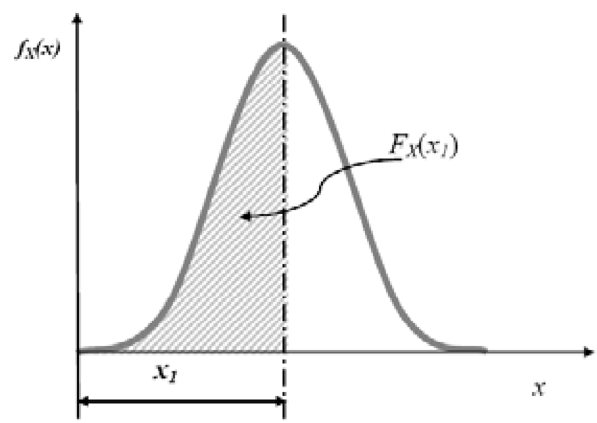
Probabilistic Description

Characteristics of Probability Distribution

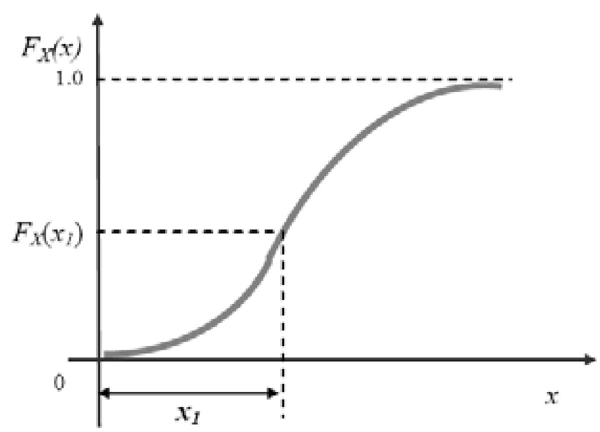
Random Variable

A random variable X takes on various values x within the range $-\infty < x < \infty$. A random variable is denoted by an uppercase letter, and its particular value is represented by a lowercase letter. Random variables are of two types: discrete and continuous. If the random variable is allowed to take only discrete values, $x_1, x_2, x_3 ...., x_n$, it is called a *discrete random variable*. On the other hand, if the random variable is permitted to take any real value within a specified range, it is called a *continuous random variable*.

Probability Density and Cumulative Distribution Function



(a) Probability Density Function



(b) Cumulative Distribution Function

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds$$

$$F_X(b) - F_X(a) = \int_a^b f_X(x)dx \qquad \text{(for all real numbers } a \text{ and } b\text{)}$$

Joint Density and Distribution Functions

$$P[a<X<b,\ c<Y<d] = \int_c^d \int_a^b f_{XY}(x,y)dxdy$$

Central Measures

The population *mean*, also referred to as the *expected value* or *average*

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x\ f_X(x)dx$$

It is called the *first moment* The mean is analogous to the centroidal distance of a cross-section.

if $g(x)$ is an arbitrary function of x,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

The *expectation operator*, $E[.]$, possesses the following useful properties: If X and Y are independent,

$$E[XY] = E[X]E[Y]$$

$$E[c] = c$$

$$E[cX] = cE[X]$$

$$E[g(X)] \neq g(E[X])$$

$$E(Z) = E(X_1) + E(X_2) ..... + E(X_n)$$

Other useful central measures are the *median* and *mode* of the data.

