Discrete-Time Markov Chains and Poisson Processes
Professor Ayon Ganguly
Department of Mathematics
Indian Institute of Technology, Guwahati
Lecture 1

Review of Basic Probability I:

Welcome to the course Discrete Time Markov Chain and Poisson Process. So, this is a MOOC NPTEL course, and duration of the course will be about 8 weeks. So, more or less, we will have at least 20 hours of lecture material for the course. And, this is the first lecture of the course.

We, myself Ayon Ganguly and Dr. Subhamay Saha are the instructors. I will teach the first part of the course, whereas the second part of the course will be taught by Dr. Subhamay Saha.

So, let us see the syllabus of the course. Brief syllabus of the course is as follows: we will start with a review of basic probability. Which is basically some formulas and definitions we are going to discuss. And then we will go quickly through this process because this is a review, is a recap. So, we will go quickly and then we will enter to the main part of this course. So, main part of the course started from the second point which is the discrete-time Markov chain, then we proceed to exponential distribution, and finally, we will conclude this course with Poisson processes. So, in discrete-time Markov chain, we will spend a lot of a time and then gradually we will proceed to other parts.

So, as I mentioned, I will start with a review of basic probability. And as I already mentioned that I will go little quick during this review part and as we entered into the main part of this course, we will reduce our pace.

So, let us start with review of probability. So, let us start with the concept of countable set. So, what is countable set, in Lehmans point of view it is nothing but a set where the elements I can count. For example, if I take the collection of all the students of this class, I can count. This is the first students, second student, third student, fourth student and so on. So that collection or the set of all the students of the class is a countable set. So, that is basically the intuitive idea behind the countable set whether I can count the element or not. On the other hand, if I think of the interval $[0, 1]$, then this is not a countable set because I cannot count the elements in the interval 0 and 1. All the real numbers in the interval 0 and 1 I cannot count. So, intuitive idea of the countable set is as follows that, if I can count the number of items in the inner set, we will call it is a countable set. If not, will not, we will call it as the uncountable set. So, let us talk about the mathematical definition of countable set now. So, most of the time I will write it is at most countable set to signify the fact that there are two possibility in that. So, what is the basic definition? Definition goes like that, if I am talking about the set S, if I have a mapping from the set $S \to \{1, 2, ..., N\}$ for some value of N or some positive integer N or if I have a mapping from S to set of all-natural numbers, so if any of this thing happens, we say that the set is at most countable. So, if I can find out a bijection from S to this set where the first set where this N is basically a finite, a positive integer, and if, or if I have a bijection from S to set of all natural numbers, we say the set S is at most countable. In the first case when I have a mapping $S \rightarrow \{1, 2, ..., N\}$ in this case this set is called finite set, because in this case I have, finite number of points in S. And in the second case when I have a mapping, I have a bijection from S to N, the set of natural numbers, then we call it a countably infinite set. So, in these two parts they are in the at most countable set. So, at most countable sets basically includes finite set as well as countably infinite sets. And, that countable set has a lot of use in probability as well as in Markov chain, in stochastic processes. So, that is why I start with this basic definition. And as I mentioned that, intuitively this means that I can count the element of the set. So, let us proceed.

