Introduction to Queueing Theory
Prof. N. Selvaraju
Department of Mathematics
Indian Institute of Technology Guwahati, India

Lecture - 00
Prerequisite: Review of Probability

Hi and hello; in this video, what we will see is a "Review of Probability theory" that is needed for understanding the course. So, this is not going to give you; the material that we are going to cover; is not going to give you the complete picture, but at least you know you will get familiarized with the notations that we will be using in the rest of the course and also like what the concepts that are most important to us will be clear from seeing this. As you know like, every stochastic model is specified in terms of what is known as a probability space. Whether you say this explicitly or not, there is always an underlying probability space. So, what do we mean by that? A probability space is nothing, but we have written down here first a sample space Ω which is nothing but the collection of all possible outcomes of a random experiment. And the next thing is a bit more technical, which is we call it an event space, but technically it is called a $\sigma-$field or a $\sigma-$algebra, which is basically a collection of events. So, one should naturally include whatever possible events one can think of with respect to the random experiment inside this set. So, the non-empty collection, whatever the condition, is given here. The third ingredient for it is a set function which is $P : \mathscr{F} \to \mathbb{R}$, which is what we call a probability measure or a probability function or simply a probability, and that satisfies these three properties:

1. $P(E) \geq 0$ for all $E \in \mathcal{F}$

2. $P(\Omega) = 1$

3. Let $E_1, E_2, \ldots \in \mathcal{F}$ be a sequence of disjoint events then $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

The first is the non-negativity property the second is the total probability must be equal to 1 and if you have a sequence of disjoint events disjoint means $E_i \cap E_j$ is empty set then the probability of their union is the sum of this probability. This property is known as a countable additivity property.

So, any set function that satisfies these three properties will be called a probability. So, $(\Omega, \mathscr{F}, P)$ is known as a probability space. So, every stochastic model is specified in terms of the underlying object of $(\Omega, \mathscr{F}, P)$. As a simple example, if you look at the tossing of two coins. As you all know, the sample space would be $\Omega = \{HH, HT, TH, TT\}$, and you can take any event

space satisfying the condition given above. So, we take $\mathscr{F} = P(\Omega)$ which is normally is taken in case Ω is finite or countable, which is nothing, but the power set of Ω, which means the set of all possible subsets of this particular set Ω is what is called the power set, and you can define probability on what we call these four elements $\{HH, HT, TH, TT\}$ which are all elementary events in this Ω. So, if you define $P(\{HH\}), P(\{HT\}), P(\{TH\}), P(\{TT\})$ such that they sum to 1, and for any other event $E \in \mathscr{F}$, if you define $P(E) = \sum_{\omega \in E} P(\{\omega\})$ you will be defining a probability. So, this is a simple example; of course, the specification of this sigma-field is important here, and on top of it, if you change $\mathscr{F}$, you will also change P, or even if you keep $\mathscr{F}$, you can define P in different ways. So, you get different probability spaces. So, there is always a probability space under every stochastic model.

Related to that are the concepts of what we call what you know already as conditional probability and theorem of total probability, and Bayes theorem. So, you know, these are all elementary stuff that you can recall, like what do we mean by a conditional probability. A total probability is one of the important results. The theorem of total probability is an important result that helps us compute probabilities of certain events very nicely. Say what you have is these particular results that you know we are pointing out here:

Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0, \forall i$. Let A be any event with $P(A) > 0$. Then $P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_j P(A|E_j)P(E_j)}, \quad i = 1, 2, \ldots$.

It is a very important concept used when you want to compute the probability of certain events. You can break it up into smaller pieces, and then you aggregate it together; that is what it would mean. Bayes theorem, of course, you will also know you can recall even for the elementary stuff. A related concept is what we call independence. Here we are defining independence in terms of probability.

So, sometimes it is also called stochastically independent or statistically independent no matter whatever it is; the definition is this. If you have two events, say the probability you would have seen in terms of

$$P(E \cap F) = P(E) \times P(F)$$

for two events E and F.

