Chapter Five

Functions of Random Variables

5.1 Introduction

A general engineering analysis model is shown in Fig. 5.1. The model output (response) Y contains the performances of a system or product, such as weight, stress, cost, etc. The input variables to the model include both design variables and design parameters. Design variables are those that are controllable by engineers, such as material types and dimensions. Design parameters are not controllable, for example, environmental temperature, usage conditions, etc. If the input variables $\mathbf{X}$ are random, the response variable Y will also be random.

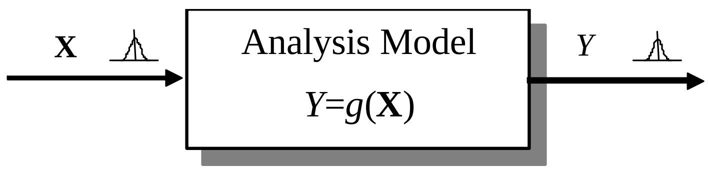

Figure 5.1 Engineering Analysis Model

It is crucial to evaluate the probabilistic characteristics of response variables given those of input variables. This helps engineers understand the impact of the uncertainty associated with the input variables on the response variables. For example, the maximum stress that a component is subject to is one of response variables, and the applied external force is one of input variables. Engineers are interested in knowing the distribution of the maximum stress when they have information about the distribution of the external force. With the knowledge about how the maximum stress varies due to the randomness of the external force, engineers will be able to make appropriate decisions to accommodate the variations in the maximum stress.

Mathematically, the task is to evaluate the distribution of a response variable given the distributions of input variables. In this chapter, we will discuss the theoretic derivation of probability distributions and statistical moments of a response variable from the distributions of input variables. Although the methodologies presented in this chapter may not be directly applicable and practical to real-world applications, the discussion will serve as a theoretic foundation to engineering uncertainty analysis and design under uncertainty.

5.2 Functions of a Single Random Variable

We will first discuss a function of only one random variable and then extend the discussion to general situations where multiple random variables are involved.

5.2.1 Linear Relationship

Assume that random variable Y is a linear function of random variable X and the functional relationship is given by

$$Y = a + bX \tag{5.1}$$

where a and b are constants.

Since Y has a linear relationship with X, Y has the *same distribution* as X, but different distribution parameters, such as mean and variance.

The *cdf* of Y is

$$F_Y(y) = P\left(Y \le y\right) = P(a + bX \le y) = P(X \le \frac{y-a}{b}) = F_X(\frac{y-a}{b}) \tag{5.2}$$

The above equation shows that the *cdf* of Y has the same functional form as X. The *pdf* of Y can also be written in terms of the *pdf* of X as

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{b} f_X(\frac{y-a}{b}) \tag{5.3}$$

Based on Eq. 5.3, the mean of Y can be derived from that of X as

$$m_Y = a + bm_X \tag{5.4}$$

and the standard deviation of Y can be derived from that of X as

$$s_Y = bs_X \tag{5.5}$$

If X follows a normal distribution, i.e. $X \sim N(m_X, s_X)$, Y will also follow a normal distribution and $Y \sim N(m_Y, s_Y) = N(a + bm_X, bs_X)$.

Example 5.1

For example, the tolerance of the length of a rectangular plate is large is assumed to be normally distributed with $X \sim N(10, 0.5)$ cm. Since the tolerance of the width is small, it is treated as a deterministic quantity without any randomness. The width is equal to 4 cm. The perimeter of the plate is $Y = 2X + 8$. Determine the distribution of Y.

The mean value of Y is

$$m_Y = a + bm_X = 8 + 2 \times 10 = 28 \text{ cm}$$

The standard deviation of Y is

$$s_Y = bs_X = 2 \times 0.5 = 1 \text{ cm}$$

Since X is normally distributed, Y is also normally distributed. Its distributed is $Y \sim N(28,\ 1) \text{ cm}$.

5.2.2. Nonlinear Relationship

If random variable Y is a nonlinear function of random variable X and $Y = g(X)$, the *cdf* of Y is given by

$$F_Y(y) = P(Y \le y) = P\left[g(X) \le y\right] = \int_{g(x)\le y} f_X(x)dx \tag{5.6}$$

The *pdf* of Y is given by

$$f_Y(y) = \frac{d}{dy}\left[F_Y(y)\right] = \frac{d}{dy}\left[\int_{g(x)\le y} f_X(x)dx\right] \tag{5.7}$$

Eqs. 5.6 and 5.7 are applicable to any continuous function $Y = g(X)$.