Dispersion Measures

The *variance*, $V(X)$, a second central moment of X,

$$\begin{aligned} V(X) &= E[(X-\mu_X)^2] \\ &= E(X^2) - 2E(X)\mu_X + \mu_X^2 = E(X^2) - \mu_X^2 \end{aligned}$$

it represents the moment of inertia of the probability density function about the mean value. A measure of the variability of the random variable is usually given by a quantity known as the *standard deviation*

$$\sigma_X = \sqrt{V(X)}$$

Nondimensionalizing the standard deviation will result in the *Coefficient of Variation* (COV), δ_X, which indicates the relative amount of uncertainty or randomness

$$\delta_X = \frac{\sigma_X}{\mu_X}$$

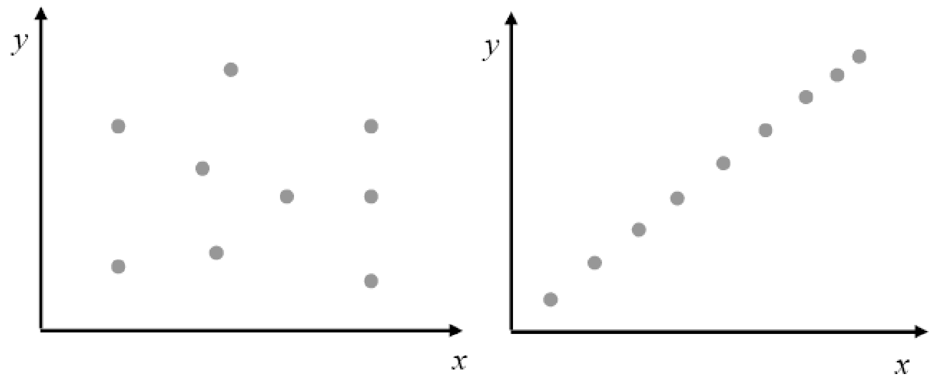
If two random variables (X and Y) are correlated, the likelihood of X can be affected by the value taken by Y. In this case, the *covariance*, σ_{XY}, can be used as a measure to describe a linear association between two random variables

$$\begin{aligned} \sigma_{XY} &= Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{XY}(x,y)dxdy \end{aligned}$$

The *correlation coefficient* is a nondimentional measure of the correlation

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

If x and y are statistically independent, the variables are uncorrelated and the covariance is 0 (a). Therefore, the correlation coefficients of ±1 indicate a perfect correlation (b)



(a) Covariance near Zero (b) Positive Covariance

If $Y = a_1 X_1 + a_2 X_2$, where a_1 and a_2 are constants, the variance of Y can be obtained as

$$Var[Y] = E[\{a_1 X_1 + a_2 X_2 - (a_1 \mu_{X_1} + a_2 \mu_{X_2})^2]$$

$$= E[\{a_1 (X_1 - \mu_{X_1}) + a_2 (X_2 - \mu_{X_2})\}^2]$$

$$= E[a_1^2 (X_1 - \mu_{X_1})^2 + 2a_1 a_2 (X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) + a_2^2 (X_2 - \mu_{X_2})^2]$$

$$= a_1^2 Var[X_1] + a_2^2 Var[X_2] + 2a_1 a_2 Cov(X_1, X_2)$$

Other Measures

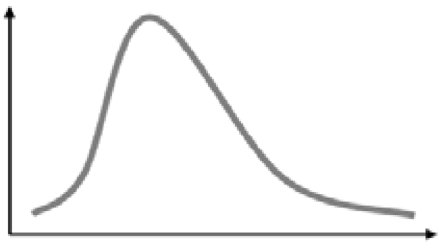
the third moment of the distribution about the mean) is taken as a measure of the skewness, or lack of symmetry

$$skewness = E[(X - \mu_x)^3] = \int_{-\infty}^{\infty} (X - \mu_x)^3 f_X(x) dx$$

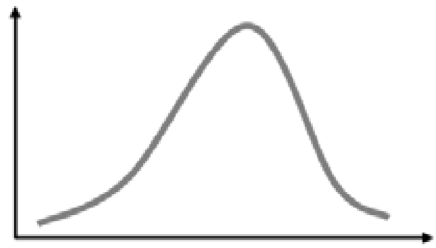
The value of $E[(X - \mu_X)^3]$ can be positive or negative.

