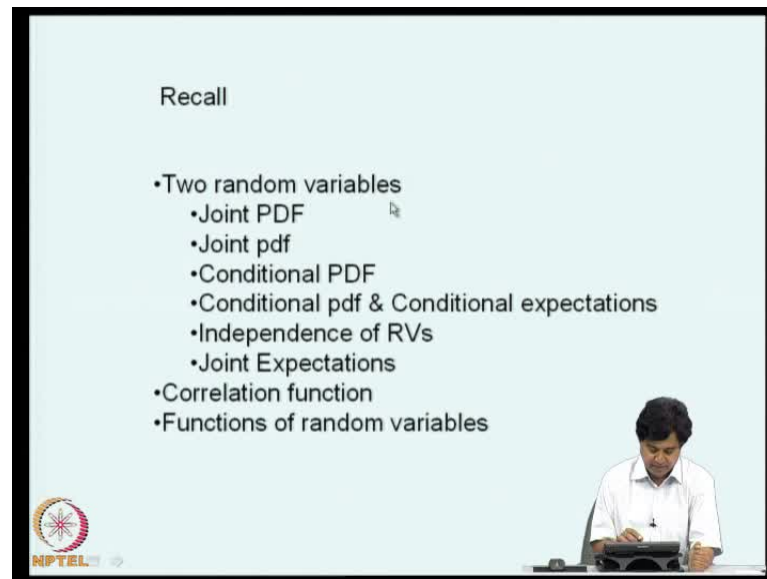


Stochastic Structural Dynamics
Prof. Dr. C. S. Manohar
Department of Civil Engineering
Indian Institute of Science, Bangalore

Lecture No. # 05
Multi-Dimensional Random Variables-2

(Refer Slide Time: 00:26)



In this lecture, we continue with discussion on two-dimensional random variables, before that, we will quickly recall what was that we discussed in the last lecture. We discuss the description of two random variables, in terms of their joint probability distribution, function joint probability density function; then, we extended the notion of conditional probability that we earlier defined for two events, to the case of two random variables, that led to the definitions of conditional probability density function, distribution function and expectations; these also helped us to define independents of two random variables.

Now, based on the properties of joint probability density function, we could also define joint expectations, joint characteristic functions, joint moment generating functions and so on and so forth. And one of the such characteristics that we discussed in the last lecture was correlation function, **we should that**, this correlation coefficient is bounded

between minus 1 and plus 1; with the bounds plus 1 and minus 1, corresponding to the case of linear relationship between X and Y, and r equal to 0 implying uncorrelatedness; and if two random variables are independent they are automatically uncorrelated. We also discuss the problem of functions of random variables, that is, given two random variables **in** and that description of their joint probability density functions, if we introduce two more random variables as functions of the original two random variables, how to obtain the joint probability density function, we will continue with that.

(Refer Slide Time: 02:16)

Example

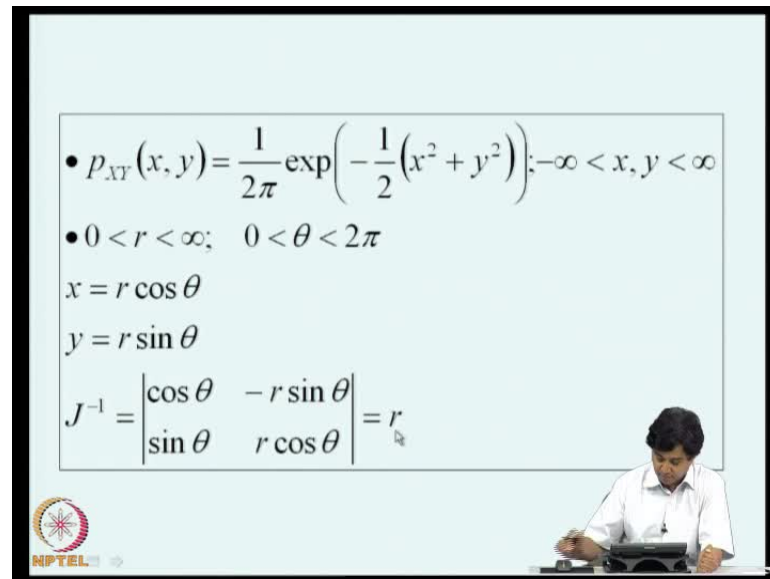
Given $R = \sqrt{X^2 + Y^2}$ & $\Theta = \tan^{-1}\left(\frac{Y}{X}\right)$

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

Find $p_{R\Theta}(r, \theta)$.

We will now consider an example where X and Y are jointly normal, they are uncorrelated, therefore they are independent; they have 0 mean and unit standard deviation. And we introduce a transformation r is square root of X square plus Y square and theta is tan inverse Y by X. So, the problem on hand is to find the joint probability density function of this random variable r and theta.

(Refer Slide Time: 02:59)



• $p_{XY}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right); -\infty < x, y < \infty$

• $0 < r < \infty; 0 < \theta < 2\pi$

$x = r \cos \theta$

$y = r \sin \theta$

$J^{-1} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r$



NPTEL

This transformation is a came to the polar coordinate transformation from Cartesian to polar coordinates, x is $r \cos \theta$ and y is $r \sin \theta$. So, given that, x and y are jointly Gaussian, and they have 0 mean and unit standard deviation, the joint probability density function is as for this equation, and the variable capital r and θ take values in 0 to infinity and 0 to 2 pi.

Now, in this example, it turns out that it is easier to determine 1 by J, then J itself that can be demonstrated, by considering this relation, x is $r \cos \theta$ and y is $r \sin \theta$; so, J will be dou x by dou r is $\cos \theta$, and dou x by dou θ is this, dou y by dou r is $\sin \theta$, dou y by dou θ is $r \cos \theta$, and this leads to the value of 1 by J as r .

(Refer Slide Time: 03:53)

Joint pdf

$$p_{R\Theta}(r, \theta) = \frac{p_{XY}(x, y)}{|J|} \Bigg|_{\substack{x=r\cos\theta \\ y=r\sin\theta}}$$
$$= \frac{r}{2\pi} \exp\left[-\frac{1}{2}(r^2 \cos^2 + r^2 \sin^2 \theta)\right]$$
$$= \frac{r}{2\pi} \exp\left[-\frac{r^2}{2}\right]; \quad 0 < r < \infty; 0 < \theta < 2\pi$$


Now, the joint density function $R\theta$ is given by $p_{XY}(x, y)$ divided by this J , evaluated it at x equal to $r \cos \theta$ and y equal to $r \sin \theta$. Now, if you make these substitutions, we get the joint density function to be r by 2π exponential minus r square by 2 , with r taking values from 0 to infinity and θ taking values from 0 to 2π .

If you are now interested in finding the marginal probability density function of r , then we have to integrate this function with respect to θ , where the limit 0 to 2π . Similarly, if you are interested in finding the marginal probability density function of θ , we have to integrate this joint density function with respect to r or the limit 0 to infinity.

(Refer Slide Time: 04:46)

Marginal pdf

$$p_R(r) = \int_0^{2\pi} p_{R\Theta}(r, \theta) d\theta$$
$$= \int_0^{2\pi} \frac{r}{2\pi} \exp\left[-\frac{r^2}{2}\right] d\theta$$
$$= r \exp\left[-\frac{r^2}{2}\right]; \quad 0 < r < \infty$$
$$p_\Theta(\theta) = \int_0^\infty p_{R\Theta}(r, \theta) dr$$
$$= \int_0^\infty \frac{r}{2\pi} \exp\left[-\frac{r^2}{2}\right] dr$$
$$= \frac{1}{2\pi}; \quad 0 < \theta < 2\pi$$

$p(r, \theta) = p_R(r)p_\Theta(\theta)$
 $\Rightarrow R \perp \Theta$

R : Rayleigh RV
 Θ : Uniformly distributed RV

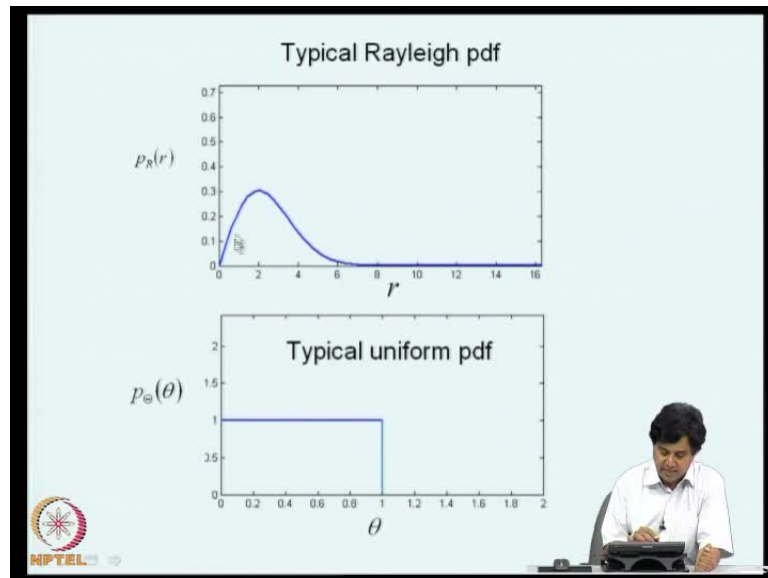
NPTEL

So, if you do that, the marginal density function of R is obtained as $\int_0^{2\pi} p_{R\Theta}(r, \theta) d\theta$, and if this integration is performed, we get the probability density function of r to be of this form. And this type of random variables, a random variable with this type of probability density function is known as a Rayleigh random variable.

Now, the probability density function of theta can be evaluated by finding this integral, that is integrate with respect to r the joint density function; if you do that, we get p theta of theta to be $\frac{1}{2\pi}$ and theta taking values from 0 to 2π , that means, the probability density function is constant over this interval and such random variables are known as uniformly distributed random variables.

Now, if you multiply the marginal density functions p_R of r given by this and p_Θ of theta given by this, we recover the joint probability density function $p_{R\Theta}$ of r and theta that would mean, the random variables R and theta are independent.

(Refer Slide Time: 05:59)



Now, a typical plot of a Rayleigh random probability density function of a Rayleigh random variable is as shown here; it has one peak, it takes values from 0 to infinity. And similarly, a typical plot of a typical uniformly distributed random variable is shown here; this is distributed uniformly between 0 and 1. So, in a way, we are talking about Cartesian to polar coordinates, so if we have Gaussian models in Cartesian space, in the polar coordinates space the amplitude that is capital R will be Rayleigh distributed in the face, will be uniformly distributed.