So, what we have given here is:
A finite collection of events $E_1, E_2, \ldots, E_n$ are said to be independent (or mutually independent) if for any sub-collection $E_{n_1}, \ldots, E_{n_k}$ of $E_1, E_2, \ldots, E_n$, $P\left(\bigcap_{i=1}^{k} E_{n_k}\right) = \prod_{i=1}^{k} P(E_{n_i})$.
Then we can extend that to countable and also conditionally independent what we mean. So, if $P(E \cap F)$ is given as a product of the probabilities of the individual ones, then we say these

two are independent. That is what you know we mean here. So, this is the basic probability idea, but what generally people work on is something you define on a probability space that you use further in all your modeling purposes. A quantity of such interest is what we call this random variable. So, what is a random variable here? That you have a probability space and a function, a real-valued function $X : \Omega \to \mathbb{R}$ is called a random variable if $(X^{-1}(-\infty, a]) \in \mathscr{F}$ for all $a \in \mathbb{R}$.

So, this is technical, which is required. Suppose if Ω is uncountable, then this makes much more sense, but whenever ω is finite or countable and $\mathscr{F}$ you take it to be the power set of that particular set omega, then any real-valued function would satisfy this. So, but otherwise, for other cases, whenever your $\mathscr{F}$ is not the power set of this Ω, you need to specify or check this condition for this function to be a random variable.

So, what it does is that is what is given here. The random variable X is used to push forward the measure P, which we define on $(\Omega, \mathscr{F}, P)$, to a measure on $\mathbb{R}$, which is what is called the probability distribution ok, that probability distribution. So, it is also giving another probability space only, but you know you are defining it on $\mathbb{R}$, and a $\sigma-$field or an event space on $\mathbb{R}$ and associated probability measure. That is what this random variable does. The purpose is to make our life easier. You would have seen in the basic probability examples in a basic course that the reason why you know you define a random variable and so on. So, that probability measure is typically defined in terms of what we call a distribution function that is defined to be this

$$F_X(x) = P(X \leq x)$$

This you all know that the

$$F_X(x) = P(X \leq x)$$

which is nothing, but $P(X^{-1}(-\infty, a])$. So, in order that you know, you can assign a probability to $(X^{-1}(-\infty, a])$, is the reason we want $(X^{-1}(-\infty, a])$, to be inside this $\mathscr{F}$, the event space, and you can recall what are the properties of this distribution function and so, on. So, in our sense, a random variable is just a function that does not depend on the probability P, but the distribution of the random variable depends on probability P.

So, for most realistic purposes, the stochastic or probabilistic model is defined in terms of a random variable or its distribution. Random variable means the values the particular random variable can take and the associated probability measure, but how does it come? It comes from the basic construct of the probability space. There, you define X that gives rise to this. So, one can see that this X, along with F_X, this F_X, can be defined in equivalently many different ways, then we call a stochastic model or a probabilistic model. Of course, you could have multiple random variables defined from the same probability space. Like in our queueing situations like there could be an inter-arrival time, there could be a service time, like that two random qualities that you need to define in order to define a complete model, but

all these things have an underlying probability space that is what you know you need to remember.

Now, as I said, F_X can be represented in many different ways; one of the ways is very simple in this particular case of the discrete random variable case. What do we mean by a discrete random variable? A random variable is said to be a discrete distribution if there exists an at most countable set $S_X \subseteq \mathbb{R}$, which we call support of the random variable X such that there is a mass at each of these points; $P(X = x) > 0$; and $\sum_{x \in S_X} P(X = x) = 1$. So, such p_X that function $p_X : \mathbb{R} \to [0, 1]$ is what is called the probability mass function as you know, and the relationship between F_X and p_X is given through

$$F_X(x) = \sum_{\substack{y \in S_X \\ y \leq x}} p_X(y)$$

and p_X and the F_X in a reverse way by $p_X(x) = F_X(x) - F_X(x-)$. So, as a whole like this line gives you the representation from or the relationship between F_X and p_X here.

As opposed to this, another class that is also very commonly occurring is a continuous random variable. So, a random variable is said to have continuous distribution if there exists a non-negative integrable function, say $f_X : \mathbb{R} \to [0, \infty)$ such that $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ for all $x \in \mathbb{R}$.

So, this particularly f_X is called the PDF of the random variable. So, this $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ will be the support. So, you can represent. So, random variable instead of the distribution function, if the random variable is discrete, then the random variable and its probability mass function, and if the random variable is continuous, the random variable and its the corresponding PDF function would define the model like the so-called whatever we could call as a model for it.

