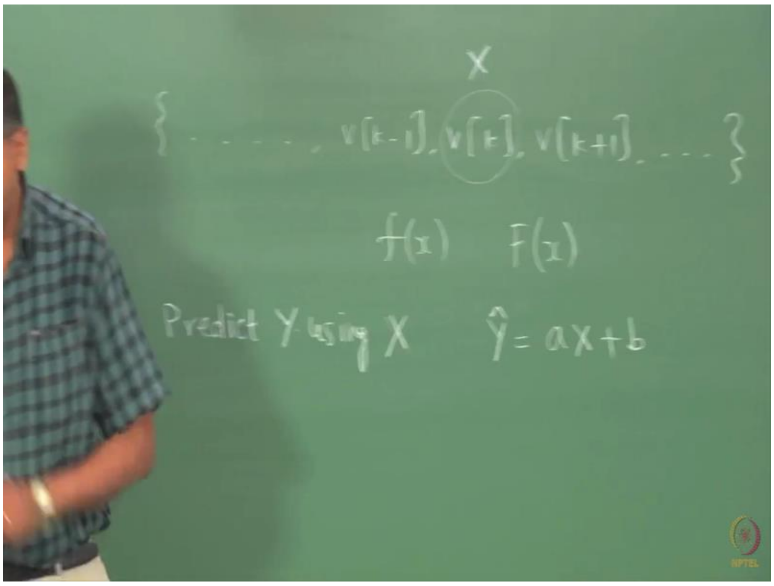
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 09
Lecture 05A - Probability and Statistics Review (Part 1)-3

So, yesterday what we began with is understanding how to characterize the random signal, at any kth instant. So, we first understand learn how to statistically describe the random signal at any instant and we denote let us say at any instant this observation by random variable x; then gradually we will learn how to analyze any two observations and then the entire signal, so that is the course on which we are.

(Refer Slide Time: 00:59)



And what we learned yesterday is this random variable x is characterized by the probability distribution function and because we are dealing with continuous valued random variables, we will worry about only the density functions. Occasionally we will refer to the distribution function whenever necessary and we went through a few density commonly used density functions and so on.

Now, all of that is nice in terms of computing probabilities and so on, but when you look at a practical scenario or real life situation the pdfs are perhaps not so handy as much as what we call as moments and that is the subject of todays at all. When you; let us take an example, suppose I am visiting let us say a new city I do not know anything about that city and I want to really I want to make a decision on what kind of clothes to pack. So, it is a very classic example that I keep giving and; obviously, then the decision on what clothes to pack depends on the climatic conditions in that city. Now for all practical purposes or realistic purposes, the temperature in any city can be treated as a random variable. So, I am looking at the temperature not only at the time I land in the city, but also you know during my stay and so on.

Now, if you look at today's resources that you have whether it is Wikipedia or any other resource that you have or even a book you lonely planet and so on, the city apart from the description of city, you will also see a description on the climatic conditions. Anywhere have you come across, this probability density function for the temperature of a city. So, you say go into Paris or you go into some other place let us say Vienna or you know even Guwahati, this is the probability density function of the temperature during the season; have you seen that say statement anywhere? No, why do you think that is not given because theory says that, the probability density function of a random variable gives me complete description.

So, why is not the density function given I mean there is so much data, there is so much research that has gone into it, we can compute the probabilities, we can do a lot of stuff why do you think these density functions are not even talked about any time in our lives and their entire life goes in without talking about pdf and I am visiting cities.

Why do you think that is the reason; that is the case, any idea?

Student: Mostly that is the (Refer Time: 03:41) quantity. And also purpose of your (Refer Time: 03:43) it is pointless to know the whole (Refer Time: 03:45).

The second part of it is alright, to say that I would rather have may be pointless you will probably disappoint the many proud people working in probability and statistics. So, you are saying as far as decision making is concerned it is to know the min and max; not a bad answer any other answer from the other hall, assuming that you can hear me, no answers that is it.

So, let us take another example suppose you are you have landed up with a job that pays you daily salary and you have to make a decision to whether to accept that offer or not. Imagine you going and ask an employer can you give me the pdf of salaries, how would that sound do not you think he will take back the offer letter. Why, why do not they ask for the density function there; any it is closer to what you are aspiring for, so you may be able to answer better.

Student: (Refer Time: 05:19).