So, next point is basically the axiomatic definition of probability. So, all of us know the classical definition of probability. And classical definition of the probability goes like this. It is nothing but the favorable number of cases divided by total number of cases. So, if I try to find out the probability of a set A then it is nothing but number of favorable cases to the set A divided by total number of cases that may be possible. So, that is basically the classical definition but we further find out that there are some problem in case of the classical definitions. So, people go for more general definition and which is called the axiomatic definition. In case of the axiomatic definition, we try to define probability as a function. Now, suppose you say that when I am trying to define a function, I have to define its domain properly. Domain as well as codomain are two very important thing when we try to define our probability or when I try to define any function. Now, if I try to define our probability as a function, then of course domain is important in this case also. So, first we will talk about the domain of the probability function and then I will go to the axiomatic definition of the probability function. So, now to give you the definition of domain, let us start with a Ω which is a non-empty set, and this Ω is basically nothing but, you see that, we generally use probability to model some random experiment. So, having a experiment where the all possible outcomes are known to me, but in particular, the outcome of a particular trial or particular performance of the experiment is not known to me. So, that kind of experiments are called random experiment and normally we will use probability to model the uncertainty of that kind of experiments. So, this Ω you can think of as the all possible outcomes of random experiment. So, Ω you can think of a all possible outcome of a random experiment and then over omega we try to define something which is called σ-field or σ algebra, which is basically nothing but will be used as the domain of probability function. So, now let us talk about that what is the domain of probability function. So, we take a collection of subsets of Ω, probabilities are basically defined for a set, so I take the subset of Ω and I put them in F, but all the subsets may not be in Ω, so I take a collection of subsets of Ω in the collection $\mathbb{F}$ or $\mathcal{F}$, we say the σ-field or σ algebra has to satisfy these three points. So, if the collection, $\mathcal{F}$ satisfy these three points, then we call the collection as a σ-field or σ algebra. So, what are these three points? First point is that the null set has to belong to $\mathcal{F}$ that is the first point. Then the second point is that, if A belongs to $\mathcal{F}$, then A complement has to belong to $\mathcal{F}$. And the final point is that, if I have a sequence of set from $\mathcal{F}$, if I have a countable sequence of set from $\mathcal{F}$, then if I take the union of all the sets, it has to belong to $\mathcal{F}$. So, what are the intuition behind these three points? Intuition is as follows. See, first point says that $\emptyset$ has to belong to F. So, $\emptyset$ belongs to $\mathcal{F}$ that means empty set belongs to $\mathcal{F}$ and the intuition is very clear. We know from our classical definition that probability of $\emptyset$ has to be 0. So, that means we try to define probability of $\emptyset$. And if we try to define probability of $\emptyset$, $\emptyset$ has to be in the domain. So, we keep $\emptyset$ in the domain. Then again from the classical probability we know that if I know the probability of a set A, then we also know the probability of A complement which is $1 - \mathbb{P}(A)$. So, clearly if A is in the domain, A complement has to be in the domain. So that actually talk about the second condition. Now, come to the third condition. Third condition basically comes from the fact that if I know the probability of different sets, then I should know probability of all of them occurring together. And, that means, if all the $A_i \in \mathcal{F}$, that means I know the probability of Ai's because they are in domain, so I can find the probability of Ai. And in that case, I should know all of them together that basically mean, union of all of them either this or that kind of thing, and that basically mean that I should know the probability of union of Ai and so $\cup A_i \in \mathcal{F}$. So, these are basically the three points, three conditions that a collection of $\mathcal{F}$ of subsets of Ω has to satisfy to become $\mathcal{F}$ to be a σ-field or σ algebra. And now, we are going to define probability as a function and where the domain of the function is $\mathcal{F}$. So, we define like that a function $\mathbb{P}$ defined on $\mathcal{F}$ to positive part of the real line is called a probability if two conditions are satisfied. The first condition is that probability of Ω has to be 1. And you see that Ω has to belong to $\mathcal{F}$. The reason is basically first two condition here, because $\emptyset$ belongs to $\mathcal{F}$. And if $\emptyset$ belongs to $\mathcal{F}$, then $\emptyset$ complement has to belong to $\mathcal{F}$ and $\emptyset$ complement is nothing but Ω. So, Ω belongs to $\mathcal{F}$ and I can talk about the probability of Ω, and in this case, if $\mathbb{P}$ is a probability function, then $\mathbb{P}$ of Ω has to be 1. And the second condition is that if I have a sequence of disjoint sets in $\mathcal{F}$, then $\mathbb{P}(\cap A_i) = \sum \mathbb{P}(A_i)$. So, if I have a sequence of disjoint sets that means if I have countable number of disjoint sets in $\mathcal{F}$, then $\mathbb{P}(\cap A_i) = \sum \mathbb{P}(A_i)$. And you will see that the range is from 1 to infinity in both the case. So, basically, a function defined on F will be probability function it satisfies three conditions; one is that it has to be greater than equals to 0, second is that probability of $\emptyset$ has to be 0, probability of Ω has to be 1, and final thing is that if I take a sequence of disjoint sets, then probability of union is same as summation of the probabilities. So, that is basically the definition of probability.

