

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

Lecture No. # 25
Random Processes-4
Monte Carlo Simulation Approach-1

In the previous lecture, we studied the question on how to characterize first passage times in the non-linear oscillators, which are driven by white noise.

(Refer Slide Time: 00:35)

Summary : $p \equiv p(\tilde{x}; t | \tilde{x}_0; t_0)$

$$dX(t) = f[t, X(t)]dt + G[t, X(t)]dB(t); t \geq 0; X(0) = X_0$$

$$\frac{\partial p}{\partial t} = -\sum_{j=1}^n \frac{\partial}{\partial x_j} [f_j(x, t)p] + \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [(GDG^t)_y p]$$

$$\frac{d}{dt} \langle h[X(t), t] \rangle = \left\langle \frac{\partial h}{\partial t} \right\rangle + \sum_{j=1}^n \left\langle f_j(X, t) \frac{\partial h}{\partial x_j} \right\rangle + \sum_{i=1}^n \sum_{j=1}^n \left\langle (GDG^t)_y \frac{\partial^2 h}{\partial x_i \partial x_j} \right\rangle$$

$$\langle X_1^m(t) X_1^m(t_0) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^m v(x_1, x_2; t, t_0) dx_1 dx_2$$

$$v(x_1, x_2; t, t_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \eta_1^n p(x_1, x_2; t, t_0 | \eta_1, \eta_2; t, t_0) p(\eta_1, \eta_2; t, t_0) d\eta_1 d\eta_2$$

$$\frac{\partial v}{\partial t} = -\sum_{j=1}^n \frac{\partial}{\partial x_j} [f_j(x, t)v] + \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [(GDG^t)_y v]$$

RELEVANT BCS & ICS

NPTEL

And to summarize what we have done till now, the basic results of Markov process theory has applied to dynamical system is displayed here. So, this is the general stochastic differential equation with multi-dimensional; this is the drift, this is the diffusion terms, and associated with this, we get the so call forward Kolmogorov equation or the Fokker Planck Kolmogorov equation - f p k equation.

As based on this, we can derive the moments, equation governing the moments, that is expected value of h of X of t comma t and this equation can be derived. This is equation for moments at a single time t, but you can generalize this and get equations for moments involving two times t and t naught, and that equation again is another partial differential

equation. We introduce an intermediate function call nu, which is function of x 1 x 2 t comma t naught and that is satisfied by this equation; this again the backward Kolmogorov, this is forward Kolmogorov operator.

(Refer Slide Time: 01:47)

Summary (Continued):

$$\frac{\partial p}{\partial t_0} = -\sum_{j=1}^n f_j [t_0, \tilde{x}_0] \frac{\partial p}{\partial x_{0j}} - \sum_{i=1}^n \sum_{j=1}^n [GDG^i(t_0, \tilde{x}_0)]_{ij} \frac{\partial^2 p}{\partial x_{0i} \partial x_{0j}}$$

$$R(t, \Gamma; t_0, \Omega) = \int_{\Omega} p(\tilde{x}; t | \tilde{x}_0; t_0) d\tilde{x} = P[T > t - t_0 | X(t_0) = \tilde{x}_0]$$

$$\frac{\partial R}{\partial t_0} = \left[-\sum_{j=1}^n f_j [t_0, \tilde{x}_0] \frac{\partial}{\partial x_{0j}} - \sum_{i=1}^n \sum_{j=1}^n [GDG^i(t_0, \tilde{x}_0)]_{ij} \frac{\partial^2}{\partial x_{0i} \partial x_{0j}} \right] R = LR$$

$(n+1)M_n + LM_{n+1} = 0; n = 1, 2, \dots$ $M_n = \langle T^n \rangle$

PLUS: RELEVANT BCS & ICS

So, along for each of these equations, for the Fokker Planck equations, moment equations and this we need to write the appropriate boundary condition. Now, by taking t naught as a independent variable, we can derive an equation known as the backward Kolmogorov equation, which is useful in the study of first passage times. And **we** if we define a survival reliability function R, which is actually probability distribution, probability of first passage time T greater than t minus t naught condition on X of t naught is x naught tilde; this function is again satisfied by this partial differential equation, based on which, we get a sequence of equations for the nth order moments of the first passage time.

So, M n is T n, where T is the first passage time. So, the Markov process approach essentially provides us with shoot of tools, which can be used to characterize stochastic response of non-linear dynamical systems. In many situations or in at least in a few situation, you can get an exact solution, either for the governing Fokker Planck equation or for the moments. In when other situations, more general situations, this forms the basis on which we can develop approximate methods for solving the problems.

(Refer Slide Time: 03:12)

Illustration of deterministic averaging procedure

$\ddot{u} + \omega_0^2 u = \varepsilon f(u, \dot{u}); t \geq 0; u(0) \text{ \& } \dot{u}(0) \text{ specified.}$

$u(t) = a(t) \cos[\omega_0 t + \beta(t)]$



$\dot{u}(t) = -a(t) \omega_0 \sin[\omega_0 t + \beta(t)]$

$\dot{a}(t) = -\frac{\varepsilon}{\omega_0} \sin \phi(t) f[a(t) \cos \phi(t), -a(t) \omega_0 \sin \phi(t)]$

$\dot{\beta}(t) = \frac{\varepsilon}{\omega_0 a(t)} \cos \phi(t) f[a(t) \cos \phi(t), -a(t) \omega_0 \sin \phi(t)]$

$\dot{a}(t) \approx -\frac{\varepsilon}{\omega_0} \frac{1}{2\pi} \int_0^{2\pi} \sin \phi f[a \cos \phi, -a \omega_0 \sin \phi] d\phi$

$\dot{\beta}(t) \approx -\frac{\varepsilon}{\omega_0 a} \frac{1}{2\pi} \int_0^{2\pi} \cos \phi f[a \cos \phi, -a \omega_0 \sin \phi] d\phi$

One such approach we were discussing, namely, how to develop models for envelope and phase of the response processes, using what are known as stochastic averaging principles. So, I briefly outline the averaging concept, as applied to deterministic systems in the previous talk. And here, we have a non-linear free vibration, $\ddot{u} + \omega_0^2 u = \varepsilon f(u, \dot{u})$, and we made a transformation a displayed here, where a and β is the amplitude or the envelope; β is the phase, there taken to be slowly varying functions of time.

And based on that, this is the exact transformations. we derived the governing differential equations for evaluation of a and β . And then, we replace the right hand side by their average over cycle, and while doing that averaging, we treated a and β as constants; and that assumption was based on the fact, that when ε is small, the change in a and β over a cycle will be quite small compare to change in trigonometric functions like, $\cos \phi$, $\sin \phi$. So, that helps us to simplify, eliminate this oscillatory terms deterministic oscillatory terms from our model for envelope and phase, and we may be able to get simpler models for system response.

(Refer Slide Time: 04:37)

Extension to randomly driven systems


$$\ddot{x} + \varepsilon^2 h(x, \dot{x}) + \omega_0^2 x = \varepsilon z(t)$$

$$\langle z(t) \rangle = 0, \langle z(t) z(t + \tau) \rangle = R_{zz}(\tau) \Leftrightarrow S(\omega)$$

$z(t)$ is taken to be broad banded.

- Characteristic time constant of excitation $\gg \ll$ characteristic time constant of the system
- Time duration over which $R_{zz}(\tau)$ decays to 10% of $R_{zz}(0)$ $\ll \gg$ Time duration over which the impulse response of the system decays by 90%.
- ε is small parameter.

$\varepsilon = 0 \Rightarrow x(t) = a \cos(\omega_0 t + \phi) = a \cos \Phi$ *No excitation*

$$\dot{x}(t) = -a\omega_0 \sin(\omega_0 t + \phi) = -a\omega_0 \sin \Phi$$


Now, we will consider the question of extending this argument to randomly driven systems. So, I consider now, $\ddot{x} + \varepsilon^2 h(x, \dot{x}) + \omega_0^2 x = \varepsilon z(t)$; this $z(t)$ is not a white noise; it is however taken to be process is zero mean and auto covariance $R_{zz}(\tau)$, and it is taken to be broad banded. When do you say an excitation is broad banded? It can be specified either in terms of the band width of the power spectral density; we saw the band width of the system transfer function or in terms of the time constants associated with auto covariance functions and impulse response of the system.

(Refer Slide Time: 06:34)

$\varepsilon = 0 \Rightarrow$
 $x(t) = a \cos(\omega_0 t + \phi) = a \cos \Phi; \Phi = \omega_0 t + \phi$
 $\dot{x}(t) = -a\omega_0 \sin(\omega_0 t + \phi) = -a\omega_0 \sin \Phi$
 $\varepsilon \neq 0 \Rightarrow$
 $x(t) = a(t) \cos[\omega_0 t + \phi(t)] = a(t) \cos \Phi(t)$
 $\dot{x}(t) = -a(t)\omega_0 \sin[\omega_0 t + \phi(t)] = -a(t)\omega_0 \sin \Phi(t)$
 \Rightarrow
 $\dot{a} = \frac{\varepsilon^2}{\omega_0} h [a \cos \Phi, -a\omega_0 \sin \Phi] \sin \Phi - \frac{\varepsilon z(t)}{\omega_0} \sin \Phi$
 $\dot{\phi} = \frac{\varepsilon^2}{a\omega_0} h [a \cos \Phi, -a\omega_0 \sin \Phi] \cos \Phi - \frac{\varepsilon z(t)}{a\omega_0} \cos \Phi$
 • No approximations till now.

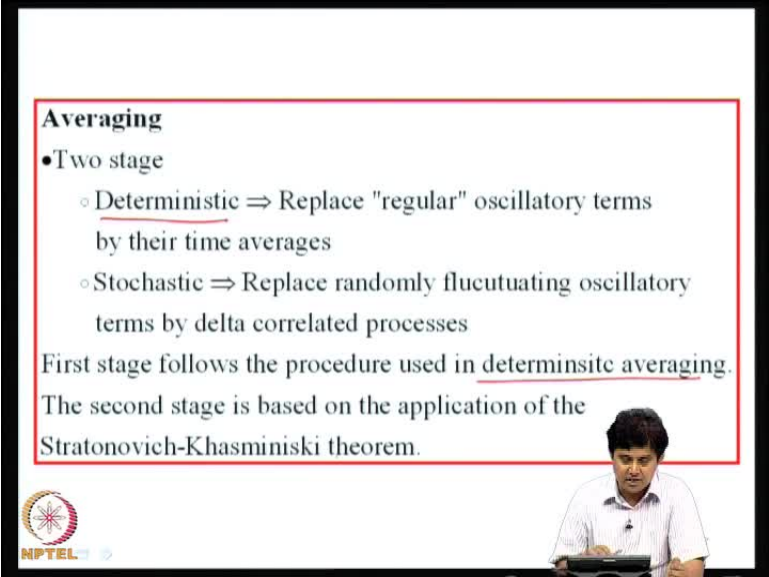
So, suppose if you say characteristic time constant of excitation is much greater than the characteristic time constant of the system, then we can say that it is a broad band process; I think it's much less. Now, time duration over which $R_z z$ of τ decays to 10 percent of its maximum value, can be taken as the characteristic time constant of excitation. For system, it can be taken as, time duration over the impulse response decays over by 90 percent's.

Now, we are taking this parameter epsilon that appears here to be small; and if indeed it is 0, and if there is no excitation, we get the solution as a cos phi a omega naught sin phi. So, this is the case for no excitation and no damping also. But when there is light damping and weak excitations, we stick to the functional form of this solution namely, x of t a cos phi and x dot of t is a omega naught sin phi, where capital phi is t omega naught plus small phi. For epsilon naught equal to 0, when epsilon is naught equal to 0, two things happen; one is system become nonlinearly damped and other is it get driven by a broad band excitation.

So, we still assume that the functional form of this equations remain the same, except that the quantities a and phi are now taken to be slowly varying functions of time. So, based on this, **I will now there is no again...** let me emphasis, there is no approximation as of now. You can implement this transformation, and get these equations for a dot and phi dot. So, z of t is now multiplied by sin phi and cos phi; these are actually quite

intimidating set of non-linear differential equations. You can see there are trigonometric terms and there is 1 by a sitting here, and this a multiplies our random excitation, and so on and so for. So, by introducing this transformation, we seem to have made the problem more complicated. But if we now bring in the fact that epsilon is small and a and beta are slowly varying, then certain simplifications become possible.