(Refer Slide Time: 06:45)

Example (Box - Muller Transformation)
Let X and Y be independent, and uniformly distributed random variables in 0 to 1. Define

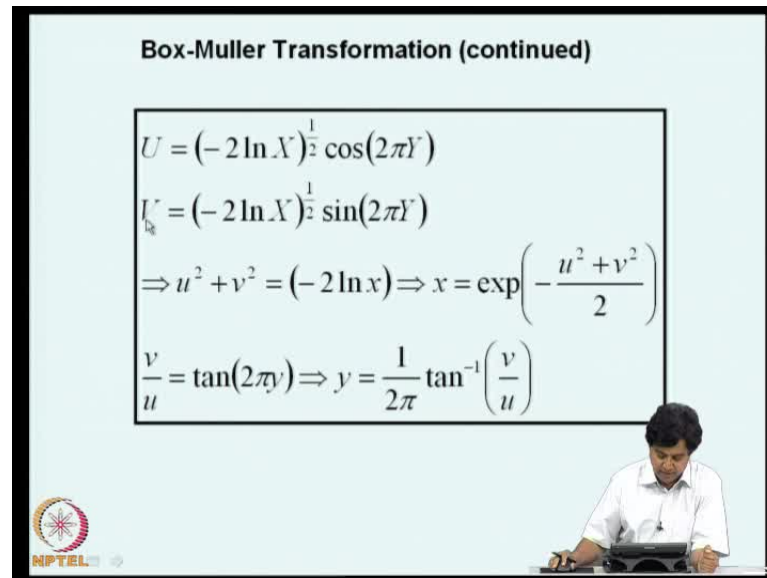
$$U = (-2 \ln X)^{\frac{1}{2}} \cos(2\pi Y)$$
$$V = (-2 \ln X)^{\frac{1}{2}} \sin(2\pi Y)$$

Determine $p_{UV}(u, v)$.

The slide features an NPTEL logo in the bottom-left corner and a small inset image of a person at a desk in the bottom-right corner.

Next, we consider one of the classical transformations, widely used in simulation work, namely the Box Muller transformation. Here, we consider two random variables X and Y which are independent and uniformly distributed in 0 to 1. We introduce two random variables capital U and V , given by these relations, and the problem on hand, is to find out the joint probability density function of U and V , \ln is the logarithm.

(Refer Slide Time: 07:23)



Box-Muller Transformation (continued)

$$U = (-2 \ln X)^{\frac{1}{2}} \cos(2\pi Y)$$
$$V = (-2 \ln X)^{\frac{1}{2}} \sin(2\pi Y)$$
$$\Rightarrow u^2 + v^2 = (-2 \ln x) \Rightarrow x = \exp\left(-\frac{u^2 + v^2}{2}\right)$$
$$\frac{v}{u} = \tan(2\pi y) \Rightarrow y = \frac{1}{2\pi} \tan^{-1}\left(\frac{v}{u}\right)$$

The slide also features the NPTEL logo in the bottom left corner and a small image of a person sitting at a desk in the bottom right corner.

So, based on the given functional relationship between U and V , and X and Y , we can derive the roots of the equation; firstly, we will write, square this, and square this, and add it, we get x in terms of U and V as exponential of minus u square plus v square by 2. Similarly, if we divide v by u , I get $\tan 2\pi y$, from which it y emerges as $\frac{1}{2\pi} \tan^{-1} \frac{v}{u}$; that means, we have now found out x and y in terms of U and V that is one of the first steps in implementing the problem of transformations.

(Refer Slide Time: 08:06)



Box-Muller Transformation (continued)

$$J^{-1} = \begin{vmatrix} \exp\left(-\frac{u^2+v^2}{2}\right) & \exp\left(-\frac{u^2+v^2}{2}\right)(-v) \\ \frac{1}{2\pi} \frac{1}{1+\left(\frac{v}{u}\right)^2} \left(-\frac{v^2}{u}\right) & \frac{1}{2\pi} \frac{1}{1+\left(\frac{v}{u}\right)^2} \left(\frac{1}{u}\right) \end{vmatrix} = -\frac{1}{2\pi} \exp\left(-\frac{u^2+v^2}{2}\right)$$

$$p_{UV}(u, v) = \frac{1}{2\pi} \exp\left(-\frac{u^2+v^2}{2}\right); -\infty < u, v < \infty$$

$$\Rightarrow U \perp V \text{ \& } U \sim N(0,1) \text{ \& } V \sim N(0,1)$$

This transformation could be used in simulation of Gaussian random numbers on a computer (more on this later).

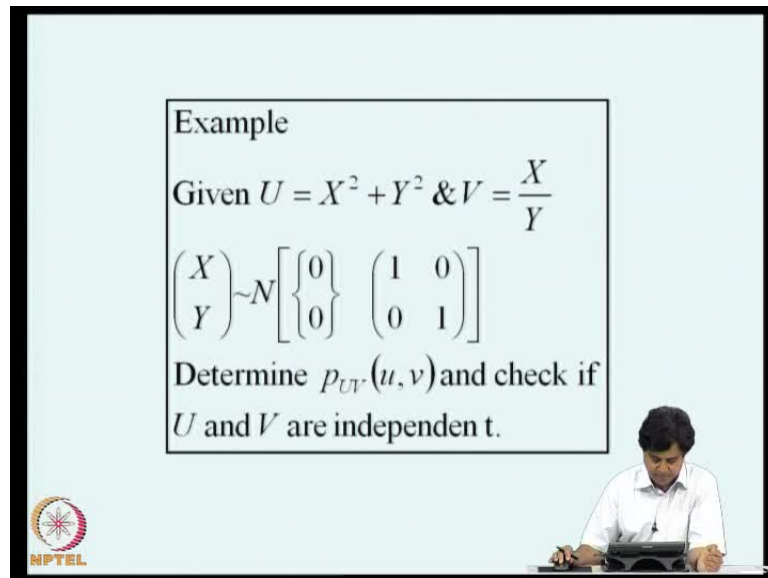



Next, in this example also, it turns out that it is easier to find 1 by J , then J , and this computation of the derivatives, that is $\frac{\partial x}{\partial u} \frac{\partial x}{\partial v} \frac{\partial y}{\partial u}$ and $\frac{\partial y}{\partial v}$, lead to this expression; from this we get 1 by J as this function.

Now, from this, we can deduce since x and y are uniformly distributed, it follows that $P_{UV}(u, v)$ is nothing but this J minus and we get this as a joint density function. Again to find marginal density of u , we can integrate over v , and to find marginal density of v , we can integrate over u , and we can verify that U and V are standard normal random variables; each has 0 mean and unit standard deviation and they are independent. From this, it also follows that U and V are independent.

This transformation has an important application in simulation of random numbers on computers, where using certain algorithms we can simulate uniformly distributed random numbers digitally on a computer and these transformations can be used to find simulate Gaussian random numbers, from the basic uniform distributed random numbers. We see more of this later, when we consider problems, we consider application of Monte Carlo simulation methods, to solve uncertainty propagation problems.

(Refer Slide Time: 09:52)



Example

Given $U = X^2 + Y^2$ & $V = \frac{X}{Y}$

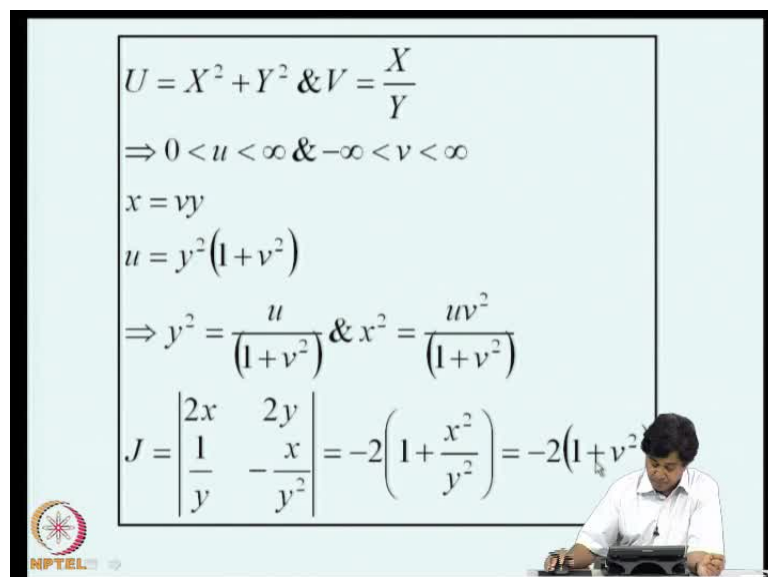
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left[\begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right]$$

Determine $p_{UV}(u, v)$ and check if U and V are independent.

The slide features the NPTEL logo in the bottom left corner and a small inset image of a man sitting at a desk with a laptop in the bottom right corner.

Another example, here I consider again X and Y to be standard normal 0 mean unit, standard deviation and uncorrelated, therefore independent also. And we introduce, now the random variables U is equal to X square plus Y square and V is equal to X by Y . The problem on hand is to determine the joint probability density function of U and V and check if U and V are independent.

(Refer Slide Time: 10:31)


$$U = X^2 + Y^2 \text{ \& } V = \frac{X}{Y}$$
$$\Rightarrow 0 < u < \infty \text{ \& } -\infty < v < \infty$$
$$x = vy$$
$$u = y^2(1 + v^2)$$
$$\Rightarrow y^2 = \frac{u}{(1 + v^2)} \text{ \& } x^2 = \frac{uv^2}{(1 + v^2)}$$
$$J = \begin{vmatrix} 2x & 2y \\ \frac{1}{y} & -\frac{x}{y^2} \end{vmatrix} = -2 \left(1 + \frac{x^2}{y^2} \right) = -2(1 + v^2)$$

The slide features the NPTEL logo in the bottom left corner and a small inset image of a man sitting at a desk with a laptop in the bottom right corner.

Now based on the given functional relationship between U and V , and X and Y , we can infer that U takes values from 0 to infinity and V takes value from minus infinity to plus

infinity. Now, we can see here, that x is given by $v y$, and u is, if you substitute that, here we get y^2 into $1 + v^2$. Therefore, Y^2 emerges as u by $1 + v^2$ and x^2 is $u v^2$ by $1 + v^2$. Now, we can find the Jacobean, from this expression we get Jacobean as this, and in terms of v , we are getting here.