Now, let us move on probability space. In the previous slide, we have concept of three things; one of them is Ω, which is a non-empty set, then a σ-field $\mathcal{F}$, and then a probability function $\mathbb{P}$. So, if I consider these three things together, this triplet, this triplet is called probability space. So, that some name is given to this triplet and the name is that it is called probability space. So, what this mean in practice? As I mentioned that normally we use probability to model a random experiment. And in that point of view, Ω is nothing but the collection of all possible outcomes of a random experiment. And sometimes this Ω is called sample space. For example, if I am tossing a coin, so if I toss a coin, I know either head or tail will come. So, in this case, Ω must be equals to either head or tail, all possible outcome of a random experiment. If I am tossing a coin till I get the first head then Ω will change and Ω turns out to be head, if it comes in the first toss, it is the head, if first tail comes and then head comes, so head tail head, then maybe two tail comes and then one head and so on and so forth. So, notice that in the first example I have a finite set, but the second example I have a infinite set, though it is a countable, but it is a countably infinite set. So, Ω is basically collection of all the outcomes of a random experiment. So, it is called a sample space. Then $\mathcal{F}$ is basically collection of all possible events in this case and $\mathbb{P}$ is basically the probability which basically model the chance of occurrence of those events. So, it basically model the chance of occurrence of events which belongs to $\mathcal{F}$. So, that is basically the practical meaning of the probability or probability space. The first one is collection of all possible outcome of a random experiment. The next one is the collection of all the events on which I try to define my probability and probability, as you know, it is nothing but the chance of occurrence of events in $\mathcal{F}$. So, that is basically the probability space.