A nondimensional measure of skewness known as the *skewness coefficient* is denoted as

$$\theta_X = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}$$



(a) Positively Skewed



(b) Negatively Skewed

The *kurtosis*, the fourth central moment of *X*, is a measure of the flatness of a distribution

$$kurtosis = \frac{E[(X-\mu_X)^4]}{\sigma_X^4}$$

An alternative definition of the kurtosis is given by

$$kurtosis = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu_X)}{\sigma_X}^4 - 3$$

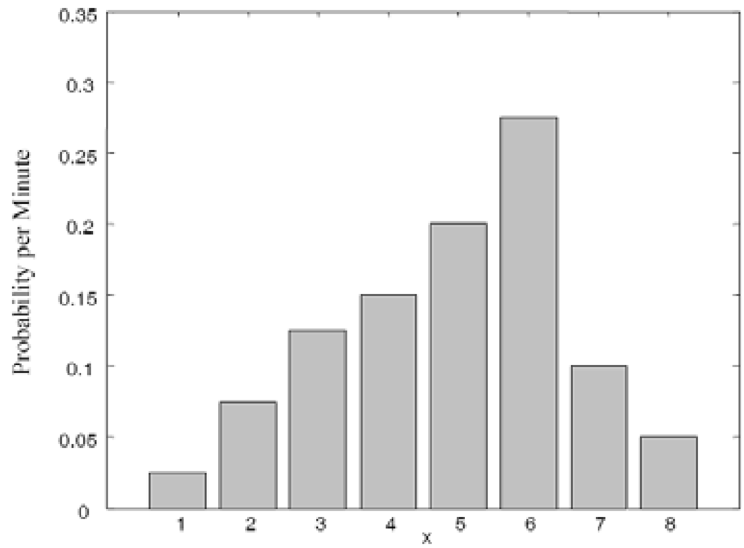
In this definition, the kurtosis of the normal distribution is zero, a positive value of the kurtosis describes a distribution that has a sharp peak, and a negative value of the kurtosis indicates a flat distribution compared to the normal distribution.

Example 1

The probability that a given number of cars per minute will arrive at a tollbooth is given in the table below. (a) Sketch the probability distribution as a function of X and find the mean, median, and mode. (b) Determine $E(X^2)$ and $E(X^3)$, the standard deviation, and the skewness coefficient.

No. of cars arriving per minute (X)	1	2	3	4	5	6	7	8
Probability per minute	0.025	0.075	0.125	0.150	0.200	0.275	0.100	0.050

(a)



Mean:

$$\mu = \sum_{i=1}^{8} x_i P_i = 1(0.025) + 2(0.075) + 3(0.125) + 4(0.150)$$
$$+5(0.2) + 6(0.275) + 7(0.10) + 8(0.05) = 4.9$$

Mode: The peak in the probability density function is at $x = 6$, therefore this is the mode.

Median: Examination of the data shows that the cumulative probability of 0.5 lies between 4 and 5 cars per minute. A quadratic interpolation of CDF using 4, 5, and 6 provides a value of 4.75.

(b) $$E(X^2) = \sum_{i=1}^{8} x_i^2 P_i = 1(0.025)+4(0.075)+9(0.125)+16(0.150)$$
$$+25(0.2)+36(0.275)+49(0.10)+64(0.05) = 26.85$$

$$E(X^3) = \sum_{i=1}^{8} x_i^3 P_i = 1(0.025)+8(0.075)+27(0.125)+64(0.150)+125(0.2)$$
$$+216(0.275)+ 343(0.10)+512(0.05) = 157.9$$

$$\sigma_X^2 = \sum_{i=1}^{8} (x_i - \mu)^2 P_i = E(X^2) - \mu_X^2 = 26.85 - 4.9^2 = 2.84$$

$$\sigma_X = \sqrt{2.84} = 1.69$$

$$\theta_X = \frac{E[(x-\mu)^3]}{\sigma_X^3} = \frac{\sum_{i=1}^{8} (x_i - \mu)^3 P_i}{\sigma_X^3} = \frac{-1.497}{1.69^3} = -0.313$$