(Refer Slide Time: 11:21)

$$p_{UV}(u, v) = \frac{1}{2\pi\sigma^2} \frac{1}{2(1+v^2)} \exp\left(-\frac{u}{2\sigma^2}\right);$$

with $0 < u < \infty, -\infty < v < \infty$

$$\Rightarrow p_U(u) = \int_{-\infty}^{\infty} p_{UV}(u, v) dv = \frac{1}{2\sigma^2} \exp\left(-\frac{u}{2\sigma^2}\right); 0 < u < \infty$$

$$\& p_V(v) = \int_0^{\infty} p_{UV}(u, v) du = \frac{1}{2\pi(1+v^2)}; -\infty < v < \infty$$

$$\Rightarrow p_{UV}(u, v) = p_U(u)p_V(v)$$

$$\Rightarrow U \perp V$$

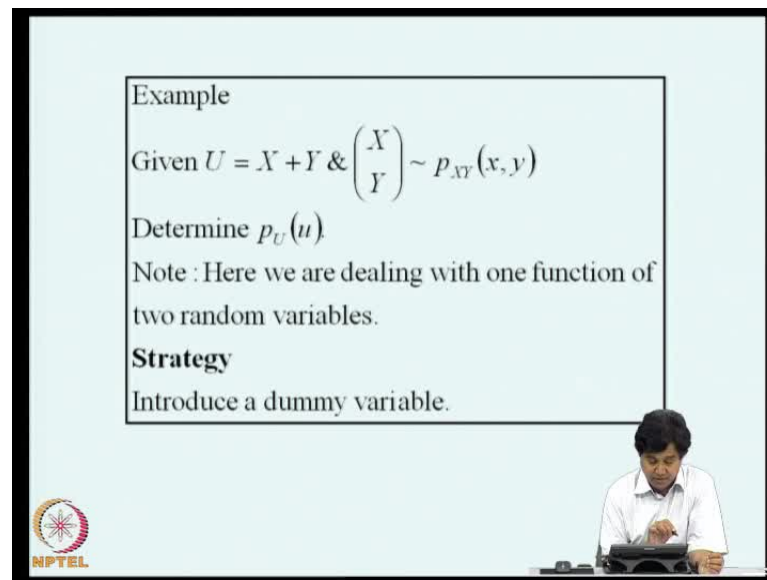
U is exponentially distributed and V is Cauchy d.

And now, if we consider the joint density function of u and v , in terms of the joint density function of X and Y , it can be shown that, this is the requisite joint density function.

Now, the marginal density function of u can be obtained by integrating this over v , and we get an exponential distribution, actually σ here is 1, that was a data, but here I have included a general case, so σ can be taken as 1 here. Similarly, if we now find the marginal density function of v , we integrate from 0 to infinity, because u takes values from 0 to infinity and we get this density function; this density function is known as Cauchy, I mean, this variable is known as a Cauchy random variable.

Again, if I now multiply p_U of u and p_V of v , we see that it turns out to be the same as the joint density function of u and v , from which we infer that U and V are independent; so, U is exponential distributed and V is Cauchy distributed.

(Refer Slide Time: 12:38)



Example

Given $U = X + Y$ & $\begin{pmatrix} X \\ Y \end{pmatrix} \sim p_{XY}(x, y)$

Determine $p_U(u)$.

Note : Here we are dealing with one function of two random variables.

Strategy
Introduce a dummy variable.

NPTEL

Next, we consider another example, where there is a slight variation in the way the problem is posed, here we consider two random variables X and Y , with a given probability density function $p_{XY}(x, y)$, and we introduce the transformation U is equal to X plus Y , and the problem on hand is to determine the probability density function of u ; this does not fit into the type of problems, that we have been discussing so far, because here we are dealing with one function of two random variables; so, the method that we have outlined just now, is not immediately applicable. Now, the strategy that we adapt is we introduce a dummy variable, that means, we will introduce another function of X and Y , and consider the joint density function of the given random variable, and this newly introduced dummy variable.



(Refer Slide Time: 13:42)

$$U = X + Y$$
$$V = Y$$
$$J = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$$
$$p_{UV}(u, v) = p_{XY}(x, y) \Big|_{\substack{x=u-v \\ y=v}}$$
$$p_U(u) = \int_{-\infty}^{\infty} p_{XY}(u - v, v) dv$$

Note : If $X \perp Y \Rightarrow$

$$p_U(u) = \int_{-\infty}^{\infty} p_X(u - v) p_Y(v) dv$$

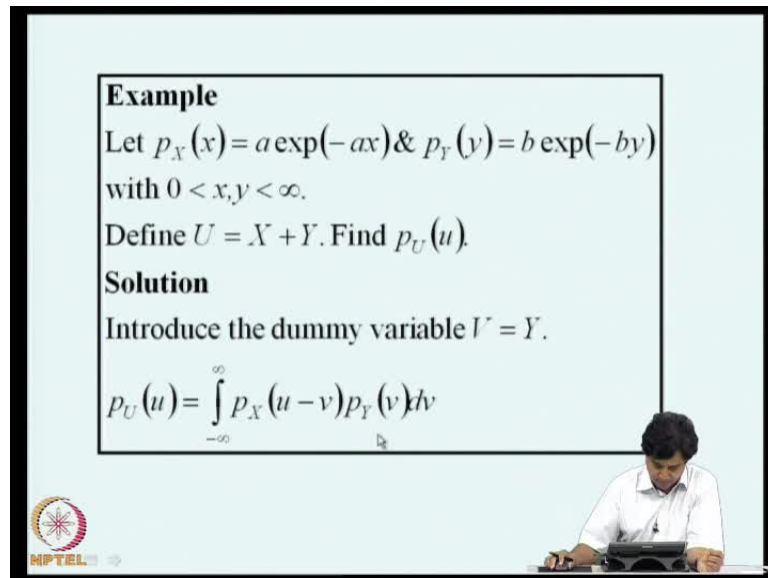
Note : The above integral represents the convolution operation.



So, this is a given function U is equal to X plus Y , now I introduce a dummy random variable V is equal to Y ; so, this J for this is 1 and $p_{UV}(u, v)$ is actually $p_{XY}(x, y)$, $y = x$ evaluated at u minus v and y evaluated at v .

Now, once I determine the joint probability density function, I can now find out the marginal density function of U , which is indeed the function that we are looking for this is minus infinity to infinity $p_{XY}(u - v, v) dv$, this is the answer to the question that we originally post. Now, additionally if you assume now, that X and Y are independent, this joint density function can be expressed as product of the marginal. So, this integral can now be written as $p_U(u) = p_X(u - v) p_Y(v) dv$; this kind of integrals are known as convolution integrals and they can easily be solved using transform techniques like Laplace transforms. This is also will be seeing, when we encounter vibration problems will be seeing that, this type of integrals appear there also.

(Refer Slide Time: 15:05)



Example
Let $p_X(x) = a \exp(-ax)$ & $p_Y(y) = b \exp(-by)$
with $0 < x, y < \infty$.
Define $U = X + Y$. Find $p_U(u)$.

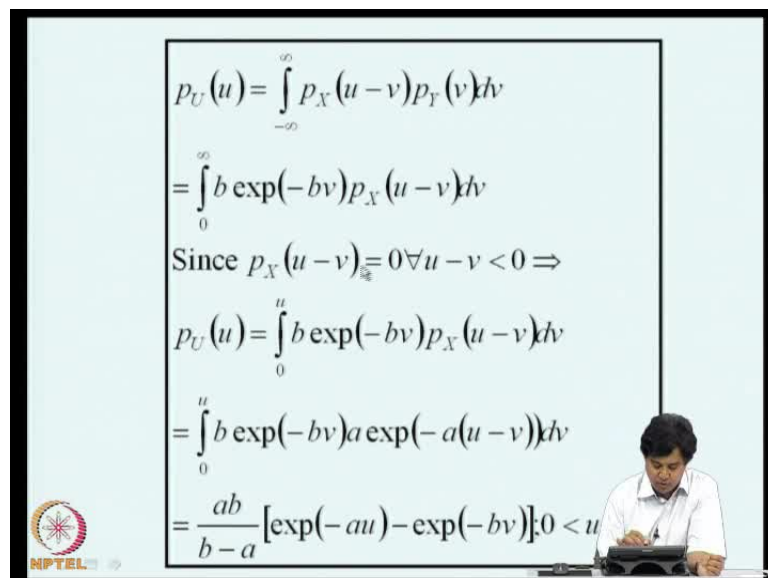
Solution
Introduce the dummy variable $V = Y$.

$$p_U(u) = \int_{-\infty}^{\infty} p_X(u-v)p_Y(v)dv$$

The slide also features the NPTEL logo in the bottom left and a small image of a person at a desk in the bottom right.

As an example of this particular case of summing two random variables, we now consider two random variables X and Y which are exponentially distributed, that means, $p_X(x)$ is $a e^{-ax}$ and $p_Y(y)$ is $b e^{-by}$ with x and y taking values from 0 to infinity. Now, we define the function U is X plus Y . So, the problem on hand is to find the probability density function of U . Again, we introduce the dummy variable V equal to Y and this is the expression that we got just now.

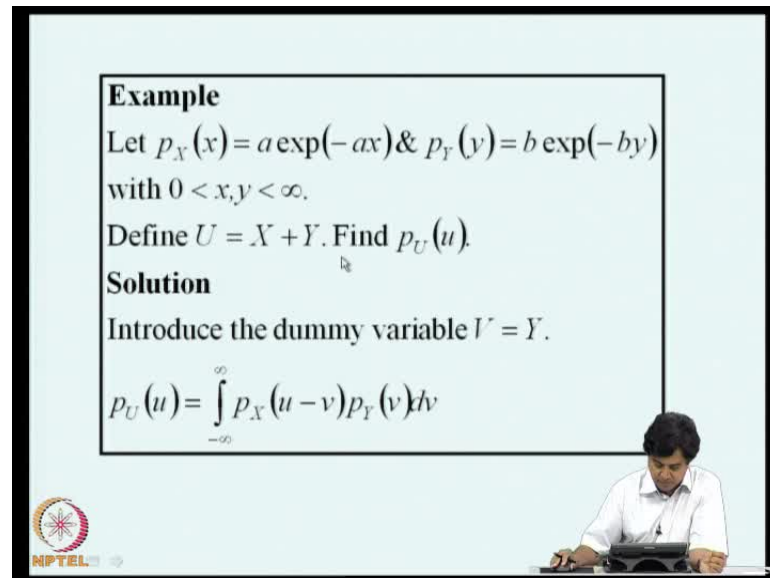
(Refer Slide Time: 06:01)


$$\begin{aligned} p_U(u) &= \int_{-\infty}^{\infty} p_X(u-v)p_Y(v)dv \\ &= \int_0^{\infty} b \exp(-bv)p_X(u-v)dv \\ \text{Since } p_X(u-v) &= 0 \forall u-v < 0 \Rightarrow \\ p_U(u) &= \int_0^u b \exp(-bv)p_X(u-v)dv \\ &= \int_0^u b \exp(-bv)a \exp(-a(u-v))dv \\ &= \frac{ab}{b-a} [\exp(-au) - \exp(-bu)]; 0 < u \end{aligned}$$

The slide also features the NPTEL logo in the bottom left and a small image of a person at a desk in the bottom right.

Now, p_Y of v is given by $b \exp(-bv)$ into p_X of $u - v$ dv .
 Now, we have to fix this limit of upper limit, bearing in mind the property of this density function. Since p_X of u is an exponential distribution, that is p_X of x is exponential minus a x for x between 0 to infinity, it implies that p_X of x is 0 for negative values of x .