And now let us see another very, very important topics in probability, which is called the conditional probability. And I know that all of you have an idea of the conditional probability. So conditional probability basically mean that, if I have some information in hand and under the information I try to find out what is the probability of some event. So, in those cases, we use the conditional probability and the definition of the conditional probability goes like this. So, we know that the definition of the conditional probability given by this. This is nothing but the fact that, well, I have the information that H has already been occurred and now I try to find out what is the probability that A occur. So, that notationally we will write in this form that, $\mathbb{P}(A|H)$ and that conditional probability is given by $\frac{\mathbb{P}(A\cap H)}{\mathbb{P}(H)}$. Now, you see that we know that when I am dividing a number by another number, then the denominator has to be positive otherwise the ratio is not defined. So, in this case, if I try to define the conditional probability, I have to have that probability of H has to be greater than 0. So, the $\mathbb{P}(A|H)$ is only defined if the $\mathbb{P}(H)$ is strictly greater than 0 and in that case for a arbitrary event A, the $\mathbb{P}(A|H)$ is defined by $\frac{\mathbb{P}(A\cap H)}{\mathbb{P}(H)}$. So, now come to the concept of mutually exclusive and exhaustive events. This concept will be very, very useful in the next point. So, what is the mutually exclusive means? Mutually exclusive means that, if I have a collection of events $E_1, E_2, ...$, and if I take any two events from here E_i and E_j such that $i \neq j$ and if I take the $E_1 \cap E_2$, if this is $\emptyset$, then we will call the collection of events $E_1, E_2, ...$ is mutually exclusive. So, mutually exclusive means that if I take any two events from the collection, they have to be disjoint. So, a collection of events $E_1, E_2, ...$ is said to be mutually exclusive if $E_1 \cap E_2 = \emptyset$ for all $i \neq j$. Now, what is exhaustive? Exhaustive basically mean that, if I have this collection and if I take the union of all the sets in the collection and if the union is equals to the Ω, the sample space, then we call that the collection is exhaustive. So, we have two concept. One is mutually exclusive, which basically mean that all the events in the collection are disjoint. And then we have another concept which is exhaustive, which basically mean that the collection actually covers the whole Ω, whole sample space is being covered by the collection of events. So, now we have the next theorem, which is a very, very well known and very, very useful theorem in probability which is called the theorem of total probability. So, what does theorem of total probability says? Theorem of total probabilities goes like that. Suppose I have a collection of events and this is mutually exclusive and exhaustive, that means if I take the intersection, it will be $\emptyset$ and if I take the union that union will be Ω, with the fact that $\mathbb{P}(E_i) > 0$ for all i, then if I take any event E, the $\mathbb{P}(E)$ can be written in this particular form and this form actually can be also written in this way also that it is basically same as $\sum_i \mathbb{P}(E_i \cap E)$. So, that diagrammatically things are quite simple. So, basically suppose I have E is basically the set. And I have the partition of this one so that are basically my E_is. So, this is E_1, this is E_2, this is E_3, this is E_4 and finally this one is E_5. So, suppose I have these events E_1, E_2, E_3, E_4, E_5, and then basically this statement says that to find out the $\mathbb{P}(E)$, I will calculate the $\mathbb{P}(E)$ along with several different parts and then I add them up. So, that is why this is called the theorem of total probability that I find out the partwise probability and then add them up to find out the final probability. And you see that when I am writing this probability in this manner, I actually do not need this one, this particular event to be true. But when I am writing this in probability of the intersection in terms of the conditional probability, then of course this condition has to be true, otherwise, I cannot define this conditional probability. So, that is basically the idea or basically the intuitive thing of the theorem of total probability that, to find out the probability of a, of an event I can see the mutually exclusive and exhaustive events in that Ω and then basically I will find out the probability of each part and then I will add them up to find out the final probability. So, let us proceed.

