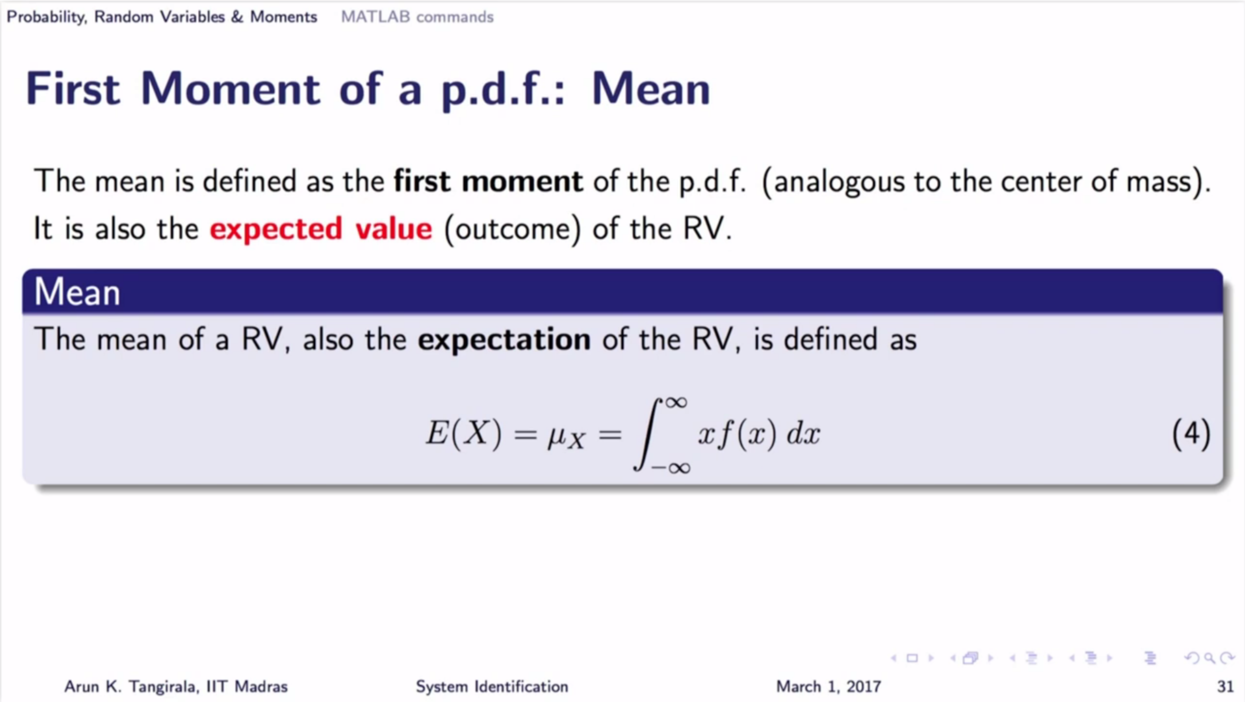
CH5230: System Identification

Probability, Random variables and moments: Review 4

Arun K. Tangirala

Department of Chemical Engineering
IIT Madras

Arun K. Tangirala, IIT Madras   System Identification



Probability, Random Variables & Moments   MATLAB commands

First Moment of a p.d.f.: Mean

The mean is defined as the **first moment** of the p.d.f. (analogous to the center of mass). It is also the **expected value** (outcome) of the RV.

Mean

The mean of a RV, also the **expectation** of the RV, is defined as

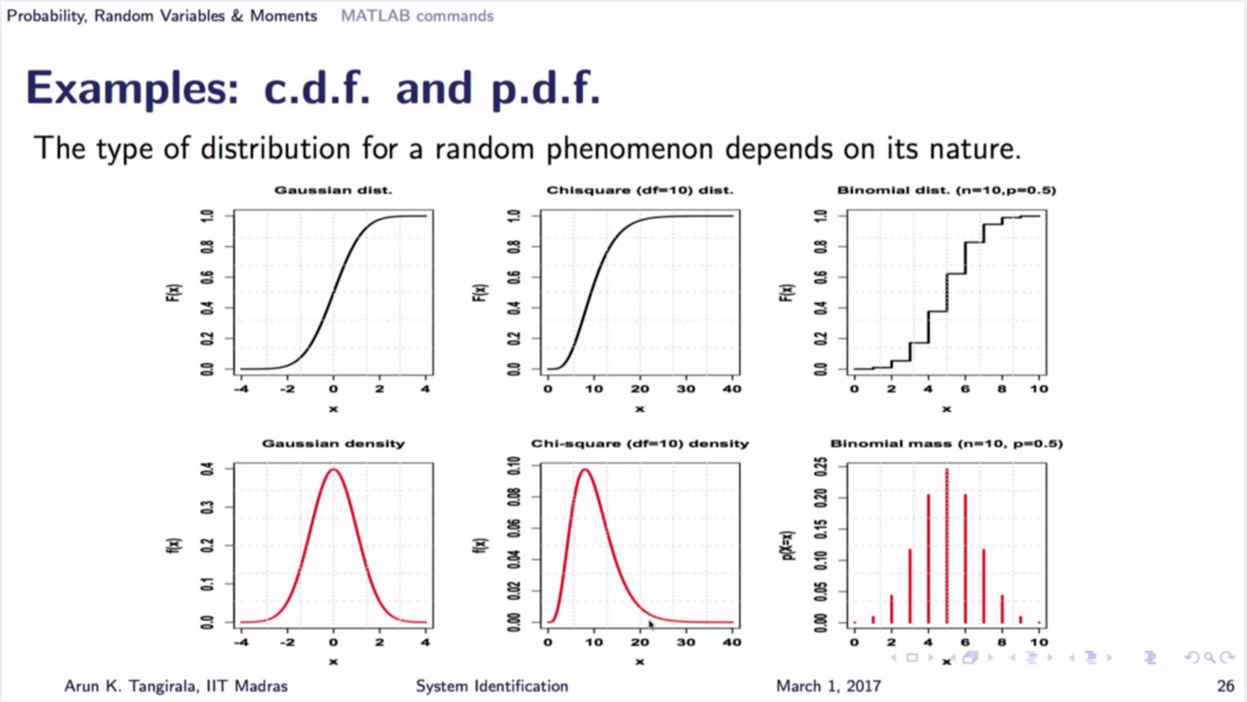
$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f(x)\, dx \tag{4}$$

Arun K. Tangirala, IIT Madras   System Identification   March 1, 2017   31

Arun K. Tangirala: So the first moment of the PDF denoted by expectation. There’s a reason why it’s called expectation, but let’s understand now what is this first moment, which is the mean, and this mean is a statistical center. Look at the integral carefully. You can look at it as a moment, fine, ∫ xf(x) dx, but I would also like you to view I in a different way. Think of this integral approximately as a summation, right. So you have ∫ xf(x) dx, approximately it is, I mean if I were to discretize the outcomes, this is what it’s going to be, right. I’d just quantize the outcomes. What you see now here, it is in some sense an average of the outcomes, but it is not a simple average, it’s a weighted average.

What are the weights? No, the xi themselves are the values. Probability, so we are seeking average of outcomes, outcomes are xi, these are the weights, right, $f(x_i)\ \Delta x_i$ is a weight. What is the interpretation of that? It’s already written on the board there. It’s a probability that x will take on a value within the vicinity of x_i. It’s a crude approximation but it’s not a bad approximation.

So this is the probability that x will take on values within the vicinity of xi. Therefore, what you are doing is by computing mean, you’re actually seeking an average of the outcomes, but a weighted average, so that the ones that have higher probability have a bigger say, obviously, which means for Gaussian, which is symmetric and so on, you should expect the mean to be here. In fact, it does turn out to be that, because it’s symmetric. Why did I write the mean there? In fact, the point where I’ve written μ is also the geometric center. Because it is symmetric with respect to that point, the geometric center and the statistical center coincide for the Gaussian, but for chi^2, you’ve seen the chi^2 distribution, it’s an asymmetric distribution if you look at chi^2. Look at here.