So, you can again recall certain properties of this PMF PDF, which is what you can recall and some of the common examples that we might encounter or you would have encountered are these distributions say, for example,

- (Binomial Distribution: $Bin(n, p)$) $S_X = \{0, 1, \ldots, n\}$, $p_X(k) = \binom{n}{k}p^k(1-p)^{n-k}$.

- (Geometric Distribution: $Geo(p)$) $S_X = \{0, 1, \ldots\}$, $p_X(k) = p(1-p)^k$.

- (Poisson Distribution: $Poi(\lambda)$ $(\lambda > 0)$) $S_X = \{0, 1, \ldots\}$, $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

these are all forms under the class of discrete distributions. So, there is

- (Exponential Distribution: $Exp(\lambda)$) $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$

- (Uniform Distribution: $U(a, b)$) $F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b. \end{cases}$

- (Normal Distribution: $\mathcal{N}(\mu, \sigma^2)$) $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$ if $x \in \mathbb{R}$.

and, say, Gamma distribution these are all in the continuous distributions; of which the most relevant if we are doing the discrete-time analysis then you will require this discrete-time distribution of geometric and, of course, Poisson distribution also would occur even in continuous time.

And in continuous time, mainly, we will be concerned with the exponential and gamma distribution as the most important distribution, and a Poisson distribution also would arise naturally. So, some of the well-known distributions you can recall and correspondingly like what the case is. So, this is the notation whenever we write in this form. These are just notations that we are going to use to represent such distributions, which you can understand here. Now, once we have these distributions in mind, there are always certain quantities associated with the distribution, which is important. Suppose I want to know the complete picture of the random variable X, then I need to know the distribution function equivalently, either PMF or PDF. If I do not know that completely, then it does not mean that I cannot say anything about the random variable X.

Nevertheless, you can say something if you know some more or some other quantities other than this, but if I know the distribution function, I can answer any question related to the random variable. That is what it would mean, but there are certain quantities in that sense that would be of importance or interest whenever you are dealing with any random variable, even if you have a distribution function, but you want again to summarize certain information in a very compact form. So, quantities that are used to do that business will also help us in the analysis part, as well as certain quantities, which are called expectations. Suppose if I have a discrete random variable with a probability mass function $p_x(.)$ and support S_X, then we say expectation or mean of X is defined by $E(X) = \sum_{x \in S_X} x p_X(x)$ provided $\sum_{x \in S_X} |x| p_X(x) < \infty$.; that means, this absolute summability, as you know we require it so, that this is defined in a meaningful way.

Similarly, the expectation is defined as $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$ for the continuous random variable with a PDF $f_X(.)$, provided $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$. Now both this form: expectation here we are writing it in terms of sum; here we are writing it in terms of integral can be written using Stieltjes integral. For those of you who are not familiar with Stieltjes integral, it does not matter much, but it means this, $\int_{-\infty}^{\infty} x dF_X(x)$, notation is what is then would be used. So, what does $\int_{-\infty}^{\infty} x dF_X(x)$ mean? In the discrete case, it would mean $\sum_{x \in S_X} x p_X(x)$, and in the continuous case, this would mean $\int_{-\infty}^{\infty} x f_X(x) dx$. Of course, if you know the theory about Stieltjes integral, it will be convenient, but otherwise, you do not need to worry too much about this part. So, this is and also like random variable there are mixtures what we call mixtures meaning that it has a continuous part it has a discrete part both of them are available as in a single random variable case. So, that is what you call mixtures or not mixers really, mixed random variable or mixers means like it is in certain proportions you are mixing the different distributions and so on. For all those things, you can write in a similar way, and also, if X is a random variable and

$g(X)$ is a function of this random variable where you know $g : \mathbb{R} \to \mathbb{R}$ is some nice function, we will assume that some continuous function or some increasing function and things like that. So, this is a technical term we need to introduce; we will not do that. So, if this is a nice function, its expectation can be computed as $E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)dF_X(x)$. This is an important result; otherwise, what would require? You would require a function of $g()$ and then which is we call Y, obtain its distribution function Y and apply $E(X) = \sum_{x \in S_X} x p_X(x)$, but rather you can directly apply this; for example, $E(g(X))$ would be $\sum_{x \in S_X} g(x)p_X(x)$ that is how it will be. It will look like when you write it in this form, in this case.

