6 Discrete Random Variables, P.M.F, and C.D.F

Random variables are the numerical data gathered from experiments. From a previous note recall that the definition of a random variable X is *a real function on sample space S*

X is **Discrete** if its possible values are finite or countably infinite set of real numbers.

X is **Continuous** if possible values are an interval on the real number line

The **probability mass function**(p.m.f) of a discrete random variable X is the function

$$f(x) = P(X = x) \tag{1}$$

$\mathcal{R}_X$ denotes the range of X can be either $\{x_1, x_2, .., x_n\}$ or $\{x_1, x_2, ...\}$ depending if X is finite or countably infinite.

The p.m.f has the following properties

1. $f(x_i) \geq 0$: The probability mass function given any x will be greater than zero
2. $\sum_i f(x_i) = 1$: The sum of the probability mass function over every random variable is 1
3. $f(x_i) = P(X = x_i)$: The probability mass function at x_i is always equal to the probability of x_i

It follows from the addition of probability that for any set of real numbers A

$$\sum_{x_i \in A} f(x_i) = P(X \in A) \tag{2}$$

This simply states that for any set of real numbers $A = \{0, 1, 2, ..., N\}$ the probability of A is the sum of the probability of every real number in A.

Example 3.5 Let $X = \#$ of semiconductor wafers that need to be analyzed to first detect a large particle of contamination. Wafers are independently contaminated each with a probability of 0.01. Find $P(X > 10)$.

Ans. We know that $P(X > 10) = 1 - P(X \leq 10)$. So we need to find $P(X \leq 10)$. Notice that we look at $x - 1$ wafers of non-contaminated wafers, until we reach 1 which is contaminated. Thus our equation follows a simple geometric distribution.

$$P(X = x) = (0.99)^{x-1}(0.01)$$

Now we needed to solve for $P(X \leq 10)$

$$\begin{aligned}
P(X = 10) &= (0.99)^9(0.01) = 0.009135172 & P(X = 9) &= (0.99)^8(0.01) = 0.009227446 \\
P(X = 8) &= (0.99)^7(0.01) = 0.009320653 & P(X = 7) &= (0.99)^6(0.01) = 0.009414801 \\
P(X = 6) &= (0.99)^5(0.01) = 0.009509900 & P(X = 5) &= (0.99)^4(0.01) = 0.009605960 \\
P(X = 4) &= (0.99)^3(0.01) = 0.009702990 & P(X = 3) &= (0.99)^2(0.01) = 0.009801000 \\
P(X = 2) &= (0.99)^1(0.01) = 0.009900000 & P(X = 1) &= (0.99)^0(0.01) = 0.010000000
\end{aligned}$$

So $P(X \leq 10) = 0.09561792499$. Now we need to get $P(X > 10)$

$$P(X > 10) = 1 - 0.09561792499 = 0.904382750$$

The **Cumulative Distributive Function**(c.d.f) of X is the function

$$F(x) \equiv P(X \leq x) = \sum_{x_i \leq x} f(x_i) \tag{3}$$

The c.d.f follows the following properties

1. $0 \leq F(X) \leq 1$: The c.d.f can only range between numbers from 0 to 1
2. $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$: The c.d.f is the sum of the p.m.f from every x_i less than or equal to the given x
3. If $x \leq y$, then $F(x) \leq F(y)$: The c.d.f is non-decreasing as it is sum of the p.m.f's $x_i \leq x$ and p.m.f ≥ 0

Example 3.6 Change the following values gathered from a p.m.f to a c.d.f

x	0	1	2
$f(x)$	0.1	0.4	0.5

Ans. The c.d.f is as follows

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.1 & 0 \leq x < 1 \\ 0.5 & 1 \leq x < 2 \\ 1.0 & 2 \leq x \end{cases}$$

We can also see that it is easy to revert from a c.d.f to a p.m.f

$$f(x) = F(x) - \lim_{t \to x^-} F(t) \tag{4}$$

or for any $a < b$

$$f(a \leq x \leq b) = F(b) - \lim_{t \to a^-} F(t) \tag{5}$$