Introduction to Uncertainty Quantification
Module 2.2: Random Variables

1 Random Variables

In the previous lesson, we defined a probability space $(\Omega, \mathcal{F}, P)$. In this lesson, we operate in this probability space and specifically define random variables that allow us to operate mathematically on random events by assigning real numerical values (or ranges of real numbers) to outcomes in the sample space, Ω.

1.1 Definition

A **random variable**, X, is a *function*, or mapping, from the sample space of events (Ω) to the real numbers such that for every real number x there exists a probability $P(X(\omega) \leq x)$, where ω indexes Ω.

By convention, random variables are almost always denoted with a capital letter while an associated real number value is denoted with a lower case letter. Hence, we define X as a random variable and x as a specified numerical value. We further note that the index on the sample space (i.e. the random event $\omega \in \Omega$) is typically ignored in standard notation. That is, the random variable is typically expressed as X rather than $X(\omega)$ and the associated probability is given by $P(X \leq x)$ where the dependence on ω is implied.

1.2 Types of Random Variables

It is common to categorize random variables based on the nature of their sample space. Recall that we defined three types of sample spaces: discrete and finite, discrete and infinite, and continuous. Here, we will distinguish between two types of random variables – discrete random variables and continuous random variables – where both finite and infinite discrete sample spaces are considered together. To summarize we have:

- *Discrete Random Variables*: The sample space is discrete and therefore probabilities are associated with discrete events. The probability measure is discontinuous.
- *Continuous Random Variables*: The sample space is continuous and probabilities are associated with ranges of values with a continuous probability measure.

We will formally define the probability measure in each case soon. But first let's look at some simple examples.

1.3 Examples

Discrete Random Variables:
Two concrete cylinders are loaded in a testing machine. Each cylinder can either fail (F) or pass (P) the

test. Here, our sample space is discrete and comprised of the following possible outcomes:

$$\begin{aligned}
O_1 &= PP \\
O_2 &= PF \\
O_3 &= FP \\
O_4 &= FF
\end{aligned}$$

Let us define the following random variable:

$X =$ Number of failures observed from two cylinder tests.

The random variable X maps each possible outcome to a discrete integer value (i.e. 0, 1, or 2) corresponding to the number of cylinders that fail. Looking at each outcome, we have:

$$\begin{aligned}
X(O_1) &= 0 \\
X(O_2) &= 1 \\
X(O_3) &= 1 \\
X(O_4) &= 2
\end{aligned}$$

Continuous Random Variables:
Rather than simply determining pass or failure for each test, consider that the strength of each cylinder is measured. Here, each element of the sample space, $\omega \in \Omega$, (i.e. tested cylinders) can be mapped through the random variable $Y(\omega)$ to a real number (strength) in the continuous range $(0, \infty)$. That is, we can define the following random variable:

$Y =$ Strength of a given cylinder.

The random variable Y maps each possible outcome in the sample space (each possible tested cylinder) the range $(0, \infty)$. That is:

$$Y(\omega) \in (0, \infty)$$

We will discuss the measure of probability for both discrete and continuous random variables in the next section.

2 Probability Distributions

In this section, we will introduce the measure of probability for both discrete and continuous random variables through their respective distribution functions.

2.1 Discrete Random Variables

For discrete random variables, probabilities are assigned to each possible discrete value, $x_i, \quad i = 1, \ldots, n$, of the random variable through it's **Probabilty Mass Function**. The probability mass function is defined, for each value of x_i as:

$$f_X(x_i) = P(X = x_i) \tag{1}$$

where P is the associated probability measure. According to the Axioms of Probability, the probability mass function must possess the following properties:

$$
\begin{aligned}
&f_X(x_i) \geq 0 \quad \forall i = 1, \ldots, n \\
&\sum_{i=1}^{n} f_X(x_i) = 1
\end{aligned}
\tag{2}
$$