The next thing is that random variable. And as you know that random variable is very, very important thing in probability and we will see that this is also very, very important in stochastic process, in Markov chain, in Poisson process everywhere this is very, very important thing. So, now, let us talk about random variable. So, what is random variable? Random variable is basically a function of Ω, function $X : \Omega \to \mathbf{R}$. And before going further into it, please keep this thing in mind that this is a partial definition of random variable. But for our course, this definition we will take, because this is a preliminary course. And to know the complete definition of the random variable I suggest you to go for advance course in probability. So, for our course, the definition of the random variable is it is a function from Ω to $\mathbf{R}$. Now, the question is that, why we talk about random variable. The first point is that, say most of the cases when we talk about some random experiment, the outcomes are some numerical values. For example, if I talk about heights of the students of some class, that is a numerical value. If I talk about weight of the students of the class, that is a numerical value. If I talk about income of households of India, then that is a numerical value. If I talk about air pollution in a particular region, that is also a numerical value. So, most of the cases when we talk about some random experiment, we say that we finally talk about some numerical values. So that means X, random variable is a meaningful thing in this case. Another point is that, in some cases, of course, Ω is, the outcome of the experiment are not directly numerical values. For example, if I take Ω to be head and tail that I toss a coin, the random experiment is a tossing a coin, and then in that case basically Ω is consists of head and tail. And in this case of course outcomes are not numerical values. But I can define a function X from Ω to $\mathbf{R}$ such that X of head is equals to 1 and here X of tail is equals to 0. Now, I have a function from Ω to $\mathbf{R}$, so Ω is mapped to $\mathbf{R}$, and I have a random variable. So, this way I can map a arbitrary Ω to real line and once I map it to real line I have a function and that function is basically our random variable. Now, second benefit of the random variable is as follows that we know a sophisticated tool to analyze real line which is basically our real analysis. So, if I have a mapping from arbitrary Ω to $\mathbf{R}$, then I can use all such sophisticated tools of real analysis to analyze the probability structure to analyze the relationship between several random variables using those real analysis stuffs, real analysis tools. So, that twofold benefit we have if we talk about random variables, and that is why the random variable is very, very important in probability, in stochastic processes, in statistics and in many fields of research, many field of practical uses. So, now, let us go to the next concept, in case of the random variable which is a very, very important thing which is called Cumulative Distribution Function or in short CDF. Sometime we will also call the Cumulative Distribution Function as distribution function. So, we will call it Cumulative Distribution Function or distribution function and say in short CDF or either DF. So, what is the definition of CDF? The definition of the CDF goes like this. It is nothing but a function F which maps $\mathbf{R}$ to the interval $[0,1]$, and the function definition is like that $F(x) = \mathbb{P}(X \leq x)$. So, the distribution function at the point small x is nothing but the probability that the random variable takes values less than or equals to small x. So, it is basically if this is my point x it is the probability of this side. The distribution function is the probability of the side including the point x. The point x is included there. Now, if I have a distribution function or cumulative distribution function it is going to satisfy these three conditions. So, what are these conditions? F is non-decreasing, that is obvious, because if I increase my x from this to that side, the probability will also, will not decrease. So, that is why this F has to be a non-decreasing function. F has to be right continuous function. And finally, if I take $x \to -\infty$, then the limit of $F(x)$, the CDF is 0. And if I take $x \to +\infty$, then the limit of the CDF is 1. And this is very intuitive from this graph, because if $x \to -\infty$ that means this point actually goes to the side. When this point goes this side that means nothing will be covered in the limiting sense so the probability is 0, and when $x \to +\infty$ this is basically goes to $-\infty$ and x goes to $+\infty$ basically mean that this point goes decide and when it goes this side whole $\mathbf{R}$ will be covered and so the whole Ω will be covered, finally, I have probability 1. So, these points are quite intuitive from the definition of the cumulative distribution function. One point I should mention here is that, see, if I have a CDF then these three properties has to be true. On the other hand, if any function satisfies these three properties, then that function has to be a CDF of some random variable. So, this is kind of a if and only if condition. If I have a CDF, then it has to satisfy these three conditions. If I have any function which satisfy these three conditions, then that function has to be a CDF of some random variable. This point is very useful because you see that this basically can be written as $\mathbb{P}(X = x)$ is same as $F(x) - F(x-)$. And what is

$x-$? $x-$ is basically left-hand side limit. So, here we have seen that F is right continuous, so F is right continuous as well as F is non-decreasing. So, that too actually tells us that F can have only jump discontinuity. These two points that F is non-decreasing and F is right continuous these two together tells us that F can have only jump discontinuity. And in this case the jump discontinuity will come from left hand side. So, this one basically tells us that, if I try to find out the probability that is $\mathbb{P}(X = x)$, then that is nothing but the amount of jump, the distance of jump at the point small x. So, if the distribution function is continuous at the point small x, then this probability has to be 0. If the distribution function is not continuous, I have a jump from the left side, then the length of the jump is basically the $\mathbb{P}(X = x)$. So, that shows that if I know the distribution function, I can find out what is the $\mathbb{P}(X = x)$. Let us proceed to the next slide.