No I am not asking first ones well see just want pdf; I do not really bother, I do not care really how the employer gets me the pdf, I just want the pdf man I mean, I am being taught in time series analysis that pdf characterizes a random variable.

Student: (Refer Time: 05:41).

Ok.

Student: (Refer Time: 05:47).

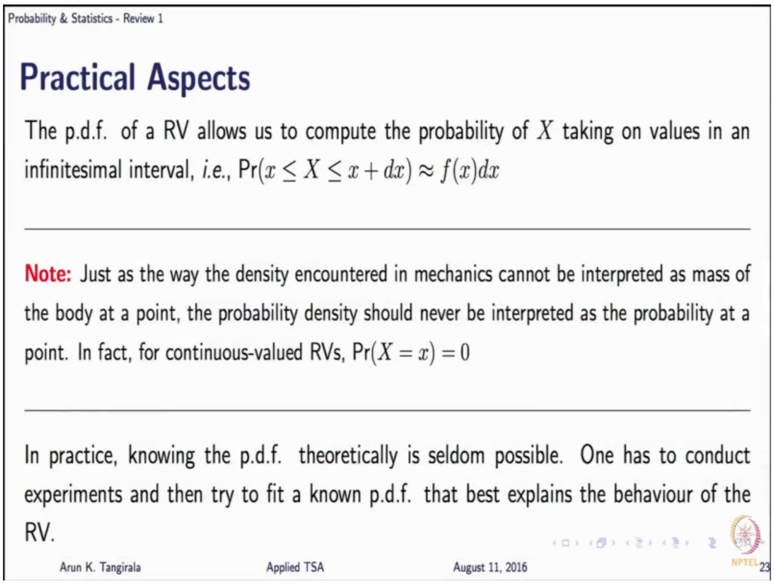
Not such a satisfied answer, so now let us understand there are at least a couple of reasons why practically you do not see the pdfs being mentioned about. The first fact is that it is not easy to obtain an accurate and reliable estimate of pdf. Although it is not the top reason, but that is a very significant reason as to why not only in the examples that I have given, but in any other literature on data analysis at least some; let us say in the linear world or big body of literature on data analysis, you do not hear anybody estimating pdf, simply because estimating pdfs is not so easy; easy in the sense it takes a lot of data to estimate the pdf accurately and reliably.

In the examples that I have given agreed that the pdf can change with time, so that is an additional complexity, even where you have situations that the pdf remains invariant with time so called stationary processes, there as well you have a situation that you require lot of data to accurately and reliably estimate the pdf. Now equally perhaps more important reason is in the linear world that is when we are looking at fitting linear models and when we are making so called basic predictions. Typically all our predictions are begin with linear models and then if they do not meet they do not serve the purpose then we move to the non-linear world.

In the linear world, it is not necessary to know the pdf as far as predictions are concerned and that you will know through today’s class as well as tomorrows lecture that it is sufficient to know what are known as the first and second moments of the pdf. And it is a lot easier to estimate those quantities rather than the pdf. The other answers that you have given will plague at every stage of the data analysis that is the presence of uncertainty or changing quantities with time and so they are very generic, those are not the reasons why we do not work with pdfs, they taught two reasons why we do not need will not work with pdfs and need to work with pdfs is simply because one pdfs cannot be estimated accurately and reliably as good as you can actually estimate the moments.

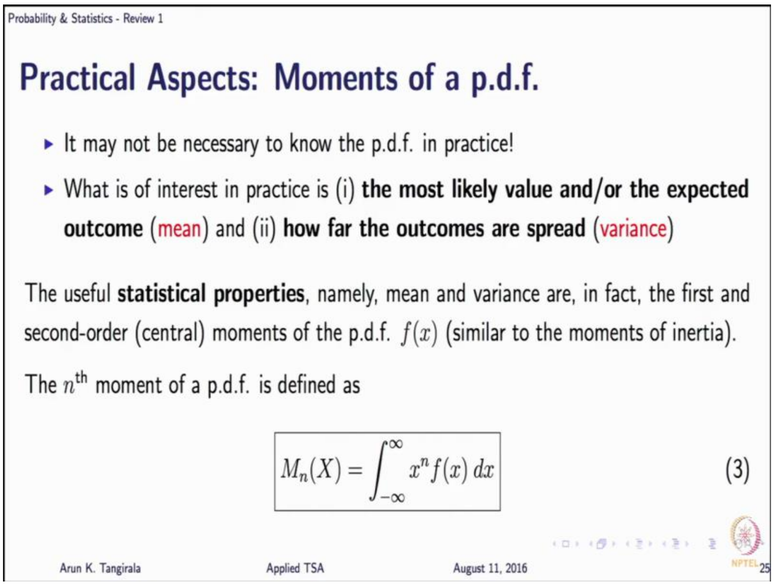
Now, having said that there are methods in the literature it is not that there is no method out there at all, there are methods out there to estimate pdfs, but we do not need them in a linear world. So, when there is no need necessity is a mother of all invention, so we do not need so we are not really paying so much attention. When the need arises then we look into estimates how to estimate pdfs and that typically occurs in the non-linear world alright.