The **Cumulative Distribution Function** (CDF) is defined as:

$$
F_X(x) = P(X \leq x) \tag{3}
$$

and can be determined from the probability mass function by:

$$
F_X(x) = \sum_{x_i \leq x} P(X = x_i) \tag{4}
$$

2.2 Continuous Random Variables

For continuous random variables, the **Cumulative Distribution Function** is defined by Eq. (3), with the difference being that x now belongs to a continuous set of real numbers. Therefore, the CDF cannot be defined through a discrete summation over the probabilities of individual events as in Eq. (4). Instead, the CDF for continuous random variable is a continuous function having the following properties:

- $F_X(b) \geq F_X(a)$ if $b \geq a$.
- $F_X(-\infty) = 0$, $F_X(\infty) = 1$
- $P(a < X \leq b) = F_X(b) - F_X(a)$

The first property states that the probability is non-decreasing, while the second property states that it is bounded on the interval [0, 1]. The CDF is therefore a valid probability measure.

For continuous random variables, the **Probability Density Function** (PDF) is defined by

$$
f_X(x) = \frac{dF_X(x)}{dx} \tag{5}
$$

This implies that the CDF can be obtained from the PDF by:

$$
F_X(x) = \int_{-\infty}^{x} f_X(\xi) d\xi \tag{6}
$$

Like the CDF, the PDF is a continuous function having the following properties:

- $f_X(x) \geq 0, \quad \forall x$
- $F_X(\infty) = \int_{-\infty}^{\infty} f_X(\xi) d\xi = 1$
- $P(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(\xi) d\xi - \int_{-\infty}^{a} f_X(\xi) d\xi = \int_{a}^{b} f_X(\xi) d\xi$

The PDF and CDF for a continuous random variable are illustrated graphically in Figure 1.

We conclude this introduction to random variables with a few important notes:

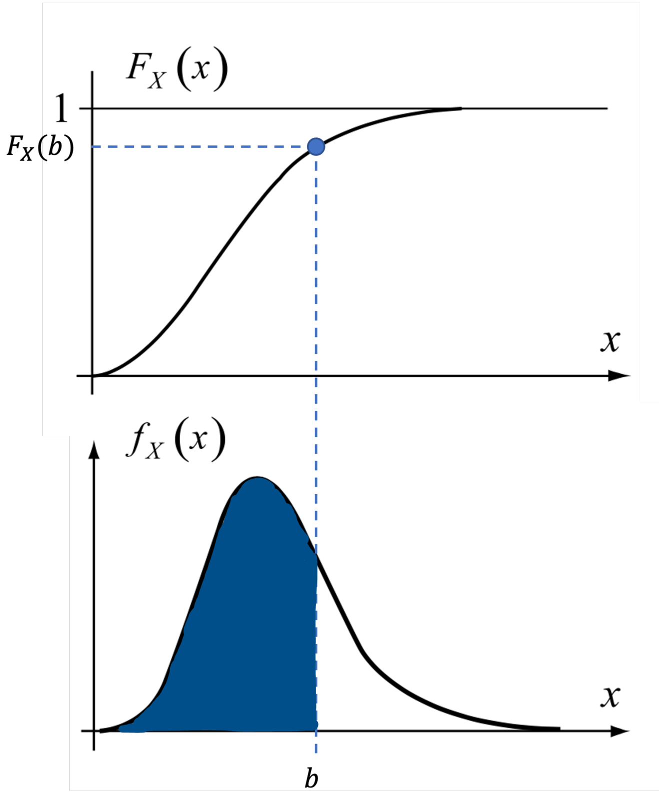

Figure 1: Cumulative Distribution Function (top) and Probability Density Function (bottom)

1. $F_X(x)$ and $f_X(x)$ are non-negative functions. This has previously been stated, but is reiterated here.

2. Although $0 \leq F_X(x) \leq 1$, the PDF can take *any non-negative value*. That is, the PDF is not bounded on [0, 1] and is therefore *not* a probability measure.