So, next slide talk about discrete random variable. So, this is a classification of random variable. And in this course we are going to see mainly two classification, one is discrete random variable, another is continuous random variable. And intuitively speaking the discrete random variable means that when the random variable can take discrete values, like if I talk about marks of the students and, if the marks are given in integer values only, then marks of the students, is a discrete random variable, because it takes the discrete values. On the other hand, the continuous random variable means that it takes continuous values. For example, if I am talking about the weight of a student, the weight can take any value maybe in some range. So, that means, weight is a, the random variable, in this case, it takes some continuous values. So, that is a continuous random variable. So, this is basically the intuition behind the discrete random variable and continuous random variable. Now, let us see the mathematical definition. So, the mathematical definition of the discrete random variable goes like this. If I have a atmost countable set $S \subset \mathbf{R}$, so S is a subset of R here, if I have a atmost countable set S such that $\mathbb{P}(X \in S)$ equals to 1, then we call it is a discrete random variable and the corresponding distribution we call the discrete distribution. So, the definition goes like that. A random variable X is said to have a discrete distribution if there exist an atmost countable set S such that $\mathbb{P}(X \in S)$ equals to 1. So, in the previous case when we talk about that head and tail thing and if I map it to 0 and 1, if I take S equals to 0 and 1, then $\mathbb{P}(X \in S)$ equals to 1. So that random variable is a discrete random variable. And in case of the discrete random variable we define a function F which is nothing but $\mathbb{P}(X = x)$ and that one is called probability mass function or PMF. And then using the PMF I can define the CDF of a discrete random variable which is nothing but you need to take the sum up to the point x. You need to take the sum of PMF up to the point x. And finally, I give here is a definition of a function which is a basically PMF and it will have some use in stochastic processes so that is why this definition is given here. It is basically nothing but called Kronecker delta which basically means that for some real number c this function takes value 1 if x is same as c, otherwise it is 0. And this is basically a PMF of a random variable X which takes constant value c. So, $\mathbb{P}(X = c)$ equals to 1 that random variable has this PDF. So, this notation we will going to use later. So, keep this thing in mind and it is called Kronecker delta.

Now, continuous random variable. The definition of the continuous random variable goes like this. If I can find out a non-negative function f such that $\int_{-\infty}^{x} f(u)du$ is same as the CDF for all x then we call the corresponding random variable has continuous distribution. And just keep in mind that this x is here. So, whatever x value I take here, that will come in the upper bound of the integration and it has to be true for all $x \in \mathbf{R}$. So, the definition goes like that. A random variable X said to have continuous distribution if there exists a non-negative function f on $\mathbf{R}$ such that $F(x) = \int_{-\infty}^{x} f(u)du$. And in this case this function is called PDF that is Probability Density Function. This function is called Probability Density Function. And of course, because I can write the CDF as an integration, the CDF has to be continuous in this case. And because CDF is continuous the $\mathbb{P}(X = x)$ will be equals to 0 for all x belongs to $\mathbf{R}$. So, this is basically the continuous random variable. Just keep one thing in mind, see that PDF, Probability Density Function is only defined for a continuous random variable whereas probability mass function is only defined for discrete random variable. So, probability density is only for continuous variable and it is not defined for discrete random variable. Similarly, probability mass function is only defined for discrete random variable and not for continuous random variable.

Now, we will go for the expectation. The expectation is basically kind of average. It is the average of all possible values that can be taken by a random variable. So, expectation for a discrete random variable is defined by this particular formula that expectation of x is same as summation value of x multiplied by the PMF. And you just keep in mind the PMF is basically the probability. So, it is basically $(X = x)$ and $(X \in S)$. So, this is the expectation provided by that if this condition is true. Why this condition is kept? This condition is kept here so that this summation is finite. And so, I have this definition that expectation of a discrete random variable X is defined by this provided this one is finite. If this condition is not true, we call that the expectation do not exist or expectation is infinite. For the continuous random variable what we have. We basically need to replace the summation with a integration sign. So, for a continuous random variable with PDF f, the expectation is defined by $\mathbb{E}(X =) \int_{-\infty}^{\infty} xf(u)du$. And in this case this condition has to be true. Again, this condition is kept so that the expectation is finite. And in this case also if the condition is not true we say that the expectation do not exist or expectation is infinite. So, with that, I stop in this particular lecture. In the next couple of lecture, we will see some more review of probability. Thank you for listening.