(Refer Slide Time: 08:03)



Averaging

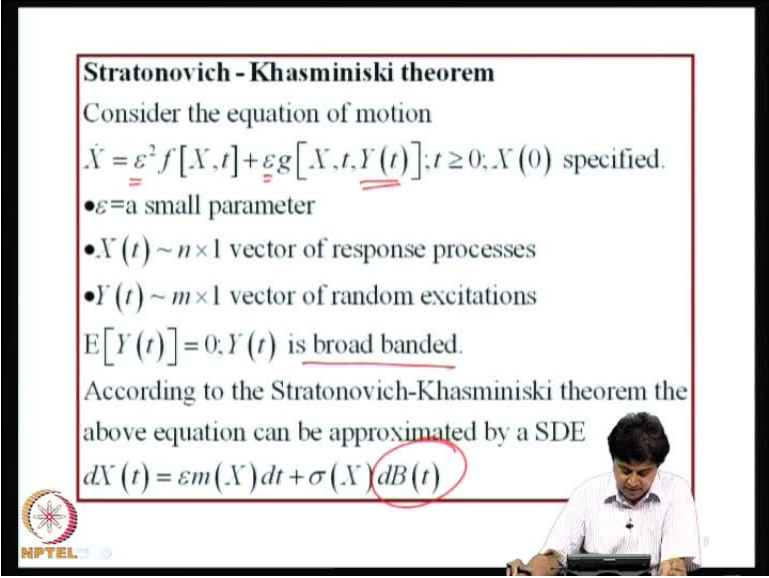
- Two stage
 - Deterministic \Rightarrow Replace "regular" oscillatory terms by their time averages
 - Stochastic \Rightarrow Replace randomly fluctuating oscillatory terms by delta correlated processes

First stage follows the procedure used in deterministic averaging.
The second stage is based on the application of the Stratonovich-Khasminski theorem.

NPTEL

The averaging here is done in two steps: one is the deterministic averaging, where we replace regular oscillatory terms by their time averages, as we did in case of deterministic averaging; we also do another component of averaging known as stochastic averaging, where we replace randomly fluctuating oscillatory terms by delta correlated processes. That means, excitations which are not truly white noise excitations, where replacing them by equivalent white noise excitations; what is that equivalent, is what this method, the theory behind this method establishes. The first stage follows the procedure used in deterministic averaging. The second stage is based on the application of what is known as stratonovich khasminiski theorem.

(Refer Slide Time: 08:48)



Stratonovich - Khasminiski theorem

Consider the equation of motion

$$\dot{X} = \varepsilon^2 f[X, t] + \varepsilon g[X, t, Y(t)]; t \geq 0; X(0) \text{ specified.}$$

- ε = a small parameter
- $X(t) \sim n \times 1$ vector of response processes
- $Y(t) \sim m \times 1$ vector of random excitations

$E[Y(t)] = 0$; $Y(t)$ is broad banded.

According to the Stratonovich-Khasminiski theorem the above equation can be approximated by a SDE

$$dX(t) = \varepsilon m(X) dt + \sigma(X) dB(t)$$

The slide includes an inset image of a man sitting at a desk with a laptop. In the bottom left corner, there is a logo for NPTEL (National Programme on Technology Enhanced Learning).

So, what this theorem says is, if you have a non-linear system, which is excited by a broad band excitation Y of t , and this parameter ε is small, this X of t is n cross of vector of response processes and Y of t m cross one vector of random excitations; and we assume that, mean of this excitation is 0 and Y of t is broad banded. According to the stratonovich khasminiski theorem, the above equation can be approximated by a stochastic differential equation, where we are now replacing the band limited random excitations by equivalent increments of Brownian process, and also, we are simplifying the functions f and g by doing certain time averaging.

(Refer Slide Time: 09:39)


$$dX(t) = \varepsilon m(X) dt + \sigma(X) dB(t)$$

$$m = T^{av} E\{f\} + \int_{-\infty}^0 E\left\{\left(\frac{\partial g}{\partial X}\right)_t (g^t)_{t+\tau}\right\} d\tau$$

$$\sigma\sigma^t = T^{av} \int_{-\infty}^0 E\left\{(g)_t (g^t)_{t+\tau}\right\} d\tau$$

$$T^{av}\{\bullet\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t_0}^{t_0+T} \{\bullet\} dt$$

Reference :
 J B Roberts and P D Spanos, 1986, Stochastic averaging: an approximate method of solving nonlinear random vibration problems, Invited Review, International Journal of Nonlinear Mechanics, 21(2), 111-134.



So, that time averaged and stochastically averaged equations are given here; this drift is given by this and diffusion coefficient is given in terms of this. So, while carrying out this averaging, the quantities a and this f , these quantities X here are treated as constants. So, there is this paper I reference in the previous lecture; so, this is a review paper, which explains the basics of this method.

(Refer Slide Time: 10:17)

Exact

$$\dot{a} = \frac{\varepsilon^2}{\omega_0} h [a \cos \Phi, -a\omega_0 \sin \Phi] \sin \Phi - \frac{\varepsilon \xi(t)}{\omega_0} \sin \Phi \checkmark$$


$$\dot{\phi} = \frac{\varepsilon^2}{a\omega_0} h [a \cos \Phi, -a\omega_0 \sin \Phi] \cos \Phi - \frac{\varepsilon \xi(t)}{a\omega_0} \cos \Phi \checkmark$$

Averaging \Rightarrow

$$da(t) \approx -\frac{\varepsilon^2}{\omega_0} F(a) dt + \frac{\pi S(\omega_0)}{2a\omega_0^2} dt - \frac{[\pi S(\omega_0)]^{\frac{1}{2}}}{\omega_0} dB_1(t)$$

$$d\phi(t) \approx -\frac{\varepsilon^2}{a\omega_0} G(a) dt - \frac{[\pi S(\omega_0)]^{\frac{1}{2}}}{a\omega_0} dB_2(t) \checkmark$$

$$F(a) = -\frac{1}{2\pi} \int_0^{2\pi} h [a \cos \Phi, -a\omega_0 \sin \Phi] \sin \Phi d\Phi \checkmark$$

$$G(a) = -\frac{1}{2\pi} \int_0^{2\pi} h [a \cos \Phi, -a\omega_0 \sin \Phi] \cos \Phi d\Phi \checkmark$$


Now, first level of exercise is to make the transformation; these two equations are equations for amplitude and phase, in terms of the random excitations, and the non-linear

dissipation and other terms; these are exact. Now, if you do the averaging, that is the two stage averaging, the resulting stochastic differential equation will have this form, da of t is some F of a and this is a constant. And now, I get a increment of Brownian motion process with an equivalent multiplier. Similarly, the phase equation is govern by this, $s d$ this function F of a and G of are given by this equation, which are quite similar to what we did it in deterministic averaging.

(Refer Slide Time: 11:11)

$$da(t) \approx -\frac{\varepsilon^2}{\omega_0} F(a) dt + \frac{\pi S(\omega_0)}{2a\omega_0^2} dt - \frac{[\pi S(\omega_0)]^{1/2}}{\omega_0} dB_1(t)$$

$$d\phi(t) \approx -\frac{\varepsilon^2}{a\omega_0} G(a) dt - \frac{[\pi S(\omega_0)]^{1/2}}{a\omega_0} dB_2(t)$$

$\begin{Bmatrix} a(t) \\ \phi(t) \end{Bmatrix}$ is a Markov vector; **more interestingly** $\{a(t)\}$ is Markov.

- Forward equation : transient and steady state solutions
- One and two time moment equations
- Backward equation
- Reliability function
- GPV equations

So, you now look at these two equations; you can recognize that, the vector a and ϕ have Markovian vector - Markovian character. We look carefully at equation for a of t , there is no ϕ of t in this; that means, a of t itself is a Markov process, but you look at ϕ of t , there is a here. So, ϕ of t needs on its own is not Markov, but you look now at the vector a and ϕ together, that is Markovian. So, this is the case in which the response vector has a components, which has Markovian property; another components which does not have Markovian property, but taken together, they have Markovian property.

Now, once this simplified equations is set of equations are obtained, I can take two routes now. One is to look at only the envelope and write the forward equations, forward Kolmogorov equations, backward Kolmogorov equations, the equation for first passage times, the moments, first passage times, and so on and so forth. Alternatively, I can consider the joint density function between a and ϕ , and develop the forward equation,

the backward equation, the moments equations, the generalized punter-ion-wit equation, etcetera, for those problems.

So, that would mean this method of stochastic averaging has now enhanced; our ability to study non-linear oscillators, which are now driven not necessarily by white noise, but by broad band excitations. So, they method establishes a equivalent Markovian model for the system behavior, which is not in reality Markov.

(Refer Slide Time: 13:08)

FPK equation governing $p(a, \phi, t | a_0, \phi_0, t_0)$

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial a} \left[\left\{ -\frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p \right] - \frac{\partial}{\partial \phi} \left[\left\{ -\frac{\varepsilon^2}{a\omega_0} G(a) \right\} p \right] + \frac{\pi S(\omega_0)}{2\omega_0^2} \left[\frac{\partial^2 p}{\partial a^2} + \frac{1}{a^2} \frac{\partial^2 p}{\partial \phi^2} \right]$$

Steady state

$$0 = -\frac{\partial}{\partial a} \left[\left\{ -\frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p \right] - \frac{\partial}{\partial \phi} \left[\left\{ -\frac{\varepsilon^2}{a\omega_0} G(a) \right\} p \right] + \frac{\pi S(\omega_0)}{2\omega_0^2} \left[\frac{\partial^2 p}{\partial a^2} + \frac{1}{a^2} \frac{\partial^2 p}{\partial \phi^2} \right]$$

NPTEL

(Refer Slide Time: 13:32)

FPK equation governing $p(a, t | a_0, t_0)$

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial a} \left[\left\{ -\frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p \right] + \frac{\pi S(\omega_0)}{2\omega_0^2} \frac{\partial^2 p}{\partial a^2}$$

Steady state

$$-\frac{d}{da} \left[\left\{ -\frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p \right] + \frac{\pi S(\omega_0)}{2\omega_0^2} \frac{d^2 p}{da^2} = 0$$

$$\Rightarrow \left\{ \frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p + \frac{\pi S(\omega_0)}{2\omega_0^2} \frac{dp}{da} = 0$$

$$p(a) = Ca \exp \left\{ -\frac{2\varepsilon\omega_0}{\pi S(\omega_0)} \int_0^a F(s) ds \right\}; 0 < a < \infty$$

NPTEL

So, this is the FPK equation for the joint density of a and phi for this function transition PDF. And if you are interested in steady state, $\frac{\partial p}{\partial t}$ vanishes; if it is admissible and I get the reduced equations properly. This you may be able to solve lot more easily than this equation. Or you can consider now a of t itself as a scalar Markov random process, and write the FPK equation governed by p of a comma t a naught t naught; and in steady state, this becomes a ordinary differential equations, because now there are only two independent variables t and a.

So, in steady state, it becomes ordinary differential equation and I can solve for the envelope of the response process in steady state. So, this is one of the major achievements now; I have got now probability distribution function of the envelope process of a non-linear systems, which is driven by broad band excitation. So, our ability to model envelopes has now considerably enhanced; this an exact solutions within the framework of approximation, that the method of averaging introduce.