3. For continuous random variables, $P(X = x) = 0$. This may seem counterintuitive, but can be proven quite easily by considering $P(a < X \leq b) = F_X(b) - F_X(a)$ where $b = a + \Delta a$ and letting $\Delta a \to 0$. More importantly, we make the distinction that the probability density function does not have the same interpretation as the probability mass function, i.e. that of representing the probability of a specific outcome.

4. $P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)$. Again, this can be shown quite simply using the mathematical relations above. But, intuitively we can make sense of this based on what we've seen before because $\{X > x\}$ and $\{X \leq x\}$ are complementary events and we've already seen that $P(E) = 1 - P(E^*)$.

2.3 Examples

In this section, we will return to the examples from Section 1.3 and discuss the associated discrete and continuous probability distributions.

Discrete Random Variables:
Consider again that we test two concrete cylinders. Further consider that the probability of failure of

any single cylinder is 0.2 ($P(F) = 0.2$). What is the probability that *at most* one cylinder will fail, i.e. $P(X \leq 1)$? What is the probability that at most x cylinders will fail, i.e. $P(X \leq x)$?

Let's begin by assessing the probability of each outcome:

$$\begin{aligned} P(O_1) &= 0.8 \times 0.8 = 0.64 \\ P(O_2) &= 0.8 \times 0.2 = 0.16 \\ P(O_3) &= 0.2 \times 0.8 = 0.16 \\ P(O_4) &= 0.2 \times 0.2 = 0.04 \end{aligned}$$

Next, we can see that

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= P(O_1) + P(O_2 \cup O_3) \\ &= P(O_1) + P(O_2) + P(O_3) \\ &= 0.64 + 0.16 + 0.16 = 0.96 \end{aligned} \tag{7}$$

We can further generalize this to any value of x and define the cumulative distribution function as:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 0.64 & 0 \leq x < 1 \\ 0.96 & 1 \leq x < 2 \\ 1.0 & x \geq 2 \end{cases} \tag{8}$$

Continuous Random Variables:
In the continuous case, we can determine the probability that the strength is less than some value x by evaluating the cumulative distribution function $F_X(x) = P(X \leq x)$ at this value. Here, we need to know the form of the CDF (or equivalently the PDF), which is a continuous non-decreasing function bounded on [0,1]. We will explore common forms for the CDF later in this lesson.

3 Moments of Random Variables

Random variables have probabilities defined through their cumulative distribution function. The distribution of probability over the range of values of x, as illustrated in Figure 1, can be further described by the moments of the distribution, which we will define here. We note that the moments of a distribution are simple descriptors of the distribution and *do not* define the distribution in general. To this point, we explain the meaning or interpretation of the the first several moments. Moreover, we will focus moving forward on continuous random variables and note that, very often, we can simply replace integration with summation to obtain equivalent expressions for discrete random variables.

3.1 Expectation

To define the moments of a random variable, we first need to introduce the **Expectation Operator**. The **Expectation** or **Expected Value** of a random variable X having PDF $f_X(x)$ is given by the following integral:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx \tag{9}$$

The expectation is a linear operation with the following properties. Let a, b, c be constants

$$\begin{aligned}\mathbb{E}[c] &= c \\ \mathbb{E}[cX] &= c\mathbb{E}[x] \\ \mathbb{E}[a + bX] &= a + b\mathbb{E}[X] \\ \mathbb{E}[g_1(X) + g_2(X)] &= \mathbb{E}[g_1(X)] + \mathbb{E}[g_2(X)]\end{aligned} \tag{10}$$

3.2 Moments about the Origin

The n^{th} moment about the origin of the random variable X having PDF $f_X(x)$ is given by the expectation of X^n. That is

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx \tag{11}$$

Moments can be of arbitrary order, n, and describe different properties of the distribution. Here, we will specifically explore moments about the origin of first and second order.