(Refer Slide Time: 16:19)



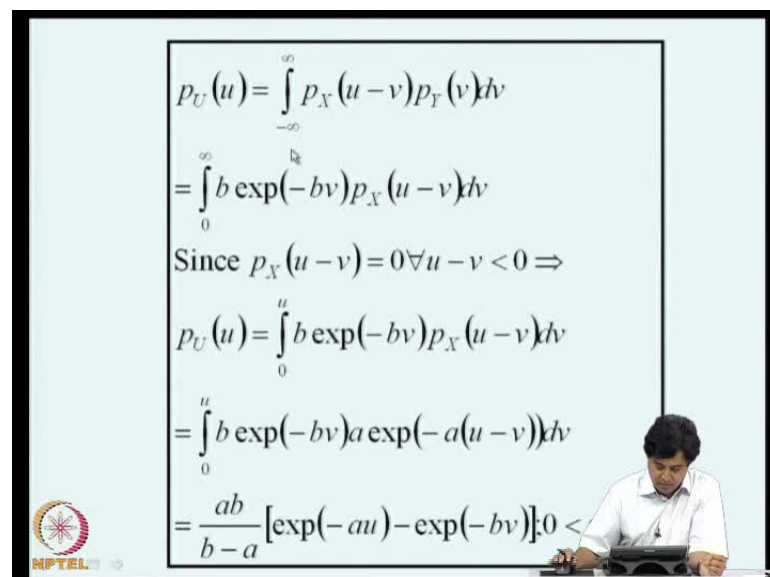
Example
 Let $p_X(x) = a \exp(-ax)$ & $p_Y(y) = b \exp(-by)$
 with $0 < x, y < \infty$.
 Define $U = X + Y$. Find $p_U(u)$.

Solution
 Introduce the dummy variable $V = Y$.

$$p_U(u) = \int_{-\infty}^{\infty} p_X(u-v)p_Y(v)dv$$

The slide also features the NPTEL logo and a small image of a person at a desk in the bottom right corner.

(Refer Slide Time: 16:33)



$$p_U(u) = \int_{-\infty}^{\infty} p_X(u-v)p_Y(v)dv$$

$$= \int_0^u b \exp(-bv)p_X(u-v)dv$$

Since $p_X(u-v) = 0 \forall u-v < 0 \Rightarrow$

$$p_U(u) = \int_0^u b \exp(-bv)p_X(u-v)dv$$

$$= \int_0^u b \exp(-bv)a \exp(-a(u-v))dv$$

$$= \frac{ab}{b-a} [\exp(-au) - \exp(-bu)] ; 0 < u < \infty$$

The slide also features the NPTEL logo and a small image of a person at a desk in the bottom right corner.

So, consequently p_X of $u - v$ will be 0 , whenever $u - v$ is less than 0 . So, this implies that in the upper limit here, I should write u instead of infinity, because this

function is 0 for negative values of u minus v . So, with that modification, now if we proceed and evaluate this integral, this is a straight forward exercise, we get the marginal density function of u to be given by this expression.

(Refer Slide Time: 17:10)



So, I have shown a simple plot of the marginal density of X and Y , these are exponential density functions and this is the density function of U . So, summation of random variable, if you are dealing with summation of random variables, you have to use convolution integrals, to evaluate the resulting probability density functions.

(Refer Slide Time: 17:42)

More than one random variables

Let $\tilde{X} = \{X_i\}_{i=1}^n$ be a set of random variables.

Definitions

n-th order Joint PDF of \tilde{X}

$$P_{\tilde{X}}(\tilde{x}) = P\left(\bigcap_{i=1}^n (X_i \leq x_i)\right) \text{ where } \tilde{x} = \{x_i\}_{i=1}^n.$$

n-th order Joint pdf of \tilde{X}

$$p_{\tilde{X}}(\tilde{x}) = \frac{\partial^n P_{\tilde{X}}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

Expectation of $g(\tilde{X})$

$$\langle g(\tilde{X}) \rangle = \int_{-\infty}^{\infty} g(\tilde{x}) p_{\tilde{X}}(\tilde{x}) d\tilde{x}$$

Note : The above is an *n*-fold integral.

Now, so far we have discussed the problem of describing two random variables; now, this can easily be generalized to characterize more than two random variables. Suppose, if you have a now a set of random variables $x_1, x_2, x_3, \dots, x_n$, now we introduce the definitions here, on what are known as n-th order joint probability distribution function of X . This is defined as probability of intersection of the events $X_i \leq x_i$, this function is known as the n-th order joint probability distribution function; associated with that, we can also define in n-th order joint probability density function. This is the n-th derivative of this probability distribution function with respect to the states $x_1, x_2, x_3, \dots, x_n$.

Again we can generalize the notion of expectation for n-dimensional random variables, and if we now consider expected value of g of X , where g of X is a non-linear function, in general of the random variables $x_1, x_2, x_3, \dots, x_n$; this is given by the n fold integral, this an n fold integral $\int \dots \int g(x) p(x) dx$. Here again we can give specific meanings to g of X and obtain requested characteristic of this vector of random variables given, again n-th order joint density function, we can get lower order joint density functions by integrating the n-th order density function with respect to the pertinent states. So, a set of n random variables is completely specified in terms of the n-th order joint probability distribution function or n-th order joint probability density function. Again, we can define moment generating functions characteristic functions in n-dimensions that is a reasonably straight forward exercise.

(Refer Slide Time: 20:23)



Multi - dimensional Gaussian random variable

Let $X = \{X_1 \ X_2 \ \dots \ X_n\}^t$ be a vector of random variables.
 Let $m_i = \langle X_i \rangle$ such that the mean vector is given by
 $m = \{m_1 \ m_2 \ \dots \ m_n\}^t$.

The elements of covariance matrix C is given by
 $C_{ij} = \langle (X_i - m_i)(X_j - m_j) \rangle$; clearly $C_{ij} = C_{ji}$ and hence $C^t = C$.
 Assume that C^{-1} exists.

Let $x = \{x_1 \ x_2 \ \dots \ x_n\}^t$ be a realization of $X = \{X_1 \ X_2 \ \dots \ X_n\}^t$.

Definition
 $X = \{X_1 \ X_2 \ \dots \ X_n\}^t$ is said to be Gaussian distributed if
 $P_X(x) = P_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$

$$\frac{1}{(2\pi)^{\frac{n}{2}} |C|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \{ (x - m)^t C^{-1} (x - m) \} \right]; -\infty < x_i < \infty$$



As an example of this idea, we will consider multidimensional Gaussian random variables. Let us consider a vector random variable X of dimension n , that is $X_1, X_2, X_3, \dots, X_n$, there are n random variables. Let m_i be the mean of each of these random variables. And by taking two random variables at a time, we can define the covariance of these random variables, and we define C_{ij} as the covariance between the random variable X_i and X_j . This is given by expectation $(X_i - m_i)(X_j - m_j)$ and it is clear that $C_{ij} = C_{ji}$. And therefore, if we now assemble this as a matrix C transposed would be same as C .

Now, if we assume that C^{-1} exists and if we consider x to be a realization, this vector x_1, x_2, \dots, x_n to be a realization of these random variables X_1, X_2, \dots, X_n , we now introduce a notion of a n -dimensional Gaussian probability density function. This vector random variable x_1, x_2, \dots, x_n is said to be Gaussian distributed, if it is joint probability density function, it takes this form. Notice, here n is appearing, here C is the covariance matrix, that just which we defined just now; this is determinant of C to the power of half and the mean vector, that is, this m is appearing here and these states x_1, x_2, \dots, x_n takes values from minus infinity to plus infinity.

So, the parameters in this joint probability density function are this mean vector m and the covariance matrix C . So, a n -dimensional Gaussian random variable, now can be completely described in terms of its mean vector and covariance matrix, because if you are given the mean vector and covariance matrix, I will be in a position to write down the n -th order probability density function, which constitutes the complete description of n -dimensional Gaussian random variable.

(Refer Slide Time: 23:05)

Remarks

(1) $n = 1$

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right), -\infty < x < \infty$$

(2) $n = 2 \Rightarrow X = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}, m = \begin{Bmatrix} m_1 \\ m_2 \end{Bmatrix}, x = \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}$

$$C = \begin{bmatrix} \langle (X_1 - m_1)^2 \rangle & \langle (X_1 - m_1)(X_2 - m_2) \rangle \\ \langle (X_1 - m_1)(X_2 - m_2) \rangle & \langle (X_2 - m_2)^2 \rangle \end{bmatrix}$$


$$= \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\Rightarrow |C| = (1 - \rho^2)\sigma_1^2\sigma_2^2$$

$$C^{-1} = \begin{bmatrix} \frac{1}{(1 - \rho^2)\sigma_1^2} & \frac{-\rho}{(1 - \rho^2)\sigma_1\sigma_2} \\ \frac{-\rho}{(1 - \rho^2)\sigma_1\sigma_2} & \frac{1}{(1 - \rho^2)\sigma_2^2} \end{bmatrix}$$

$$\Rightarrow p_{x_1, x_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1 - \rho^2)}} \exp\left[-\frac{1}{2(1 - \rho^2)}\left(\frac{(x_1 - m_1)^2}{\sigma_1^2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} - \frac{2\rho(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2}\right)\right]$$

$x_1, x_2 < \infty$




We could see as specific cases of this n-dimensional joint density function, what happens if n equal to 1, so as we should, we recover the first order probability density function, that means, a marginal density function of a multidimensional Gaussian random variable continues to be Gaussian.

Similarly, for n equal to 2, we consider two random variables X 1, X 2, and we can write down the covariance, can be written down in this form, and this is sigma 1 square is the variance of X 1, rho is the correlation coefficient, sigma 2 is the standard deviation of random variable X 2; so, based on this, I can evaluate the determinant of C and the inverse of C. And if I now substitute this into the expression for n-th order density function, this is the form of the two-dimensional joint probability density function of X and Y. There are five parameters, here sigma 1, sigma 2 standard deviation of the x 1 and x 2, rho is the correlation coefficient, m 1 and m 2 are the mean of x 1 and x 2 respectively, and x 1 x 2 takes values from minus infinity to plus infinity.

(Refer Slide Time: 24:29)

Given a set of correlated Gaussian random variables, how to Transform them into a set of uncorrelated Gaussian random variables?