Probability, Random Variables & Moments MATLAB commands

Examples: c.d.f. and p.d.f.

The type of distribution for a random phenomenon depends on its nature.

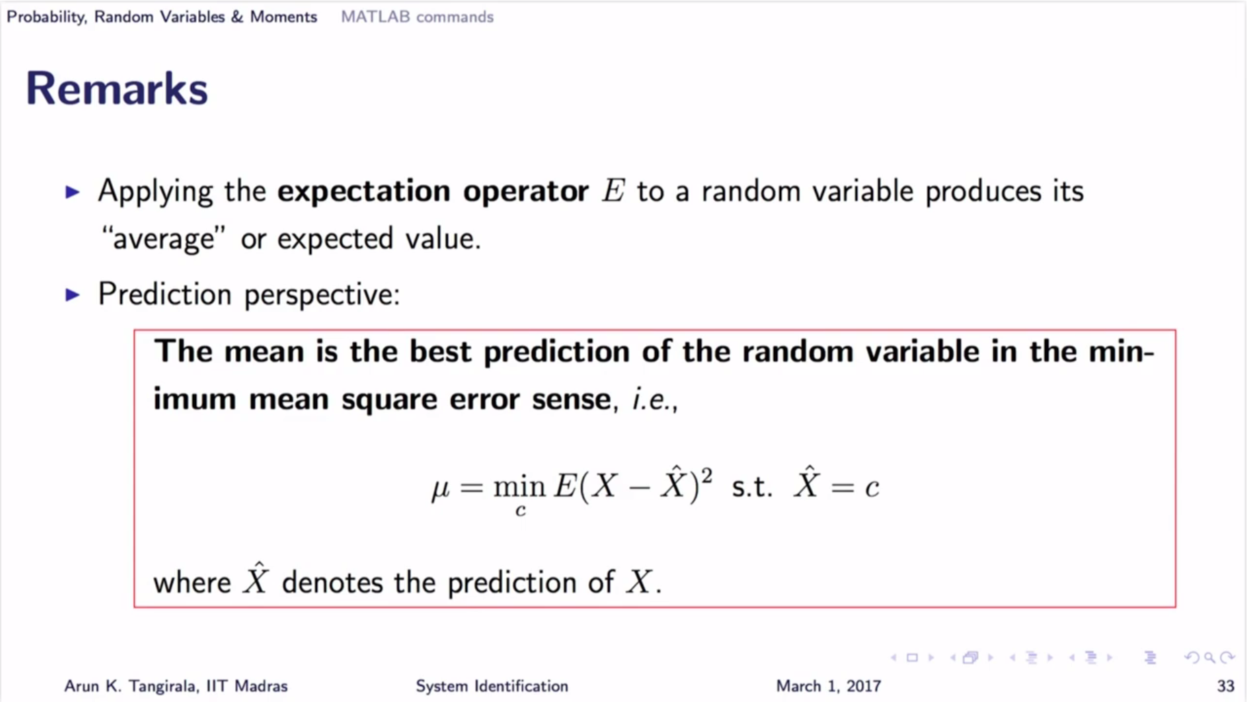
Arun K. Tangirala, IIT Madras System Identification March 1, 2017 26

This is an asymmetric distribution, in fact, you should notice chi2 distribution is only defined for non-negative value of random variables. In this case, the geometric center is somewhere in the middle here, all right, but the statistical center, roughly, where would it be, to the left, right. So they don't coincide. So there's a big difference between geometric center and statistical center, μ is the statistical center. You can say, it's a center in the probabilistic sense. If you don't like the word statistics, you can say it's a center in a probabilistic sense not in a geometrical sense.

Why is this piece of information important? Because we say it's expected value, it is expectation that I have for this random variable. What is this expectation? This is where the prediction perspective is extremely helpful. Many a times you don't see this kind of explanation in basic statistic textbook, but the fact is that if I were to seek the best prediction of this random variable, let's say the temperature in the new city. And in what sense am I seeking best? Best in a minimum mean square error sense.

So there are many possibilities for the temperature, so let's say, these are all the possibilities here. There are many possibilities for the temperature, or if you don't, obviously this is unidimensional, so maybe it's not a good idea to draw a region like this. Let's draw the line. So there are many possibilities for the temperature. I don't know which one the temp will occur on the day I land in the city, but I still want a prediction.

So let's say, prediction is somewhere here. I don't know, I mean my prediction could be here or it could be here and so on, but I want it in such a way that it is at a minimum distance from the outcomes, but minimum in what sense, in a probabilistic sense. The moment you have expectation of anything, you go back to the definition, we are saying expectation of x is the probabilistic average. It's not your regular average.



Remarks

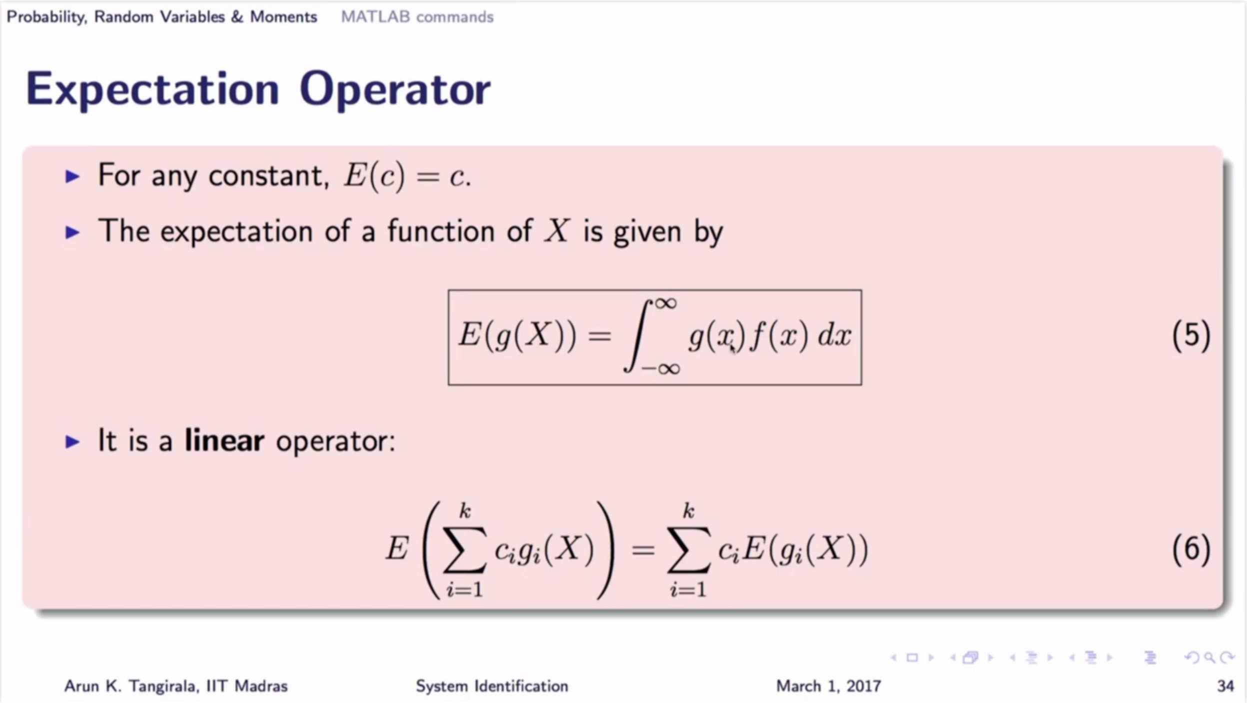
- Applying the **expectation operator** E to a random variable produces its "average" or expected value.
- Prediction perspective:

The mean is the best prediction of the random variable in the minimum mean square error sense, *i.e.*,

$$\mu = \min_{c} E(X - \hat{X})^2 \text{ s.t. } \hat{X} = c$$

where $\hat{X}$ denotes the prediction of X.