(Refer Slide Time: 14:24)

Remarks

$$0 = -\frac{\partial}{\partial a} \left[\left\{ -\frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p \right] - \frac{\partial}{\partial \phi} \left[\left\{ -\frac{\varepsilon^2}{a\omega_0} G(a) \right\} p \right] + \frac{\pi S(\omega_0)}{2\omega_0^2} \left[\frac{\partial^2 p}{\partial a^2} + \frac{1}{a^2} \frac{\partial^2 p}{\partial \phi^2} \right]$$

$p(a, \phi | a_0, \phi_0)$ $\left\{ \begin{matrix} a \\ \phi \end{matrix} \right\}$

$$0 = -\frac{\partial}{\partial a} \left[\left\{ -\frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p \right] + \frac{\pi S(\omega_0)}{2\omega_0^2} \frac{\partial^2 p}{\partial a^2}$$

$p(a | a_0)$ $\leftarrow \frac{\partial}{\partial a^2}$

$$\Rightarrow p(a, \phi) = \frac{p(a)}{2\pi} \quad 0 < a < \infty; 0 < \phi < 2\pi$$

$$\Rightarrow p(u, \dot{u}) = \frac{p(a)}{2\pi a} \Big|_{a=\sqrt{u^2 + \left(\frac{\dot{u}}{\omega_0}\right)^2}}; -\infty < u, \dot{u} < \infty$$

NPTEL

Now, you look at the two equations in steady state; this is equation for p of a phi a naught phi naught, whereas this equation this p is a condition on a naught; this is steady state. Now, although I use same notations p, you must understand at this equation is for a of t and this is equation is for a and phi. Now, if you now look at a possible solution which is in variable separable form, you can verify that, the first of these equation is satisfied by p of a divided by 2 phi, where p of a solution of this equation; that would

mean, **the amplitude**, the phase is almost characterless here; it is uniformly distributed 0 into 2π . So, one moment you solve this ordinary differential equation, actually this ordinary differential equation, because a is the only independent variable, other problem is completely solved, because if you substitute this and look for phase being uniformly distributed in 0 to 2π , you substitute this solution into the first of this equations; this equation is satisfied.

So, therefore, you can also get an approximation to the joint density function between the process and its time derivatives plus when velocity, because I have now model for a and ϕ ; I can use method of transformation of random variables, and $1/a$ will be a Jacobean and I can get this. And this also enables us to get the model for the parents process and its derivatives. So, again, I have a handle on the response of the non-linear system to broad band excitations, I have solved envelope process - phase process; and from them, I am now getting the probability density of the parents process, although if I directly try to write the, I mean, originally the parent process and its time derivatives do not constitute a Markov process.

(Refer Slide Time: 16:32)

Remarks (continued)

- The transient solution of

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial a} \left[\left\{ -\frac{\varepsilon^2}{\omega_0} F(a) + \frac{\pi S(\omega_0)}{2a\omega_0^2} \right\} p \right] + \frac{\pi S(\omega_0)}{2\omega_0^2} \frac{\partial^2 p}{\partial a^2}$$

p(a,t|a_0,t_0)

$$p(a,0|a_0,0) = \delta(a-a_0)$$

can be obtained by using eigenfunction method.

- Similar approximate solutions for first passage times can also be obtained.
- The formulaion can be generalized to deal with systems with random parametric excitation such as in

$$\ddot{u} + \varepsilon^2 h(u, \dot{u}) + \omega_0^2 [u + \varepsilon \zeta(t)] = \varepsilon \xi(t); t \geq 0; u(0) \& \dot{u}(0) \text{ specified.}$$

$\zeta(t)$ & $\xi(t)$: broad band, zero mean random excitations

NPTEL

Now, however, transient solution, the transient solution for p of a t conditions on a naught t naught can be developed based on a method of variable separation; we will get, we will have to solve on the way an Eigen value problem, which leads to Eigen functions

in terms of hypergeometric functions, and we can like laguerre hermite and that type of functions, and we will be able to solve this problem.

(Refer Slide Time: 17:27)

Special case

$$\epsilon^2 h(x, \dot{x}) = 2\eta\omega_0 \dot{u}$$

$$p(a) = \frac{a}{\sigma^2} \exp\left(-\frac{a^2}{2\sigma^2}\right); 0 < a < \infty$$

$$p(\phi) = \frac{1}{2\pi}; 0 < \phi < 2\pi$$

$$\sigma^2 = \frac{\pi S(\omega_0)}{2\eta\omega_0^3}$$

Note: compare this with results on envelope and peak distribution obtained earlier.

Handwritten red equation: $\ddot{u} + 2\eta\omega_0 \dot{u} + \omega_0^2 u = z(t)$

NPTEL 18

Now, in the illustration that I have done, I have consider only external excitation. But suppose there is a parametric excitation, that two random excitations, zeta of t and s i of t; both are broad banded, zero mean, may be dependent, could be independent; actually, the method of averaging can be used for this also. Now, in the special case, where system is linear; so, what I am taking about, u double dot plus 2 eta omega u dot plus omega square u is equal to z of t. Now, I know that by a independent study, that the envelope of this process is rarely distributed. If f of t is broad banded, excitation is broad banded, the response is narrow banded; and for narrow banded process, the envelopes are rayleigh distributor.

Now, the same result I get through averaging; I indeed get a rayleigh density function for this system, and the phase is uniformly distributed between 0 to 2 phi, and this sigma square is equivalent strength of the white noise; there is no white noise in this problem, because z of t is the band limited excitation. But the method of averaging replaces the excitation - band limited excitation - by an equivalent you know white noise excitation. So, the strength of that white noise is provided in the calculations of this method. Now, you can compare this with the envelope and peak distribution obtained earlier, by using you know convolution integral approach, and so on and so forth.

(Refer Slide Time: 18:46)

Remarks (continued)

- The method can also be generalized to deal with systems with nonlinear stiffness:


$$\ddot{u} + \varepsilon^2 h(u, \dot{u}) + \omega_0^2 \Lambda(u) [1 + \varepsilon \zeta(t)] = \varepsilon \zeta(t)$$

$t \geq 0; u(0) \& \dot{u}(0)$ specified

The definition of the envelope here needs to be modified suitably as

$$V(t) = \frac{\dot{x}^2}{2} + \omega_0^2 \int_0^u \Lambda(s) ds$$

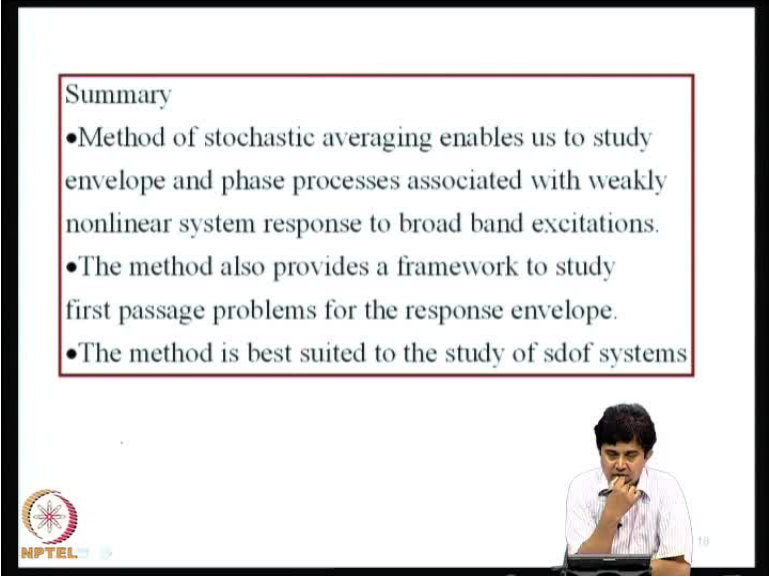
The method leads to a Markov approximation to the process $V(t)$.



17

Now, what happens if stiffness is also non-linear? I have been talking about non-linear damping; stiffness is non-linear and of course, I also added for variety parameter excitation term also, the definition of envelope now will not be in terms of trigonometric function. You saw this, that the envelope V of t can be defined in terms of the energy, that is kinetic energy plus the potential energy in the system, and we can develop a Markov model for this envelope. I am not going into get into details, I already given a reference. I am trying to provide certain key ideas, which would help you to read the relevant literature.

(Refer Slide Time: 19:34)



The image shows a video frame from a lecture. In the center, there is a white rectangular box with a red border containing the following text:

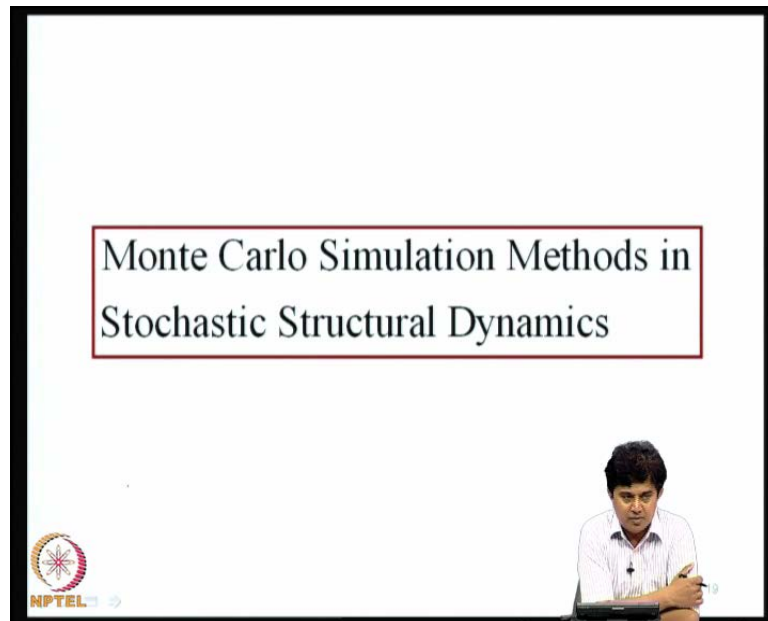
Summary

- Method of stochastic averaging enables us to study envelope and phase processes associated with weakly nonlinear system response to broad band excitations.
- The method also provides a framework to study first passage problems for the response envelope.
- The method is best suited to the study of sdof systems

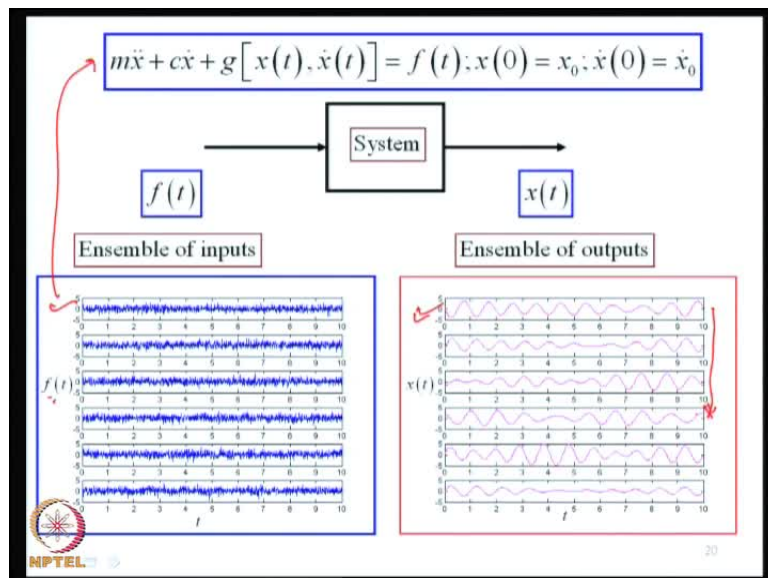
In the bottom right corner of the video frame, a man in a light-colored shirt is visible, sitting at a desk and looking towards the camera. In the bottom left corner of the video frame, there is a logo for NPTEL (National Programme on Technology Enhanced Learning) featuring a stylized sun or starburst design.

So, in summary, what we can say is, the method of stochastic averaging enables us to study envelope and phase processes associated with weakly non-linear system response to broad band excitations. The method also provides a framework to study first passage problems for the response envelope. The method is best suited to the study of single degree freedom systems. It leads to many interesting results, which provide very useful inside into non-linear system behavior. Although there applicable to small scale systems, the insides that we gain from this are quite valuable in understanding behavior of larger systems. with this, I conclude the discussion on the use of Markov process theory in the evaluation analysis.