(Refer Slide Time: 08:53)



So, in practice knowing the pdf is a much more cumbersome and herculean task as compared to estimating what are known as the moments, alright. Today what we are going to focus and henceforth in the course, we are going to work with the moments of pdf and of course, will convince ourselves to why this moment knowing these moments is sufficient as far as the linear modeling is concerned. So, how are the moments of the pdf defined? As I said this is a review so will kind of breeze through some of the definitions quickly. You must have come across this definition of the moment and it is not that this is the first time your encountering moments of a density function. As I mentioned yesterday, a good analogy is a density function that you encounter mechanics, there also we run into moments of inertia and so on you know we talk about center of mass and so on.

(Refer Slide Time: 09:53)



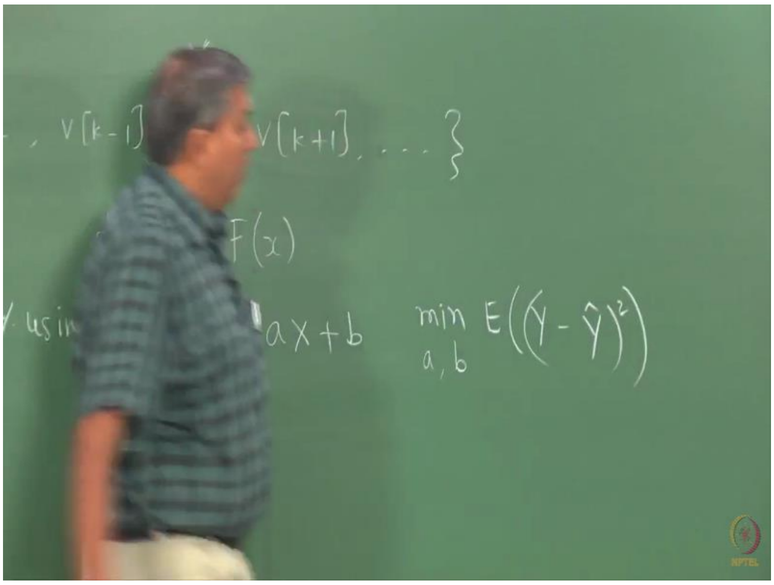
So, a general moment is defined as you see on the slide there it is an nth moment of the pdf and typically we work with two different moments; one is this pure moment as you see and the other is a central moment alright. The central moment differs from this ordinary moment in the sense that the moments are calculated about the center, but to calculate the center itself you need the moment and that is the first thing that will talk about, but in passing again I want to reiterate that.

As far as linear random processes are concerned or linear models are concerned what we mean by linear model is, suppose I am interested in predicting, let us say a random variable y using x. So, I want to predict a random variable y using excess random variable y could be the weight of an individual and x could be the height or y could be the temperature x could be the pressure of a gas there are numerous examples that one can give. So, this is a general scenario, I would like to predict y using x and this is a classic problem that one runs into in any data analysis exercise. Now when I want to do that typically I start with what is known as a predictor equation and predictions are denoted with a hat and this is a notation that will follow, I will talk about this notation later on as well.

So, suppose I say that the prediction or the predictor of y is a linear function of x, it is not exactly linear function it is an affine function, but now let us use linear in loose sense here. So, you have y hat equals a x plus b that is given x, I am going to plug in x into what I mean by given x is outcome. So, given pressure I am going to plug in that pressure value into this equation and obtain a prediction of the temperature knowing a and b, but a priori I do not know a and b, I have to estimate this from the historical data of temperature and pressure.