So expectation of $x - \hat{x}$, $\hat{x}$ is a prediction to the who square is the mean squared error, but this mean is in the probabilistic sense. It's not your regular -- it's not your simple suppose $\hat{x}$ x_i are all the outcomes, your expectation is not simply $(x_i - \hat{x})^2$, this is not your expected thing. There is an additional factor there, which is that it is a weighted average, right. What is expectation of $(x - \hat{x})^2$ roughly? Again, this is an approximation. Strictly speaking, it should be $\int (x - \hat{x})^2 f(x)\, dx$, this is the correct expression, because expectation of any function of x, we'll go back to that.



Probability, Random Variables & Moments MATLAB commands

Expectation Operator

- For any constant, $E(c) = c$.
- The expectation of a function of X is given by

$$\boxed{E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx} \qquad (5)$$

- It is a **linear** operator:

$$E\left(\sum_{i=1}^{k} c_i g_i(X)\right) = \sum_{i=1}^{k} c_i E(g_i(X)) \qquad (6)$$

Arun K. Tangirala, IIT Madras System Identification March 1, 2017 34

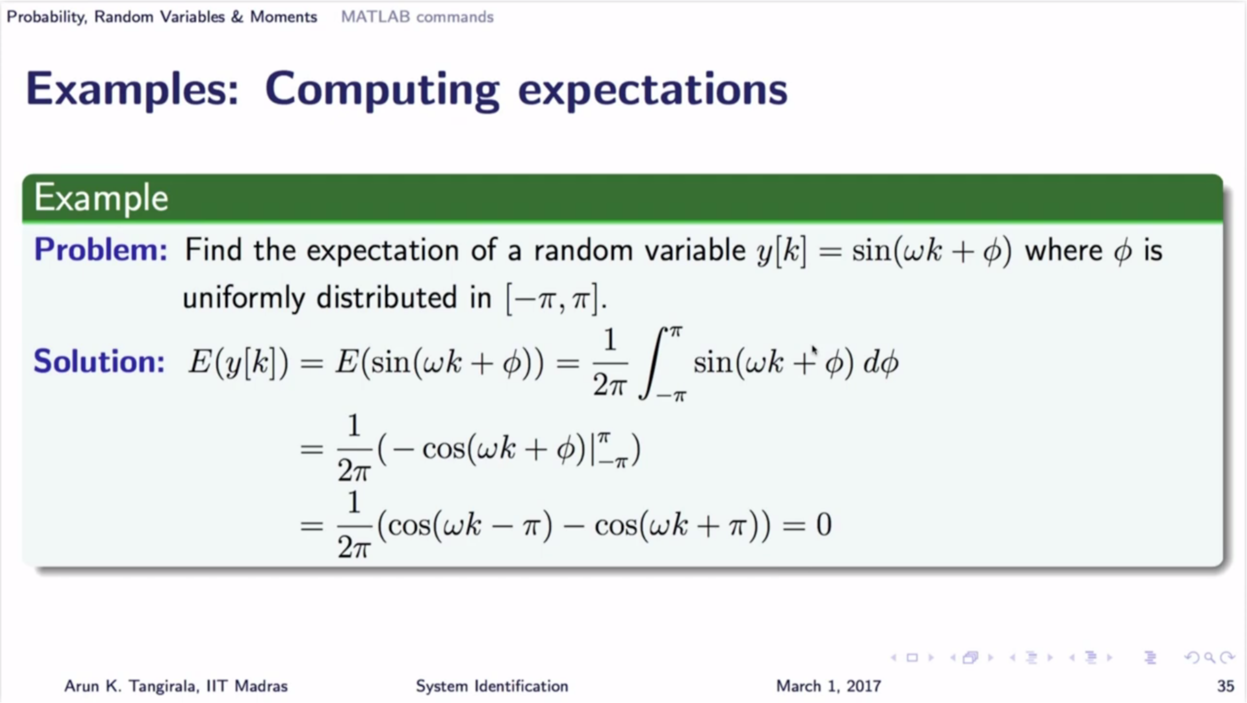
Expectation of any function of x is simply g(x) f(x) dx. Notice that very carefully. Many people think that is g(x) times f(g(x)) dx. That f(x) dx is a probability that x will take on some value within that, g is the deterministic function. So unless x occurs, you will not be able to calculate g. So the moment x occurs, the transformation is deterministic, which is d(x), x occurs with a probability f(x) dx roughly within that vicinity. Therefore, the rating that you give always is f(x) dx. It’s not f(g(x)) dx. There is nothing -- once has occurred g(x) is going to happen for sure, because you’re going to transform it. So there is nothing uncertain about that. Therefore, the weightage always have to be given to the occurrence of x not the occurrence of g(x).

So what we are essentially asking here is for a prediction that is best in the minimum mean square error sense, also known as the MMSE. It’s a very standard acronym that you should get used to, minimum mean square error. Square error because you are looking at squared distance between x and x̂, mean because you will compute an expectation, minimum because you are seeking an optimal value. And when you solve this problem, simply you’ll get the solution as expected value of x. Just remember that expectation operator is a linear operation, so you’ll have three terms in the quadratic, right, and only two terms will depend on x̂ and you differentiate and set it to 0, you’ll get the optimal x̂ as being expectation of x, which is nothing but mu. And that is why it is called expectation, because you expect that. It is not that it will happen that day, but in the absence of any other information, what is this any other information, I’ve not looked at any correlations, I’ve not looked at what has been the temperature yesterday, I don’t know what the temperature was the day before I land and so on. I am going to land maybe three months of four months later. So I am just looking at -- it has a random variable but has a random signal yet. The moment I look at it, the temperature, as a random signal, I will probably bring in correlations also, the time correlations, but at the moment, we are not looking at it that way, we are just treating it as a random variable.

So I am only looking at the prediction in that sense and that best prediction happens to be the mean. That is why we tend to work with averages. Now the big, like in the biggest font size possible caution, I want to give you is typically the moment you hear the word average, you tend to think of averaging in time, which is absolutely wrong. Does time figure out in any of this definitions, time has not occurred anywhere, appeared anywhere, right. We have said already when we talk of random variables, we should let go the notion of time. So you should never ever think by default that average is averaging in time. These are theoretical definitions, these are defined in outcomes space, not in the time space. The moment you see expectation, you should think of the outcome space.

So random variable or a random signal always has this additional dimension called outcome, which should not be confused with these other dimension or time. So we have frozen time here, standing somewhere where, and we are analyzing in this direction. So we are asking what is the average in the outcome space and what kind of averages we are computing, probabilistic averages, remember that. So any time you see expectation, it is actually an averaging in the outcome space not in the time at all, never, okay. So that is something to remember.

So we've already talked about the properties of expectation. Now general thing to remember is expectation of anything is average. What kind of average? Probabilistic average.



Probability, Random Variables & Moments MATLAB commands

Examples: Computing expectations

Example

Problem: Find the expectation of a random variable $y[k] = \sin(\omega k + \phi)$ where ϕ is uniformly distributed in $[-\pi, \pi]$.