$\{X_i\}_{i=1}^n$ = correlated Gaussian rvs.
 $\langle X_i \rangle = \mu_i; C_{ij} = \langle (X_i - \mu_i)(X_j - \mu_j) \rangle$
 $X'_i = \frac{X_i - \mu_i}{\sigma_i} \Rightarrow \langle X'_i \rangle = 0$
 $C'_{ij} = \langle X'_i X'_j \rangle = \begin{bmatrix} 1 & C'_{12} & C'_{13} \cdots & C'_{1n} \\ C'_{21} & 1 & C'_{23} \cdots & C'_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C'_{n1} & C'_{n2} & C'_{n3} \cdots & 1 \end{bmatrix}$
Transform
 $Y = T^T X' \Rightarrow \langle Y \rangle = 0; \langle Y Y^T \rangle = \langle T^T X X^T T \rangle = T^T C^T T$
 Select T such that $T^T C^T T = I$



22

We will need some properties of vector Gaussian random variables in due codes, so we could look into those issues right now. Suppose, we are given a set of correlated Gaussian random variables, the question is how to transform them into a set of uncorrelated Gaussian random variables? It turns out that, there exist linear transformations which achieve this requirement.

So, to illustrate that, we will consider the n correlated Gaussian random variables and the expected values are here written as μ_i and C_{ij} is the covariance between X_i and X_j . Now, associated with each X_i , now I introduce a standard normal random **vary** i , that is X_i minus μ_i divided by σ_i . Clearly, expected value of X_i prime is 0 and variance of X_i prime is 1. So, if I now write the covariance matrix for X_i primes, sees mean is 0, that are not to be same as correlation functions, so this will be a matrix which is symmetric, and along the diagonal, I get 1, because these are standard normal variables and this is a fully populated square symmetric matrix.

Now, if we now transform, introduce a transformation Y as some T transpose X prime, where T is a transformation matrix, the expected value of Y would be still 0, and you want covariance of Y , it is $Y Y^T$ which will be given by this T transpose C prime T . So, the question that we are looking for is, we are trying to answer is, what is this transformation T , so that, this matrix T transpose C prime T is diagonal; if that happens, then we would have transform the correlated Gaussian random variables into

uncorrelated Gaussian random variables. This transformation is linear and linear transformation of Gaussian random variables preserves the Gaussian property.

(Refer Slide Time: 27:04)

How to select T?

Consider the eigenvalue problem
 $[C']\{\alpha\} = \lambda\{\alpha\}$
 eigenvalues: $|C' - I\lambda| = 0$
 $\Rightarrow n$ eigenvalues $\{\lambda_i\}_{i=1}^n$
 C' is positive definite \Rightarrow
 $\lambda_i > 0 \forall i = 1, 2, \dots, n$
 eigenvectors: Φ
 $C'\phi_i = \lambda_i\phi_i$
 $C'\phi_j = \lambda_j\phi_j$
 $\phi_j^t C'\phi_i = \lambda_i\phi_j^t\phi_i$
 $\phi_i^t C'\phi_j = \lambda_j\phi_i^t\phi_j$
 $\phi_j^t C'\phi_i = \lambda_j\phi_j^t\phi_i \quad (\because C' = C'^t)$
 $\Rightarrow (\lambda_i - \lambda_j)\phi_j^t\phi_i = 0 \Rightarrow \phi_j^t\phi_i = 0 \forall i \neq j$
 $\Rightarrow \phi_i^t C'\phi_i = 0 \forall i = j$
 Select Φ such that $\Phi^t C'\Phi = I$.
 Take $T = \Phi$

Now, how to select this T, to do this way borrow the result from linear algebra, we consider the Eigenvalue problem associated with the matrix C prime, C prime alpha is lambda alpha. As we know that the Eigenvalues are given by roots of this so-called characteristic equation, C prime minus I lambda equal to 0; this originates from the argument, that we do not want inverse of C prime to exist, for example, if C inverse exist, I can pre-multiply this equation by C inverse, and C inverse C prime inverse into C prime will be identity matrix, and from that it follows alpha equal to 0 is the only solution possible.

We do not want that, that is actually C prime minus i lambda, not C prime matrix C prime minus i lambda, its inverse should not exist, if it exist alpha equal to 0 is a trivial solution **which in were not interested**; therefore, we do not want that inverse to exist and the condition for that is this determinant should be 0.

So, this leads to n Eigenvalues i equal to 1 to n, C prime is symmetric real valued and it is positive definite; so, we get n Eigenvalues and associated with each of the Eigen value, there will be an Eigen vector and the ith Eigen value and Eigen vector, the so-called ith Eigen pairs satisfy this equation; similarly, the jth Eigen pairs satisfy this equation. Actually, this Eigen vectors have a property known as orthogonality property to show

that, what we do is we pre-multiply this equation by ϕ_j^T and we get this; and we pre-multiply this equation by ϕ_i^T and I get this.

Then I transpose both sides of this equation, so $a^T b$ is $b^T a$, using that relation, I get $\phi_j^T C^T \phi_i$ is this. Actually, here we will be getting C^T , but C being symmetric, C^T is same as C . Now, subtracting these two equations, I get the relation $(\lambda_i - \lambda_j) \phi_j^T \phi_i = 0$.

Now, if $\lambda_i \neq \lambda_j$, it follows that $\phi_j^T \phi_i = 0$ for every $i \neq j$. From this, it follows $\phi_i^T C \phi_j = 0$, for $i \neq j$. Therefore, we select now this ϕ_i , such that $\phi_i^T C \phi_i$ is actually an identity matrix, we can normalize the Eigen vectors; so that, the matrix of Eigen vectors is capital Φ is the matrix of all the Eigen vectors, and the normalization of the Eigen vectors is done in such a way, that $\Phi^T C \Phi = I$. This Φ matrix is what we take to be the T matrix that we are looking for, if you do this, then we would have uncorrelated, the correlated Gaussian random variables.

(Refer Slide Time: 30:26)

Given a Gaussian random variable, how to transform it into a specified non-Gaussian random variable?

Example
X be a Rayleigh RV

$Y = g(X)$

$$p_X(x) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right], 0 \leq x < \infty$$

$$P_X(x) = \int_0^x \frac{s}{\sigma^2} \exp\left[-\frac{s^2}{2\sigma^2}\right] ds = 1 - \exp\left[-\frac{x^2}{2\sigma^2}\right]$$



Let $Z \sim N(0,1)$ and consider the transformation

$$1 - \exp\left[-\frac{X^2}{2\sigma^2}\right] = \Phi[Z]$$

with $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u \exp\left[-\frac{s^2}{2}\right] ds$

$$\Rightarrow \exp\left[-\frac{X^2}{2\sigma^2}\right] = 1 - \Phi[Z]$$

$$\Rightarrow X = \sqrt{-2\sigma^2 \log[1 - \Phi[Z]]}$$

In a similar way, we can consider certain inverse problem associated with transformation of random variables, so far what we have done is, we consider a random variable X and a transformation g of X , Y is equal to g of X , given g and specification of X we want the probability density function of Y . Here, the question I am asking is, I have Y is equal to

g of X, we have given the probability density function of X and the definition of this function g, what is probability density function of Y, this is a question we have consider till now.

We can now ask the question, given the probability density function of X and given the probability density function of Y, how to select this function g, this is a problem that we will be facing, when we have to simulate random numbers on a computer following prescribe probability density function, so this is an inverse problem here. So, we will consider a simple problem; now, given a Gaussian random variable, how to transform it into a specified non Gaussian random variable, for example, let X be a Rayleigh random variable, this is a probability density function of a Rayleigh random variable, this is a probability distribution function.

Now, if we now take a Gaussian random variable Z, 0 mean and unit standard deviation, what transformation I should apply on Z, so that, the resulting random variable will have this density function. Now, the proposition is that function should be $1 - \exp(-X^2 / 2\sigma^2)$ is ϕ of X; please notice here, that this function is similar to the probability distribution function of X and this capital ϕ denotes a probability distribution function of Z, that means, ϕ of u is $1 / \sqrt{2\pi} \exp(-u^2 / 2)$. This is actually a relationship between two random variables and it is not relationship between two probability distribution functions. Now, the question now is given the probability density function of Z, what is the probability distribution function or density function of X; so, to do that, we have to find root of X in terms of z and I get this.

(Refer Slide Time: 33:13)

Claim : X is Rayleigh

$$1 - \exp\left[-\frac{x^2}{2\sigma^2}\right] = \Phi(z)$$
$$\Rightarrow \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right] \frac{dx}{dz} = \phi(z)$$
$$\Rightarrow \frac{dx}{dz} = \frac{\phi(z)}{\frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right]}$$
$$\Rightarrow p_X(x) = \frac{\phi(z)}{\left|\frac{dx}{dz}\right|}$$
$$= \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right]; 0 \leq x < \infty; \text{QED}$$

$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
 $-\infty < z < \infty$

NPTEL

Now, we claim that if you follow this transformation X is going to be Rayleigh distributed; this is the transformation that I am looking for.

Now, we can now verify quickly, you have to find the derivatives, we have found the root, now we have to find the derivatives; now, if you differentiate these with respect to Z , we get, first you differentiate this with respect to X , and then dx by dz must be equal to the derivative of probability distribution function with respect to Z , which I denote as a lower case phi, and this is the probability density function of the unit standard normal random variable; that means, phi of Z is $\frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}$, this is a normal probability density function. So, if you use that relation, it turns out that the probability density function of X is in the Rayleigh, **this, what we are trying to demonstrate.**

(Refer Slide Time: 34:19)

In general: Let X be a RV with PDF $P_X(x)$.

Consider

$$P_X(X) = \Phi(Z) \Rightarrow X = P_X^{-1}[\Phi(Z)]$$
$$\Rightarrow P_X(x) \frac{dx}{dz} = \phi(z)$$
$$\Rightarrow P_X(x) = \frac{\phi(z)}{\left| \frac{dx}{dz} \right|} = P_X(x).$$

So, generalization of this is let X be a random variable with a probability distribution function P_X of x . Now, what transformation I should apply on Z , which is normal random variable, with 0 mean and unit standard deviation; so that, the resulting random variable X will have this requisite target probability distribution function. The answer to this question is a transformation, that we have to consider is P_X of X is ϕ of Z . Notice again, that this is not a relationship between two probability distribution functions, it is relationship between two random variables. Z is standard normal; X is a target random variable with distribution function given by P_X of x as given here.

Now, you differentiate this with respect to x , **and this is a**, this is a density function, normal density function, and use it in the formula for probability density function of the transform random variable, we recover the target probability density function. So, on a computer, it is easy to generate uniformly distributed random numbers, and from that, we can always simulate other forms of random numbers, using this kind of rules of transformation of random variables.

(Refer Slide Time: 35:46)

Rosenblatt transformation
Case of two random variables

Let X_1 & X_2 be two non-Gaussian RVs.
JPDF: $P_{12}(x_1, x_2)$
jpdf: $p_{12}(x_1, x_2)$
MPDF: $P_1(x_1)$ & $P_2(x_2)$
mpdf: $p_1(x_1)$ & $p_2(x_2)$
Let $U_1 \rightarrow N(0,1)$ & $U_2 \rightarrow N(0,1)$ with $U_1 \perp U_2$
Define
 $P_1(X_1) = \Phi(U_1)$
 $P_2(X_2 | X_1) = \Phi(U_2)$

How about this problem for two-dimensional random variables; now, let X_1 and X_2 be two non-Gaussian random variables, with a joint probability distribution function $P_{12}(x_1, x_2)$ and joint probability density function lower case $p_{12}(x_1, x_2)$. The marginal probability distribution functions are here, the marginal density distribution functions are here and these are the density functions.

