to cover up all these concepts. So, the summary is that we defined what the sigma field is, we defined the sigma field generated by a collection, then we defined what a probability space is, right? And then we said how we could, you know, come up with this probability space. So, we looked at two examples: one where the sample space was a discrete set, and there was another example or scenario where the sample space was an uncountable set. In the discrete case, which can either be when the sample space is finite or countable, you can first define the probability on the singleton elements and then extend it to the, you know, power set in a very straightforward fashion, right? On the other hand, this recipe cannot be followed when the sample space is uncountable. In that case,

one of the standard approaches is to first identify a semi-algebra, assign values to the semi-algebra, and then appeal to what is called the Carathéodory Extension Theorem to, you know, enable the extension of the p-function that has been defined on the semi-algebra to the sigma field that is generated by that semi-algebra. So, if you are hearing these words for the first time, you know, it may appear a bit non-trivial, but once you start reading over these concepts of probability from this Probability Path textbook, you will see that these are very basic definitions. And once you are familiar with these definitions, you will soon be able to talk about conditional probabilities, conditional expectations, and, you know, martingale differences. And you can rigorously prove some of the properties of conditional expectations and martingale differences that we would need for the analysis of stochastic approximation algorithms. With this, let me stop this lecture.

Thank you and Namaste.