# **1. COMMON PROBABILITY DISTRIBUTIONS IN ENGINEERING**

# **1.1 Objectives**

- To understand probability distributions relevant to engineering applications
- To investigate statistical fundamentals of probability distributions
- To make use of Matlab statistical toolbox

#### **1.2 Types of Probability Distributions**

Let X be a random variable in an engineering application. The probability density function (pdf) and cumulative distribution function (cdf) of X are denoted by  $f_X$  and  $F_X$ , respectively.

,

# 1.2.1 Normal Distribution

$$y(x;\mu,\sigma) = f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
  
>> x=[-10:0.1:10];  
>> y=normpdf(x,0,1);  
>> plot(x,y)  
• Symmetric distribution, skewness=0, kurtosis=3  
• Central limit theorem states that any  
distribution with finite mean and standard  
deviation tends to follow normal distribution  
• Special case of chi-squared distribution and  
gamma distribution

• Dimension of fabricated part

# 1.2.2 Lognormal Distribution

$$y(x; \mu, \sigma) = f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$





#### 1.2.3 Weibull Distribution

$$y(x;v,k,a) = f_X(x) = \frac{k-a}{v-a} \left(\frac{x-a}{v-a}\right)^{k-1} e^{-\left(\frac{x-a}{v-a}\right)^k}, \quad 2 \text{ parameter Weibull if } a$$

>> x=[0:0.1:10]; >> y=weibpdf(x,1,2); >> plot(x,y)

- Originally proposed for fatigue life
- Used in analysis of systems with weakest link
- Wear, fatigue, and fracture

#### 1.2.4 Exponential Distribution

$$y(x; \mu) = f_X(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$$

>> x=[0:0.1:10]; >> y=exppdf(x,1);

>> plot(x,y)

- Amount of time between occurrences
- Called as "memoryless random distribution"
- Continuous version of Poisson distribution to describe the number of occurrences per unit time
- Mean time between failures



= 0



#### 1.2.5 Poisson Distribution (Discrete)

$$y(x;\lambda) = f_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}$$

>> x=[0:0.1:10];

>> y=poisspdf(x,1);

>> plot(x,y)

- An event occurrence in a given interval
- The occurrences are independent
- Average number of occurrence is fixed
- Product defections in a given batch

#### 1.2.6 Uniform Distribution

$$y(x;a,b) = f_X(x) = \frac{1}{b-a}$$

>> x=[0:0.1:10];
>> y=unifpdf(x,3,7);
>> plot(x,y)

- Symmetric, skewness=0
- Equal occurrence
- Random number generator

#### 1.2.7 Beta Distribution

$$y(x;a,b) = f_X(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}, \quad B(a,b)$$
: Beta function

>> x=[-10:0.1:10];
>> y=betapdf(x,3,6);
>> plot(x,y)







- Bounded distributions
- Related to Gamma distribution
- Manufacturing tolerance

#### 1.2.8 Other Distributions in Engineering

Rayleigh distribution, Gamma distribution, Extreme Type I, II distributions, etc. Refer to <u>http://mathworld.wolfram.com/topics/ProbabilityandStatistics.html</u>.

Perform a parametric study for normal, lognormal, and Weibull distributions! (Homework)

#### **<u>1.3 Multi-variate Random Vector</u>**

Suppose  $X_1$  and  $X_2$  are jointly distributed, and joint event is defined as  $X_1 \le x_1$  and  $X_2 \le x_2$ . The corresponding bi-variate distribution of random vector is defined as

Joint CDF: 
$$F_{X_1X_2}(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2)$$
  
Joint PDF:  $f_{X_1X_2}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1X_2}(x_1, x_2)$ 
(1)

Assume that two random variables are normally distributed. To define the joint PDF of a multivariate distribution, five parameters are required, namely, the mean values of X and Y,  $\mu_{X_1}$  and  $\mu_{X_2}$ , their standard deviations  $\sigma_{X_1}$  and  $\sigma_{X_2}$ , and the correlation coefficient  $\rho_{X_1X_2}$ . The PDF of the bivariate normal distribution can be expressed as

$$f_{X_{1}X_{2}}(x_{1}, x_{2}) = \frac{1}{2\pi\sigma_{X_{1}}\sigma_{X_{2}}\sqrt{1-\rho_{X_{1}X_{2}}^{2}}} \exp\left\{-\frac{1}{2(1-\rho_{X_{1}X_{2}}^{2})}\left[\left(\frac{x_{1}-\mu_{X_{1}}}{\sigma_{X_{1}}}\right)^{2} -2\rho_{X_{1}X_{2}}\frac{(x_{1}-\mu_{X_{1}})(x_{2}-\mu_{X_{2}})}{\sigma_{X_{1}}\sigma_{X_{2}}} + \left(\frac{x_{2}-\mu_{X_{2}}}{\sigma_{X_{2}}}\right)^{2}\right]\right\}$$
(2)

If  $X_1$  and  $X_2$  are correlated, namely,  $\rho_{X_1X_2} \neq 0$ ,  $f_{X_1X_2}(x_1, x_2)$  is not symmetry.