(Refer Slide Time: 20:46)



(Refer Slide Time: 20:46)



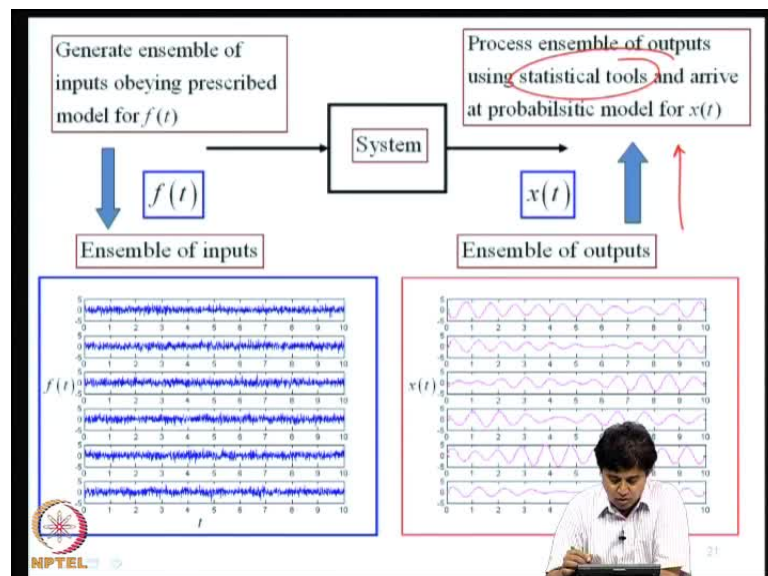
We now move on to a new topic, which is Monte Carlo simulation methods in stochastic structural dynamics. So, as the name suggests, it essentially a numerical approach to a random evaluation studies. So, what it does? So, imagine the equilibrium equation of a dynamical system of $m\ddot{x} + c\dot{x} + g[x(t), \dot{x}(t)] = f(t)$; this f of t let it be a random excitation is a random process. Since excitation is the random

process, the response process is also a random process. So, we have f of t going to the system and producing x of t ; if f of t is a random process, x of t is also the random process. What we have been doing till now, is to characterize x of t in terms of the properties of f of t , in terms of for example, if the system is linear, if I know the power spectral density function of excitation, can I get the power spectral density function of the response in steady state? Or if I know the covariance function of f of t , auto covariance of f of t , can I get an equation which governs auto covariance of x of t ?

So, here, I will be basically using the governing equation of motion; the transmission of uncertainty in f of t to the response is essentially through the mechanics of the problem. Now, in Markov vector approach, what we did; we assume excitation to be white in nature. So, we were aiming to get the probability density functions, first passage times, and its probability density function, moments, etcetera, directly from the governing equation and find that they have Markov properties.

Now, in Monte Carlo simulation approach, we take an alternative route. So, here, what we do is, we treat excitation is the random process, that would mean, this random process is actually ensemble of time histories.

(Refer Slide Time: 23:25)



So, this continues; all though I shown if you here, this is an infinite ensemble. For any single realization of this ensemble, suppose if I take this realization, I can solve this equation numerically and get the sample of the response; that means, I am essentially

carrying out a deterministic analysis, as far as the realization of excitation is concerned. I take a realization of excitation, solve this equation using standard methods of solving ordinary differential equation - numerical methods, get the time history of the response. I repeat that for every realization of f of t ; for corresponding to these ensembles of inputs, I get an ensemble of response time histories. From this, what I do? So, one of the task that I should do is, I should be able to produce an ensemble of inputs, obeying prescribed models for f of t ; see, f of t is a random process, means, it is completely specified by its n th order joint density function.

When I say that I have an ensemble of inputs, which corresponds to realization f of t , means, that I am able to produce time histories of f of t , which obey the prescribed probabilistic model for f of t . So, if f of t is say a gaussian random process with a specified power spectral density function, zero mean stationary random process with the specified power spectral density function, the problem would be, how do you simulate samples of such a random process? That would mean, if I were to simulate these ensemble of time histories, and on these ensemble of time histories, if I were to do statistical processing of these time histories and estimate the power spectral density function. I should get back the target power spectral density function, which I had at the outside, fine; that mean that is one of the capabilities that we have to develop. How to generate samples of random quantities, which obey the prescribed probabilistic loss; that is one of the important components of Monte Carlo simulation study and this has to be done digitally on a computer.

Then, we simulate the system dynamics through the solution of the system behavior, which could itself be a, for example, a finite element model. This could be a wind on a cooling tower, and I have a finite element model made up of beams plates shells and soil, and more media, etcetera. And I can do a sample calculation and get the response of the systematic given location. So, at the end of the this exercise, I get ensemble of outputs, but I am interested in statistical properties of x of t , I am interested in knowing what is power spectral density function of x of t , what is first passage time associated with x of t , what is the probability distribution of first passage time associated with x of t ; that would mean, after getting this ensemble, I have to perform one more exercise, where I have to process the ensemble of outputs using statistical tools and arrive at probabilistic model

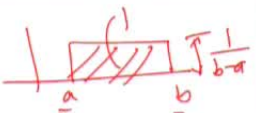
for x of t . Based on these probabilistic models, I will be able to take a decision on the system behavior, right.

So, task one is to simulate samples of f of t obeying prescribed probabilistic loss. How do you test that you actually succeeded in doing? We have to use methods of statistics. You simulate samples of time histories and get ensemble of **response** time histories, and you have to get a probabilistic model for this. How do you do? You have to do statistical processing of this ensemble. So, the subject of statistics underlies forms one of the main components of Monte Carlo simulation studies. So, we will split this discussion Monte Carlo simulations into four aspects: one is how to simulate realizations of random quantities, which obey prescribed probabilistic loss; second, after having simulated those realizations, how do you test that you are succeeded in doing so; Third, corresponding to ensemble of excitations and system parameters, how do you solve the governing equation for the system behavior and get an ensemble of time histories of the response; if this system behavior is characterized in terms of say a set of stochastic differential equations, how do numerically simulate samples of response of a system which is governed by an SDE.

(Refer Slide Time: 27:38)

Another perspective


Consider the problem of evaluation of the definite integral $I = \int_a^b f(x) dx$.



This can be re-written as

$$I = (b-a) \int_a^b f(x) \left(\frac{1}{b-a} \right) dx = (b-a) \int_a^b f(x) p_X(x) dx$$

where $p_X(x) = \left(\frac{1}{b-a} \right); a < x < b$ is now interpreted as the pdf of a random variable that is uniformly distributed in a to b .


22

So, that is third aspect that we need to investigate. The fourth aspect is of course, after getting the ensemble of response, how do we get the probabilistic models for the response process? So, I will begin this discussion by considering questions on statistics.

Before that, we could also look at the problem of simulation from a slightly different perspective. Consider the problem of evaluation of a definite integral $\int_a^b f(x) dx$; this I can rewrite as, $\int_a^b f(x) \frac{1}{b-a} dx$ and I will write it as $\int_a^b f(x) p_X(x) dx$. Now, if you interpret $p_X(x)$, what is this function? this is a, this is b and this height is $\frac{1}{b-a}$. So, what is this area? This is 1; this function between a and b represents the probability density function of a random variable, which is uniformly distributed between a and b.

(Refer Slide Time: 28:40)

Following this, the integral I is now interpreted as an expectation

$$I = (b-a) \langle f(X) \rangle \quad \text{wrt } p_X(x) \sim U(a,b)$$

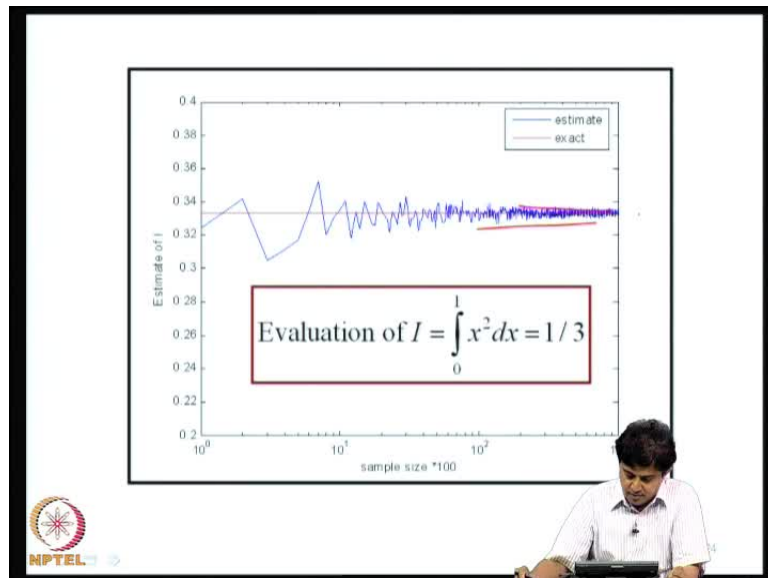
where the expectation is evaluated with respect to $p_X(x)$. Furthermore, I is now approximated by

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N f(X_i)$$

where X_i -s are uniformly distributed random numbers samples from $p_X(x)$.

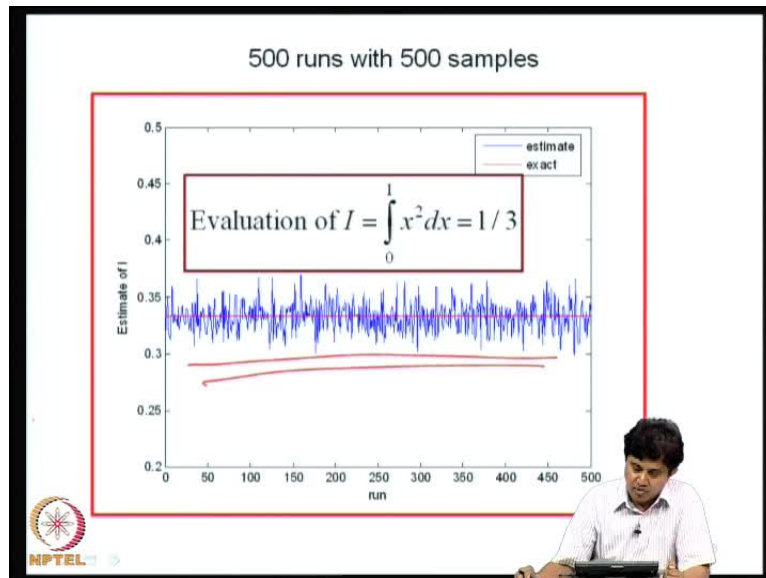
So, if that probability density function I write it as $p_X(x)$. Then, this integral can be interpreted as expected value of f of x with respect to this density function; that is, I is $\int_a^b f(x) p_X(x) dx$. Now, this expectation is nothing but an average and that average can be return as, $\frac{1}{N} \sum_{i=1}^N f(X_i)$; that would mean an integral which is a deterministic quantity now is being evaluated as an expectation of a random variable which is f of X , where X is uniformly distributed in between a and b.

(Refer Slide Time: 29:21)



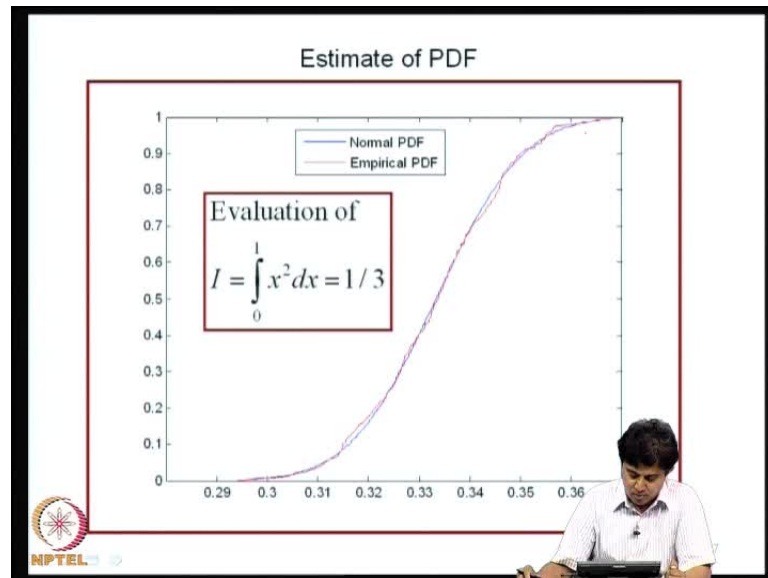
Now, let me just to illustrate, suppose I am interested in evaluating $\int_0^1 x^2 dx$, I know the answer is $1/3$. Now, for a moment, let us assume that, I have a method to simulate samples of x of t on a computer such that, $p(X=x)$ is uniformly distributed in a and b , that I have to still demonstrate, but let us assume that capability exist with us. So, what I was shown here? There red line is the exact solution; this value of the integral on y axis and these samples that we use in evaluating this expectation through simulation is shown on x axis. So, as you see the smaller samples, as a fluctuation and samples become large 1000 samples, we are hovering very closely around the exact value which is $1/3$.