Now how do I generally do that, there are several ways of force fitting this predictor onto the data that you have, typically be force fed they are not going to listen to you otherwise.

(Refer Slide Time: 12:15)

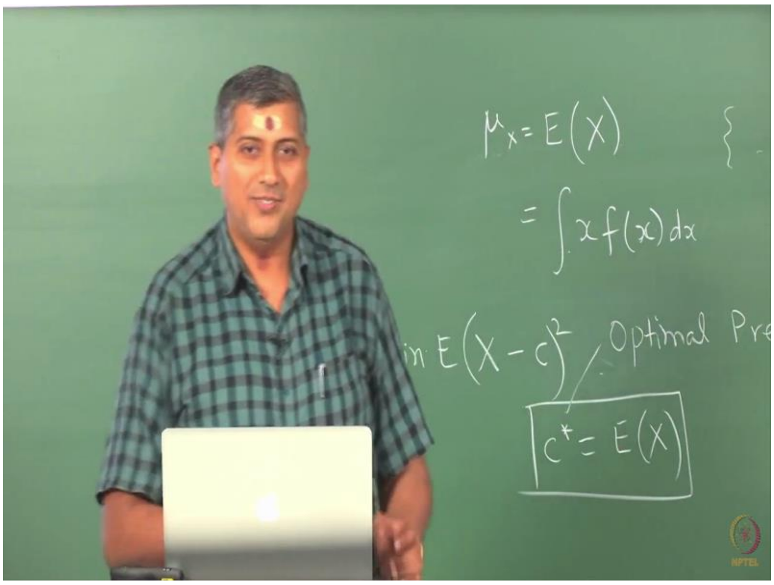


So, what we do is we say for example, find a and b such that some cost function is minimized, it is an essentially an optimization problem and we will come across this problem later on, but I am just giving you a glimpse up front. So, what we say is fine a and b such that we have not introduced expectation operator yet, but assuming that you have sat through the NPTEL course on intro to statistical hypothesis testing, you must be familiar with the expectation operator.

So, this is typically the optimization problem that we solve to estimate a and b we will show later on that the optimal estimates of a and b that is a solution to this problem does not require the knowledge of the pdf that is it I mean theoretically yes, but practically does not require because we can estimate the moments of a pdf without estimating the pdf. So, that is something that we should keep in mind alright. So, the point that I am making here is in when working with linear models it is sufficient to know the first and second order moments of course, we will see a similar situation even when we extend this idea to random signals alright.

So, let us go through quickly the first moment of a pdf which is known as the mean and this mean of course, the first moment would mean plugging in n equals 1 in the general definition of the moment and this mean also known as the average is called the expected value of the random variable.

(Refer Slide Time: 14:00)



So, the first time we are formally talking of expectations and as. I mentioned yesterday in the notation we use E for expectation operation and we denote this mean with the Greek symbol mu which is also the expectation of x. The name expectation is not without a reason, there is a strong reason for this terminology what do you think is a reason?

Student: (Refer Time: 14:28).

That is not correct most likely is different. See when you learn maybe expectations and so on in pure statistics courses this aspect is perhaps not dwelt upon, but the moment you come into the world of time series analysis then we are talking of predictions and the reason for this name is best understood from a prediction viewpoint and we will talk about that briefly in a short while. But you should always understand that expectation essentially is calculated this way of a random variable shortly will also look at the expression for computing expectations of a function of a random variable.

What is this expectation? This is a formula that I have written on the board the conceptual understanding is that, it is the average that is the first understanding. The second understanding is it is that single number that in some sense represents, the center of the outcomes you have a set of outcomes for the random variable X, it is like finding out who is the MLA or who is the MP. So, this is actually mean is some kind of a representative of all of the outcomes; obviously, there is more than one way in which you can represent the entire outcome space with a single number, this is one possible way. There are several other ways of representing your outcome space with a single number; an alternative is median, right. How is the median defined theoretically please do not tell me that you will sort the outcomes in ascending and descending order and so on that is a sample median, how is the median defined.

Student: (Refer Time: 16:22).