Mean Value

The *mean value* (or expected value) is the first moment, given by

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx \tag{12}$$

The mean value is mathematically identical to the centroid of the probability density function $f_X(x)$.

Mean Square

The *mean square* is the second moment about the origin, given by

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx \tag{13}$$

The mean square is mathematically equivalent to the moment of inertia of the probability density function about the origin. Here, we will specifically explore the first four central moments and their interpretations.

3.3 Moments about the Mean (Central Moments)

The n^{th} moment about the mean (central moment) of the random variable X having PDF $f_X(x)$ is given by the expectation of $(X - \mathbb{E}[X])^n$. That is

$$\mathbb{E}[(X - \mathbb{E}[X])^n] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^n f_X(x)dx \tag{14}$$

Moments can be of arbitrary order, n, and describe different properties of the distribution.

First Moment

By definition, the first central moment of a random variable is zero.

Variance

The *variance* is the second moment about the mean, given by

$$\mathrm{Var(X)} = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) dx \tag{15}$$

Variance is a measure of scatter around the mean value. It is mathematically equivalent to the moment of inertia of the distribution about the mean value.

Using the properties of the expectation, we can show that

$$\begin{aligned}\mathrm{Var(X)} = \sigma_X^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] &= \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f_X(x) dx - 2\mathbb{E}[X] \int_{-\infty}^{\infty} x f_X(x) dx + \mathbb{E}[X]^2 \int_{-\infty}^{\infty} f_X(x) dx \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned} \tag{16}$$

Standard Deviation

The *standard deviation* (σ_X) is defined by:

$$\sigma_X = \sqrt{\mathrm{Var}(X)} \tag{17}$$

Standard deviation gives a measure of scatter around the mean, but it preserves the dimensions of the random variable X.

Coefficient of Variation

The *coefficient of variation* (COV) is given by

$$\nu_X = \frac{\sigma_X}{\mu_X}, \quad \mu_X \neq 0 \tag{18}$$

The COV provides a dimensionless measure of scatter of the distribution.

Skewness

The *skewness* is the normalized third central moment, given by

$$\gamma_1 = \frac{1}{\sigma_X^3} \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^3 f_X(x) dx \tag{19}$$

Skewness gives a measure of the symmetry of the distribution. The skewness may take values that are:

- *Positive* – The distribution has a heavier right tail and peak that lies to the left of the mean value. The distribution's density is more concentrated on the left and more dispersed on the right. It may be referred to as right-skewed, right-tailed, or skewed to the right. This can be confusing the because the distribution appears to lean to the left.

- *Negative* – The distribution has a heavier left tail and peak that lies to the right of the mean value. The distribution's density is more concentrated on the right and more dispersed on the left. It may be referred to as left-skewed, left-tailed, or skewed to the left. This can be confusing the because the distribution appears to lean to the right.

- *Zero* – The distribution is symmetric.

This is illustrated in Figure 2.

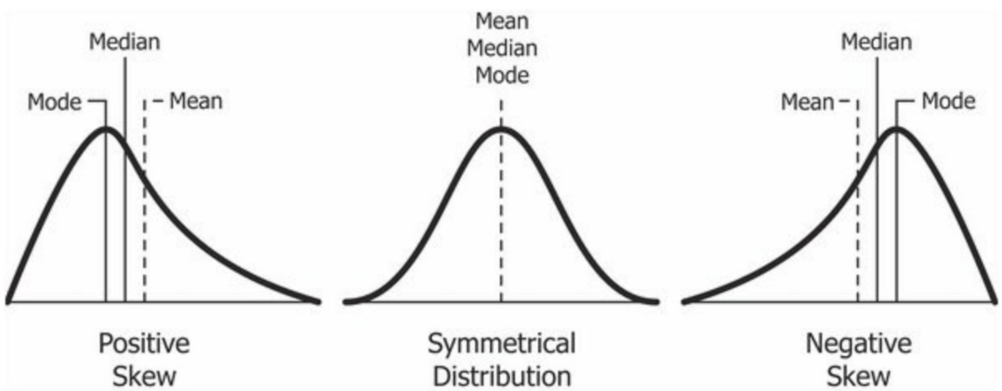

Figure 2: Illustration of skewness. Note that we have not defined the median and mode. Image obtained from: https://commons.wikimedia.org/wiki/File:Relationship_between_mean_and_median_under_different_skewness.png with the following attribution. Diva Jain, CC BY-SA 4.0 https://creativecommons.org/licenses/by-sa/4.0, via Wikimedia Commons.