And again, you can recall the properties of this expectation operator. This is an operator whether it is linear or so. So, that you know, you can do certain operations. You can write the $E[X + Y]$ as the $E[X] + E[Y]$. So, the relevance for us is that you can break it if you want to compute the expectation of a certain quantity, you can break it up into pieces, and then you can compute individually, and you can sum it up; that is what it would mean. So, what does it mean for us? All those properties are fine, but what is their significance? What it means for us is what is more important for us. Now, there are certain special expectations like $E[X^r]$. If you take the g function to be X^r, then this is called the rth raw moment of X, and the first one, μ_1, we call as mean μ. We will often denote it is by μ rather than μ_1.

Similarly, $\mu_r^{'}$, which is this quantity, is called the rth central moment; we are assuming that all this expectation exists that is what it would mean, and then $mu_2^{'}$ is called the variance, which is simply we will denote it by σ^2, or variance of X. One quantity that is of use in our studies which is dimensionless is this coefficient of variation, which talks about how variable is a random variable compared to another random variable is given by C_X which is called the coefficient of variation, which is $C_X = \dfrac{\sigma}{\mu}$.

Some other expectations that are again useful when we analyze the models are what we call moment generating function or simply MGF, which is defined as $M_X(t) = E\left(e^{tX}\right)$ provided the expectation exists in some neighborhood of the origin. This expectation should exist for t in some intervals around the origin. So, if it exists, then the derivative of all orders exists, and its raw moments can be obtained through this relationship for all positive integers k. Whenever this MGF does not exist, of course, as you know that the MGF need not exist for every random variable, people have to work based upon the characteristic function that always exists, which is defined as the $E(e^{itx})$. So, that always exists, but this is more complex to work on; however, we will be working with only MGFs whenever MGF exists. Another generating function, or you know which is relevant to our case when we have non-negative integer-valued random variables, is this probability generating function, which is defined as $P(z) = E\left(z^X\right)$ that is nothing, but $\sum_{n=0}^{\infty} P\{X = n\}z^n$.. So, essentially it means that the coefficient of z^n will give you the corresponding probability.

So, that is why it is called probability generating function is what then you call this one. Now, as opposed to a univariate case, you could also have a scenario where on the same probability space, which means the one that describes the original model, you could define multiple random variables. So, for example, if you want to define a population's certain characteristics like, for example, height and weight. So, height could be one random variable; weight could be one random variable. Then you have a relationship as height and weight are related. So, some relationships will be there, or there will be no relationship. All those things will be studied through jointly distributed random variables, which are defined through a random vector, what we call a multidimensional or a multivariate random variable or a simply random vector, which is nothing, but $\mathbf{X} = (X_1, ..., X_n)$ where each of these is random variables on the same probability space as each other, which is the same probability space is what we are looking at it here. Now, just like for a univariate random variable, we have defined a probability function, either distribution function or PMF or PDF equivalently; here, also for any random vector, you have what we call joint distribution function, which is defined to be

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = P\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

which is the extension from $1-$dimensional to $n-$dimensional as you see here. Now from this, one can obtain what we call marginal distribution functions, which is defined $F_X(x) = \lim\limits_{y \to \infty} F_{X,Y}(x, y)$ for X. Suppose it is two random variables you are considering, and you are looking at the marginal distributions of X and Y, then, $F_X(x) = \lim\limits_{y \to \infty} F_{X,Y}(x, y)$ and $F_Y(y) = \lim\limits_{x \to \infty} F_{X,Y}(x, y)$. this is how you know you can obtain them.

And you can now then recall certain properties of these joint distribution functions and so on. Again, in a random vector case, we say that this random vector is discrete if there is a countable set $S_{X,Y}$ on which this jointly (X, Y), and this (x,y) the number of points (x, y) in $\mathbb{R}^{\not=}$ in this particular case, if this is countable and you have $P\left((X, Y) = (x, y)\right) > 0$ for all $(x, y) \in S_{X,Y}$ such that $P\left((X, Y) \in S_{X,Y}\right) = 1$, then you say this one is a discrete distribution. This random vector is a discrete random vector in this particular case. And then, you can define it just like a univariate case joint PMF, which is represented by

$$p_{X,Y}(x, y) = \begin{cases} P(X = x, Y = y) & \text{if } (x, y) \in S_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

And from which you can get the marginal PMF $p_X(x) = \sum\limits_{(x, y) \in S_{X,Y}} p_{X,Y}(x, y)$, for fixed $x \in \mathbb{R}$.