Right, so essentially the quantile you can say at which that partitions the cdf into two halfs that is the theoretical definition of median and then there is this most likely value which is the mode not most likely value you can say that the one that occurs with kind of maximum probability you can say that. So, there is mode which is an alternative to the mean there is median that is an alternative to the mean, yet we work with mean a lot of times, but it does not mean that median is not used, median has its own place, but this is one of the most widely used measure of the center of the outcome and it is very simple you have an analytical expression here to calculate. There are other ways of writing expectations for example, you can write this expectation using the cdf also we do not get into that. And there are situations where this expectation itself is not defined for a random variable, we will not get into such pathological situations they are usually reserved for assignments and exams, but in the lectures we assume everything is nice.

Anyway I am just joking, but we will assume that the expectations exist for that random variable and it is given the pdf, I can calculate it this way, so this E is the ensemble average and if you look at the integral a bit closely after all integrals or summations. So, what you are essentially doing is; you are computing some kind of a weighted average what you do not see here is the invisible denominator which is the integral f of x d x; these are all actually definite integrals; I do not mention the limits its understood that you are going to integrate over the entire outcome space.

So, the denominator is also integrated across the entire outcome space by definition of the pdf the denominator is 1. So, given a density function before you compute the mean or variance and other kind of moments, you should make sure that your density function is a legitimate one in the sense that the area under the density is unity because that represents a probability that something is going to something will happen. So, normally this is not mentioned assuming that you have already done that. So, we will omit this denominator.

But if you look at the expression full expression with the denominator it reminds you of the center of mass definition or center of gravity definition and so on and in that sense this is a weighted average, what are the weights probability, within a small in the vicinity of that observation that you are the outcomes sorry that you have. So, the expectation is in some sense a weighted average with the weights being the probabilities. Now there are other facts about expectation which and most important of all is the prediction perspective is which also tells us why this expectation operator acquires its name.

So, let us understand that very quickly. So, I have a random variable x and it has you know set of outcomes I would like to make a prediction or let us look at this way I would like to find that single number that can represent the outcomes in the best possible manner and the I will use a single number in many different ways, in many different decision making like we talked about the visiting a new city or the daily salary example and so on.

One of the first pieces of information that I seek is; what is the first piece of information that I would see? Let us say in the temperature, temperature example when I am visiting a new city?

Student: Average.

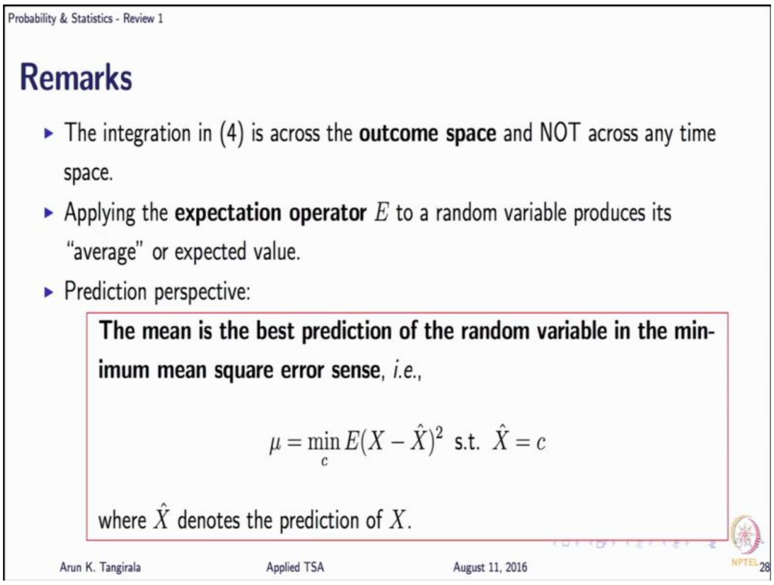
Average, likewise in the daily salary example I would like to know what is the average salary. So, that is the first piece of information although it is not sufficient it is necessary to have that information. So, let us say I want to now use look at all the outcomes and find that single number which will help me in some kind of decision making, essentially in terms of for predicting what can happen.

So, if you call that single number as c; I would like to choose the single number c in such a way. There are many different ways of choosing c, suppose I say that I will find the single number that will represent all the outcomes such that it is at a minimum squared distance from all the outcomes in a statistical weight; why I say statistical weight that we are using the expectation operator here. So what you are saying; if you look at this cost function carefully x minus c to the whole square is the distance you clear squared Euclidean distance from the number that you are looking for and any outcome x alright.