Kurtosis

The *kurtosis* is the normalized fourth central moment, given by

$$\text{Kurt[X]} = \frac{1}{\sigma_X^4}\int_{-\infty}^{\infty}(x - \mathbb{E}[X])^4 f_X(x)dx \tag{20}$$

Kurtosis is a measure of the heaviness of the tails of the distribution.

Kurtosis is often computed relative to the normal distribution, which always has a kurtosis = 3. We define *excess kurtosis* simply by $\gamma_2 = \text{Kurt}[X] - 3$. Excess kurtosis is often used as a measure of how non-Gaussian the distribution is.

- $\gamma_2 < 3$: *platykurtic*. Tails approach zero faster than the normal. The distribution produces fewer extreme values than the normal

- $\gamma_2 > 3$: *leptokurtic*. Tails approach zero slower than the normal. The distribution produces more extreme values than the normal.

We will discuss the normal, or Gaussian, distribution next.

References

[1] S. Ferson, L. R. Ginzburg, Different methods are needed to propagate ignorance and variability, Reliability Engineering & System Safety 54 (2-3) (1996) 133–144.

[2] M. H. Faber, On the treatment of uncertainties and probabilities in engineering decision analysis (2005).

[3] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, Structural safety 31 (2) (2009) 105–112.

[4] P. Walley, Statistical reasoning with imprecise probabilities (1991).

[5] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, in: Classic works of the Dempster-Shafer theory of belief functions, Springer, 2008, pp. 57–72.

[6] G. Shafer, A mathematical theory of evidence, Vol. 42, Princeton university press, 1976.

[7] D. Dubois, H. Prade, Possibility theory: an approach to computerized processing of uncertainty, Springer Science & Business Media, 2012.

[8] L. A. Zadeh, Fuzzy sets, Information and control 8 (3) (1965) 338–353.

[9] E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Machine Learning 110 (3) (2021) 457–506.

[10] S. Ferson, C. A. Joslyn, J. C. Helton, W. L. Oberkampf, K. Sentz, Summary from the epistemic uncertainty workshop: consensus amid diversity, Reliability Engineering & System Safety 85 (1-3) (2004) 355–369.

[11] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mechanical systems and signal processing 37 (1-2) (2013) 4–29.

Nomenclature

Functions

$P(\cdot)$ Probability measure

$f_X(\cdot)$ Probability density function (PDF) of a random variable X

$F_X(\cdot)$ Cumulative distribution function (CDF) of a random variable X

$\mathbb{E}[\cdot]$ Expected value of a random variable. Also denoted $\mu_X \triangleq \mathbb{E}[X]$

$\mathbb{E}[X^n]$ The n^{th} moment about the origin of a random variable X

$\mathrm{Var}(\cdot)$ Variance of the random variable. Also denoted $\sigma_X^2 \triangleq \mathrm{Var(X)}$

$\mathrm{Kurt}[\cdot]$ Kurtosis of a random variable

Variables

O Outcome from an experiment

E	Event
Ω	Sample space
ω	A random event from the sample space Ω
$\mathcal{F}$	Event space
$(\Omega, \mathcal{F}, P)$	Probability space, also known as a probability triple
X	A random variable
σ_X	Standard deviation of the random variable X
ν_X	Coefficient of Variation (COV) of the random variable X
γ_1	Skewness of the random variable X
γ_2	Excess kurtosis of the random variable X