If random variable Y is a monotonically increasing or decreasing function of random variable X, Eqs. 5.6 and 5.7 can be evaluated conveniently. As shown in Fig. 5.2, since Y is a monotonically increasing or decreasing in terms of X, $X = g^{-1}(Y)$ will be a single-valued function of Y, and $Y \le y$ is equivalent to $X \le x$. Therefore,

$$F_Y(y) = P(Y \le y) = P(X \le x) = P\left[X \le g^{-1}(y)\right] = F_X[g^{-1}(y)] \tag{5.8}$$

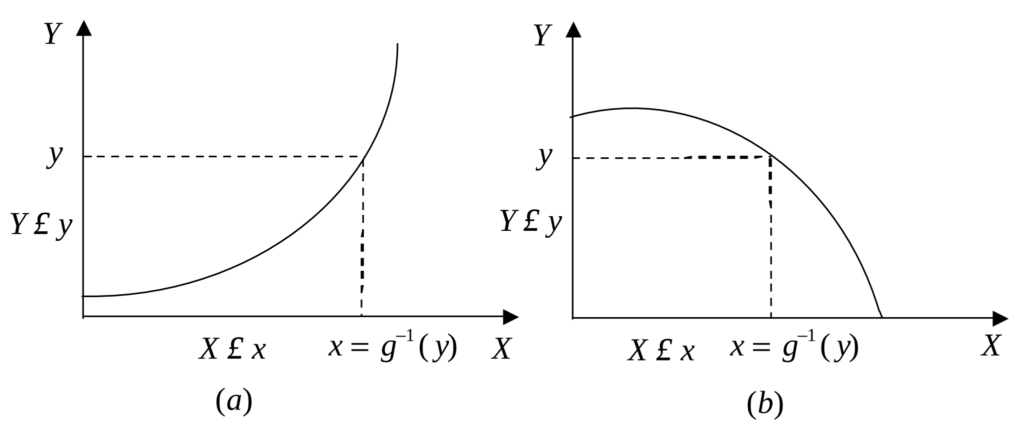

Figure 5.2 (a) Monotonically Increasing and (b) Decreasing Functions

The *pdf* of Y is derived as

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx}\frac{dx}{dy} = f_X(x)\frac{dx}{dy} = f_X[g^{-1}(y)]\frac{dg^{-1}(y)}{dy} \tag{5.9}$$

Since a *pdf* is nonnegative and the derivative $\frac{dg^{-1}(y)}{dy}$ can be negative, the absolute value of $\frac{dg^{-1}(y)}{dy}$ is used. Eq. 5.9 is then rewritten as

$$f_Y(y) = f_X[g^{-1}(y)]\left|\frac{dg^{-1}(y)}{dy}\right| \tag{5.10}$$

Example 5.2

If the diameter of the circular cross section of a transmission shaft is $X = D \sim N(\mu_X, \sigma_X)$ (see Fig. 5.3), what is the probability density function of the cross sectional area $Y = A = g(X) = \frac{\pi}{4}X^2$?

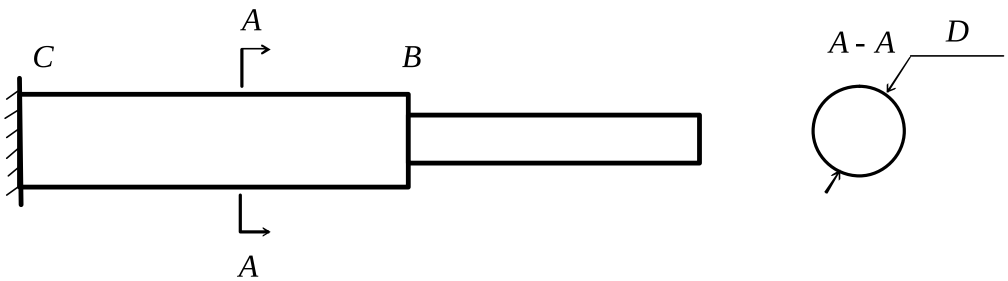

Figure 5.3 A Transmission Shaft

The function $Y = g(X) = A = \frac{\pi}{4}X^2$ is shown in Fig. 5.3.