And then you are averaging it across the outcome space and you are not doing a simple averaging, you are doing a statistical averaging; that means, you are looking at the probability associated with that outcome rather the vicinity of that outcome and then you say find that number c such that this is minimized and you can straight away apply your optimization principles, essentially take the derivative and set to 0 and so on, this is a convex function and therefore, the solution is expectation of x, you can simply work it out. So, expand this cost function expectation operators is a linear operator, you can use that property, so expectation of sum of terms is sum of expectations and simply differentiate with respect to c.

In this slide I use x hat; that is a more formal way of writing it, essentially x hat is a prediction or you can say approximation whatever term you want to use your free to do that, but essentially I am going to use that to predict what is going to happen. Now; obviously, whatever happens is going to be different from what I am predicting, but this prediction is such that among all the possible outcomes, this particular prediction minimizes what is known as the mean square error. So, this is nothing, but the minimum mean square error prediction you can say approximation; whatever you want to call it.

(Refer Slide Time: 22:55)



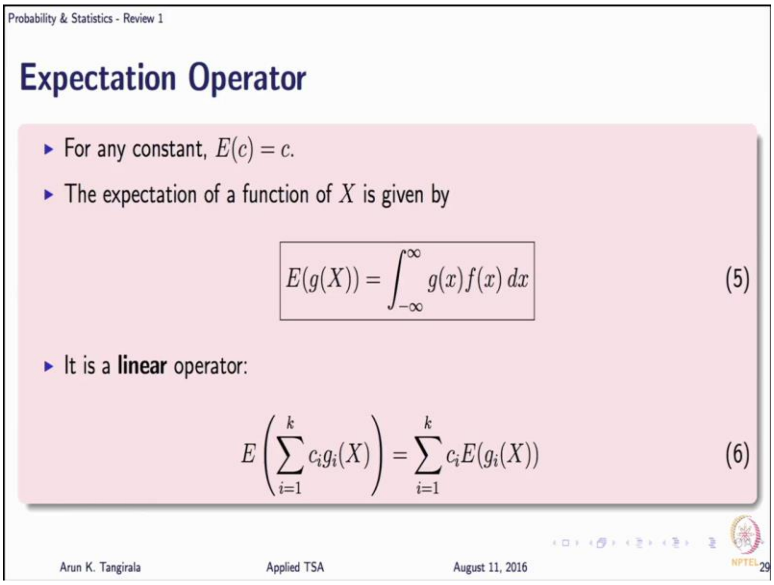
Therefore the mean has another interpretation from a prediction viewpoint, this is the best prediction that you can get in the minimum mean square error sense and therefore, you expect that the average will occur when no other information is given; I do not have any history at all, I am just looking at all possible outcomes at any instant and I am going to predict what is going to happen, no history is available. At the moment we are not looking at any random signal, we are not talking about the signal we are just talking about the random variable.

This is what we mean by a basic prediction, even you look at the random signal if you can improve this prediction then there is a scope for developing a model based on the history, but if the random signal is such that any amount of history is not going to improve upon this prediction then we say that is an ideal random signal. In the linear world we call we call this as a white noise signal, we will talk about that later; yes who is that yeah sorry.

Student: (Refer Time: 24:03).

Sorry just optimal solutions, see, come on you by now you should know in Bollywood you have stars and so on what are they? They are all optimal people. Now generally this is the notation that is used in optimization. So, this is a star c among all the c is a star, I do not know if it is mega star or superstar and so on, but it is star; depends on the font size I suppose.

(Refer Slide Time: 24:36)



So, let us move along; now as I said the expectation operator is a linear operator and the expected value of a constant is a constant. There is nothing much to think about it, the expectation of a function of X; you have to remember the you have to note the expression carefully, when you look at equation 5 here, you see that expectation of g of X is integral g of X; f of x d x, we are not going; many a times students think that it should be integral g of X; f of g of X; d x that is not correct.

When you are calculating expectation of function of a random variable, remember expectation is once again a statistical average. So some weights have to be associated attached to the outcome and when the outcome occurs, you are actually going to take a transformation of that through g. So, the weights that you are going to attach is still going to be the probabilities of X; not of g of X because once X occurs, then g is just a deterministic transformation, there is no randomness in g. Therefore, we still compute averages using the density function of X and not the density function of g of X, so do not make that mistake.