Solution:

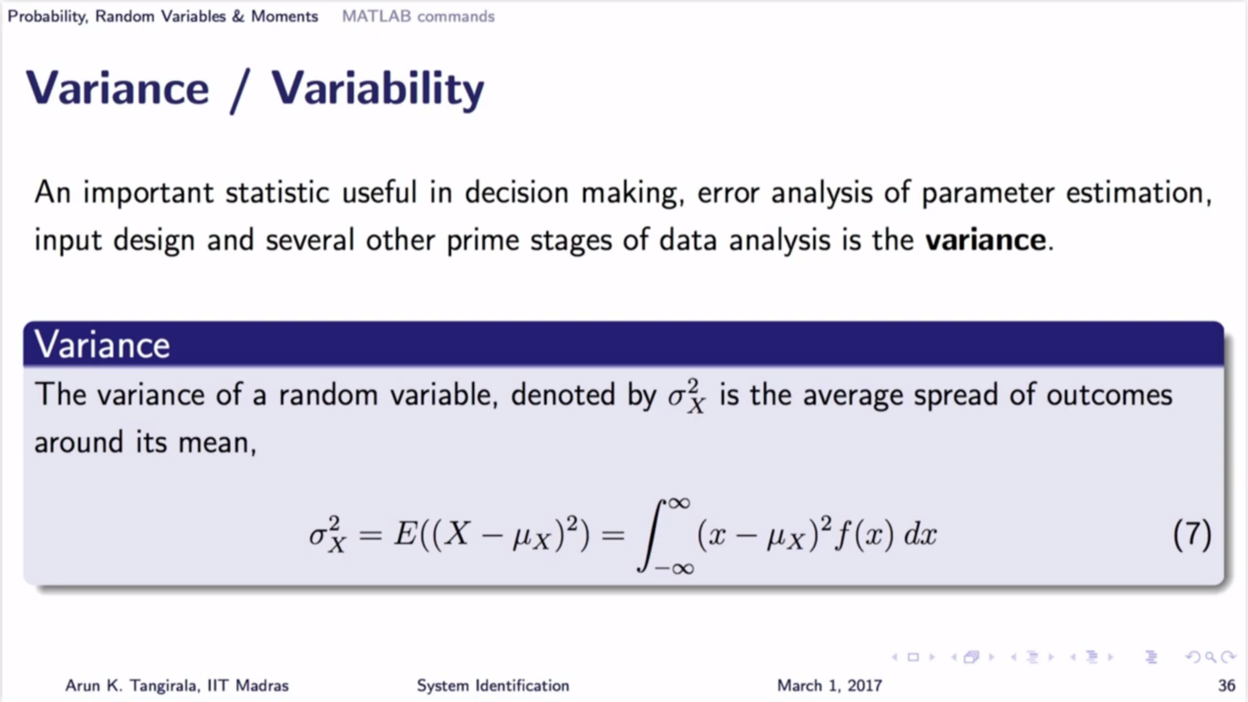
$$\begin{aligned} E(y[k]) = E(\sin(\omega k + \phi)) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sin(\omega k + \phi)\, d\phi \\ &= \frac{1}{2\pi}(-\cos(\omega k + \phi)|_{-\pi}^{\pi}) \\ &= \frac{1}{2\pi}(\cos(\omega k - \pi) - \cos(\omega k + \pi)) = 0 \end{aligned}$$

Arun K. Tangirala, IIT Madras System Identification March 1, 2017 35

So the next -- I mean just to reinforce this fact that expectation works across outcomes and not across time, I have a very simply example here, sine(ωk + φ). Now I am saying that any instant k, this signal, although I am writing it as a function of time, at any instant k is a random variable. Again, I have to find out what is the source of randomness here. Sin is not a random function, it’s not going to generate a random value, it’s going to generate only one value. Omega is fixed, I am fixing it. It is the phase that has randomness in it, that is when the sin began I don’t know. Phase is a measure of when the sin began with respect to some reference or the signal began. And I am assuming this φ to be random variable, uniformly distributed in interval [-π to π], because that is usually the case. The phase of signals are uniformly distributed.

Given this information, what is expected value of yk? Now very quickly, people look at sin and say, oh, average of sin, that is 0, but that’s in time. You’re supposed compute freeze time, stand at any instant in time and look at the outcomes. What are the reasons for these many outcomes at any instant in time, φ, φ is the random variable. So if you related to the notation that we’ve used, x is a random variable, g(x) is transformation. So here the g(x) is sin. Well, there’s an ωk, right. So expectation of g(x) is g(x) f(x) dx. What is the density function here for φ? Random variable is φ 1/2 π, because it’s uniformly distributed. So f(φ) is 1/2 π, the integral is not in time, it’s in outcomes space - π to π, that’s the possibilities for φ, and once you work out the integral, you get 0. Yeah that coincides with time average, but it coincides with many things. It also coincides perhaps with my marks and so on, but that doesn’t mean it is the same. It just happens to be that it’s the same as time average.

If you change, for example, as a simple homework, change the interval of φ instead of - π to π say it is 0 to π, uniformly distributed, and see if you get the same average, very simple, right. You just change the possibility for φ from - π to π to 0 to π. Rest of the story remains the same and see what kind of average you get. That will be the simple exercise.



Probability, Random Variables & Moments MATLAB commands

Variance / Variability

An important statistic useful in decision making, error analysis of parameter estimation, input design and several other prime stages of data analysis is the **variance**.

Variance

The variance of a random variable, denoted by σ_X^2 is the average spread of outcomes around its mean,

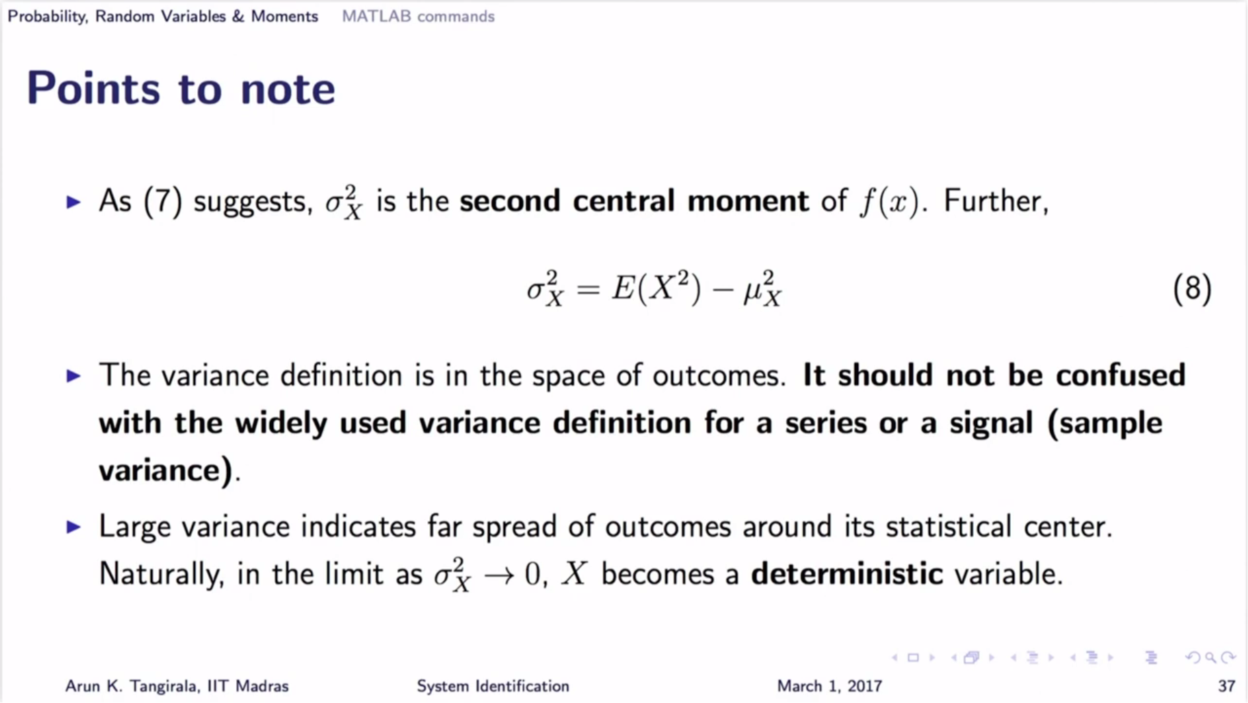
$$\sigma_X^2 = E((X - \mu_X)^2) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx \qquad (7)$$

Arun K. Tangirala, IIT Madras System Identification March 1, 2017 36

Okay, so the second moment that’s of interest is variance, which gives you an idea of the spread of outcomes. One gives the center of outcomes, other gives you the spread of outcomes, which is this variance. It is defined as a second central moment of the PDF, and once again here, the expectation comes in expectation of (x - mu)^2. So you’re looking at how the outcomes are spread with respect to the mean, which is the center and that’s why it’s called central moment, and you’re looking at a squared spread, squared distance, average of the squared distance of each outcome with respect of mu, average in a probabilistic sense, okay.