(Refer Slide Time: 30:23)



Now, if I repeat this exercise with another set of a samples, I get a similar curve; not exactly the same curve, because I am using different random numbers. Thus, they also show a similar behavior. Now, if I take 500 samples and do 500 runs, in each run, I will take 500 different samples, and I perform the evaluation of I and plot it as a function of different runs; there are 500 runs, and each runs, there are 500 samples, I get this kind of trajectory; that would mean, integrals can be described in probabilistic terms and we may have a way of evaluating integrals through Monte Carlo simulations; indeed one of the major application of Monte Carlo simulations lies in evaluation of multi fold integrals. And since reliability is essentially in some sense evaluation of multi-dimensional integrals, you can easily see that Monte Carlo simulations have important role to play in these exercises.

(Refer Slide Time: 31:11)



So, here, what I have shown is, I would treated this, I that have got here as a random variable and plotted its probability distribution function. It turns out that, the blue line is the normal probability distribution function and red line is the empirical probability distribution function of a realizations of I; and it turns out that the nearly Gaussian. These are empirical observations, I have to put all these matters on a proper footing.

(Refer Slide Time: 31:44)

Ingredients of MCS

- Methods for generating samples of excitations and system parameters compatible with the prescribed probabilistic models
- Test statistically if the generated samples indeed obey the prescribed probabilistic laws.
- A computational model for the system dynamics which accepts samples of inputs and system parameters produced above and generates an ensemble of response quantities.
- Statistical processing of ensemble of response time histories and inferences on system behavior

We will begin with a review of elements of statistical methods

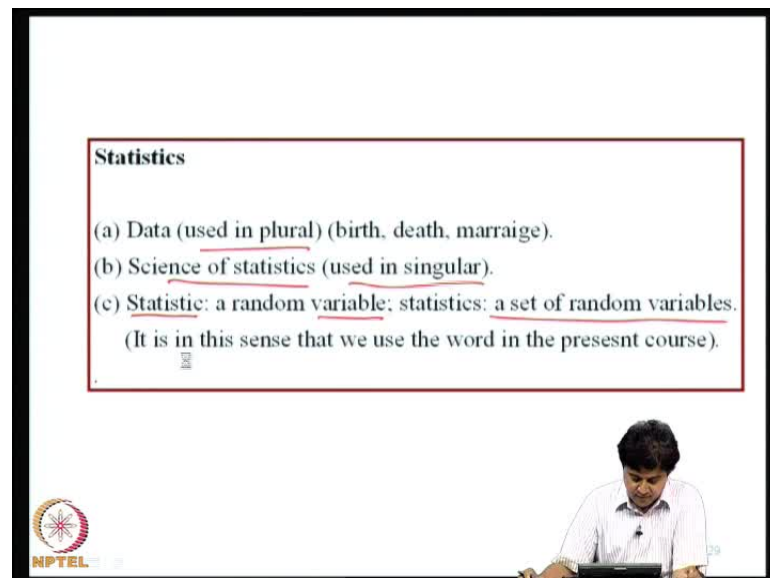
MPTEL

20

So, we will begin that discussion now. So, I will again re-capsulate what are the major ingredients of Monte Carlo simulation. First, methods for generating samples of

excitations and system parameters compatible with the prescribed probabilistic models. Next, to test statistically if the generated samples indeed obey the prescribed probabilistic laws. How do you know what you are doing is right? So, that have to do. Then, third one is a computational model for the system dynamics, which accepts samples of inputs and system parameters produced in the previous steps and generates an ensemble of response quantities. This as its already set could be a finite element model or a numerical model for solution of a stochastic differential equation; it could be any model that is based on the physical laws, that govern the system behavior. The last step is a statistical processing of ensemble of response time histories and inferences on system behavior, where based on which we make decisions. So, there are two places, that is, namely this step and this step, where we need to understand how to process data. So, we will begin with a review of elements of statistical methods, just to place all the terms and techniques in a unified framework.

(Refer Slide Time: 33:08)



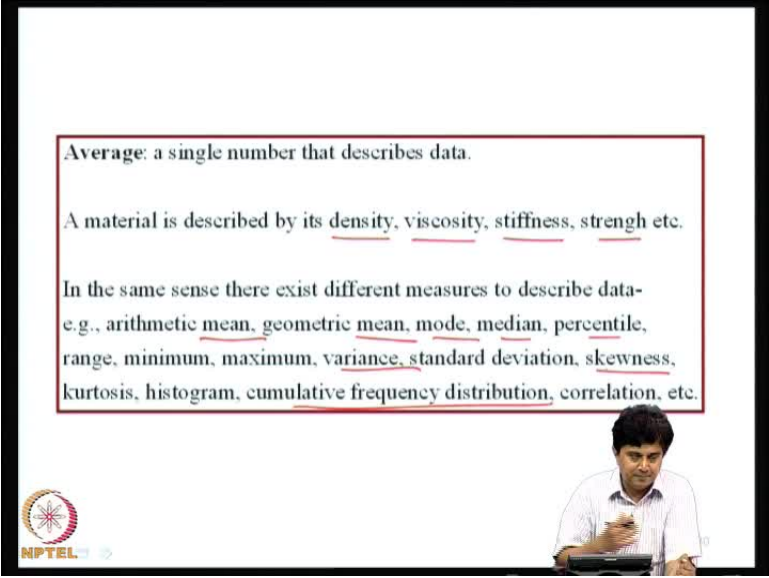
Statistics

- (a) Data (used in plural) (birth, death, marriage).
- (b) Science of statistics (used in singular).
- (c) Statistic: a random variable; statistics: a set of random variables.
(It is in this sense that we use the word in the present course).

NPTEL

So, the word statistics can be used in different senses; it can be used in plural, to mean data like data on birth, death and marriage so called vital statistics or it is a subject - science of statistics; it is used in singular, but what we do is, we use statistic in singular to denote a random variable statistics is a set of random variables.

(Refer Slide Time: 33:58)



Average: a single number that describes data.

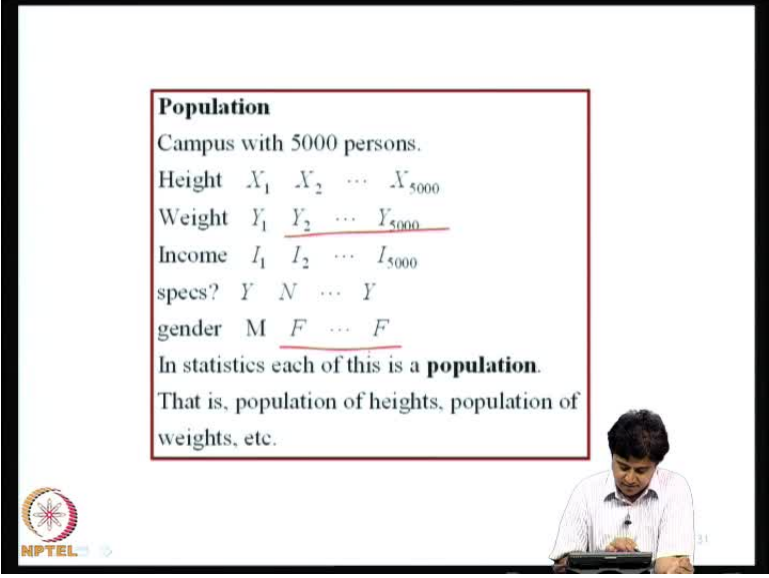
A material is described by its density, viscosity, stiffness, strength etc.

In the same sense there exist different measures to describe data-
e.g., arithmetic mean, geometric mean, mode, median, percentile,
range, minimum, maximum, variance, standard deviation, skewness,
kurtosis, histogram, cumulative frequency distribution, correlation, etc.

NPTEL

So, it is in this sense that we are going to use this word. Statistics is a subject is a singular, but this form a, it is a set of random variables; so, what it mean? What does that mean? We will have to build a set of terminologies, vocabulary to describe these issues. So, I will briefly introduce them; the word average, we consider average is single number that describes the data. So, the data is a collection of numbers, which describes some property of certain physical phenomena; for example, if we have a material, we will describe it by its density, viscosity, stiffness, strength, etcetera. So, if we have a data, what you do? You talk about its mean, geometric mean, arithmetic mode, median, percentile, minimum, maximum, variance, skewness, kurtosis, cumulative distribution, correlation, etcetera. So, these are descriptors of data, which conveys certain specific meaning; for example, if I say material as viscosity of certain units, certain number, it means something in the same sense, if I say skewness of this data is so much, it should convert some specific meaning.

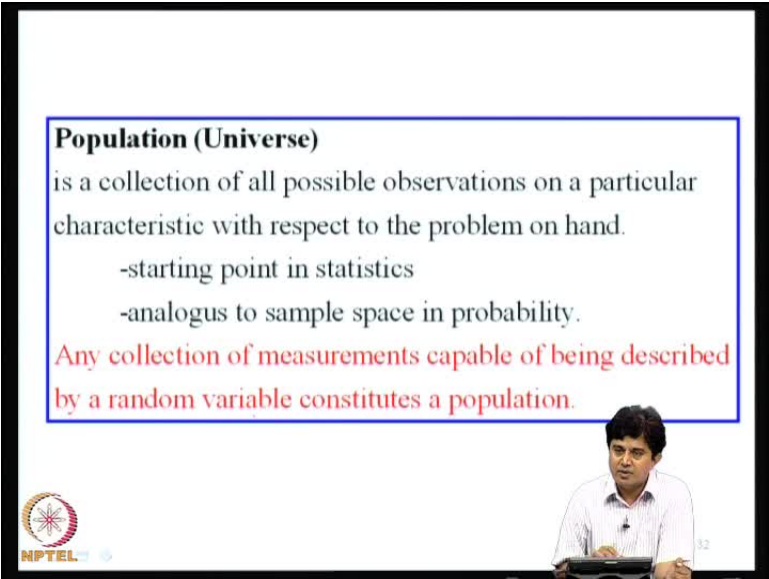
(Refer Slide Time: 35:00)



Population
Campus with 5000 persons.
Height $X_1 \ X_2 \ \dots \ X_{5000}$
Weight $Y_1 \ Y_2 \ \dots \ Y_{5000}$
Income $I_1 \ I_2 \ \dots \ I_{5000}$
specs? $Y \ N \ \dots \ Y$
gender $M \ F \ \dots \ F$
In statistics each of this is a **population**.
That is, population of heights, population of weights, etc.

MPTEL 3.1

(Refer Slide Time: 35:42)



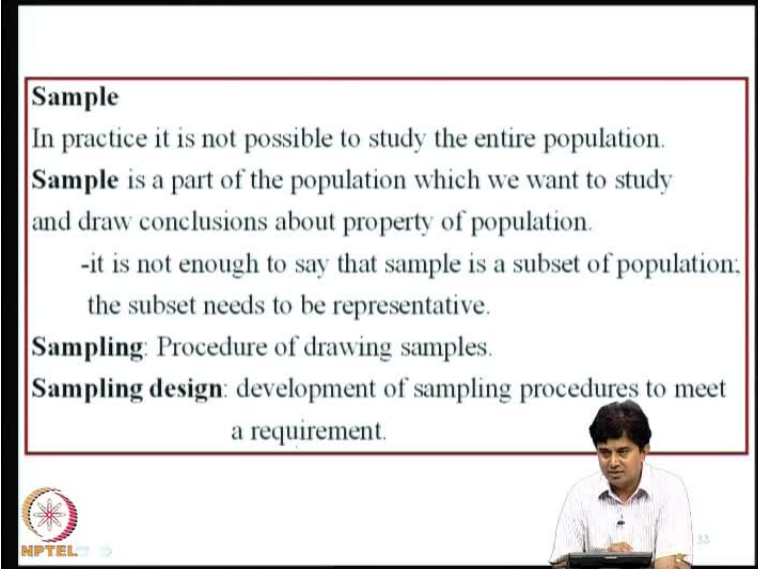
Population (Universe)
is a collection of all possible observations on a particular characteristic with respect to the problem on hand.
-starting point in statistics
-analogous to sample space in probability.
Any collection of measurements capable of being described by a random variable constitutes a population.