Bivariate distribution of random vector can be generalized for *n*-dimensional random vector,  $\mathbf{X}: \Omega \to \mathbb{R}^n$ . Joint CDF and PDF for *n*-dimensional random vector are written as

Joint CDF: 
$$F_{\mathbf{X}}(\mathbf{x}) = P\left(\bigcap_{i=1}^{n} \{X_i \le x_i\}\right)$$
  
Joint PDF:  $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(\mathbf{x})$  (3)

A multi-variate normal random vector is distributed as

$$f_{\mathbf{X}}(\mathbf{x}) = \left(2\pi\right)^{-\frac{\pi}{2}} \left| \boldsymbol{\Sigma}_{\mathbf{X}} \right|^{-\frac{1}{2}} \exp\left[ -\frac{1}{2} \left( \mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}} \right)^T \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \left( \mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}} \right) \right]$$
(4)

where  $\mu_X$  and  $\Sigma_X$  are mean and covariance matrix of X.

# 2. MOMENTS OF A RANDOM VECTOR

# 2.1 Objectives

- To extend statistical moments of a random variable to a random vector
- To apply statistical moments to an uncertain response
- To prepare an uncertainty propagation through system in Section 3

### **2.2 Definition of Moments of a Random Vector**

Let  $\mathbf{X} = \{X_1, \dots, X_n\}^T$  be an *n*-dimensional random vector and  $g(\mathbf{X})$  be a function of  $\mathbf{X}$ . In general, the  $N^{\text{th}}$  statistical moment of  $g(\mathbf{X})$  is defined as

$$E[g(\mathbf{X})]^{N} \equiv \int_{\Omega} g^{N}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$
(5)

where  $f_{\mathbf{X}}(\mathbf{x})$  is the joint PDF of **X** and  $\Omega$  is a random space.

#### **2.3 Statistical Moments of a Random Vector**

First, one special case is considered to find out statistical moments of input random variable, that is,  $g(\mathbf{X}) = X_i$ ,  $i = 1, \dots, n$ .

#### 2.3.1 Mean of a Random Vector

Let  $g(\mathbf{X}) = X_1$  and set N=1. The first moment of random variable  $X_1$  is defined as

$$E[X_1]^{1} \equiv \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1$$
  
=  $\mu_{X_1}$  (6)

Similarly,

$$E[X_2]^1 \equiv \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2 = \mu_{X_2}$$
  
$$\vdots$$
  
$$E[X_n]^1 \equiv \int_{-\infty}^{\infty} x_n f_{X_n}(x_n) dx_n = \mu_{X_n}$$
  
$$\mu_{\mathbf{X}} = \left\{ \mu_{X_1} \quad \cdots \quad \mu_{X_n} \right\}^T$$

#### 2.3.2 Covariance of a Random Vector

Let  $g(\mathbf{X}) = (X_i - \mu_i)(X_j - \mu_j)$ . The statistical moment is defined as

$$E\Big[(X_i - \mu_i)(X_j - \mu_j)\Big] = \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f_{X_i X_j}(x_i, x_j) dx_i dx_j$$
  
=  $\Big[\Sigma_{ij}\Big] = \Sigma$  (7)

010.141

where  $f_{X_iX_j}(x_i, x_j)$  and  $\Sigma_{ij}$  are the joint PDF and the covariance matrix of  $X_i$  and  $X_j$ , respectively.