The square root of this variance is known as a standard deviation. So what we are giving you here are theoretical definitions of mean, variance and so on, never ever confuse these definitions with the expressions that we use or the formulae that we use for estimating them. Sample mean for example is an estimator of the mean. Sample variance is an estimator of variance. What is the different between sample mean and theoretical mean? Sample mean works with time data [Indiscernible 00:15:11] in time, or across experiments, likewise sample variance, whereas the true mean, true variance are looking at the outcomes, right. So one is a theoretical definition, other is an estimator. When we enter the world of estimation, we will talk of sample mean, sample variance and so on.

Okay, so once again here, it's a second center moment, it's an average expectation -- you should be extremely comfortable with working with expectation, and that is expectation that I have of you, okay. You can rewrite this variance as difference between the second moment and mu^2. Expectation of x^2 - mu^2. Often this expression is useful.



Probability, Random Variables & Moments MATLAB commands

Points to note

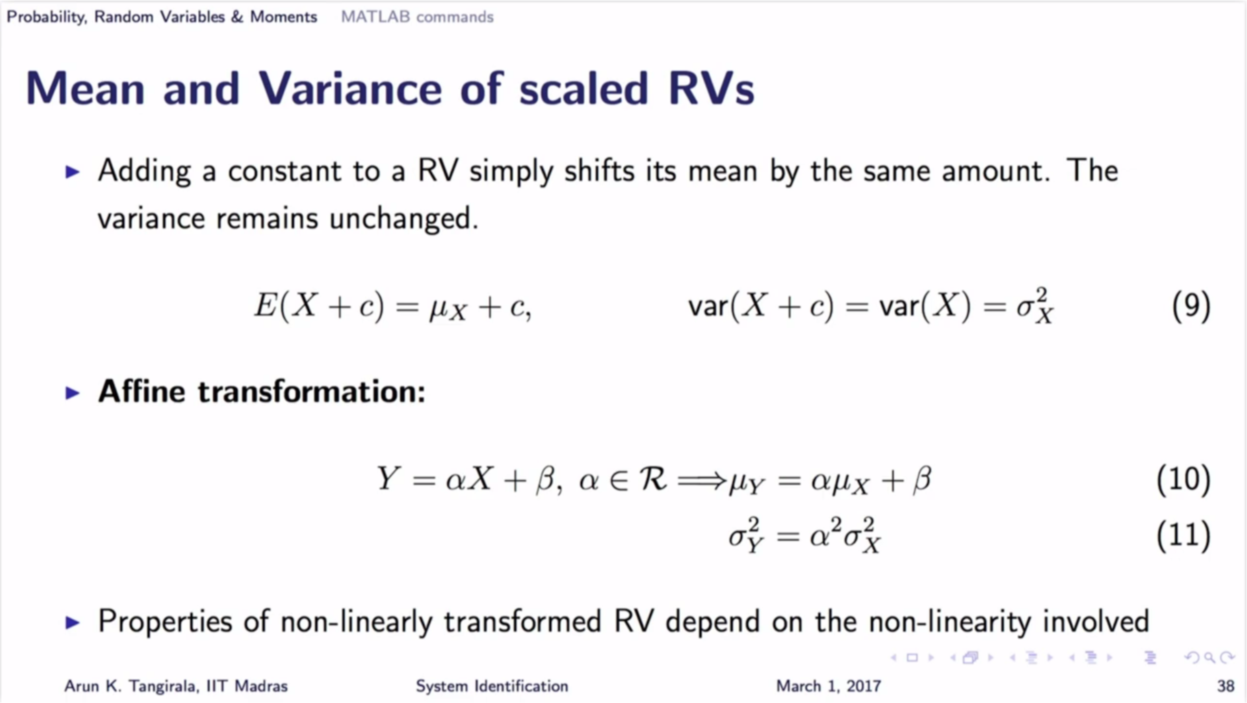
- As (7) suggests, σ_X^2 is the **second central moment** of $f(x)$. Further,

$$\sigma_X^2 = E(X^2) - \mu_X^2 \tag{8}$$

- The variance definition is in the space of outcomes. **It should not be confused with the widely used variance definition for a series or a signal (sample variance)**.
- Large variance indicates far spread of outcomes around its statistical center. Naturally, in the limit as $\sigma_X^2 \to 0$, X becomes a **deterministic** variable.

Arun K. Tangirala, IIT Madras System Identification March 1, 2017 37

Remember that any random variable by definition will have a non-zero variance, and non-zero and non-negative by definition. When the variability shrinks to zero, what does it mean? The outcomes are shrinking to one point, and therefore, it becomes a deterministic variable. For this reason, variance is considered a measure of uncertainty. Higher the variance, more the spread, so in some sense, more the uncertainty, larger the uncertainty is, okay. That's a very important thing to remember.



Mean and Variance of scaled RVs

- Adding a constant to a RV simply shifts its mean by the same amount. The variance remains unchanged.

$$E(X+c) = \mu_X + c, \qquad \text{var}(X+c) = \text{var}(X) = \sigma_X^2 \tag{9}$$