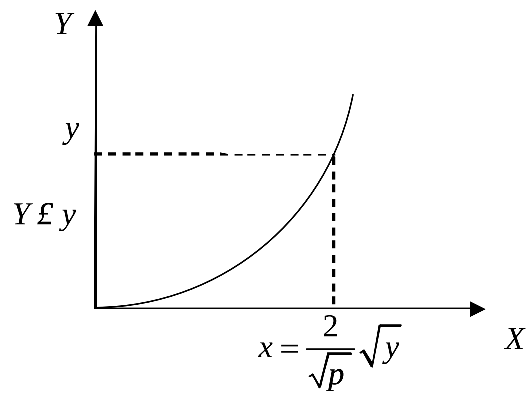

Fig. 5.3 Function $Y = g(X) = \frac{\pi}{4} X^2$

Fig. 5.3 graphically suggests that $Y \le y$ is equivalent to $X \le \frac{2}{\sqrt{\pi}}\sqrt{y}$, and therefore,

$$F_Y(y) = P(Y \le y) = P\left(X \le \frac{2}{\sqrt{\pi}}\sqrt{y} \right) = F_X\left(\frac{2}{\sqrt{\pi}}\sqrt{y} \right)$$

Differentiating the *cdf* gives the *pdf*

$$f_Y(y) = \frac{1}{\sqrt{\pi y}} f_X\left(\frac{2}{\sqrt{\pi}}\sqrt{y} \right)$$

Since

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

the *pdf* of Y is then given by

$$f_Y(y) = \frac{1}{\pi \sigma_X \sqrt{2y}} \exp\left[-\frac{1}{2}\left(\frac{\frac{2}{\sqrt{\pi}}\sqrt{y} - \mu_X}{\sigma_X} \right)^2 \right]$$

If the distribution of the diameter is $X \sim N(50, 1)\ \text{mm}$, the above equation gives the following distribution of the cross sectional area.

$$f_Y(y) = \frac{1}{p\sqrt{2y}} \exp\left[-\frac{1}{2}\left(\frac{2}{\sqrt{p}}\sqrt{y} - 50\right)^2\right]$$

The *pdf*s of *X* and *Y* are depicted in Fig. 5.4.

The same result can be obtained by using Eq. 10 directly.

$$g^{-1}(y) = \frac{2}{\sqrt{p}}\sqrt{y}$$

$$\left|\frac{dg^{-1}(y)}{dy}\right| = \frac{1}{\sqrt{p\,y}}$$

Using Eq. 10 yield the same *pdf*

$$f_Y(y) = \frac{1}{\sqrt{p\,y}} f_X\left(\frac{2}{\sqrt{p}}\sqrt{y}\right)$$

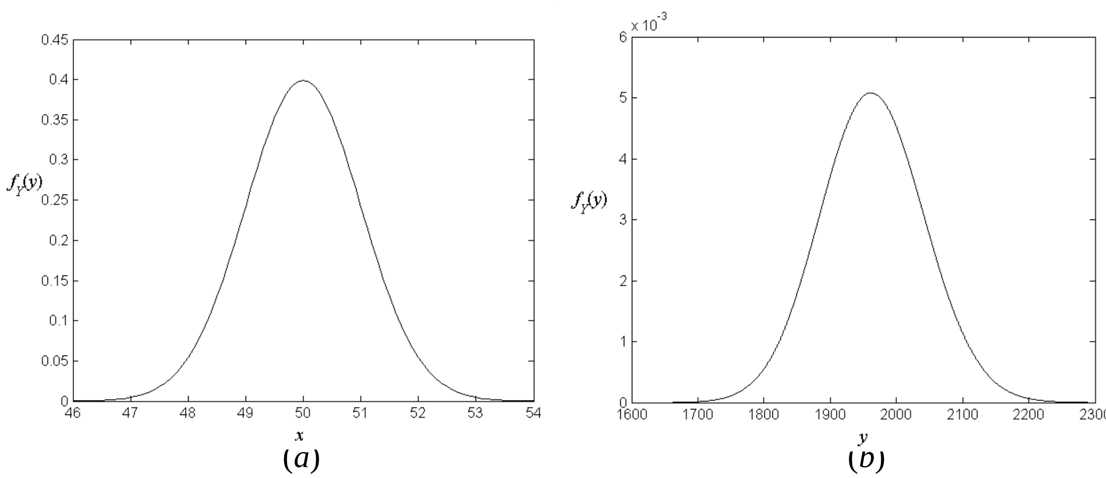

Figure 5.4 (a) *pdf* of *X* and (b) *pdf* of *Y*

5.3 Functions of Several Random Variables

Consider a function of random variables $(X_1, X_2, \cdots, X_n)$

$$Y = g(X_1, X_2, \cdots, X_n) \tag{5.11}$$

If the joint *pdf* of $(X_1, X_2, \cdots, X_n)$ is $f_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n)$, the *pdf* of the function is given by

$$F_Y(y) = P(Y \le y) = \underset{g(x_1, x_2, \cdots, x_n) \le y}{\int \cdots \int} f_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n) dx_1 dx_2, \cdots dx_n \tag{5.12}$$

For a general nonlinear function in engineering applications, it is very difficult or even impossible to use the above equation to obtain the *cdf* of the response variable. In uncertainty analysis that will be presented later in this book, we will discuss approximation methods to the probability integration in Eq. 5.12.

