Random variables

Let S denote the sample space underlying a random experiment with elements $s \in S$. A **random variable,** X**,** is defined as a function $X(s)$ whose domain is S and whose range is a set of real numbers, i.e., $X(s) \in \mathbf{R}^1$.

Example A: Consider the experiment of tossing a coin. The sample space is $S = \{H, T\}$. The function

$$X(s) = \begin{cases} 1 & \text{if } s = H \\ -1 & \text{if } s = T \end{cases}$$

is a random variable whose domain is S and range is $\{-1, 1\}$.

Example B: Let the set of all real numbers between 0 and 1 be the sample space, S. The function $X(s) = 2s - 1$ is a random variable whose domain is S and range is set of all real numbers between -1 and 1.

A **discrete random variable** is one whose range is a countable set. The random variable defined in example A is a discrete randowm variable. A **continuous random variable** is one whose range is not a countable set. The random variable defined in Example B is a continiuos random varible. A **mixed random variable** contains aspects of both these types. For example, let the set of all real numbers between 0 and 1 be the sample space, S. The function

$$X(s) = \begin{cases} 2s - 1 & \text{if } s \in (0, \frac{1}{2}) \\ 1 & \text{if } s \in [\frac{1}{2}, 1) \end{cases}$$

is a mixed random variable with domain S and range set that includes set of all real numbers between -1 and 0 and the number 1.

Cummulative Distribution Function

Given a random variable X, let us consider the event $\{X \leq x\}$ where x is any real number. The probability of this event, i.e., $\Pr(X \leq x)$, is simply denoted by $F_X(x)$:

$$F_X(x) = \Pr(X(s) \leq x), \quad x \in \mathbf{R}^1.$$

The function $F_X(x)$ is called the **probability or cumulative distribution fuction (CDF)**. Note that this CDF is a function of both the outcomes of the random experiment as embodied in $X(s)$ and the particular scalar variable x. The properties of CDF are as follows:

- Since $F_X(x)$ is a probability, its range is limited to the interval: $0 \leq F_X(x) \leq 1$.
- $F_X(x)$ is a non-decreasing function in x, i.e.,

$$x_1 < x_2 \longleftrightarrow F_X(x_1) \leq F_X(x_2).$$

- $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.
- For continuous random variables, the CDF $f_X(x)$ is a unifromly continuous function in x, i.e.,

$$\lim_{x \to x_o} F_X(x) = F_X(x_o).$$

- For discrete random variables, the CDF is in general of the form:

$$F_X(x) = \sum_{x_i \in X(s)} p_i u(x - x_i), \ \ x \in \mathbf{R}^1,$$

where the sequence p_i is called the probability mass function and $u(x)$ is the unit step function.

Probability Distribution Function

The derivative of the CDF $F_X(x)$, denoted as $f_X(x)$, is called the **probability density function (PDF)** of the random variable X, i.e.

$$f_X(x) = \frac{dF(x)}{dx}, \quad x \in \mathbf{R}^1.$$

or, equivalently the CDF can be related to the PDF via:

$$F_X(x) = \int_{-\infty}^{x} f_X(u) du, \quad x \in \mathbf{R}^1.$$

Note that area under the PDF curve is unity, i.e.,

$$\int_{-\infty}^{\infty} f_X(u) du = F_X(\infty) - F_X(-\infty) = 1 - 0 = 1$$

In general the probability of a random variable $X(s)$ taking values in the range $x \in [a, b]$ is given by:

$$\Pr(x \in [a, b]) = \int_a^b f_X(x) dx = F_X(b) - F_X(a).$$

For discrete random variables the PDF takes the general form:

$$f_X(x) = \sum_{x_i \in X(s)} p_i \delta(x - x_i).$$

Specifically for continuous random variables:

$$\Pr(x = x_o) = F_X(x_o^+) - F_X(x_o^-) = 0.$$