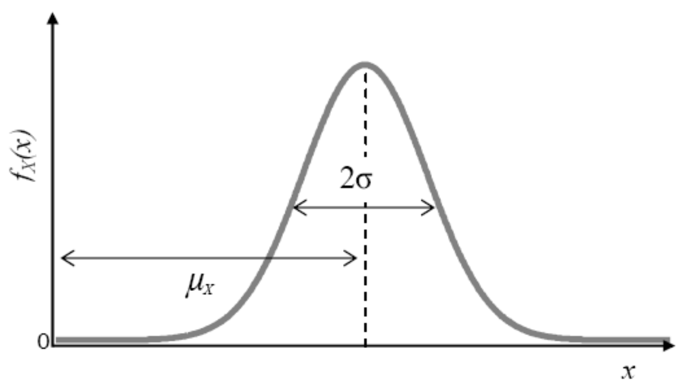
Common Probability Distributions

Gaussian Distribution

is used in many engineering and science fields due to its simplicity and convenience. often used for small coefficients of variation cases, such as Young's modulus, Poisson's ratio, and other material properties.

The Gaussian distribution is given by

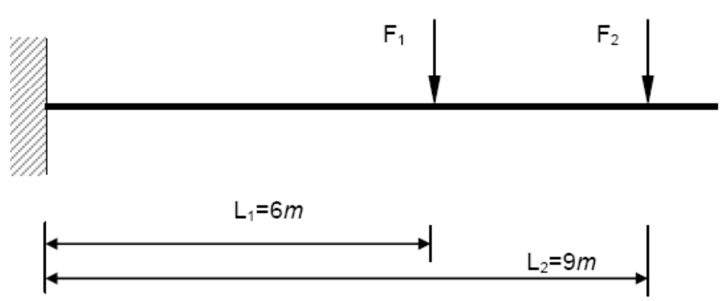
$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right], \quad -\infty < x < \infty$$



The areas under the curve within one, two, and three standard deviations are about 68%, 95.5%, and 99.7% of the total area, respectively.

Example 2

If a cantilever beam supports two random loads with means and standard deviations of $\mu_1 = 20\,\text{kN}, \sigma_1 = 4$ kN and $\mu_2 = 10$ kN, $\sigma_2 = 2$ kN as shown in the accompanying drawing, the bending moment (M) and the shear force (V) at the fixed end due to the two loads are $M = L_1F_1 + L_2F_2$ and $V = F_1 + F_2$, respectively.



(a) If two loads are independent, what are the mean and the standard deviation of the shear and the bending moment at the fixed end?

$$V = F_1 + F_2 \Rightarrow E[V] = \mu_V = E[F_1] + E[F_2] = 20 + 10 = 30 \text{ kN}$$

$$Var[V] = Var[F_1] + Var[F_2] + 2Cov(F_1, F_2)$$
$$= 4^2 + 2^2 + 0 = 20 \text{ kN}^2$$

$$\therefore \sigma_V = \sqrt{20} = 4.47 \text{ kN}$$

$$M = L_1 F_1 + L_2 F_2 \Rightarrow$$
$$E[M] = \mu_M = L_1 E[F_1] + L_2 E[F_2] = 6 \times 20 + 9 \times 10 = 210 \text{ kNm}$$
$$Var[M] = L_1^2 Var[F_1] + L_2^2 Var[F_2] + 2L_1 L_2 Cov(F_1, F_2)$$
$$= 6^2 \times 4^2 + 9^2 \times 2^2 + 0 = 900 \text{ kNm}^2$$
$$\therefore \sigma_M = \sqrt{900} = 30 \text{ kNm}$$

(b) If two random loads are normally distributed, what is the probability that the bending moment will exceed 235 kNm?