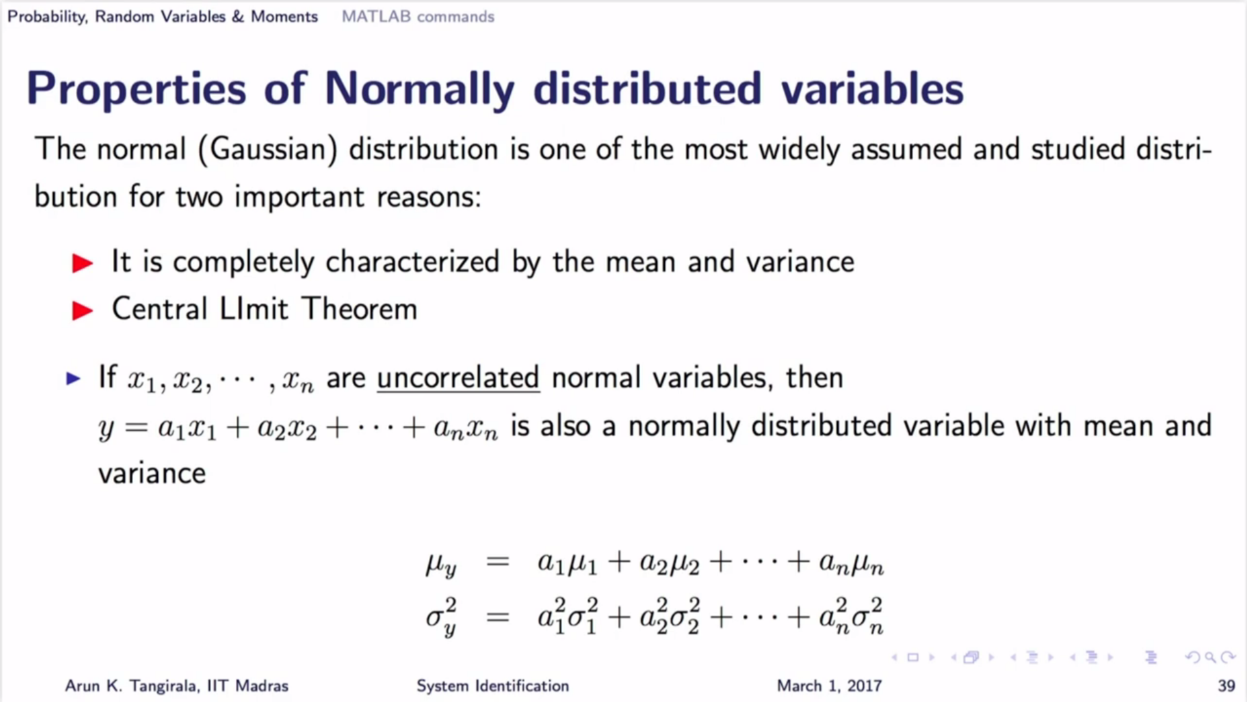
- **Affine transformation:**

$$Y = \alpha X + \beta,\ \alpha \in \mathcal{R} \Longrightarrow \mu_Y = \alpha\mu_X + \beta \tag{10}$$

$$\sigma_Y^2 = \alpha^2\sigma_X^2 \tag{11}$$

- Properties of non-linearly transformed RV depend on the non-linearity involved

Okay, so over often, we scale random variables, for example, we may do a change of units and so on just to show you what happens to their means and variances when you shift x by a constant, the expected value also shifts. The variance remains unperturbed, because variance is a central moment. The center has shifted but the spread may not change, and that’s a same story with this so-called affine transformation. Affine transformation is a big more than linear. So you have y = αX + β, many a times you do this kind of transformation of random variables. When you do that, the mean is altered by both α and β, but the variance is only influenced by α. So these are simple formulae to remember.



Probability, Random Variables & Moments MATLAB commands

Properties of Normally distributed variables

The normal (Gaussian) distribution is one of the most widely assumed and studied distribution for two important reasons:

- It is completely characterized by the mean and variance
- Central LImit Theorem

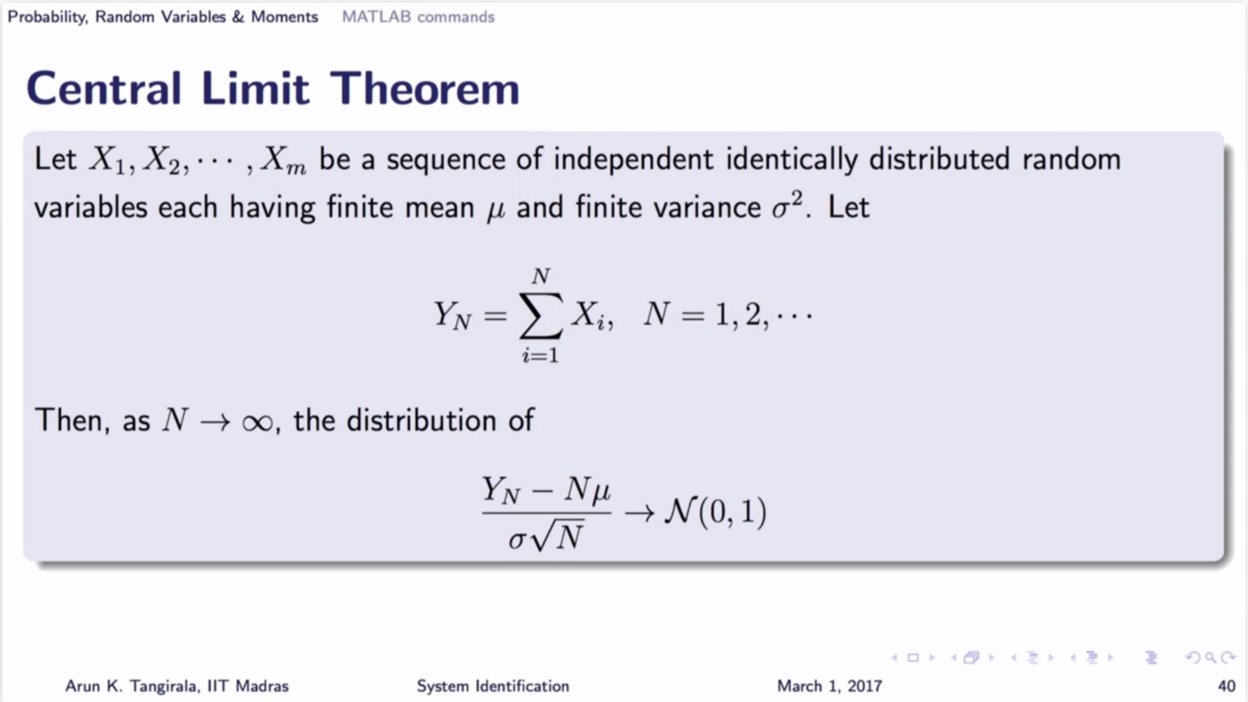
- If $x_1, x_2, \cdots, x_n$ are uncorrelated normal variables, then $y = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ is also a normally distributed variable with mean and variance

$$\begin{aligned}\mu_y &= a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \\ \sigma_y^2 &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2\end{aligned}$$

Arun K. Tangirala, IIT Madras System Identification March 1, 2017 39

When it comes to Gaussian distributed random variables, there are some special properties. One of the things that you should remember is when I mix linearly mix Gaussian distributed random variables, I will always get a Gaussian distributed random variable. It need not be true of any other distribution. Furthermore, if these random variables. In fact, there’s a syntax error there, notational error. These x that you see here should be uppercase, I’ll fix that. So if I take n uncorrelated random variable, what is uncorrelated, we’ll discuss soon, either very briefly today or on Friday. When I am looking at n uncorrelated random variables, I am going to linearly mix them. When am I going to run into this linear mixing? When I am taking for example simple averaging of observations in time. Every observation is a random variable. Remember, that holds.

So in such situations, and I am linearly mixing n random variables that are uncorrelated, then the mean expression is given here, the mean of the resulting mixed variable is given. That has got nothing to do with uncorrelated, whether it’s uncorrelated or not, the mean expression is the same. It is the variance expression that you see at the bottom that makes use of this uncorrelated property. What uncorrelated means is that these random variables do not linearly influence each other, that’s all. That is as simple as you can remember. The moment we say two random variables are uncorrelated, what it means is that those two random variables do not share a linear relation. Maybe they are neither the sibling nor the first cousin, maybe they are the second cousin or third cousin, there maybe some non-linear dependencies, but definitely there is no linear relation. That is what uncorrelated means. So these are some results that we’ll use, so there’s nothing great to discuss about this here.