MPTEL 3.2

We use the term population; suppose, we consider a campus with 5000 persons, we can measure all their heights, and I get 5000 numbers: $X_1, X_2, \dots, X_{5000}$; I can measure the weight, I get $Y_1, Y_2, \dots, Y_{5000}$; I can look at the income, I get another set of $I_1, I_2, \dots, I_{5000}$ numbers. I can look if they wear specs or not, I get a sequence of yes or no 5000; gender: male, female, 5000 number. So, in study of statistics, each of this is a population; this a population of heights; this a population of weights; this a population of gender, so on and so forth. So, population is synonym somewhat analogous to the sample space, that we

used in probability and defining probability. So, it is a collection of all possible observations on a particular characteristic with respect to the problem on hand. This is starting point in statistics; it analogues to sample space in probability.

(Refer Slide Time: 36:25)



Sample
In practice it is not possible to study the entire population.
Sample is a part of the population which we want to study and draw conclusions about property of population.
-it is not enough to say that sample is a subset of population; the subset needs to be representative.
Sampling: Procedure of drawing samples.
Sampling design: development of sampling procedures to meet a requirement.

Any collection of measurements capable of being described by a random variable constitutes a population. This can be taken as a working description of a population; that means, **the outcomes that we talk about should be...** the numbers we talk about, we should be able to constitute them as outcome of a random experiment and hence a random variable; that should be born in mind.

A sample in our studies, it is not practically to studied all members of the population; therefore, we talk about sample. Sample is a part of the population which we want to study and draw conclusions about property of population. We want to study average income of people in this country; one approaches to ask everyone get that number or sample take 10000 samples, a sample of 10000 persons and get that an estimate for the average, and see, how good you are in your estimate. It is not enough to say that sample is a subset of population; the subset needs to be representative. The word sampling denotes the procedure of drawing samples. And the word sampling design describes a development of a sampling procedures to meet a specific requirement.

(Refer Slide Time: 37:31)

Random sample


Let X be a random variable with pdf $p_X(x)$.

Let $\{X_i\}_{i=1}^n$ be a set of iid random variables with common pdf $p_X(x)$.

The set of random variables $\{X_i\}_{i=1}^n$ is called a random sample of size n of X .

Consider the real valued function $S(X_1, X_2, \dots, X_n)$.

This function is called a statistic. It is a random variable.

34

So, we use these terms sample sampling, sample design, etcetera. So, we need to have some working definitions or descriptions of these words. Now, the word random sample, we can give a fairly rigorous definition for that. So, let X be a random variable with probability density function $p_X(x)$, and we consider an iid sequence $X_1, X_2, X_3, \dots, X_n$, with a common probability density function $p_X(x)$; this set of random variables X_1, X_2, \dots, X_n is called a random sample of size n of X .

(Refer Slide Time: 38:23)



Let the pdf $p_X(x)$ be of the form $p_X(x; \theta)$ where $\theta =$ unknown parameter.

The joint pdf of $\{X_i\}_{i=1}^n$ is of the form

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p_X(x_i; \theta)$$

$\{x_i\}_{i=1}^n =$ values of observed data taken from the random sample.

An estimator of θ is a statistic $S(X_1, X_2, \dots, X_n)$ denoted as $\hat{\theta} = S(X_1, X_2, \dots, X_n)$





Now, if you now consider a transformation S , on these n random variables S of X_1, X_2, X_n , that is the function of X_1, X_2, X_n ; this function is called statistic. This is a random variable, because it is a transformation on X_1, X_2, X_3, X_n . We will write the probability density function of X in the form p_x of x colon θ , where θ are parameters of the probability density function; for example, in a normal random variable, we have μ and σ ; in poisson, we have λ , so on and so forth. The joint probability density function of X_i, i running from 1 to n is of the form, this is X_i 's are all independent; it is i running from 1 to n p_x of x_i θ .

(Refer Slide Time: 39:25)

For a particular set of observations
 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, the value of the estimator $S(x_1, x_2, \dots, x_n)$ is called as an estimate of θ and is denoted by $\hat{\theta}$.

Estimator: a random variable
 Estimate: the realization of the estimator.






(Refer Slide Time: 39:58)

Consider a population of four numbers $x=[1\ 2\ 3\ 4]^T$.
 Population mean=2.5

N=1		N=2		N=3		N=4	
Samples	Mean	Samples	Mean	Samples	Mean	Sample	Mean
1	1	1,2	1.5	1,2,3	2.0	1,2,3,4	2.5
2	2	1,3	2.0	1,2,4	7/3	Mean	2.5
3	3	1,4	2.5	2,3,4	3.0	Std dev	0.0
4	4	2,3	2.5	1,3,4	8/3		
Mean	2.5	2,4	3.0	Mean	2.5		
Std dev	1.29	3,4	3.5	Std dev	0.4303		
		Mean	2.5				
		Std dev	0.7071				

Sampling PDF

Now, this states X_i can be interpreted as values of observed data, taken from the random sample; that means, it is a realization of these n random variables, is actually what can be interpreted as a measurement that you make. An estimator of θ is a statistics S , θ is a parameter; so, essentially, the estimator is a random variable, which is a function of your random sample, and that we denote as θS of X_1 comma X_2 comma X_n ; so, I am describing the word estimator.

For a particular set of observations, say, X_1 is small x_1 , lower case x_2 , lower, etcetera, the value of this estimator is called an estimate of θ and is denoted by $\hat{\theta}$. So, estimator therefore is a random variable, but estimate is realization of the estimator; that means, it is a realization of this random variable. So, we will be using these terms, so you need to have some clarity on what exactly they mean. Just to fix the idea, we will consider a hypothetical problem, where there is a population with only 4 numbers, say, 1, 2, 3, 4.

Now, if you now take samples, suppose I draw only one sample, you are allowed to take only one sample and tell me what is average of this; what is average of this? This is 2.5; this is 1 plus 1 plus 3 plus 4 divided by 4 is 2.5. If you draw one sample, if you get one, you will report one as an answer; if you where to draw two report, two as an answer, so on and so forth. Now, if you are allowed to take two samples, how many such samples we can draw? 1 2, 1 3, 1 4, 2 3, 2 4, 3 4; depending on which of these samples are actually drawn, you get the estimate for mean as 1.5, 2, 2.5, 2.5, 3, 3.5, etcetera, right.

So, the estimate that we are getting is therefore random variable and the probability distribution of that is known as sampling distribution; that is, first term is sampling distribution. Now, if N equal to 3, you can now the number of ways in which sample reduces, you can draw 1 2 3, 1 2 4, 2 3 4, 1 3 4, and we get these numbers as a answers; of course, if you take N equal to 4, there is only one sample that you can draw and you get an answer 2.5; and we will now look at the last two entries in these tables.

Now, look at what happens if I take two samples, I can draw samples in 1, 2, 3, 4, 5, 6 number of x and I get these answers; these are all estimator answer to the question that we are asking; 3.5 as good as 1.5, which is as good as 2.5, because you are drawn two samples. Now, if you take average of this, I get 2.5. And you look at its standard deviation, if you are taking only one sample, the average is still 2.5, but the standard

deviation will be 1.29. I will explain how this standard deviation is computed in due course, but you can assume that it is based on these numbers. Here, the sample size has increased and standard deviation has reduced; the samples are not become 3, the mean still remains to be 2.5, but the standard deviation has becomes 0.43. In the last case, the sample size reaches its possible maximum value, and mean is 2.5 and standard deviation is 0.

(Refer Slide Time: 42:42)



Estimator: $T = \frac{1}{n} \sum_{i=1}^n X_i$

The PDF of T is known as the sampling distribution of T .
 A realization of T is known as an estimate.

The estimator is said to be unbiased if $\langle T \rangle = \text{population mean}$.

The estimator is said to be consistent if $\lim_{n \rightarrow \infty} \text{Var}(T) \rightarrow 0$.

Estimation : Finding a realization of T as an approximation to a population parameter.






(Refer Slide Time: 43:20)

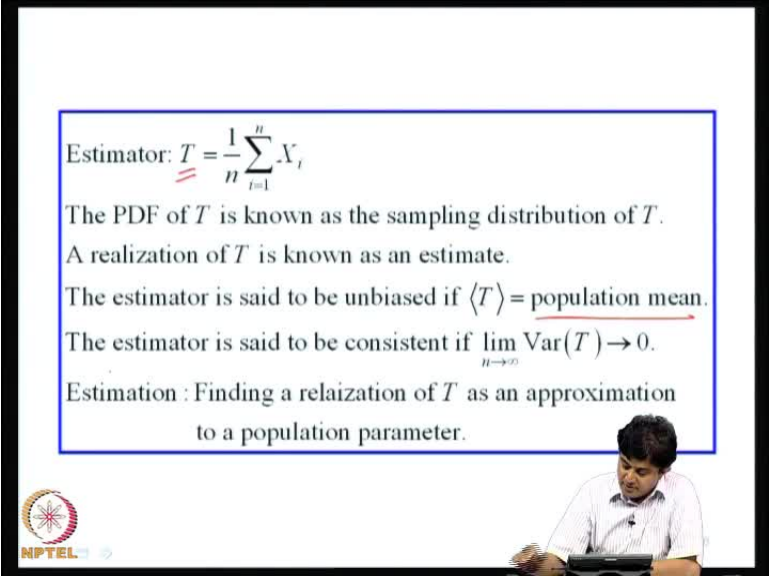
Consider a population of four numbers $x=[1\ 2\ 3\ 4]^T$.
 Population mean=2.5

N=1		N=2		N=3		N=4	
Samples	Mean	Samples	Mean	Samples	Mean	Sample	Mean
1	1	1,2 ✓	1.5 ✓	1,2,3	2.0	1,2,3,4	2.5
2	2	1,3 ✓	2.0 ✓	1,2,4	7/3	Mean	2.5
3	3	1,4 ✓	2.5 ✓	2,3,4	3.0	Std dev	0.0
4	4	2,3 ✓	2.5 ✓	1,3,4	8/3		
Mean	2.5	2,4 ✓	3.0 ✓	Mean	2.5		
Std dev	1.29	3,4 ✓	3.5 ✓	Std dev	0.4303		
		Mean	2.5				
		Std dev	0.7071				

Sampling PDF

(Refer Slide Time: 42:42)



Estimator: $T = \frac{1}{n} \sum_{i=1}^n X_i$

The PDF of T is known as the sampling distribution of T .

A realization of T is known as an estimate.

The estimator is said to be unbiased if $\langle T \rangle = \text{population mean}$.

The estimator is said to be consistent if $\lim_{n \rightarrow \infty} \text{Var}(T) \rightarrow 0$.

Estimation : Finding a realization of T as an approximation to a population parameter.

So, what is happening here? We will summarize now; if you say capital T is a estimator $\frac{1}{n} \sum_{i=1}^n X_i$, the probability distribution function of T is known as sampling distribution of T . The fact that the T is random variable is very clear, because X_1, X_2, X_3, X_n are an random variables. A realization of T is known as an estimator; this is an estimator; estimator is a random variable; estimate is a realization. The estimator is said to be unbiased, if the expected value of this random variable T is actually equal to the population mean. Is it true here? Yes. As we saw here, if you take the mean of this random variable, I am getting 2.5, which is indeed the population mean which is not known to me. We say that, such an estimator is unbiased; an estimator is unbiased, if its expected value is exactly equal to the population quantity that you are trying to estimate.

