1 Random Variables

If a real variable X is associated with an outcome of a random experiment, it is called a *random variable* or a *stochastic variable* or simply a *variate*.

Types of Random Variables:

- Discrete Random Variables
- Continuous Random Variables

1.1 Probability Distribution Function (pdf)

This is a function that denotes the probability of a given event as a continuous/discrete function of $f(x)$ where $x \in \mathbb{R}$.

1.2 Cumulative Distribution Function (cdf)

This is a function that denotes the sum of probability of a given event as a continuous/discrete function of $F(X)$ where X will be $\leq x$.

1.3 Statistical Terminologies

- Mean (Expectation of x): Denoted by $\boxed{E(x)}$
- Variance: Denoted by $\boxed{V(x) \text{ or } \sigma^2}$
- Standard Deviation: Denoted by $\boxed{\sigma}$

	Discrete Random Variables	Continuous Random Variables
μ or $E(x)$	$\sum\limits_{i=1}^{n} x_i P(x_i)$	$\int\limits_{-\infty}^{\infty} xP(x)\,dx$
$E(x^2)$	$\sum\limits_{i=1}^{n} x_i^2 P(x_i)$	$\int\limits_{-\infty}^{\infty} x^2 P(x)\,dx$
$E(x-\mu)^2$ or σ^2	$V(x) = E(x^2) - E(x)^2$	

1.4 Chebyshev's Inequality

Let x be a random variable with $E(x) = \mu$ and c be any real number, then if $E(x-c)^2$ is finite and is any positive number,

$$\boxed{P\{|x-c| \geq \varepsilon\} \leq \frac{E(x-c)^2}{\varepsilon^2}}$$

OR

$$\boxed{P\{|x-c| \leq \varepsilon\} \geq 1 - \frac{E(x-c)^2}{\varepsilon^2}}$$

If $\boxed{c = \mu}$ then,

$$P\{|x-c| \geq \varepsilon\} \leq \frac{V(x)}{\varepsilon^2}$$

If $\boxed{c = \mu}$ & $\boxed{\varepsilon = k\sigma}$ then,

$$P\{|x-c| \geq \varepsilon\} \leq \frac{1}{k^2}$$

1.5 Markov's Inequality

For $a > 0$,

$$\boxed{P\{x \geq a\} \leq \frac{E(x)}{a}}$$

1.6 Uniform Distribution

If X is a continuous random variable defined over an interval $[a, b]$ and having probability distribution function

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$

then we say X has uniform distribution. Denoted as follows: $X \sim \mathbf{U}(a, b)$

We define the mean,variance as follows:

- $E(x) = \frac{a+b}{2}$
- $E(x^2) = \frac{1}{3}(a^2 + b^2 + ab)$
- $V(x) = \frac{(b-a)^2}{12}$

1.7 Two Dimensional Random Variables

Let x, y be 2 random variables distributed in a 2 dimensional space S.

$$x, y \rightarrow \text{random variable}$$

$$x(S) = x_1, x_2 \ldots x_n \qquad y(S) = y_1, y_2 \ldots y_m$$

then we define P(x=x_i,y=y_j) $= P_{ij}$ such that,

- $P_{ij} \geq 0$
- $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} P_{ij} = 1$

1.7.1 Joint Probability Function

also known as Joint Probability Mass Function is function on the set (x_i, y_j, P_{ij}).

$x_i \backslash y_j$	y_1	y_2	$\ldots$	y_m	
x_1	P_{11}	P_{12}	$\ldots$	P_{1m}	$f(x_1)$
x_2	P_{21}	P_{22}	$\ldots$	P_{2m}	$f(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
x_n	P_{n1}	P_{n2}	$\ldots$	P_{nm}	$f(x_n)$
	$g(y_1)$	$g(y_2)$	$\ldots$	$g(y_m)$	1

We define a few terms such as $f(x_i)$ and $g(y_j)$ for the probability function of two variables $f(x, y)$.

$$\boxed{f(x_i) = \sum_{j=1}^{m} P_{ij} \quad ; \quad f(y_j) = \sum_{i=1}^{n} P_{ij}}$$

Based on the terms mentioned above, we have following formulae,

	Discrete Random Variables	Continuous Random Variables
$E(x)$	$\sum_{i=1}^{n} x_i P f x_i)$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy$
$E(y)$	$\sum_{j=1}^{m} y_j g(y_j)$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dx\,dy$
$E(xy)$	$\sum_{1\leq i\leq n, 1\leq j\leq m}^{n} x_i\, y_i P_{ij}$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f(x,y)\,dx\,dy$
$E(x^2)$	$\sum_{i=1}^{n} x_i^2 P f x_i)$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\,dx\,dy$
$E(y^2)$	$\sum_{j=1}^{m} y_j^2 g(y_j)$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f(x,y)\,dx\,dy$

For a crv, (x, y) is associated with function $f(x, y)$ such that,

- $f(x, y) \geq 0$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

$f(x, y)$ is known as the joint probability density function.

1.7.2 Covariance and Correlation Coefficient

The relation of the two variables x and y can be defined by covariance which when $+ve$ means that they are directly proportional and when $-ve$ means inversely proportional. When the covariance is 0, it means that the 2 variable are completely unrelated.

$$Cov(x, y) = E(xy) - E(x)\,E(y)$$

This is called Measure of Correlation.

Correlation Coefficient

The numerical measure of correlation is called the coefficient of correlation and is defined by the relation:

$$r(x,y) = r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{E(xy) - E(x)\,E(y)}{\sqrt{V(x)\,V(y)}}$$