$$P(M > 235) = P\left(\xi > \frac{235 - 210}{30}\right)$$
$$= P(\xi > 0.8333) = 1 - \Phi(0.83) = 0.2023$$

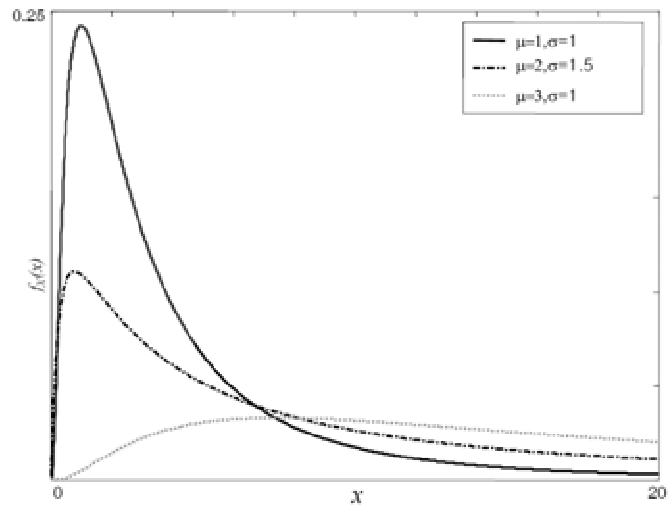
(c) If two loads are independent, what is the correlation coefficient between V and M ?

$$\sigma_{VM} = Cov(V, M)$$
$$= E[(V - \mu_V)(M - \mu_M)] = E[VM] - \mu_V \mu_M$$
$$= E[(F_1 + F_2)(L_1 F_1 + L_2 F_2)] - (\mu_1 + \mu_2)(L_1 \mu_1 + L_2 \mu_2)$$
$$= L_1 E[F_1^2] + L_2 E[F_2^2] + (L_1 + L_2) E[F_1 F_2]$$
$$- L_1 \mu_1^2 - L_2 \mu_2^2 - (L_1 + L_2) \mu_1 \mu_2$$

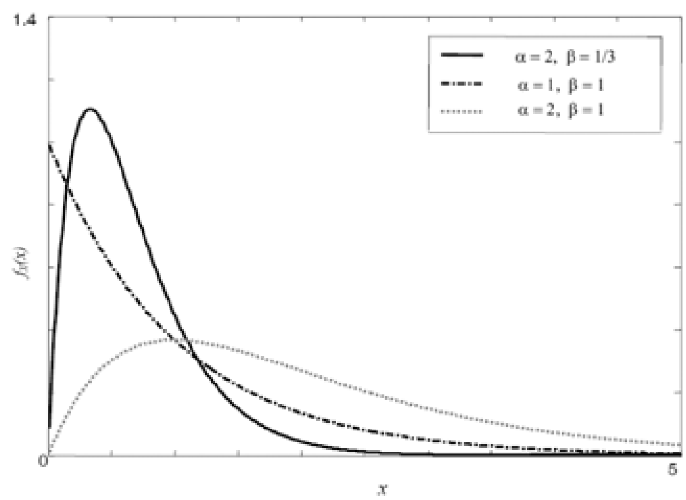
$$= L_1 \sigma_1^2 + L_2 \sigma_2^2$$

$$\rho_{VM} = \frac{\sigma_{VM}}{\sigma_V \sigma_M} = \frac{L_1 \sigma_1^2 + L_2 \sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}\sqrt{L_2^2 \sigma_1^2 + L_2^2 \sigma_2^2}} = 0.98387$$