Now, we introduce two random variables U_1 and U_2 , which has standard normal and independent. According to Rosenblatt transformation, if you now make the transformation P_1 of X_1 is $\Phi(U_1)$, and P_2 of X_2 condition on X_1 is $\Phi(U_2)$, and using this prescription for given values of U_1 and U_2 , if you compute X_1 and X_2 . The joint density function between X_1 and X_2 will be conforming to the specified joint distribution function that you are looking for.

(Refer Slide Time: 36:54)

$$p_1(x_1) \frac{dx_1}{du_1} = \phi(u_1); \frac{dx_1}{du_2} = 0$$

$$p_2(x_2 | x_1) \frac{dx_2}{du_2} = \phi(u_2)$$

$$\Rightarrow p_{12}(x_1, x_2) = \frac{\phi(u_1, u_2)}{|J|} @ u_1$$

$$= \Phi^{-1}\{P_1(x_1)\} \& u_2 = \Phi^{-1}\{P_2(x_2 | x_1)\}$$

$$\Rightarrow p_{12}(x_1, x_2) =$$

$$\frac{\phi(u_1)\phi(u_2)}{\phi(u_1)\phi(u_2)} p_2(x_2 | x_1)p_1(x_1) = p_{12}(x_1, x_2)$$

NPTEL

(Refer Slide Time: 37:00)

Models for limits of sums, products and extremes

Central limit theorem
 Let $\{X_i\}_{i=1}^n$ be an iid sequence (iid=independent, identically distributed) of random variables.
 Let $\langle X_i \rangle = \mu$ & $\langle (X_i - \mu)^2 \rangle = \sigma^2$.
 According to the central limit theorem

$$\lim_{n \rightarrow \infty} P \left[\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a \exp\left(-\frac{x^2}{2}\right) dx = \Phi(a)$$

Central limit theorem for products
 Let $\{X_i\}_{i=1}^n$ be an iid sequence (iid=independent, identically distributed) of random variables that take only positive values.
 Define

$$Y = X_1 X_2 \dots X_n$$

$$\lim_{n \rightarrow \infty} p_Y(y) \rightarrow \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (\ln y - \eta)^2\right) U(y)$$
 with $\eta = \sum_{i=1}^n \langle \ln X_i \rangle$ & $\sigma^2 = \sum_{i=1}^n \text{Var}(\ln X_i)$

NPTEL

The proof of this I leave as an exercise is outlined here. Now, we move on to an important set of results, in the context of sequence of random variables; these are questions on limits of sums, products and extremes of sequence of random variables. So, we will begin with discussion on models for sums. There is a theorem known as central limit theorem, which helps us to characterize the probability density function of sums of random variables, as a number of random variables being sum become large.

So, according to this theorem, if $X_1, X_2, X_3, \dots, X_n$ form a sequence of independent identically distributed random variables, this phrase independent identically distributed, we call it as IID sequence; if, so X_i is form an IID sequence with expected value of X_i being μ and variance being σ^2 ; this μ and σ^2 are common to each one of this $X_1, X_2, X_3, \dots, X_n$ because there identically distributed and they are independent.

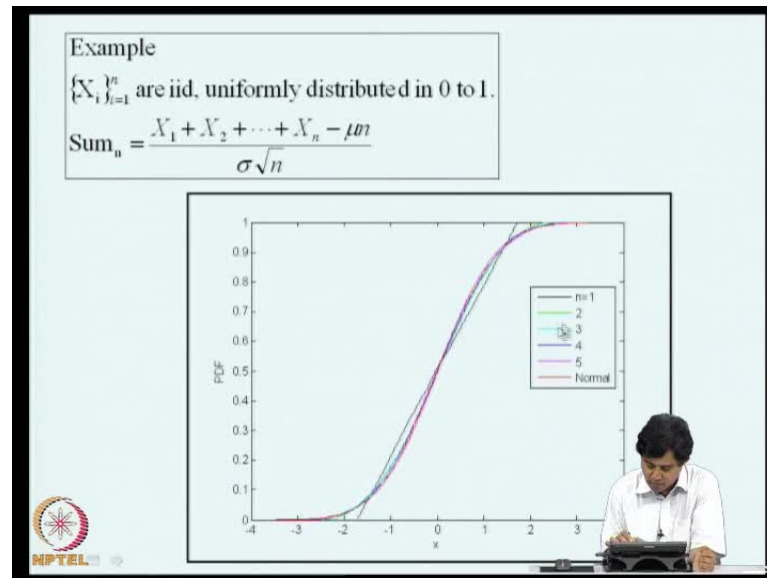
According to the central limit theorem, if you now consider the random variable X_1 plus X_2 plus, so up to X_n minus $n\mu$, normalize with respect to $\sigma\sqrt{n}$; this is a new random variable that we introduce, and if we ask the question, what is probability of this random variable being less than equal to a of this event as n becomes large. According to central limit theorem, the probability distribution function, this is nothing but the probability distribution function of this ratio evaluated at a approaches the probability distribution function of a Gaussian random variable, with 0 mean and unit standard deviation.

What this, the secret tells us is, if there is a random phenomenon which is a consequence of addition of several causes, in such a way that none of the individual causes dominate over the other. The resulting phenomena can be modeled as being Gaussian in nature; this is a loose statement, loose translation of what this theorem tells.

Now, just as we have a limit theorem for sums, we can also have a theorem for products under certain situations; again, let us consider the sequence of random variables $X_1, X_2, X_3, \dots, X_n$, again there IID sequences, but these random variables take only positive values, that mean, probability of any of these X_i being less than or equal to 0, is 0.

Now, we define the product $Y = X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n$ according to this theorem for products, the probability density function of this product as n becomes large, approaches the probability density function of a log normal random variables; this U of y is the U side step function, which is 0 for negative arguments of y and it is 1 for y being positive; this is not surprising, because if you were to take logarithm of Y , we will get \log of X_1 plus \log of X_2 plus \log of X_n , which is again a summation mechanism coming into play. So, on the logarithms, if you apply the central limit theorem, we show that, that logarithms are normally distributed, and consequently, the density function will have the so-called log normal density functional form.

(Refer Slide Time: 41:22)

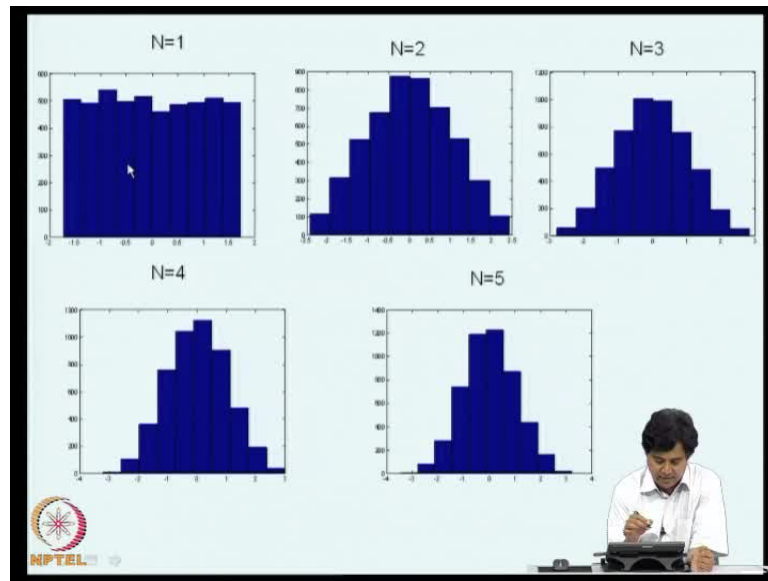


To illustrate this, we can consider a simple example, where I consider this, I call this as sum as sum subscript n of X 1 plus X 2 plus X n minus mu n by sigma square root n, I will take that this exercise are uniformly distributed between 0 to 1 and their independent and identically distributed. If you were to perform this transformation, and plot the probability distribution function of this sum as a function of n, which is shown here; the rate line here is the probability distribution function of a Gaussian random variable.

So, black one is the uniform distribution straight line, with n equal to 2, we see that this black line is gentling moving towards the normal probability distribution function curve; so 2, 3, 4, 5 already were pretty close to the Gaussian curve and mind you as starting with uniform distribution.

So, in this particular case, although it says this n the central limit theorem says that, the n should go to infinity, we see that even with n equal to 5, we have a reasonable match between the Gaussian probability distribution function and the probability distribution function of this sum.

(Refer Slide Time: 43:02)



So, this is also illustrated in terms of histograms, I will explain this later when we do lecture on Monte Carlo simulations, but we can you quickly see what it means. Imagine I have randomly distributed numbers, say uniformly distributed, that is N equal to 1, it look, it appears in this form; this histogram appears in this form. With N equal to 2, if I plot the histogram of the sum of the two uniform distributed random variables, it assumes a kind of triangular shape; with N equal to 3, it starts moving towards the Gaussian density function, and with N equal to 5, we are pretty close to a Gaussian density function; so, this is a kind of a visual illustration of what is central limit theorem is trying to tell us.

(Refer Slide Time: 43:52)

Extreme value distributions: distribution of maxima

Example
Let X and Y be two random variables with specified jpdf.
Define $Z = \max(X, Y)$. What is the pdf of Z ?

$$P_Z(z) = P(Z \leq z)$$
$$= P(\max\{X, Y\} \leq z)$$
$$= P(X \leq z \cap Y \leq z) = P_{XY}(z, z)$$
$$= P_X(z)P_Y(z) \text{ (if } X \perp Y)$$
$$= [P_X(z)]^2 \text{ (if } X \text{ and } Y \text{ are iid).}$$
$$p_Z(z) = 2P_X(z)p_X(z)$$

There exist an another class of, important class of limit theorems and their associated with extremes. As engineers, we are always interested in highest and the lowest, highest loads maximum wind velocity, highest magnitude of earthquake, highest value of peak ground acceleration. Similarly, when we characterized strength, we are interested in minimum, what is the lowest strength, what is the lowest stiffness, so extremes always of interest to engineers; in fact, earthquakes and cyclones and waves etcetera, are all manifestation extremes of certain environmental processes.

Now, we will consider what happens to extremes of sequence of random variables, and consider the question, what happens to the highest value as the number of random variables becomes large, and the lowest value where random variables becomes large. We begin by considering two random variables X and Y , with a specified joint probability density function. I now, define the function Z as maximum of X , Y , so we will now consider what is the probability density function of Z . This function is not differentiable, so we cannot use a rule of transformation of random variables that we discussed, we cannot find j and invert this relationship etcetera; so, we have to use certain other arguments.