(Refer Slide Time: 44:20)

Consider a population of four numbers $x=[1\ 2\ 3\ 4]^T$.
Population mean=2.5

N=1		N=2		N=3		N=4	
Samples	Mean	Samples	Mean	Samples	Mean	Sample	Mean
1	1	1,2	1.5	1,2,3	2.0	1,2,3,4	2.5
2	2	1,3	2.0	1,2,4	7/3	Mean	2.5
3	3	1,4	2.5	2,3,4	3.0	Std dev	0.0
4	4	2,3	2.5	1,3,4	8/3		
Mean	2.5	2,4	3.0	Mean	2.5		
Std dev	1.29	3,4	3.5	Std dev	0.4303		
		Mean	2.5				
		Std dev	0.7071				

Sampling PDF

Now, one more thing that we saw was, as n was increasing, a sample size was increasing, the standard deviation of the estimator was reducing. So, the estimator in that case is said to be consistent. So, the estimator is said to be consistent, if as limit n tending to infinity variance of T goes to 0. So, that means, you draw more samples, you are more sure of your answer; that dispersion in your answer goes on reducing. Estimation is a finding a realization of T as an approximation to a population parameter; that would mean, what it involves, this the population, suppose with N equal to 2, I take the number say 2 and 4 and compute this 2 plus 4 divided by 2 which is 3 as an approximation to 2.5 which is not known to me; I take 3 as an approximation to 2.5. This process of drawing a sample and computing, this is known as estimation.

(Refer Slide Time: 44:44)

Estimation of mean

Let X be a random variable with PDF $P_X(x)$, pdf $p_X(x)$, mean μ , and standard deviation σ .

Let $\{X_i\}_{i=1}^n$ be an iid sequence with common pdf $p_X(x)$.

That is, $X_i \perp X_j \forall i \neq j \in [1, n]$.

$\langle X_i \rangle = \mu$, $Var [X_i] = \sigma^2$, $p_{X_i}(x) = p_X(x) \forall i \in [1, n]$.

Let $\Theta = \sum_{i=1}^n a_i X_i$ be an estimator of μ .

The estimator is said to be unbiased if $\langle \Theta \rangle = \mu$.

$$\langle \Theta \rangle = \left\langle \sum_{i=1}^n a_i X_i \right\rangle = \mu \sum_{i=1}^n a_i = \mu$$

$\sum_{i=1}^n a_i = 1 \Rightarrow \Theta$ is an unbiased estimator of μ .

Now, let us consider the problem of estimation of mean. So, let X be a random variable with probability distribution function $P_X(x)$, probability density function lower case $p_X(x)$, a mean μ and standard deviation σ . We will form a sequence of identical, independent and identically distributed random variables X_i , i running from 1 to n with common PDF $P_X(x)$ which is given here, in the specification of X . What this means is, X_i is independent of X_j , for all i not equal to j in 1 to n . Since X_i 's are all identical, expected value of X_i is μ , variance of X_i is σ^2 and $p_{X_i}(x) = p_X(x)$, for all i 1 to n , because there are iid's.

(Refer Slide Time: 46:30)

$$\sum_{i=1}^n a_i = 1 \Rightarrow \Theta \text{ is an unbiased estimator of } \mu.$$

$$\Rightarrow \text{The above unbiased estimator is not unique.}$$

$$\text{Var}(\Theta) = \text{Var}\left\{\sum_{i=1}^n a_i X_i\right\} = \left\langle \left\{\sum_{i=1}^n a_i (X_i - \mu)\right\}^2 \right\rangle = \sigma^2 \sum_{i=1}^n a_i^2$$
 To get an unbiased estimator with minimum variance, we minimize

$$\text{Var}(\Theta) = \sigma^2 \sum_{i=1}^n a_i^2 \text{ subject to the constraint } \sum_{i=1}^n a_i = 1.$$
 Lagrangian

$$L = \sigma^2 \sum_{i=1}^n a_i^2 + \lambda \left\{ \sum_{i=1}^n a_i - 1 \right\}$$

Now, let theta, i running from 1 to n, a_i x i be an estimator of mu; mu is a property of the population which is not known; theta is an estimator. Now, estimator is said to be unbiased, if expected value of theta is equal to mu. When that will happen? You consider expected value of theta; this is expected value of i running from 1 to n a_i X I, but expected value of X_i is mu; therefore, I get this mu i equal to 1 to n. If this has to be equal to mu, it means the summation of this a_i must be equal to 1; that means, if you select an estimator, where the weights a_i added to 1, then the resulting estimator is unbiased. Unbiased is a desirable property, we can start with that notion, but this is not unique; there are so many ways in which I can select a I, so that i running from 1 to n, a_i is equal to one. Now, what I will do now is, I will consider variance of theta; variance of theta is variance of this summation, and since X_i are all independent, variance is a_i X i minus mu summed over i to 1 square and we can show that variance indeed this.

Now, what we do is, to make the unbiased estimator unique, we will select a i show that it minimizes this variance. Since theta is being used as an approximation to the population description mu, say, mean, population mean is not random, it is a deterministic quantity. So, you are approximating a deterministic quantity via random variable. So, the best that you can do is to minimize the variance of random variable; that is the logic beyond minimizing the variance here. So, variance of theta is given this; I want to minimize variance of theta given by this; subject to the constraint that I want an

unbiased estimator. So, what we do? We form the langrangian sigma square, i equal to 1 to n a i square plus lambda; this is the constraint equation, i equal to 1 to n a i minus 1.

(Refer Slide Time: 48:14)

Necessary conditions for optima

$$\frac{\partial L}{\partial a_k} = 0 \Rightarrow 2\sigma^2 a_k + \lambda = 0; k = 1, 2, \dots, n. \Rightarrow a_k = -\frac{\lambda}{2\sigma^2}.$$

$$\frac{\partial L}{\partial \lambda} = 0 \Rightarrow \sum_{i=1}^n a_i = 1 \Rightarrow \sum_{i=1}^n \left(-\frac{\lambda}{2\sigma^2}\right) = 1$$

$$\lambda = -\frac{2\sigma^2}{n} \Rightarrow a_k = \left(-\frac{1}{2\sigma^2}\right) \left(-\frac{2\sigma^2}{n}\right) = a_k = \frac{1}{n}$$

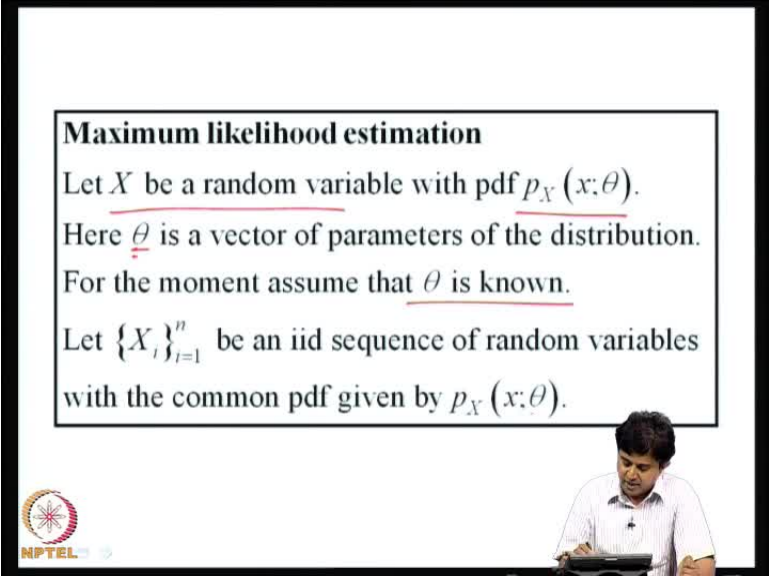
The optimal $\text{Var}(\Theta) = \sigma^2 \sum_{i=1}^n \frac{1}{n^2} = \frac{\sigma^2}{n}.$

Summary:

$\Theta = \frac{1}{n} \sum_{i=1}^n X_i$ is an unbiased estimator of μ with minimum variance and the lowest variance is $\frac{\sigma^2}{n}.$

Now, for optimality, $\frac{\partial L}{\partial a_i}$ must be equal to 0 for i equal to 1 and $\frac{\partial L}{\partial \lambda}$ must be equal to 0, and that basically that condition optimality of L with respect to λ essentially gives our constraint equation; that is the idea behind doing introducing langrangian multiplier. So, we go through these calculations, $\frac{\partial L}{\partial a_k} = 0$, would mean $2\sigma^2 a_k + \lambda = 0$, for k running from 1 to n , and from this, we get a_k to be minus λ by $2\sigma^2$; **do λ by λ** $\frac{\partial L}{\partial \lambda} = 0$, I can recover the constraint equation. And in this if I put this a_k as minus λ by $2\sigma^2$, I can eliminate, that an I can get λ in terms of σ^2 and n as shown here, and I get a_k to be $1/n$; that means, all a_1, a_2, a_3, a_n , there are all $1/n$. If n is a sample size, all a_k 's are $1/n$; therefore, **the optimal** corresponding to this optimal values, the minimum variance that I get is σ^2/n .

(Refer Slide Time: 49:45)



Maximum likelihood estimation

Let X be a random variable with pdf $p_X(x;\theta)$.

Here θ is a vector of parameters of the distribution.

For the moment assume that θ is known.

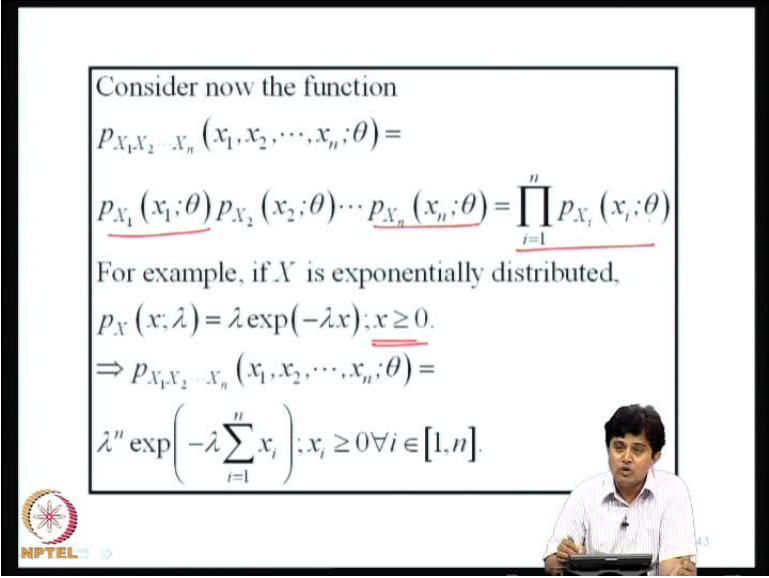
Let $\{X_i\}_{i=1}^n$ be an iid sequence of random variables with the common pdf given by $p_X(x;\theta)$.

MPTEL

So, to summarize, \bar{X} is equal to $\frac{1}{n} \sum_{i=1}^n X_i$, the X_i 's are all 1 by n , they come out of the summation; \bar{X} is an unbiased estimator of μ with minimum variance, and the lowest variance is $\frac{\sigma^2}{n}$. So, whenever we are given a set of numbers and ask to find an average, this is what we intuitively do, and that actually has these desirable properties that that procedure leads to an estimate, which is an estimator which is unbiased and has its smallest variance.

Now, another approach to parameter estimation is what is known as maximum likelihood estimation. Let X be a random variable with pdf $p_X(x;\theta)$, and θ is a vector of parameters of the distribution; for example, if X is normal, I have σ and μ as parameters. For the moment, let us assume that θ is known, and let X_i be an iid sequence of random variables with the common pdf given by $p_X(x;\theta)$; this θ is given now.

