PROBABILITY

RANDOM EXPERIMENT:

- An experiment that can be repeated as many times as wished under identical conditions
- For each trial the outcome of the experiment is not known in advance
- But for each trial, the set of all possible outcomes are known.

Eg: E_1 - Toss a die and observe the number that shows on the top

E_2 - Toss a coin three times and observe the total number of heads obtained.

SAMPLE SPACE:

The set of all possible outcomes of a random experiment. Usually denoted by `S' or . Observe that an outcome of an experiment need not be a number also that number of outcomes can be finite, countably infinite or uncountably infinite.

Eg: E_1 - {1,2,3,4,5,6}

E_2 - {1,2,3}

EVENT:

An event is simply a set of possible outcomes. In general, it is a subset of the sample space.

Eg: Events that are associated with experiment E_1

A_1: An even number occurs; that is, $A_1 = \{2,4,6\}$

A_2: A prime number occurs; that is, $A_2 = \{2,3,5\}$

MUTUALLY EXCLUSIVE EVENTS:

Two events A and B are said to be mutually exclusive if they cannot occur together. This will be expressed as A B=0, that is intersection of A and B is an empty set.

PROBABILITY FUNCTION:

Let ε be an experiment. Let S be a sample space associated with ε. With each event A we associate a real number, designated by *P(A)* and called the *probability of A* satisfying the following properties.

(1) $0 \leq P(A) \leq 1$
(2) $P(S) = 1$
(3) If A and B are mutually exclusive vents, $P(A \cup B) = P(A) + P(B)$
(4) If $A_1, A_2, \ldots, A_n, \ldots$ are pairwise mutually exclusive events, then

$$P(\cup_{i=1}^{\infty} A_i) = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots$$

Note:

- If Φ is the empty set, then $P(\Phi) = 0$.
- If $\bar{A}$ is the complementary event of A, then
  $$P(A) = 1 - P(\bar{A}).$$
- If A and B are any two events, then
  $$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
- If A, B and C are any three events, then
  $$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C).$$

FINITE SAMPLE SPACE:

A sample space S with finite number of elements i.e., S can be written as $S = \{a_1, a_2, \cdots a_k\}$. To each elementary event $\{a_i\}$, we assign a number p_i, called the probability of $\{a_i\}$, satisfying the following conditions:

(a) $p_i \geq 0, \quad i = 1, 2, \ldots, k.$
(b) $p_1 + p_2 + \cdots + p_k = 1.$

EQUALLY LIKELY OUTCOMES:

If S is a finite sample space, and if all its k outcomes are equally likely, it follows that each p_i = 1/k. For the condition $p_1 + p_2 + \cdots + p_k = 1$ becomes $kp_i=1$ for all i. This follows that for any event A consisting of r outcomes, we have

$$P(A) = r/k.$$

$$P(A) = \frac{number\ of\ ways\ in\ which\ \varepsilon\ can\ occur\ favorable\ to\ A}{total\ number\ of\ ways\ in\ which\ \varepsilon\ can\ occur}$$

CONDITIONAL PROBABILITY:

$P(A|B)$ denotes the probability of occurrence of event A, given that event B occurs. It can be determined that

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Also observe that the function $P(A|B)$ satisfies all the conditions of a probability function for $P(B) \geq 0$.

BAYES' THEOREM:

Let S be the sample space of a random experiment. Let $B_1, B_2, \cdots B_k$ be a partition of S by events i.e., they are mutually exclusive to each other. Let A be any event. Then

$$P\left(\frac{B_i}{A}\right) = \frac{P\left(\frac{A}{B_i}\right) P(B_i)}{\sum_{j=1}^{k} P\left(\frac{A}{B_j}\right) P(B_j)}$$

for all i = 1,2,….,k.

INDEPENDENT EVENTS:

Two events *A, B* are said to be independent if `the occurrence of one event does not depend on the occurrence of the other'.

So, A and B are independent $\Leftrightarrow$ $P(A/B) = P(A)$ & $P(B/A) = P(B)$.

From above relations it implies that, A and B are *independent* if and only if

$$P(A \cap B) = P(A).P(B)$$

NOTE:

- Three events *A, B, C* are *independent* if and only if

$$P(A \cap B) = P(A).P(B)$$

$$P(B \cap C) = P(B).P(C)$$

$$P(C \cap A) = P(C).P(A)$$

$$P(A \cap B \cap C) = P(A).P(B).P(C)$$

- If *A and B* are *independent*, then A, B, A^c, B^c are also independent with respect to each other.