Now, what is a probability distribution function of Z , is it probability of Z less than equal to z , that is, probability of maximum of X and Y being less than equal to z , this means, if highest value is less than z , both X and Y will be less than z or in another words, we are

considering the intersection of the event X less than or equal to z and Y less than equal to z. So, the probability distribution that we are looking for indeed is the joint probability density function of X and Y evaluated at z. Now, if X and Y are independent, this joint distribution function, can be expressed as product of the marginal distribution functions and I get this; now, if they are identically distributed, this gets squared. Now, the associated probability density function can be obtained by differentiating this with respect to z, I get $2 P_X(z) p_X(z)$; this is the probability density function of maximum of X, Y.

(Refer Slide Time: 46:39)

Generalization

Let $\{X_i\}_{i=1}^n$ be an iid sequence.

Define $Z = \max \{X_i\}_{i=1}^n$.

$$P_Z(z) = [P_X(z)]^n$$

$$p_Z(z) = n [P_X(z)]^{n-1} p_X(z)$$

Now, if we generalize this and consider n random variables and if we assume $X_1, X_2, X_3, \dots, X_n$ to be to form an IID sequence and we consider the maximum of this sequence and define that to be Z. The probability distribution function of z will be the nth power of the probability distribution function of X. And from this, I get the density function is derivative of this n into this, to the power of n minus 1 and then derivative of this is the density function.



(Refer Slide Time: 47:16)

Extreme value distributions: distribution of minima

Example

Let X and Y be two random variables with specified jpdf.

Define $Z = \min(X, Y)$. What is the pdf of Z ?

$$P_Z(z) = P(Z > z)$$
$$= P(\min\{X, Y\} > z)$$
$$= \{1 - P_X(z)\} \{1 - P_Y(z)\} \text{ (if } X \perp Y\text{)}$$
$$= \{1 - P_X(z)\}^2 \text{ (if } X \text{ and } Y \text{ are iid)}$$
$$1 - P_Z(z) = \{1 - P_X(z)\}^2$$
$$\Rightarrow P_Z(z) = 1 - \{1 - P_X(z)\}^2$$
$$\Rightarrow p_Z(z) = 2\{1 - P_X(z)\} p_X(z)$$




Same question now can be asked on minimum, we have consider maxima, now if I consider z as minimum of X, Y , now here, we consider probability of Z greater than equal to z , this must be 1 minus P_Z of z . This is minimum of X, Y greater than equal to z and this is nothing but 1 minus p_X of z into 1 minus p_Y of z , if X and Y are independent. And if they are identically distributed, it becomes square of that, so 1 minus p_Z of z is square of this and P_Z of z is therefore 1 minus this. Now, the density function can be obtained by differentiating this with respect to z and we get this expression.

(Refer Slide Time: 48:10)

Generalization

Let $\{X_i\}_{i=1}^n$ be an iid sequence.

Define $Z = \min\{X_i\}_{i=1}^n$.

$$P_Z(z) = 1 - [1 - P_X(z)]^n$$
$$p_Z(z) = n [1 - P_X(z)]^{n-1} p_X(z)$$


this again can be generalize for a sequence of n IID sequence of n random variables and we can get distribution function and the density function

(Refer Slide Time: 48:25)

Asymptotic extreme value distributions

Let $\{X_i\}_{i=1}^n$ be an iid sequence

Define

$$Z = \max\{X_i\}_{i=1}^n$$

$$Y = \min\{X_i\}_{i=1}^n$$

We have shown that

$$P_Z(z) = [P_X(z)]^n$$

$$P_Y(y) = 1 - [1 - P_X(y)]^n$$

Questions

what happens as $n \rightarrow \infty$?

$P_X(x)$ is not known?

n is not known?

Degeneracy

$$\lim_{n \rightarrow \infty} [P_X(z)]^n \rightarrow 1 \text{ if } P_X(z) = 1$$


$$\lim_{n \rightarrow \infty} [P_X(z)]^n \rightarrow 0 \text{ if } P_X(z) < 1$$

$$\lim_{n \rightarrow \infty} \{1 - [1 - P_X(z)]^n\} \rightarrow 0 \text{ if } P_X(z) = 0$$

$$\lim_{n \rightarrow \infty} \{1 - [1 - P_X(z)]^n\} \rightarrow 1 \text{ if } 0 < P_X(z) \leq 1$$

Stability hypothesis

$$F^n(x) = F(a_n x + b)$$

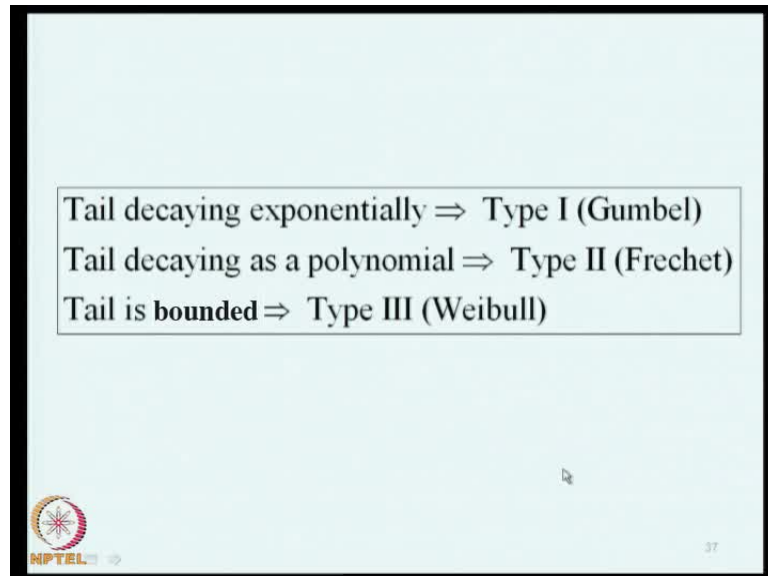

36

The question is now, what happens as n becomes large, so let us consider X_i again to be an IID sequence and I define now two random variables Z and Y, one is a maximum of X_i , another one is minimum of X_i . We already shown that P_Z of z is this and p_Y of y is, this question is what happens as n tends to infinity. More interesting questions, can we characterize Z and Y, if even, if I do not know everything about this X_1, X_2, X_3, X_n ; no matter, what are this details of X_1, X_2, X_3, X_n are, there any generic limiting forms for these extremes, and also n may not be known, you may not be knowing on what values of n, you are talking the maximum or minimum.

If you now look at these expressions as n becomes large P_X of z, takes values between 0 and 1, therefore, if you go on raising that to the power of n, as n becomes large, it goes to either 1 or 0; at the very end, it is 1 everywhere else it is 0; similarly, the minimum also would go to 0 and 1, that means, a kind of degeneracy occurs and this does not contain any information on details of these exercise, no matter what are the X_1, X_2, X_3, X_n the maximum and minimum will have this distribution function. This is a degenerate form, but it turns out that, there are under certain situations, there are non-degenerate asymptotic distributions possible for the extremes. And this based on the hypothesis

known as stability hypothesis, I will not be getting into details of that, but I will briefly mention what it is.

(Refer Slide Time: 50:23)





If you now consider the probability distribution density function of the underlying random variables, if the tails are decaying exponentially, see highest values of random variable is, you have to seek its property in the tails of the probability distribution function; the right hand tail gives a highest and low left hand tail gives a lowest. So, if the tail decays exponentially, we get Gumbol distribution, if tail decays as polynomial we get Freshet distribution, and if tail is bounded we get, what is known as Weibull distribution; so, these three are the canonical extreme value distributions.

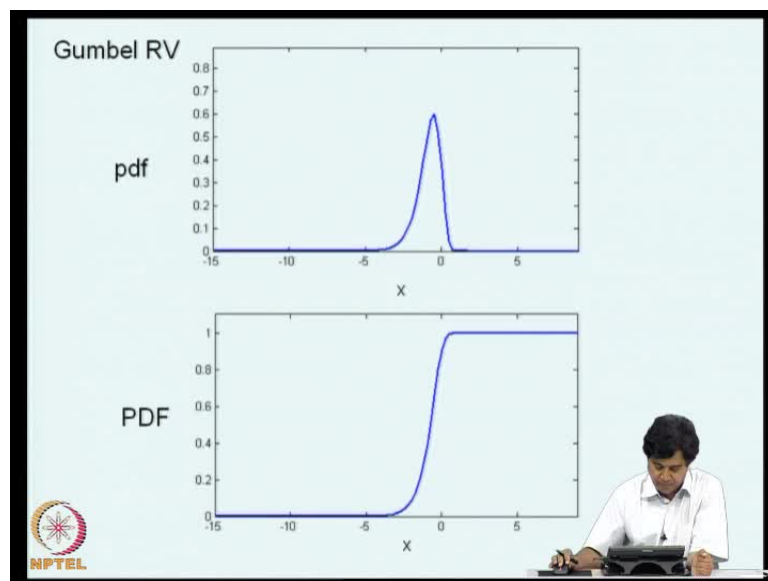
(Refer Slide Time: 51:14)

Maxima

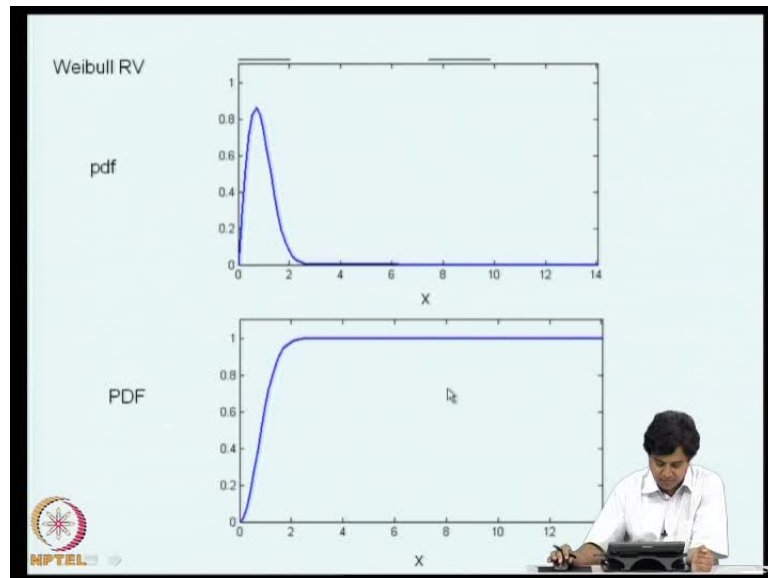
$H_{1,\gamma}(x) = \exp(-x^{-\gamma})$	if $x > 0$ otherwise
$H_{2,\gamma}(x) = 1$ $= \exp[-(-x)^\gamma]$	if $x \geq 0$ otherwise
$H_{3,0}(x) = \exp[-\exp(-x)]$	$-\infty < x < \infty$



(Refer Slide Time: 51:21)



(Refer Slide Time: 51:28)



There are some details of the probability distribution functions, of these three random variables. This is how a Gumbel random variable looks like, it is probability density function; this is a probability distribution function. This is a Weibull random variable, this is a density function; this is a distribution function.