Probability, Random Variables & Moments   MATLAB commands

Central Limit Theorem

Let $X_1, X_2, \cdots, X_m$ be a sequence of independent identically distributed random variables each having finite mean μ and finite variance σ^2. Let

$$Y_N = \sum_{i=1}^{N} X_i, \quad N = 1, 2, \cdots$$

Then, as $N \to \infty$, the distribution of

$$\frac{Y_N - N\mu}{\sigma\sqrt{N}} \to \mathcal{N}(0,1)$$

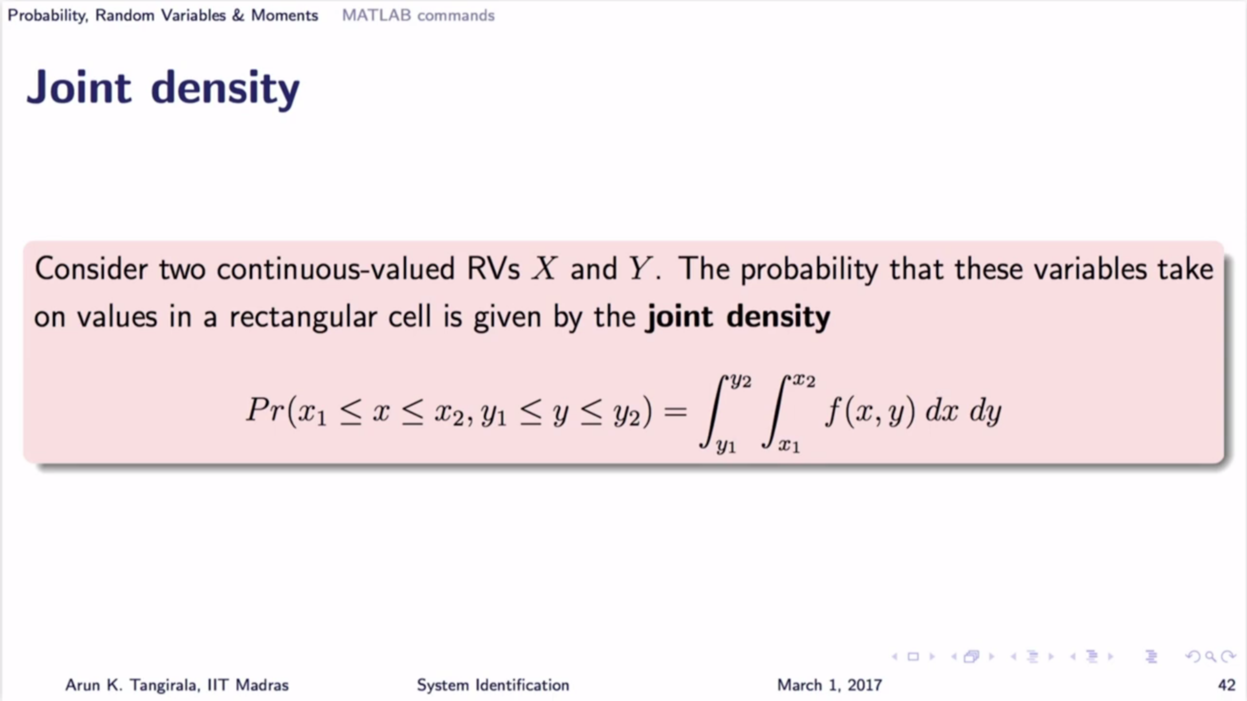
Arun K. Tangirala, IIT Madras   System Identification   March 1, 2017   40

Then you have the central limit theorem, which we use quite often in estimation theory, which basically says, when I mix linearly add up a bunch of identical and independent, earlier we used the term uncorrelated, now we are using the term independent. Independence is a much stronger requirement. When I say two random variables are independence, what it means is that, there’s absolutely no dependence at all. It’s like these random variables have no blood relation, even though the aunt, uncle, grandfather, niece and so on, there’s nothing, there’s no non-linear function that relates them. When I take bunch of such random variables that are not necessarily Gaussian distributed, all it’s saying is they should all fall out of the same distribution, and I linearly mix them up. The more I mix, and as I mix more and more of them, the resulting random variable, always remember, whenever I perform I operation on the random variable, mathematical operation, the child that has born out of that operation is also random variable, has a DNA or randomness.

So why here, which is the outcome of linearly mixing these n random variables is also a random variable? And what this theorem says is, when these variables that you are mixing are independent and identically distributed, and as you keep increasing such random variables into your summation, y will tend to have a Gaussian distribution. That is what it says, and this is one of the reasons why the Gaussian distribution is also very popular, okay. And this is a result that is used in linear estimation widely. For example, straightaway, I can say sample mean.

How do I estimate sample mean? I am going to take n observations and simply average them out. Here I am saying $\sum x_i$, but I could have $\sum x_i$, $1/\sum x_i$, it doesn’t matter, the result doesn’t change or the nature of the result. So when I am looking at sample mean, what am I doing? I am actually computing $\bar{x}$ $1/N \sum x[k]$, each observation is a random variable, 1 to n. Regardless of what distribution observations are falling out from as long as these observations are independent of each other, $\bar{x}$ will always have a Gaussian distribution as n goes to infinity. Of course, if x[k] for all sort of the Gaussian distribution, then $\bar{x}$ regardless of n will always have a Gaussian distribution. So the central limit theorem is quite useful in deriving so-called sampling distributions or distributions of estimate, which we will talk about later on.

Until now, we have discussed single random variable, but very often, we will have the need. In fact, that is what we will do, we will have to analyze more than one random variable at a time. So when I look at a random signal, now if you think of a signal, I am hoping to build a model. How am I hoping to build a model by exploiting the correlation between at least two observations. I am hoping that one observation influence another, maybe more than one observation influences, that’s okay, but even if I just take a pair of variables, I am hoping that there will be some correlation between them.



Probability, Random Variables & Moments   MATLAB commands

Joint density

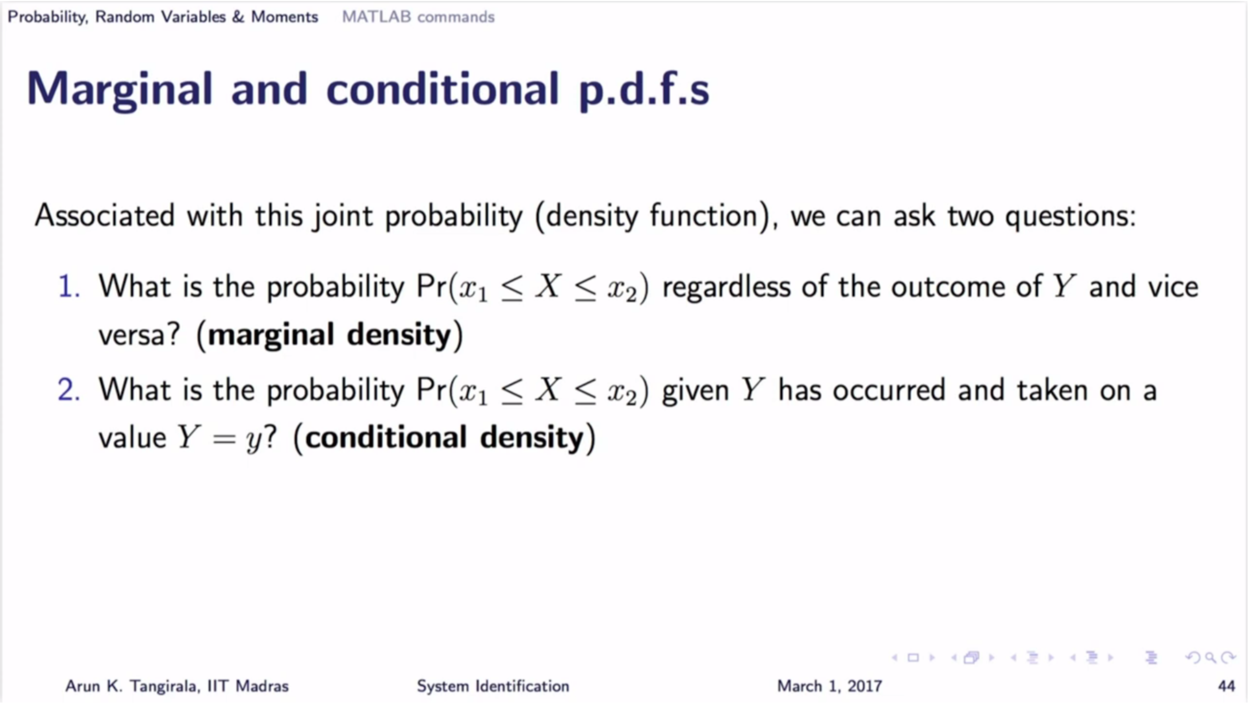
Consider two continuous-valued RVs X and Y. The probability that these variables take on values in a rectangular cell is given by the **joint density**

$$Pr(x_1 \leq x \leq x_2, y_1 \leq y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x,y)\,dx\,dy$$

Arun K. Tangirala, IIT Madras   System Identification   March 1, 2017   42

So I need a formal definition of correlation. I need a formal way of analyzing two random variables, simultaneously. So I am going to now pick some of the random variable here and I want to ask here is v[k_1] and here is V[k_2]. So I have random variable 2, random variable 1. Now I need to know how to analyze them jointly, and that is where the notion of joint density comes in, okay. So the joint density now allows me to analyze more than one random variable at a time. So now we are going into an outcome space that is more than one dimensional, and as usual joint densities have this interpretation, the standard interpretation, that is the area under the joint density will give me the probability now.