When i = j, the diagonal terms in the covariance matrix are obtained as

$$E[X_{1} - \mu_{1}]^{2} \equiv \int_{-\infty}^{\infty} (x_{1} - \mu_{1})^{2} f_{X_{1}}(x_{1}) dx_{1}$$
  
$$= \sigma_{X_{1}}^{2} = \Sigma_{11}$$
  
$$\vdots$$
  
$$E[X_{n} - \mu_{n}]^{2} \equiv \int_{-\infty}^{\infty} (x_{n} - \mu_{n})^{2} f_{X_{n}}(x_{n}) dx_{n}$$
  
$$= \sigma_{X_{n}}^{2} = \Sigma_{nn}$$
(8)

If  $i \neq j$ , the off-diagonal terms in the covariance matrix are obtained as

$$E[(X_{1} - \mu_{1})(X_{2} - \mu_{2})] \equiv \int_{-\infty}^{\infty} (x_{1} - \mu_{1})(x_{2} - \mu_{2})f_{X_{1}X_{2}}(x_{1}, x_{2})dx_{1}dx_{2}$$

$$= \Sigma_{12}$$

$$\vdots$$

$$E[(X_{n} - \mu_{n})(X_{n-1} - \mu_{n-1})] \equiv \int_{-\infty}^{\infty} (x_{n} - \mu_{n})(x_{n-1} - \mu_{n-1})f_{X_{n}X_{n-1}}(x_{n}, x_{n-1})dx_{n}dx_{n-1}$$

$$= \Sigma_{nn-1}$$
(9)

The covariance matrix is written as



# 2.3.3 Properties of Covariance Matrix, $\Sigma_{X}$

- $\Sigma_{\mathbf{X}}$  is symmetric, i.e.,  $\Sigma_{\mathbf{X}} = \Sigma_{\mathbf{X}}^T$ : <u>Prove this property!</u> (Homework)
- Variance of  $X_i$  is the *i*<sup>th</sup> diagonal element of  $\Sigma_{\mathbf{X}}$ , i.e.,  $\sigma_{X_i}^2 = \Sigma_{ii}$
- $\Sigma_{\mathbf{X}}$  is a positive semi-definite matrix, i.e.,  $\mathbf{A}^T \Sigma_{\mathbf{X}} \mathbf{A} \ge 0$ ,  $\forall \mathbf{A} \in \mathbb{R}^n$

# 2.3.4 Correlation Coefficient, $\rho_{ii}$

The correlation coefficient  $\rho_{ij}$  is defined as

$$\rho_{ij} \equiv \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}}\sqrt{\Sigma_{jj}}} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}$$
(10)

The correlation coefficient  $\rho_{ij}$  is a degree of correlation between two random variables. Note

that  $\Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$  represents the off-diagonal elements of covariance matrix,  $\Sigma_X$ .

- If  $X_i$  and  $X_j$  are independent (*i.e.*,  $f_{X_1X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ ), then  $X_i$  and  $X_j$  are uncorreleated (*i.e.*  $\rho_{ij} = 0$ ), but vice versa is not true.
- $-1 \le \rho_{ij} \le +1$
- If  $X_i = aX_j + b$ ,  $\rho_{ij} = \pm 1 = \operatorname{sgn}(a)$ : <u>Prove this property!</u> (Homework)

# **3.** UNCERTAINTY PROPAGATION IN ENGINEERING APPLICATIONS

#### 3.1 Objectives

- To understand a mechanism of uncertainty propagation through an engineering system
- To survey methods to identify and manage the mechanism of uncertainty
- To apply the probabilistic approach to design, instead of a safety factor approach

#### 3.2 What is "Uncertainty Propagation"?

- Types of Uncertainty:
  - 1. "Physical Uncertainty" is the actual variability of physical quantities, such as loads, material properties, and dimensions
  - 2. "Statistical Uncertainty" arises solely as a result of lack of statistical information
  - "Model Uncertainty" occurs as a result of simplifying assumptions, approximate mathematical modeling, unknown boundary conditions, etc.

In all engineering applications, uncertainties in inputs are propagated through a system to uncertainties in outputs, as shown Fig. 1.



Figure 1. Uncertainty Propagation through Physical System



Figure 2. Uncertainty Propagation of V6 Gasoline Engine System

One example of uncertainty propagation is shown in Fig 2. To identify and manage uncertainty propagation in an engineering system, the mechanism of uncertainty propagation must be thoroughly understood. Uncertainty propagation in engineering applications will be discussed in the following sections.

## **<u>3.3 Uncertainty Propagation</u>** for Linear System

Let the response variable *Y* be linearly related to input random vector  $\mathbf{X} = \{X_1, \dots, X_n\}^T$ , i.e.,

$$Y = a_1 X_1 + \dots + a_n X_n + b \quad \text{or} \quad Y = \mathbf{a}^T \mathbf{X} + b \tag{11}$$

where  $\mathbf{a} = \{a_1, \dots, a_n\}^T$  is an *n*-dimensional coefficient vector.