Another case, as usual, is the continuous case when you say that random vector (X, Y) is of continuous type; if there exists a function in $\mathbb{R}^{\not=}$ such that

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t, s)dsdt, \quad (x, y) \in \mathbb{R}^2,$$

satisfies then, we call this $f_{X,Y}$ as the joint PDF of (X, Y). So, again you can recall certain properties and when you will call and everything and not just that again in a similar way. What

is the meaning and definition of expectation of $g(\text{random vector} X)$, which is the similar way one can define it here.

And when you have more than one random variable on the table, you question whether they are related to each other or what the relationship is. So, one needs to study about these random variables are independent or not. So, how do we define the notion of independence in the case of random variables? It is again in terms of the distribution function. So,

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n.$$

then, we say $X_1, X_2, \ldots, X_n$ are independent.

So, again this condition can be written down in a simpler way. Equivalently whenever you have either PMF or PDF exists. Whenever that is not the case, as you know in elementary probability theory, PDF, for example, every distribution function, need not correspond to PDF. So, in such cases, one has to use

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

but whenever PMFs/PDFs exist, you can write down

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

in terms of the PMF or PDF.

This is for easy understanding. We will be using most in that form only. And similarly, when two random variables are independent and g and h again are some nice functions, then the expectation of this quantity can be written in this form. We are assuming that all the expectations would exist ok. With respect to two random variables, we have certain quantities apart from mean and variance that we define even for a univariate distribution. Of course, even in this case, for each random variable, you have mean and variance, but then you also have what we call covariance, which is

$$Cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right] = E(XY) - E(X)E(Y)$$

and equivalently the correlation coefficient which is

$$\rho_{X,Y} = \rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

and you know what the relevance is and what is the definition, what are the properties of all these notions random vector variances, covariances, correlation coefficients with and without independent what will happen and so on. Whenever $|\rho| \leq 1$, which means ρ lies between -1 and 1, both ends included. When $\rho = 0$, we say that the two random variables are uncorrelated, but then that does not imply in general that the random variables are independent, but in certain

cases, it might imply under what condition and so on. So, these are what we say with and without independent notion, the properties that we talk about.

Similar to MGF in the multivariate case also, you can define the MGF of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ to be

$$M_{\mathbf{X}}(\mathbf{t}) = E\left(\exp\left(\sum_{i=1}^{n} t_i X_i\right)\right),$$

as a simple extension of that idea and from where you can obtain $E(X_1^{r_1} X_2^{r_2} \cdots X_n^{r_n}) = \frac{\partial^{r_1+r_2+\ldots+r_n}}{\partial t_1^{r_1} \partial t_2^{r_2} \ldots \partial t_n^{r_n}} M_{\mathbf{X}}(\mathbf{t})\Big|_{\mathbf{t}=0}$, and X and Y are independent iff $M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$.

Now, apart from marginals, we also have the notion of the conditional distribution, which is very important in any probabilistic study. So, you have what we have a scenario? If you have a PMF $p_{X,Y}(.,.)$ and suppose the marginal PMF of Y, $p_Y(.)$ also exists, then the conditional PMF of X given $Y = y$ is

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \text{ provided } p_Y(y) > 0.$$

And the conditional distribution, you can write

$$F_{X|Y}(x|y) = P(X \le x|Y = y) = \sum_{\{u \le x : (u,y) \in S_{X,Y}\}} p_{X|Y}(u|y), \text{ provided } p_Y(y) > 0.$$

Similarly, you can write down the conditional distribution function for a continuous case with slight complexity. Because in a continuous random variable case, a random variable always takes a particular value with a probability of 0. So, the conditioning event $Y \in (y - \epsilon, y + \epsilon]$ cannot be equal to 0. So, what you do is that you take it in a small interval around that and then take its limit; that is how you define the conditional distribution function, and you have $f_{X|Y}(x|y)$ as the conditional density function which is related to $F_{X|Y}(x|y)$ as,

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y)dt, \qquad \forall x \in \mathbb{R}.$$

Now we know that whenever you have the joint PDF, $f_{X,Y}$, and f_Y is the marginal PDF at the points at which $f_Y(y) > 0$, the conditional PDF of X given $Y = y$ exists and is given by $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ which is something similar to what you would have seen here. So, that is the way you know in normal cases we have, but this is the general definition. But this is the normal case with which you know we might work here.