RANDOM VARIABLE:

Let ε be an experiment and S be a sample space associated with the experiment. A function *X* assigning to every element s ϵ S, a real number *X(S)*, is called a *random variable.*

The space R_x, the set of all possible values of *X*, is called the *range space.*

NOTE:

- If R_x is finite or countably infinite, then X is called *Discrete Random Variable,* otherwise it is called *Continuous Random Variable.*

PROBABILITY DENSITY FUNCTION:

X is said to be a *Continuous Random Variable* if there exists a function *f*, called the *probability density function (pdf) of X*, satisfying the following conditions:

(a) $f(x) \geq 0$

(b) $\int_{-\infty}^{+\infty} f(x)dx = 1.$

(c) For any *a, b,* with $-\infty < a < b < +\infty$, we have

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

CUMULATIVE DISTRIBUTION FUNCTION:

Let *X* be a random variable, discrete or continuous. The cumulative distribution function [*F(x)*] is defined as:

$$F(x) = P(X \leq x)$$

NOTE:

- If *X* is a discrete random variable,

$$F(x) = \sum_{j} p(x_j),$$

where the sum is taken over all indices *j* satisfying $x_j \leq x$.

- If *X* is a continuous random variable with pdf *f*,

$$F(x) = \int_{-\infty}^{x} f(s)ds$$

THEOREM:

(a) Let *F* be the cdf of a continuous random variable with pdf *f*. Then

$$f(x) = \frac{d}{dx}F(x),$$

for all *x* at which *F* is differentiable.

(b) Let X be a discrete random variable with possible values $x_1, x_2, \ldots,$ and suppose that it is possible to label these values so that $x_1 < x_1 < \cdots$

Let F be the cdf of X. Then

$$p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1})$$

THEOREM:

Let X be a continuous random variable with pdf 'f' and cdf 'F'. Let Φ be a map from 'R' onto [a,b].

Let $Y = \Phi(X)$. Assume that Φ is strictly monotone and differentiable on [a,b]. Then the pdf of Y,

$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$

where $x = \Phi^{-1}(y)$ [or $y = \Phi(x)$].

EXPECTED (MEAN) VALUE and VARIANCE:

- Let X be a discrete random variable with $R_x = \{x_1, x_2, \ldots\}$ and probability mass function $p(x_i) = P(X = x_i)$.

  Then the *expected value* of X is given by

$$E(X) = \sum_i x_i . p(x_i)$$

  And the *variance* of X is given by

$$V(X) = E(X - E(X))^2 = \sum_i \left(x_i - E(X)\right)^2 p(x_i)$$

- Let X be a continuous random variable with pdf 'f'.
  Then the *expected value* of X is given by

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

And the *variance* of *X* is given by

$$V(X) = E(X - E(X))^2 = \int_{-\infty}^{+\infty} \big(x - E(x)\big)^2 f(x)dx$$

NOTE:

- Variance can also be found by using the relation
$$V(X) = E(X^2) - E(X)^2$$
- The *standard deviation* of *X* is defined as the positive square root of the Variance i.e.,
$$\sigma_x = \sqrt{V(X)}.$$

CHEBYSHEV'S INEQUALITY:

Let *X* be a random variable and let 'c' be any constant. Then, for any $\varepsilon > 0$,

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

NOTE:

In the above theorem let $c = \mu = E(X)$, $\sigma^2 = V(X)$ and $\varepsilon = k\sigma$, then

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2$$

MOMENTS AND MOMENT GENERATING FUNCTION:

- The k^{th} moment of a random variable *X* is : $\mu_k = E(X^k)$
- The Moment Generating Function of a random variable *X* is
$$M(t) = E(e^{tX})$$

NOTES:

$$M(t) = E(1 + tX + \frac{t^2X^2}{2!} + \cdots + \frac{t^kX^k}{k!} + \cdots)$$

$$=> M(t) = 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \cdots + \frac{t^kE(X^k)}{k!} + \cdots$$

Observe that $\frac{d^kM(t)}{dt^k}|_{t=0} = E(X^t) = \mu_k$

<u>TWO-DIMENSIONAL VECTOR</u>:

A pair (*X*,*Y*) of random variables is called a Two-dimensional random variable/ vector.

NOTE:

- If the range of (*X*,*Y*) is finite and countable, it is called Discrete random variable.
- (*X,Y*) is called Continuous random variable if there exists a two-dimensional pdf *f=f(x,y)* such that

$$p(\mathrm{a} < \mathrm{X} < \mathrm{b}, \mathrm{c} < \mathrm{Y} < \mathrm{d}) = \int_{x=a}^{b}\int_{y=c}^{d} f(x,y)dxdy$$

Whenever b>a and d>c

<u>CUMULATIVE DISTRIBUTION FUNCTION</u>:

If (*X,Y*) is a two-dimensional random variable, the *joint cdf* of *X* and *Y* is:

$$F_{x,y}(x,y) = p(X \le x, Y \le y)$$

- Continuous random variable:

$$F_{x,y}(x,y) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{y} f(u,v)dudv$$

NOTE:

1) *Continuous random variable:*
   - *Marginal pdf of x:* $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$

- *Marginal pdf of x:* $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$

2) *Discrete random variable:*
   - *Marginal pmf of x:* $p_X(x_i) = \sum_j P_{X,Y}(x_i, y_j)$
   - *Marginal pmf of x:* $p_Y(y) = \sum_i P_{X,Y}(x_i, y_j)$

CO-VARIANCE:

The co-variance of the random variable (*X,Y*) is given by:

$$cov(X,Y) = E[(X - E(X)), (Y - E(Y))]$$
$$= E(X,Y) - E(X)E(Y)$$

CAUCHY-SCHWARTZ INEQUALITY:

Continuous case: $|\int \psi\Phi| \le (\int |\psi|^2)^{1/2} (\int |\Phi|^2)^{1/2}$

Discrete case: $|\sum_{k=1}^{n} a_k b_k| \le (\sum_{k=1}^{n} |a_k|^2)^{1/2} (\sum_{k=1}^{n} |b_k|^2)^{1/2}$

CO-RELATION COEFFICIENT:

The co-relation coefficient of X and Y is:

$$\rho = \frac{cov(X,Y)}{[var(X)]^{1/2}[var(Y)]^{1/2}}$$

Note: Observe that $|\rho| \le 1$

CONDITIONAL PROBABILITY:

The conditional pmf of X given $Y=y_j$ is:

$$P_X(x_i|y_j) = P(X=x_i|Y=y_j) = \frac{P(X=x_i, Y=y_i)}{P(Y=y_j)}$$

The conditional pdf of X given Y=y is:

$$f_X(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Some Standard Distributions

Distribution	PDF	μ	Var
Binomial Distribution	${}^{n}_{k}C\ p^{k}(1-p)^{n-k}$	np	np(1-p)
Poisson Distribution	$\dfrac{\lambda^{x}e^{-\lambda}}{x!}$	λ	λ
Gamma Distribution	$\frac{1}{\beta^{\alpha}\sqrt(\alpha)}x^{\alpha-1}e^{-x/\beta}$; for x>0 0; otherwise	$\alpha\beta$	$\alpha\beta^{2}$
Exponential Distribution	$\lambda e^{-\lambda x}$; $x>0$ 0; otherwise	$1/\lambda$	$1/\lambda^{2}$
Normal Distribution	$\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^{2}}$; $-\infty<x<+\infty$	μ	σ^{2}

STATISTICS:

SAMPLE DATA:

Given n-data: $x_1 \leq x_2 \leq . . . \leq x_n$

- Sample Mean, $\dot{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
- Sample Variance, $s^2 = \frac{\sum_{i=1}^{n}(x_i - \dot{x})^2}{n-1}$
- Standard Deviation, $\sigma = \sqrt{+s^2}$

RANDOM SAMPLE (from infinite population):

A set of observations $x_1, x_2, . . . , x_n$ constitute a random sample of size 'n', from a population *f(x)* if

(1) Each x_i has distribution *f(x)*
(2) $x_1, x_2, . . . , x_n$ are independent random variables

Theorem:

Let $x_1, x_2, . . . , x_n$ be independent random variables then,

$$var(x_1 + x_2 + \cdots + x_n) = var(x_1) + var(x_2) + \cdots + var(x_n)$$

LAW OF LARGE NUMBERS:

Let $x_1, x_2, . . . , x_n$ be a random sample. If 'μ' is the population mean, then

(1) $\dot{x} \rightarrow \mu$ as $n \rightarrow \infty$
(2) $P(|\dot{x} - \mu| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$, $\forall\ \varepsilon > 0$

CENTRAL LIMIT THEOREM:

Random sample $x_1, x_2, . . . , x_n$ (for each 'n') and $\dot{x}$ is the sample mean

[$\mu_{\dot{x}} = \mu$; $\sigma_{\dot{x}}^2 = \frac{\sigma^2}{n}$] $z_n = \frac{\dot{x} - \mu}{(\frac{\sigma}{\sqrt{n}})}$ [z_n has mean '0' and variance '1']

Then, $z_n \rightarrow z \sim N(0,1)$ {i.e., convergence in distribution}

STOCHASTIC PROCESS:

A Stochastic process {X(t);t∊T} is a collection of random variables

MARKOV CHAIN:

A stochastic process {x_n; n=0,1,2,….} with state space (Z, N or N U {0}) is called a Markov Chain if,

$$P(X_{n+1} = j \mid x_n = i, x_{n-1} = i_{n-1}, \cdots\cdots, x_0 = i_0) = P(x_{n+1} = j \mid x_n = i)$$

Note:

A Markov chain {x_n; n=0,1,2,….} is called *time homogenous* if

$P(x_{n+1} = j \mid x_n = i)$ is independent of n, for each i,j ϵ S