Off x and y, these are the random variables that we are looking at, taking on values within a cell. Earlier we were talking of the interval, now we are talking of a cell, and you extend this idea to three-dimension, four-dimension and so on, then you'll get different kinds of, what do you say, geometrical shapes. So once again, the joint density function has the same interpretation as your univariated density function.



Probability, Random Variables & Moments   MATLAB commands

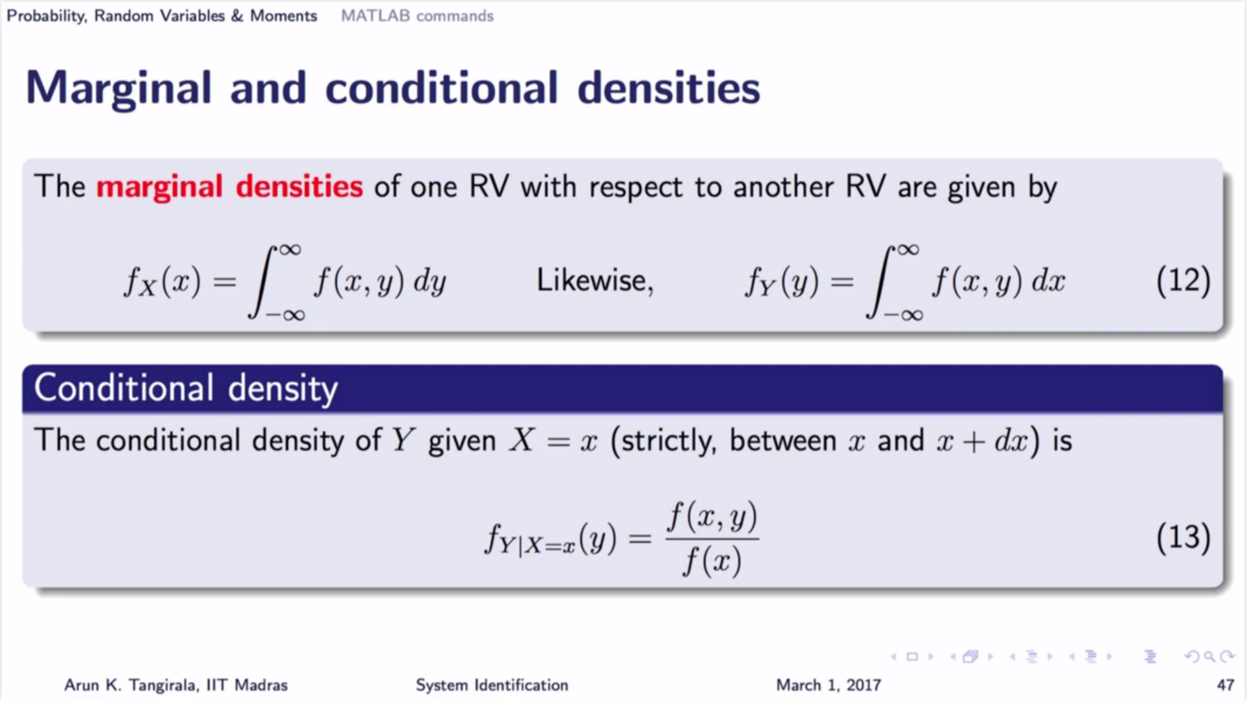
Marginal and conditional p.d.f.s

Associated with this joint probability (density function), we can ask two questions:

1. What is the probability Pr($x_1 \leq X \leq x_2$) regardless of the outcome of Y and vice versa? (**marginal density**)
2. What is the probability Pr($x_1 \leq X \leq x_2$) given Y has occurred and taken on a value $Y = y$? (**conditional density**)

Arun K. Tangirala, IIT Madras   System Identification   March 1, 2017   44

And in the context of two random variables, typically we ask two questions. One, what is the probability that x will take on some value in an interval regardless of what y takes, and the other question is what is the probability that x will take on a value within interval when y has taken on a specific value. So you can give numerous examples, a standard example is height and weight. I randomly pick an individual. Height and weight are random variables. I cannot predict. So two questions I can ask. I randomly select an individual. What is the probability that a person's weight an interval regardless of the height. The other question is, what is the probability that the individual’s weight is within the interval given that the height is this?



Probability, Random Variables & Moments MATLAB commands

Marginal and conditional densities

The **marginal densities** of one RV with respect to another RV are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad \text{Likewise,} \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx \tag{12}$$

Conditional density

The conditional density of Y given $X = x$ (strictly, between x and $x + dx$) is

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f(x)} \tag{13}$$

Arun K. Tangirala, IIT Madras System Identification March 1, 2017 47

These are two different questions. So to answer these two different questions, you need what are known as marginal PDFs and conditional PDFs. Marginal PDFS are essentially projections of this two-dimensional joint density onto a single one, because you are only concerned about one random variable. Conditional PDF has got to do with conditioning the PDF. So how is the marginal PDF defined? Well, simply the projection of the -- you can say, when you evaluate the area only along one dimension, you will get the marginal PDF in the another dimension. So you are kind of collapsing the two-dimensional joint density into a single one. So you have the respective marginal densities for x and y.

These are not necessarily the same as univariated density functions of those respective random variables, please remember that, okay. And typically, we denote marginals with the subscript x and y, whereas the conditional PDF, and this is where I will stop, the conditional PDF is defined using he base rule or the base formula as the joint density by the, in fact, strictly speaking, there should be a marginal density in the denominator here. So if I am looking at f(y) given x. Suppose I am interested in evaluating the conditional probability of y given x, then I need the conditional PDF.

For this, that is I consent the conditional PDF from the joint PDF this way, f(x, y)/f(x). There are many, many situations in which we run into conditional statements. For example, if I am looking at a game, I want to predict what is outcome before the game begins. Let's say, that is one random variable. Score is another variable, as the game starts, let's say it's a game of cricket. So if I were to look at the outcome of the game as random variable, what is the probability that the outcome is either 1 or 0 that is regardless of the score, you don't know what the score is.

Now as I keep giving you the score, I say, this is the score, this person is batting, these many overs have been balled, whatever, one of the things that I give you, pieces of information, you expect that the probabilities will keep changing, right, at least for a long time into the game, because you expected dependence between the score and the outcome, right. You said, no regardless doesn't matter, I know the team is going to lose today, then that means you are a bookie, okay. That means you have figured out who will win and who will lose, or maybe you have a very accurate prediction. You are one of those, you are either an oracle astrologer or you're a bookie, but there are many events where the outcome of y has no bearing on the outcome of x, whatsoever, in any manner, then we call them as independent events, that is independence, in which case the joint density can be simply written as a product of marginal densities. That is what is independence.

So when we meet on Friday, we will review this, we'll continue this review and move onto random signals where we'll also learn auto correlation functions.