Now connected with this is what we call conditional expectation. So, which is essentially like if I have the $p_{X|Y}(x|y)$, I am looking at the

$$E(h(X)|Y = y) = \sum_{x:(x,y) \in S_{X,Y}} h(x)p_{X|Y}(x|y),$$

much similar to the other cases that we have seen. Similarly, conditional expectation of h of X given Y = y is defined for all values y such that $f_Y(y) > 0$, by

$$E(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x)f_{X|Y}(x|y)dx,$$

, but we are assuming that you know these expectations do exist in all these cases.

So, essentially, in a simpler sense, for a fixed value of Y, this would mean the mean of the conditional distribution or the expectation of the conditional distribution. Now, if this is the case, this is what it would mean $g(y) = E(X|Y = y)$. If I replace this small y by Y, that is how you define $E(X|Y) = g(Y)$; then $E(X|Y)$ becomes a random variable with these properties:

▶ $E(X) = E(E(X|Y))$.

▶ $E(X - E(X|Y))^2 \leq E(X - f(Y))^2$ for any function f. Thus $E(X|Y)$ is the "best estimate of X given Y".

Note here $E(X) = E(E(X|Y))$ is what is relevant here for us, meaning that if you want to compute the expectation of X, you can compute it in this manner; that is what it would the implication of this is very relevant for us that you can compute it using $E(X) = E(E(X|Y))$. And then, as you know, this is the best estimate you could have, which is not the case here and similarly for the conditional variance.

So, the variance of (X, Y), suppose if you call this as $Var(X|Y) = h(Y)$, where $h(y) = E((X - E(X|Y))^2|Y = y)$
$= E(X^2|Y = y) - (E(X|Y = y))^2$.

which is basically the variance of the conditional distribution in some sense, then $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$ is what is now we will be using it to compute the variance of X by conditioning on another random variable Y but then you have to use this relationship to get this. Similarly, the conditioning argument also helps us to compute the probabilities of any particular event. So,

$$P(E) = \begin{cases} \sum_y P(E|Y = y)P(Y = y) & \text{for} \quad Y \text{ discrete} \\ \int_{-\infty}^{\infty} P(E|Y = y)f_Y(y)dy & \text{for} \quad Y \text{ continuous} \,. \end{cases}$$

exactly the same as the total probability law that we have seen earlier provided if you assume this to be an event, this to be an event finish. So, this is how you will get in discrete and in continuous, as you know.

So, this kind of argument is what is relevant in our course. Now, like again, you can talk about the conditional expectation of a random vector (X, Y) given a particular event like $Y = y$ is again an event, but $(X, Y) \in A$ is,a sub-generic event that we can tell. Sometimes we may have to do this kind of computation. You may also recall some other special expectations like characteristic function. As I said, this is essentially $E(e^{itx})$.

Modes of convergence, there are four modes, one is called weakly, almost surely, and rth mean, and in probability like these are the four types of convergence and limit theorems like this law of large numbers and the central limit theorem which has the most important results and their significance and so, on and also some of the inequalities like moment inequalities you mean say

which have Markov inequality which sometimes you may have to use, and functions of random vectors especially sums of random variable how one can handle these sums of the random variable in case of a multiple numbers of random variables that you have to have and computational expectation by conditioning again and handling of a random sum of random variables etcetera like some of these concepts if you are familiar with it will be very useful if you can understand easily.