Mean of Y

$$E[Y] = \mu_Y = E[\mathbf{a}^T \mathbf{X} + b]$$
$$= E[\mathbf{a}^T \mathbf{X}] + E[b]$$
$$= \mathbf{a}^T E[\mathbf{X}] + b$$
$$= \mathbf{a}^T \mu_{\mathbf{X}} + b$$

Variance of Y

$$Var[Y] = \sigma_Y^2 = E[(Y - \mu_Y)^2]$$
  
=  $E[(Y - \mu_Y)(Y - \mu_Y)^T]$   
=  $E[(\mathbf{a}^T \mathbf{X} + b - \mathbf{a}^T \mu_X - b)(\mathbf{a}^T \mathbf{X} + b - \mathbf{a}^T \mu_X - b)^T]$   
=  $\mathbf{a}^T E[(\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T]\mathbf{a} \leftarrow \text{From Eq. (7)}$   
=  $\mathbf{a}^T \Sigma_X \mathbf{a}$ 

Generalization

Let  $\mathbf{Y} \in \mathbb{R}^m$  be a random response vector of interest, which is related to input  $\mathbf{X} \in \mathbb{R}^n$ . The linear system is given in the following equation.

$$\mathbf{Y} = \mathbf{A}^T \mathbf{X} + \mathbf{B}$$

where  $\mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^m$  and  $\mathbf{B} \in \mathbb{R}^m$  are coefficient matrix and vector, respectively. Let  $\boldsymbol{\mu}_{\mathbf{Y}} \in \mathbb{R}^m$  and  $\boldsymbol{\Sigma}_{\mathbf{Y}} \in \mathbb{R}^m \times \mathbb{R}^m$  be the mean vector and covariance matrix of output  $\mathbf{Y}$ . Then,

$$\mu_{\mathbf{Y}} = \mathbf{A}^{T} \mu_{\mathbf{X}} + \mathbf{B}$$
  
$$\Sigma_{\mathbf{Y}} = \mathbf{A}^{T} \Sigma_{\mathbf{X}} \mathbf{A}$$
 (12)

### **Example 1: Cantilever Beam**



Given that:

$$P_{1} \sim \left[ 1000 \text{ lb}, 100^{2} \text{ lb}^{2} \right]$$
$$P_{1} \sim \left[ 500 \text{ lb}, 50^{2} \text{ lb}^{2} \right]$$
$$m \sim 10000 \text{ lb-ft, deterministic}$$

Assume  $P_1$  and  $P_2$  are uncorrelated. Calculate mean, standard deviation, and coefficient of variation (COV) of maximum moment at the fixed end.

Solution: At fixed end, the maximum moment is expressed as

$$M_{\text{max}} = 10P_1 + 20P_2 + 10000 = \mathbf{a}^T \mathbf{X} + b$$

where

$$\mathbf{a}^{T} = \{10 \ 20\}, \quad b = 10000 \text{ lb-ft}, \\ \mathbf{X} = \begin{cases} P_{1} \\ P_{2} \end{cases} \text{lb}, \quad \mu_{X} = \begin{cases} 1000 \\ 500 \end{cases} \text{lb}, \quad \mathbf{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 100^{2} & 0 \\ 0 & 50^{2} \end{bmatrix} \text{lb}^{2}$$

<u>Mean of  $M_{\text{max}}$ </u>

$$\mu_{M_{\text{max}}} = \mathbf{a}^T \mu_{\mathbf{X}} + b$$
  
= {10 20} {1000  
500} + 10000  
= 30,000 lb-ft

Variance of M<sub>max</sub>

$$\sigma_{M_{\text{max}}}^{2} = \mathbf{a}^{T} \Sigma_{\mathbf{X}} \mathbf{a}$$

$$= \{10 \quad 20\} \begin{bmatrix} 100^{2} & 0\\ 0 & 50^{2} \end{bmatrix} \{10\} \\= 2 \times 10^{6} \text{ [lb-ft]}^{2}$$

$$\sigma_{M_{\text{max}}} = \sqrt{2 \times 10^{6}} = 1414.2 \text{ lb-ft}$$

Coefficient of Variation (COV) of M<sub>max</sub>

$$COV \equiv \frac{\text{standard deviation}}{\text{mean}} = \frac{1414.2}{30000} = 0.047 \text{ or } 4.7\%$$