(Refer Slide Time: 51:34)

Domains of attraction of a few commonly occurring Random variables

RV	Domain of attraction for maxima	Domain of attraction for minima
Gaussian	Gumbel	Gumbel
Exponential	Gumbel	Weibull
Cauchy	Frechet	Frechet
Rayleigh	Gumbel	Gumbel
Log-normal	Gumbel	Gumbel
Uniform	Weibull	Weibull

NPTEL



Actually, that if you consider a sequence of say Gaussian random variables, you can show that the sequence of Gaussian random variables, the highest value maxima is attracted to a Gumbel distribution, that means, the Gaussian random variables lie in the

basin of attraction of Gumbel random variables, as far as their highest is concerned; similarly, lowest also will be associated with the Gumbel model for the lowest value. Exponentials the domain of attraction is Gumbel for maxima, for minima, it is Weibull for Cauchy, it is Frechet for both maxima and minima and so on and so forth; so, these are the theorems in mathematical literature which helps us to prove this statements.

(Refer Slide Time: 52:26)

Table 1 Properties of Gumbel distributions



Quantity	Gumbel (maximum)	Gumbel (minimum)
PDF	$G(x) = \exp\left[-\exp\left\{\frac{-(x-\lambda)}{\delta}\right\}\right]$	$\bar{G}(x) = 1 - \exp\left[-\exp\left\{\frac{-(\lambda-x)}{\delta}\right\}\right]$
Mean	$\lambda + 0.57772\delta$	$\lambda - 0.57772\delta$
Median	$\lambda + 0.3665\delta$	$\lambda - 0.3665\delta$
Mode	λ	λ
Variance	$\frac{\pi^2\delta^2}{6}$	$\frac{\pi^2\delta^2}{6}$
Kurtosis coefficient	1.1396	-1.1396

(Refer Slide Time: 52:31)

Table 2 Properties of Weibull distributions



Quantity	Weibull (maximum)	Weibull (minimum)
PDF	$G(x) = \exp\left[-\left(\frac{\lambda-x}{\delta}\right)^\beta\right]$	$G(x) = 1 - \exp\left[-\left(\frac{\lambda-x}{\delta}\right)^\beta\right]$
Mean	$\lambda - \delta\Gamma\left(1 + \frac{1}{\beta}\right)$	$\lambda + \delta\Gamma\left(1 + \frac{1}{\beta}\right)$
Median	$\lambda - \delta 0.693^{\frac{1}{\beta}}$	$\lambda + \delta 0.693^{\frac{1}{\beta}}$
Mode	$\lambda - \delta\left(\frac{\beta-1}{\beta}\right)^{\frac{1}{\beta}}; \beta > 1$ $\lambda; \beta \leq 1$	$\lambda + \delta\left(\frac{\beta-1}{\beta}\right)^{\frac{1}{\beta}}; \beta > 1$ $\lambda; \beta \leq 1$
Variance	$\delta^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right]$	$\delta^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right]$

(Refer Slide Time: 52:32)

Table 3 Properties of Frechet distributions

Quantity	Weibull (maximum)	Weibull (minimum)
PDF	$G(x) = \exp\left[-\left(\frac{x}{x-\lambda}\right)^\beta\right]$	$G(x) = 1 - \exp\left[-\left(\frac{x}{x-\lambda}\right)^\beta\right]$
Mean	$\lambda + \delta \Gamma\left(1 - \frac{1}{\beta}\right); \beta > 1$	$\lambda - \delta \Gamma\left(1 - \frac{1}{\beta}\right); \beta > 1$
Median	$\lambda + \delta 0.693 \frac{1}{\beta}$	$\lambda - \delta 0.693 \frac{1}{\beta}$
Variance	$\delta^2 \left[\Gamma\left(1 - \frac{2}{\beta}\right) - \Gamma^2\left(1 - \frac{1}{\beta}\right) \right]; \beta > 2$	$\delta^2 \left[\Gamma\left(1 - \frac{2}{\beta}\right) - \Gamma^2\left(1 - \frac{1}{\beta}\right) \right]; \beta > 2$






(Refer Slide Time: 52:36)

Bayes' theorem

Let X be a random variable and A be an event.
 Let us take that A is observable and X is not.
 Before observing A we have a model for X .
 How can we update the model for X after we have observed A ?

Recall

$$\begin{aligned}
 P(A | x < X \leq x + dx) &= \frac{P(A \cap x < X \leq x + dx)}{P(x < X \leq x + dx)} \\
 &= \frac{P(x < X \leq x + dx | A) P(A)}{P(x < X \leq x + dx)} \\
 &= \frac{p_X(x | A) P(A)}{p_X(x)}
 \end{aligned}$$



So, I have tabulated the properties of this Gumbel and Weibull and Frechet distributions, this will be available in the files. The last topic that I will be discussing in the context of random variables is the application of Bayes theorem; we have seen the Bayes theorem in the context of events. Now, we will now consider the statement of this theorem in the context of random variables.

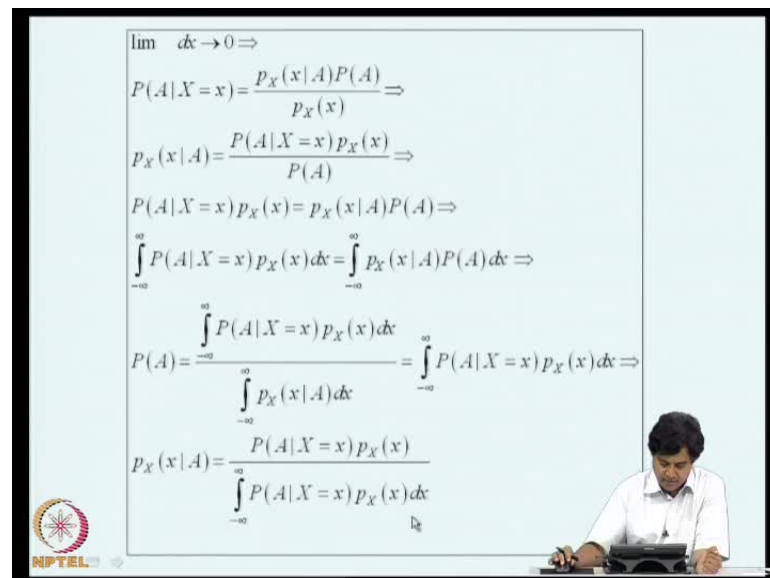
So, let X be a random variable and A be an event. Let us assume that A is an observable and X is not observable; X is a hidden variable, A is something that you can observe.

Before we make any observation on A, we can have a model for X for its probability distribution function or density function. The question is after making an observation, that is A, what can we say about the model for X, that means, how can we update the model for X, after we have observed A.

So, this is a typical question in say finite element model updating or whenever you make a model mathematical model for phenomena, and you then have the opportunity to observe the some samples of those phenomena, how do you update your mathematical model, which was made before you made any observations.

So, this is quick recall of the definition of conditional probability, where we are considering event a, and this probability this event X taking values from X to x plus d x. And we can show that, this is given by p X of x conditioned on A into probability of A divided by the probability density function of x.

(Refer Slide Time: 54:20)



$$\lim_{dx \rightarrow 0} \Rightarrow$$

$$P(A|X=x) = \frac{p_X(x|A)P(A)}{p_X(x)} \Rightarrow$$

$$p_X(x|A) = \frac{P(A|X=x)p_X(x)}{P(A)} \Rightarrow$$

$$P(A|X=x)p_X(x) = p_X(x|A)P(A) \Rightarrow$$

$$\int_{-\infty}^{\infty} P(A|X=x)p_X(x) dx = \int_{-\infty}^{\infty} p_X(x|A)P(A) dx \Rightarrow$$

$$P(A) = \frac{\int_{-\infty}^{\infty} P(A|X=x)p_X(x) dx}{\int_{-\infty}^{\infty} p_X(x|A) dx} = \int_{-\infty}^{\infty} P(A|X=x)p_X(x) dx \Rightarrow$$

$$p_X(x|A) = \frac{P(A|X=x)p_X(x)}{\int_{-\infty}^{\infty} P(A|X=x)p_X(x) dx}$$

Now, if you take in this, the limit of d x going to 0, we get and carry out few simple manipulations, we get the result that p X of x conditioned on a is given by this expression, where this is probability of a conditioned on X equal to x into p X of x. This is the so-called apriori model for X, before measurements on A were made; this is the posterior model for X, after A has been observed.

(Refer Slide Time: 54:57)

Bayes' theorem

$$p_X(x|A) = NP(A|X=x)p_X(x)$$

Remarks

If $P(A|X=x) = P(A) \Rightarrow p_X(x|A) = NP(A)p_X(x)$
 $\Rightarrow p(x|A) = p_X(x)$

If the objective of observing A is to get an improved estimate of pdf of X, then A and $[x < X \leq x + dx]$ should not be independent

NPTEL May 2009 BARC

So, this can be summarized as here, we are interested in modeling x, but where in a position to observe the event A and the objective of observing event A is to learn something about X; if that is a case, p X of x would be your apriori model, and this is a posterior model and this is where the experimental observations are contain; this is useful in structural health monitoring and some of the problems that we may consider later.

(Refer Slide Time: 55:40)

Interpretation of $P(A|X=x)$

Let Y be a random variable that is observable and dependent on X.
Let y_1, y_2, \dots, y_n be the observations made on Y.

$$P(A|X=x) = \prod_{i=1}^n p_Y(y_i | X=x) \quad (\text{likelihood function})$$
$$p_X^*(x) = NL(y_{1:n} | X=x) p_X(x)$$

$p_X^*(x)$ = Posterior pdf
 $p_X(x)$ = Apriori pdf
L = Likelihood function
N = Normalization constant

NPTEL

The interpretation for this probability of a conditioned on X equal to x, you can be offered as follows say, if Y be a random variable that is observable and dependent on X

and if y_1, y_2, \dots, y_n , are the observations made on Y , we defined this function, that is probability density function of Y evaluated at y_i which are observed, given X is equal to x and we call this product as the likelihood function.

So, these functions represent the wisdom contained in our observations and we call it as likelihood function. And gnomonical lecture is $p_X(x)$ is the posterior probability density function. This is the apriori probability distribution density function, L is the likelihood function and N is the normalization constant needed to ensure that the area under this curve is 1. With this, we will conclude the discussion on random variables.