So, what we have seen is the notations mainly that we want to introduce that we might use in our course further, and you can get familiarized with this. You can also recall the basic concepts that we have seen that you have studied in your probability theory so that you can get on with the course very nicely, and I have introduced the notations and the basic concepts and what you need to remember here. And here, what we have is a table that gives us a set of ready references for us that certain probability functions, parameters, and the mean and variance and their MGF and their PGF. As you know, MGF and PGF are related by $E(e^{tX})$ and $E(z^X)$. So, if I replace e^t by z, I will get the PGF in case of when the PGF we can talk about, which is a case only if we have a nonnegative integer-valued random variable. So, all these random variables have support in this. We did not write the support part here, but that is understood like, for example, in this binomial case, what are the support you have here, your x ranges from 0 to n. This is, you know set of all nonnegative integers like geometric $p(1-p)^x$ & $0 \leq p \leq 1$, and this is negative binomial $\binom{k+x-1}{x} p^k (1-p)^x$ and the relationship between the different distributions. So, for each of these distributions, if you can concentrate here a little bit more binomial like you know, this MGF PGF will be relevant for us whenever you know we want to identify something that we have derived is of this form we said ok. This is of this form, so this must correspond to a binomial distribution. Because you know the PGF or MGF has a one-to-one relationship with the corresponding distribution.

So, whether you talk in terms of the probability distribution or in terms of their MGF or PGF, the results will be the same. So, mainly for us, like these two distributions, Poisson and geometric are very important apart from others. Of course, negative binomial as well because that is occurring as the sum of you can think about the other one, binomial is the sum of Bernoulli and so on. So, Poisson and geometric distributions, of course, for example, then $e^{\lambda(e^t-1)}$, $e^{-\lambda(1-z)}$ & $\frac{p}{1-(1-p)e^t}$, $\frac{p}{1-(1-p)z}$ like you would see that very frequently occurring in our calculation that you might do. And this is one form of geometric distribution; as you know, there is another form of geometric distribution written $p(1-p)^{x-1}$. The difference being is that $x-1$ here and x here. One denotes the number of trials required to get the first success, and the other is the number of failures before the first success as if you have to distinguish between these two distributions. So, both these forms might be useful depending upon the scenario you might encounter here. So, these are all some of the discrete distributions and their moment-generating functions.

Similarly, we would also like to focus on these continuous-time distributions and moments and MGFs as well. So, here especially as usual, like in the discrete case. So, here we might make more use of Exponential, Gamma, Erlang-k and 2-Term Hyperexponential than normal

and uniform. So, the exponential distribution and Poisson distributions are bread and butter for many of the initial phase models that we might consider. So, this is again, in that sense, this exponential distribution of $\theta e^{-\theta x}, \theta > 0$ form, and $\frac{\theta}{\theta - t}$ is its MGF. So, the characteristic function if you want for either discrete or continuous, you can always look at replacing this t by it, then you are going to get the corresponding characteristic function.

So, gamma, which one can think of as a sum of iid exponential like this α number of times. So, here α, in general, could be any number, but suppose if α is an integer, then that is then the sum of n iid exponential with the same parameter θ. Erlang distribution is what we might call a special case of gamma in some sense, if you have to see. Suppose if I look at $\frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}$ & $\frac{(\theta k)^k}{(k-1)!} x^{k-1} e^{-k\theta x}$ what we have done is that this θ we are replaced by θk in a way. The reason is suppose if $1/\theta$ here also $1/\theta$ whereas, something else will come here. So, we want to make some adjustments here, but here k is restricted to 1, 2, and so on. So, you will understand as you go to the theory, Erlang-k is what you know that we would mean the sum of exponential, but the parameters are tweaked a bit in a way, and 2-Term Hyperexponential $p\theta_1 e^{-\theta_1 x} + (1-p)\theta_2 e^{-\theta_2 x}$ is what you call it as mixtures of two exponentials is what it is the scenario here.

So, essentially here exponential if it is in some sense if it is put it in series, you will get is Erlang-k and if it is in parallel, you will get this Hyperexponential and so on. Again much general version; this is what you called hypoexponential, which will also be used. And sometimes, a gamma distribution with certain parameters $\alpha = n/2$ where n is this set of natural numbers, and $\theta = 1/2$ is known as the χ^2 distribution, which will arise in our studies with this case. So, these are some of the distributions that we might encounter.

So, what we have done so, far is that we have reviewed the probability theory like whatever you know; this is nothing new because we are assuming this knowledge of probability theory; what I highlighted is what are the quantities of interest that are most relevant to us and what are the notation that we might use when we go further. So, this is a quick review. So, as we cover it up in this video, this complete portion that what you might go through, what you might recall, and what you need to a little bit more elaborately you need to look at. So, with this, we will end this review.

Thank you, bye.