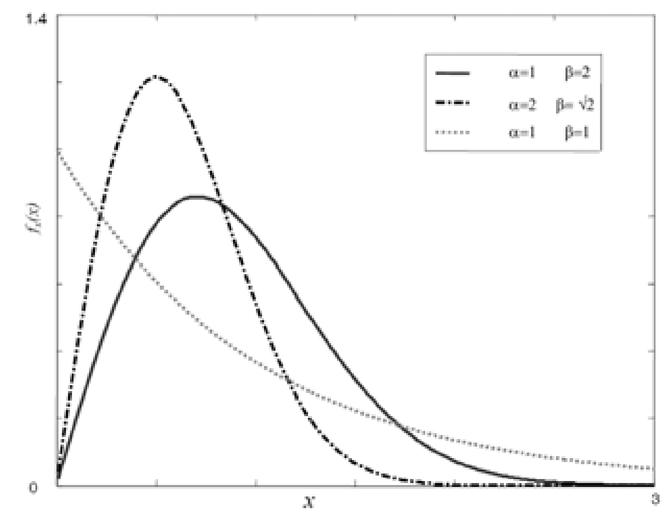
Lognormal Distribution



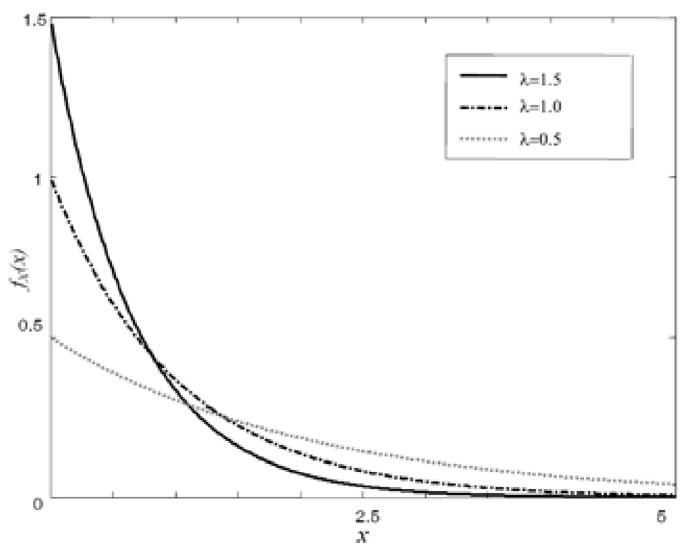
Gamma Distribution



Weibull distribution

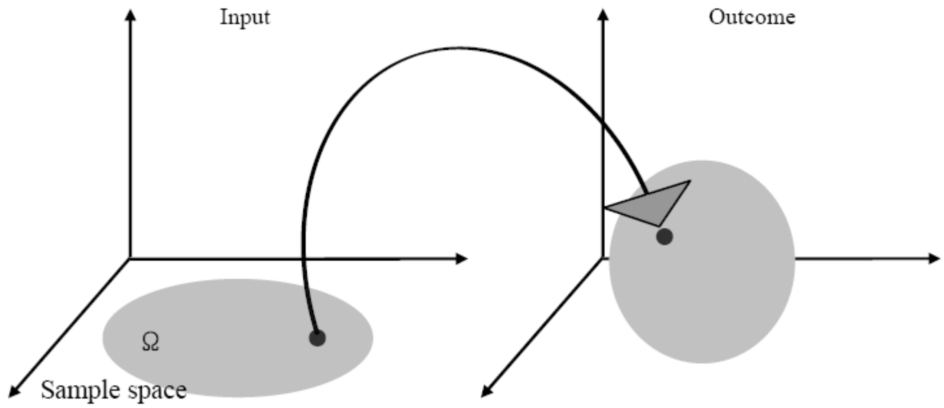


Exponential Distribution



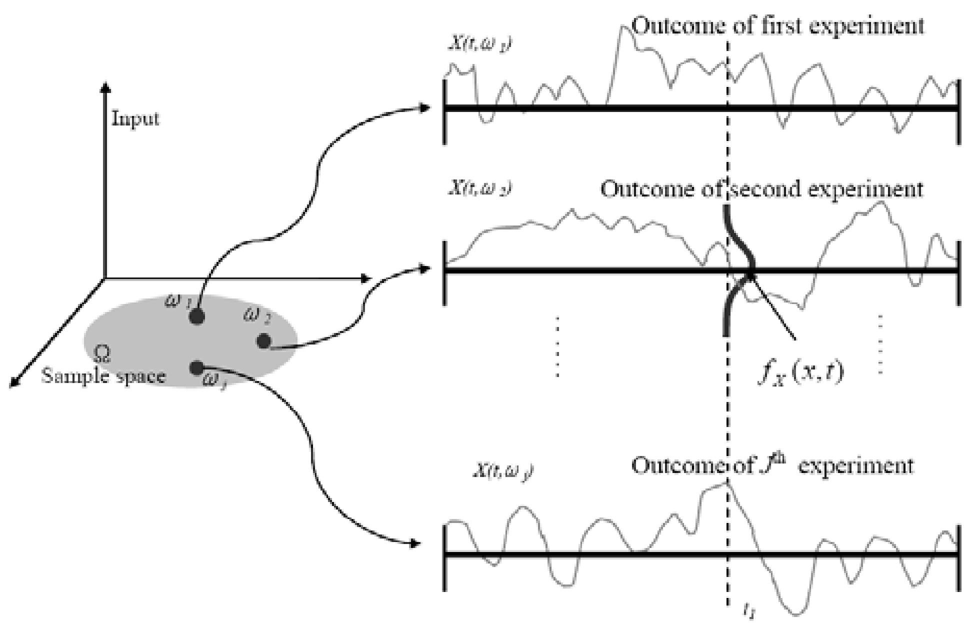
Random Field

traditional deterministic analysis, such as the finite element method, uses a single design point, considering it sufficient to represent the response



(a) Deterministic Concept

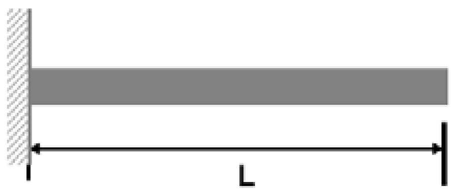
The basic idea of the random process is that the outcome of each experiment is a function over an interval of the domain rather than a single value.



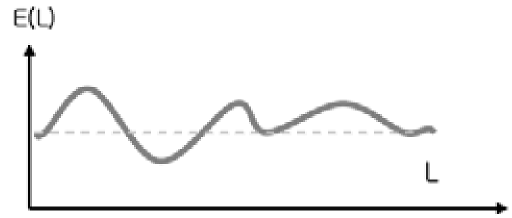
(b) Random Process Concept