It is possible to use Eq. 5.12 for some special cases. For example, if Y is a linear combination of independent normal variables $X_i \sim N(m_{X_i}, s_{X_i}),\ i = 1, 2, \cdots, n$, then

$$Y = a_0 + \sum_{i=1}^{n} a_i X_i \tag{5.13}$$

in which a_i are constants, it can be shown that Y is also normally distributed with the following mean and variance

$$m_Y = a_0 + \sum_{i=1}^{n} a_i m_{X_i} \tag{5.14}$$

and

$$s_Y^2 = \sum_{i=1}^{n} a_i^2 s_{X_i}^2 \tag{5.15}$$

Example 5.3

As shown in Fig. 5.5, three torques that exert to a transmission shaft are normally distributed with $X_1 = T_1 \sim N(m_{X_1}, s_{X_1}) = N(500, 20)$ N·m , $X_2 = T_2 \sim N(m_{X_2}, s_{X_2})$ $= N(150, 5)$ N·m , and $X_3 \sim T_3 = N(m_{X_3}, s_{X_3}) = N(300, 30)$ N·m . What is the distribution of the resultant torque?

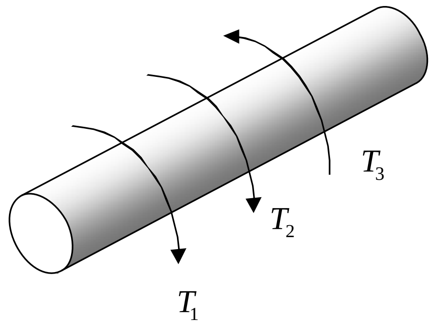

Figure 5.5 A Transmission Shaft

The total torque *T* is the sum of the three individual torques, i.e.

$$Y = T = T_1 + T_2 - T_3 = X_1 + X_2 - X_3$$

Y is also normally distributed with the following mean and standard deviation.

$$m_Y = m_{X_1} + m_{X_2} + m_{X_3} = 500 + 150 - 300 = 350 \text{ N}\cdot\text{m}$$

And

$$s_Y = \sqrt{s_{X_1}^2 + s_{X_2}^2 + s_{X_3}^2} = \sqrt{20^2 + 5^2 + 30^2} = 36.40 \text{ N}\cdot\text{m}$$

5.4 Moments of a Function of Several Random Variables

As seen in Section 5.3, it is difficult to obtain the *cdf* or *pdf* of a response variable which is a general function of random variables. However, it is relatively easy to obtain the moments of the response variable for some special functions.

5.4.1 Mean and Variance of a Linear Function

If *Y* is a linear function of $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ with the following equation

$$Y = a_0 + \sum_{i=1}^{n} a_i X_i \tag{5.16}$$

in which a_i are constants, similar to the derivation of Eqs. 5.2 and 5.3, the mean and variance of *Y* are given by

$$m_Y = a_0 + \sum_{i=1}^{n} a_i m_{X_i} \tag{5.17}$$

where m_{X_i} is the mean of X_i, and

$$s_Y^2 = \sum_{i=1}^{n} a_i^2 s_{X_i}^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} a_i a_j r_{ij} s_{X_i} s_{X_j} \tag{5.18}$$

where $s_{X_i}^2$ is the variance of X_i and r_{ij} is the correlation coefficient between X_i and X_j.

If $(X_1, X_2, \cdots, X_n)$ are mutually independent, Eq. 5.18 becomes

$$s_Y^2 = \sum_{i=1}^{n} a_i^2 s_{X_i}^2 \tag{5.19}$$

5.4.2 Other Common Equations

The moments of several common functions are provided below. g denotes the coefficient of skewness. d denotes the coefficient of variation and is given by

$$d = \frac{s}{m} \tag{5.20}$$

1) $Y = aX^2 + bX + c$

$$s_Y^2 = (2am_X + b)[2a(m_X + s_X g_X) + b]s_X^2 + \frac{1}{2}a^2 s_X^4 (4 + 3g_X^2) \tag{5.21}$$

$$g_Y = (s_Y^{-3} s_X^2 (2am_X + b)^2 [(2a(m_X + b)g_X + \frac{3}{2} a s_X (4 + 3g_X^2)) \tag{5.22}$$