# **<u>3.4 Uncertainty Propagation</u>** for Nonlinear System

Let the response variable *Y* be nonlinearly related to input random vector

$$\mathbf{X} = \{X_1, \dots, X_n\}^T$$
, i.e.,

$$Y = g(\mathbf{X})$$
, where  $Y \in R$  and  $\mathbf{X} \in R^n$  (13)

where  $g(\mathbf{X})$  is a nonlinear response. For the nonlinear response  $g(\mathbf{X})$ , it is very difficult or almost impossible to determine the mean and variance of Y exactly. There are several approximate methods available to identify uncertainty propagation for nonlinear system.

# 3.4.1 Sampling Method: Monte Carlo simulation

- Simple but numerically expensive
- Seldom used due to its computational intensiveness, but used for a benchmark study
- To estimate a failure rate,

$$p_f = \frac{N_f}{N}$$
,  $N_f$ : No of Failed Simulation  
N: No of Total Simulation

Random Number Generator

Module 8



#### 3.4.3 First-Order Second Moment (FOSM) Method or Moment Matching Method

Using the first-order Taylor series expansion, a nonlinear response  $Y = g(\mathbf{X})$  can be linearized as

$$Y = g(\mathbf{X}) = g(\mathbf{\mu}_{\mathbf{X}}) + \sum_{i=1}^{n} \frac{\partial g}{\partial X_{i}} \Big|_{\mathbf{X} = \mathbf{\mu}_{\mathbf{X}}} (X_{i} - \mathbf{\mu}_{X_{i}}) + H.O.T.$$
(14)

Assume that the first-order expansion is adequate for approximating *Y*, i.e.,

$$Y \cong \hat{Y}$$

$$= g(\boldsymbol{\mu}_{\mathbf{X}}) + \sum_{i=1}^{n} \frac{\partial g}{\partial X_{i}} \Big|_{\mathbf{X}=\boldsymbol{\mu}_{\mathbf{X}}} (X_{i} - \boldsymbol{\mu}_{X_{i}})$$

$$= \sum_{i=1}^{n} \alpha_{i} X_{i} + g(\boldsymbol{\mu}_{\mathbf{X}}) - \sum_{i=1}^{n} \alpha_{i} \boldsymbol{\mu}_{X_{i}}$$

$$= \mathbf{q}^{T} \mathbf{X} + \mathbf{b}$$
(15)

where  $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_n\}^T$  is a vector of gradients evaluated at mean of **X**. From a linear transformation theory in Section 3.3,

$$\mu_{Y} \cong E[\hat{Y}] = \boldsymbol{\alpha}^{T} \boldsymbol{\mu}_{X} + b$$
  
$$\sigma_{Y}^{2} \cong Var[\hat{Y}] = \boldsymbol{\alpha}^{T} \boldsymbol{\Sigma}_{X} \boldsymbol{\alpha}$$
(16)

Even if this method is easy to use, it has a drawback of numerical accuracy. Therefore, an advanced first-order second moment (AFOSM) method has recently been developed to identify the mechanism of uncertainty propagation.

#### 3.4.4 Advanced First-Order Second Moment (AFOSM) Method

Hasofer and Lind (1978) introduced a rotationally invariant reliability measure, which allows identifying uncertainty propagation more accurately. This is called a first-order reliability method (FORM) or an asymptotic second-order reliability method (asymptotic SORM). FORM or asymptotic SORM requires a transformation of any random space to a standard normal space. Uncertainty propagation of the nonlinear response (public force) in side impact crash is accurately identified using FORM, as depicted in Fig. 3. More detail descriptions will be presented in Lecture 6.



Figure 3. PDF and CDF of pubic force in side impact crash

# **Homework**

Consider the following simply supported beam subject to a uniform load, as illustrated in Fig. 4.



Figure 4. Simply Supported Beam

Random Vector:

$$EI = X_1 \sim N(\mu_{X_1} = 3 \times 10^7, \sigma_{X_1} = 10^5)$$
  
w = X\_2 ~ N(\mu\_{X\_2} = 10, \sigma\_{X\_2} = 1)

The maximum deflection of the beam is shown as

$$Y = g(X_1, X_2) = -\frac{5X_2L^4}{384X_1}$$

Using Monte Carlo simulation and FOSM, identify uncertainty propagation to the maximum deflection by plotting PDF or CDF, and estimating mean and standard deviation of the maximum deflection.