(Refer Slide Time: 50:20)



Consider now the function

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta) =$$
$$p_{X_1}(x_1; \theta) p_{X_2}(x_2; \theta) \cdots p_{X_n}(x_n; \theta) = \prod_{i=1}^n p_{X_i}(x_i; \theta)$$

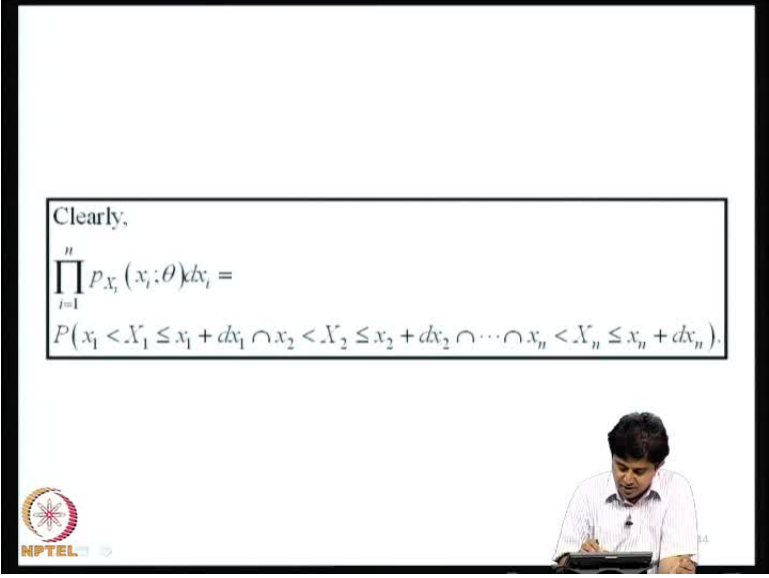
For example, if X is exponentially distributed,

$$p_X(x; \lambda) = \lambda \exp(-\lambda x); x \geq 0.$$
$$\Rightarrow p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta) =$$
$$\lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right); x_i \geq 0 \forall i \in [1, n].$$

NPTEL

Now, let us consider the joint density function of x_1, x_2, \dots, x_n and that is given by, since X_i 's are all independent, it is p_{X_1} evaluated x_1 , p_{X_2} evaluated x_2 , so on, and so forth. And using the product notation, I write it as product i equal to 1 to n $p_{X_i}(x_i; \theta)$. Now, just for an illustration, let x be exponentially distributed, so that $p_X(x; \lambda)$ is only one. Now, this is λ , I get $\lambda \exp(-\lambda x)$, this is the probability - common probability density function - and this product will be $\lambda^n \exp(-\lambda \sum_{i=1}^n x_i)$, where x_i 's are all greater than 0, and i runs from 1 to n ; here, I have already said that θ is taken to be known.

(Refer Slide Time: 51:14)

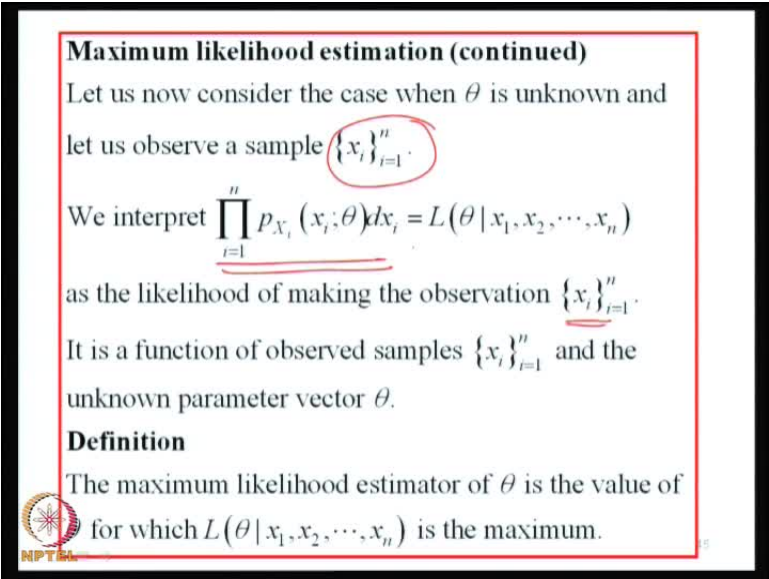


Clearly,

$$\prod_{i=1}^n p_{X_i}(x_i; \theta) dx_i = P(x_1 < X_1 \leq x_1 + dx_1 \cap x_2 < X_2 \leq x_2 + dx_2 \cap \dots \cap x_n < X_n \leq x_n + dx_n).$$

MPTEL

(Refer Slide Time: 51:35)



Maximum likelihood estimation (continued)

Let us now consider the case when θ is unknown and let us observe a sample $\{x_i\}_{i=1}^n$.

We interpret $\prod_{i=1}^n p_{X_i}(x_i; \theta) dx_i = L(\theta | x_1, x_2, \dots, x_n)$ as the likelihood of making the observation $\{x_i\}_{i=1}^n$.

It is a function of observed samples $\{x_i\}_{i=1}^n$ and the unknown parameter vector θ .

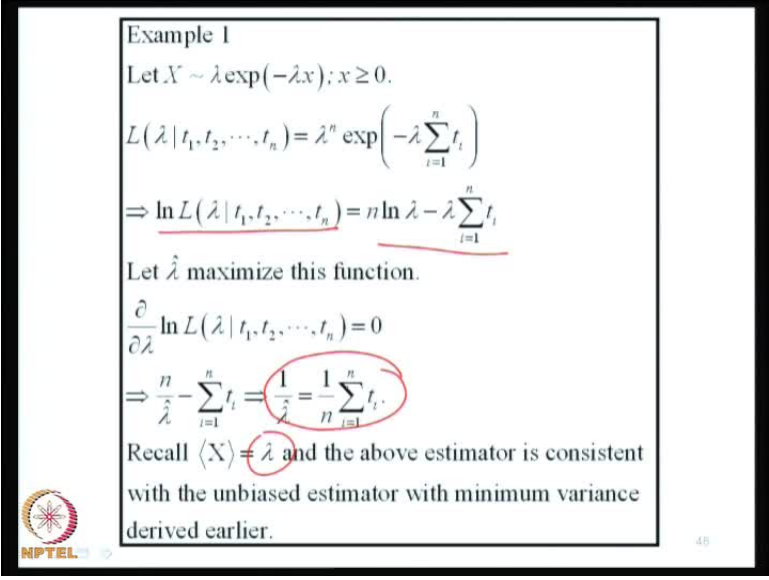
Definition

The maximum likelihood estimator of θ is the value of θ for which $L(\theta | x_1, x_2, \dots, x_n)$ is the maximum.

MPTEL

What is this product? Product of density functions, suppose I multiply this by dx_i , this product is nothing but the probability of X_1 lying between X_1 plus dx_1 intersection X_2 lying between X_2 and $X_2 + dx_2$, so on and so forth, right. Now, let us now consider the case, where θ is unknown and let us assume that we have observed a sample denoted by lower case $x_1, x_2, x_3, \dots, x_n$. Now, what I do? I interpret this product as the likelihood of making the observation x_i , because of this property, that this product is indeed the probability. So, the word likelihood is being used in the sense of a probability.

(Refer Slide Time: 52:46)



Example 1


Let $X \sim \lambda \exp(-\lambda x); x \geq 0$.

$$L(\lambda | t_1, t_2, \dots, t_n) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n t_i\right)$$
$$\Rightarrow \ln L(\lambda | t_1, t_2, \dots, t_n) = n \ln \lambda - \lambda \sum_{i=1}^n t_i$$

Let $\hat{\lambda}$ maximize this function.

$$\frac{\partial}{\partial \lambda} \ln L(\lambda | t_1, t_2, \dots, t_n) = 0$$
$$\Rightarrow \frac{n}{\lambda} - \sum_{i=1}^n t_i = \frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^n t_i$$

Recall $\langle X \rangle = \frac{1}{\lambda}$ and the above estimator is consistent with the unbiased estimator with minimum variance derived earlier.


 46

Now, theta is unknown; so, it is a function of observed samples and the unknown parameter theta. Now, for what value of theta, this likelihood is maximum is a way to find theta now. So, what we do is, we introduce the definition; the maximum likelihood estimator of theta is the value of theta, for which this L theta condition a x 1, x 2, x n is the maximum, and this L is known as a likelihood function, which is the product of the probability density function evaluated at the sample points - observed sample points. Just to give an example, let X be a exponential random variable and suppose I have made observations t 1, t 2, t n, so the L function will be the likelihood function will be lambda to the power of n exponential minus lambda, i running from 1 to n t i. We have to maximize this; so, we take logarithm; so, this is known as log likelihood function; so, this we get as this.

(Refer Slide Time: 54:02)

Example 2

Let $X \sim N(\mu, \sigma)$.

$$L(\mu, \sigma | t_1, t_2, \dots, t_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{t_i - \mu}{\sigma}\right)^2\right]$$
$$L(\mu, \sigma | t_1, t_2, \dots, t_n) = \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^n \exp\left[-\frac{1}{2}\sum_{i=1}^n \left(\frac{t_i - \mu}{\sigma}\right)^2\right]$$
$$\ln L(\mu, \sigma | t_1, t_2, \dots, t_n) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2} \sum_{i=1}^n \left(\frac{t_i - \mu}{\sigma}\right)^2$$


47

Now, let lambda had maximize this function; so, that means, do by dou lambda L n of this is 0. And if I do now the calculation, first term gives n by lambda, the other one gives summation t i; and from this, I get 1 by lambda as 1 by n, i running from 1 to n t i. Now, quickly recall, what was an expected value of X? This is consistent with our definition of expected value of X; that means, if you observed t 1 t 2 through t n and find the unbiased estimator with minimum variance, you will get the same formula for the expected value of X. So, this is consistent with that. How about normal random variable? It has two parameters: mu and sigma. So, the likelihood function is product of these density functions evaluated at these observed points t 1, t 2, t 3, t n and this is the likelihood function. And if I take logarithm, it becomes, I mean, it simplifies the matter a bit and our parameters of interest are buried here, and here and here. So, I have to minimize this maximize this with respect to mu and sigma.

(Refer Slide Time: 54:38)

$$\ln L(\mu, \sigma | t_1, t_2, \dots, t_n) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2} \sum_{i=1}^n \left(\frac{t_i - \mu}{\sigma} \right)^2$$

Let $\hat{\mu}$ & $\hat{\sigma}$ maximize the above function.

$$\Rightarrow \frac{\partial}{\partial \mu} \ln L(\mu, \sigma | t_1, t_2, \dots, t_n) = 0$$
$$\& \frac{\partial}{\partial \sigma} \ln L(\mu, \sigma | t_1, t_2, \dots, t_n) = 0$$
$$\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n t_i$$
$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - \hat{\mu})^2}$$

The slide includes an NPTEL logo in the bottom left corner and a small inset image of a person in the bottom right corner.

(Refer Slide Time: 55:03)

Sampling distribution for the estimator of mean

Consider the estimator $\Theta = \frac{1}{n} \sum_{i=1}^n X_i$

Θ is an unbiased estimator of μ with variance $\frac{\sigma^2}{n}$.

Let us consider the case in which σ^2 is known.

If X is Gaussian, it would mean that $\{X\}_{i=1}^n$ is an iid sequence of Gaussian random variables and consequently Θ would also be Gaussian distributed.

If X is not Gaussian, by virtue of central limit theorem, for large n , we may still consider Θ to be Gaussian.

It may be inferred that $\Theta \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ or, $\frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$.

The slide includes an NPTEL logo in the bottom left corner and a small inset image of a person in the bottom right corner.

So, if I do that, I use dou by dou mu equal to 0, dou by dou sigma of this L n function is 0; and I get this as the maximum likelihood estimator for the mean, maximum likelihood estimator for standard deviation. So, these are yet another approach for finding parameters of a probability density function, based on observations. Now, I have mentioned that estimator is the random variable; to be able to assess how good is our estimation, we should characterize the probability distribution of that estimator.

So, just to give an introduction to that, if you consider the estimator θ as $\frac{1}{n} \sum_{i=1}^n X_i$; θ as we know is an unbiased estimator of μ with variance as this. Now, if X is Gaussian, X_i will also be Gaussian; and you are adding gaussian random numbers, so θ will be Gaussian. So, sampling distribution for θ as per this description is gaussian with mean as μ and variance as σ^2/n . If X is not gaussian by using central limit theorem for large samples n being large, we can still make a have postulation that, sampling distribution for θ is still Gaussian.

Now, based on our knowledge of probability distribution of θ , what can we say about the accuracy of estimation? This is a question we will consider in the next lecture; that is, how to get the sampling distribution for different estimators, say, estimators for mean, estimators for variance, estimator for correlation function, estimator for power spectral density, so on and so forth? That is one exercise. After having got that, how do you use that information in assessing the accuracy of estimates? We will consider these questions in the next lecture; we will conclude this lecture at this stage.