Consider a simple cantilever beam, with its Young’s modulus, *E*, fluctuating over the length of the beam

Obviously, the fluctuation of the Young’s modulus should be considered in the analysis process.



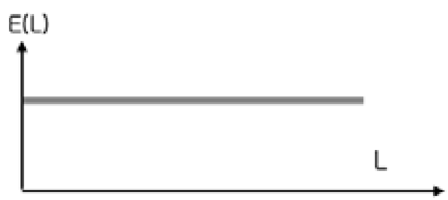
(a) Cantilever Beam



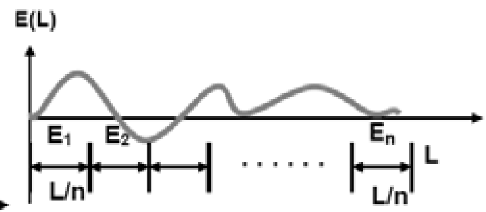
(b) Fluctuation of Young’s Modulus

To do this, the *random field discretization* is used to describe the spatial variability of the stochastic structural properties over the structure.

the mean part and the fluctuation part



(c) Mean Part



(d) Fluctuation Part

Fitting Regression Models

Regression analysis is the investigation of the functional relationship between two or more variables.

There are two types of regression analysis: linear and nonlinear.

[refer to www.eng.ukm.my/kamal/cm/Chapter3.pdf]