2) $Y = aX^n$

$$m_Y = am_X^2 [1 + \frac{1}{2}n(n-1)d_X^2 + \frac{1}{6}n(n-1)(n-2)d_X^3 g_X + \frac{1}{16}n(n-1)(n-2)(n-3)(2 + g_X^2)d_X^4] \tag{5.23}$$

$$s_Y^2 = (anm_X^n d_X^n)^2 A \tag{5.24}$$

$$g_Y = sign\left(\frac{a}{n}\right) . \frac{B}{A^{3/2}} \tag{5.25}$$

where

$$A = 1 + (n-1)d_X g_X + \frac{1}{2}(n-1)(3n-5)d_X^2 + \frac{1}{8}(n-1)(7n-11)d_X^2 g_X^2 \tag{5.26}$$

$$B = g_X + \frac{3}{4}(n-1)(4+3g_X^2)d_X \tag{5.27}$$

3) $Y = \frac{a}{X} + b$

$$m_Y = \frac{a}{m_x}(1 + d_X^2 - d_X^3 g_X + 3d_X^4 + \frac{3}{2}d_X^4 g_X^2) + b \tag{5.28}$$

$$s_Y^2 = \frac{a^2}{m_X^2} d_X^2 A \tag{5.29}$$

$$g_Y = sign(a)\frac{B}{A^{3/2}} \tag{5.30}$$

where

$$A = 1 - d_X g_X + 8d_X^2 + \frac{9}{2}d_X^2 g_X^2 \tag{5.31}$$

$$B = 6d_X - g_X + \frac{9}{2}d_X^2 g_X^2 \tag{5.32}$$

4) $Z = a\frac{X}{Y} + b$

$$m_Z = a\frac{m_X}{m_Y}(1 + d_Y^2 - d_Y^3 g_Y + 3d_Y^4 + \frac{3}{2}d_Y^4 g_Y^2) + b \tag{5.33}$$

$$s_Z^2 = a^2\left(\frac{m_X}{m_Y}\right)^2 A \tag{5.34}$$

$$g_Z = sign(a)\frac{B}{A^{3/2}} \tag{5.35}$$

where

$$A = d_X^2 + d_X^2 - 2d_X^3 g_Y + 8d_Y^4 + 3d_X^2 d_Y^2 + \frac{9}{2}d_Y^4 g_Y^2 \tag{5.36}$$

$$B = d_X^3 g_X - d_Y^3 g_Y + 6d_Y^4 + 6d_X^2 d_Y^2 + \frac{9}{2}d_Y^4 g_Y^2 \tag{5.37}$$

5) $Y = \sum_{i=1}^{n} a_i X_i + b$

$$m_Y = \sum_{i=1}^{n} a_i m_{X_i} + b \tag{5.38}$$

$$s_Y^2 = \sum_{i=1}^{n} a_i s_{X_i}^2 \tag{5.39}$$

$$g_Y = s_Y^{-3} \sum_{i=1}^{n} a_i^3 s_{X_i}^3 g_{X_i} \tag{5.40}$$

6) $Y = a \prod_{i=1}^{n} X_i + b$

$$m_Y = a \prod_{i=1}^{n} g_{X_i} + b \tag{5.41}$$

$$s_Y^2 = \left(a \prod_{i=1}^{n} m_{X_i} \right)^2 A \tag{5.42}$$

$$g_Y = sign(a) \frac{B}{A^{3/2}} \tag{5.43}$$

where

$$A = \sum_{i=1}^{n} d_{X_i}^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{X_i}{}^2 d_{X_j}{}^2 \tag{5.44}$$

$$B = \sum_{i=1}^{n} d_{X_i}^3 g_{X_i} + 6 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{X_i}^2 d_{X_j}^2 \tag{5.45}$$

5.5 Concluding Remarks

Quantifying the uncertainty of response variables given the uncertainty of input variables is one of the most important tasks in many engineering design applications, such as reliability-based design, robust design, and design for Six Sigma. This can help engineers understand the impact of uncertainty associated with input variables on response variables. Quantifying the uncertainty of response variables therefore aids engineers to make proper decisions to mitigate the effects of input uncertainty. This chapter provides a fundamental introduction about how to evaluate the randomness of response variables from the distributions of input variables.

The methods discussed in this chapter serve as a theoretic foundation for uncertainty analysis although they may not be directly applicable to real engineering problems. In engineering applications, response variables are usually nonlinear functions and involve a large number of random input variables. More practical methods for engineering applications will be discussed later in the following chapters.