UNIT-I

Discrete and General probability space- Conditional probability space-Functions and Inverse Functions - Random Variables – Induced probability space by the Random variables.

Basic Definitions:

1. **Deterministic experiment**: An experiment whose outcome can be predicted (or known) with certainty is called Deterministic experiment.
2. **Random Experiment**: An experiment whose outcome cannot be predicted (or unknown) with certainty, although all the outcomes are known, is called Random experiment. Example: When a fair die is thrown, it is known that any of the 6 possible outcomes will occur, but it cannot be predicted.
3. In a random experiment, we associate a a set Ω which is a set of all possible outcomes of the experiment is known as **Sample Space**.
4. An outcome of an experiment is called **Sample point or Event** and is denoted by ω.An event is a subset of the Sample Space, Ω.
5. **Impossible event**: an event E is said to be impossible event if E=$\emptyset$, the null set. Ex.: getting no. 7 while throwing a die.
6. If the event set contains all the points of the sample space Ω it is called **sure or certain event**. Ex.: Getting any number form 1 to 6 while throwing a die.
7. Two events A & B are said to be **mutually exclusive** if their joint occurrence is not possible. .ie., $A \cap B = \emptyset$.
8. **Field**: A class A of subsets of Ω is called a field if it contains Ω itself and finite union ie., (i) $\Omega \in A$, (ii) $A \in A,\ A^C \in A$, (iii) If $A \text{ and } B \in A$, $AUB \in A$ *implies*
9. **σ-Field**: A class A of subsets of Ω is called a σ-field if it is under the formation of countable unions.
   Let Ω be the sample space and A be the set of all subset of Ω . Then A is called σ-Field if (i) For every $A \in A,\ A^C \in A$, (ii) $A_i \in A$, where i: 1,2,3,… ; $\sum A_i \in A$ (iii) If $\emptyset \in A$, $(iv) \Omega \in A$
   In Tossing a coin experiment, the σ-Field $A = \{\Omega, \emptyset, H, T\}$ where $\Omega = \{H, T\}$.
10. **Upper Bound**: Let S be a subset of R. An element 'b' is said to be an upper bound of S if x$\leq b;\ \forall x \in S$.
11. **Lower Bound**: Let S be a subset of R. An element 'a' is said to be a lower bound of S if a$\leq x;\ \forall x \in S$.

12. **Bounded Above**: A set having an upper bound is said to be bounded above.
13. **Bounded Below**: A set having a lower bound is said to be bounded below.
14. **Bounded Set**: A set having an upper bound and a lower bound is said to be bounded set.
15. Let S be a subset of R. A real number 'b' is said to be **the least upper bound or Supremum of S** if (i) 'b' is the upper bound of S and (ii) b $\leq c$; $\forall$ *upper bound* $c \in S$.
16. Let S be a subset of R. A real number 'a' is said to be the **greatest lower bound or Infimum of S** if (i) 'a' is the lower bound of S and (ii) a $\leq c$; $\forall$ *lower bound* $c \in S$.

A set cannot have more than one least upper bound or greatest lower bound.
Example: Let S = {1,2,3,4}. The element 1 and all the numbers less than 1 are lower bound of S.

Probability Space

In probability theory, a **probability space** or a **probability triple** (Ω, **F**, P) is a triplet that provides a formal model of a random experiment.

A probability space consists of three elements:

a) A **sample space**, Ω, which is the set of all possible outcomes.
b) An **event space**, which is a set of events , **F**, an event being a set of outcomes in the sample space.
c) A **probability function**, which assigns each event in the event space a probability, which is a number between 0 and 1.

A probability space is a mathematical triplet (Ω, **F**, P) that presents a model for a particular class of real-world situations.

- The sample space, Ω, is the set of all possible outcomes. An outcome is the result of a single trail of the random experiment. Outcomes may be states of nature, possibilities, experimental results, and the like. Every instance of the real-world situation (or run of the experiment) must produce exactly one outcome. If outcomes of different runs of an experiment differ in any way that matters, they are distinct outcomes.
- The σ-algebra , **F**, is a collection of all the events we would like to consider. This collection may or may not include each of the elementary events. Here, an "event" is a set of zero or more outcomes, i.e., a subset of the sample space. An event is considered to have "happened" during an experiment when the outcome of the latter is an element of the event. Since the same outcome may be a member of

many events, it is possible for many events to have happened given a single outcome.

- The probability measure , P, is a function returning an event's probability. A probability is a real number between zero and one (the event happens almost surely, with almost total certainty). Thus , P is a function P: **F** -> [0.,1]. The probability measure function must satisfy two simple requirements: First, the probability of a countable union of mutually exclusive events must be equal to the countable sum of the probabilities of each of these events. Second, the probability of the sample space must be equal to 1.

Definition

In short, a probability space is a measure space such that the measure of the whole space is equal to one.

The expanded definition is the following: a probability space is a triple $(\Omega, \mathcal{F}, P)$ consisting of:

- the sample space Ω — an arbitrary non-empty set,
- the σ-algebra $\mathcal{F} \subseteq 2^{\Omega}$ (also called σ-field) — a set of subsets of Ω, called events, such that:
  - $\mathcal{F}$ contains the sample space: $\Omega \in \mathcal{F}$,
  - $\mathcal{F}$ is closed under complements: if $A \in \mathcal{F}$, then also $(\Omega \setminus A) \in \mathcal{F}$,
  - $\mathcal{F}$ is closed under countable unions: if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then also $(\bigcup_{i=1}^{\infty} A_i) \in \mathcal{F}$
    - The corollary from the previous two properties and De Morgan's law is that $\mathcal{F}$ is also closed under countable intersections: if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then also $(\bigcap_{i=1}^{\infty} A_i) \in \mathcal{F}$
- the probability measure $P : \mathcal{F} \to [0, 1]$ — a function on $\mathcal{F}$ such that:
  - P is countably additive (also called σ-additive): if $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ is a countable collection of pairwise disjoint sets, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$,
  - the measure of entire sample space is equal to one: $P(\Omega) = 1$.

1.2 **Discrete and General probability space**

Discrete probability space:

In a sample space **Ω**, if the class of events A is generated by a countable partition of subsets of **Ω**, then (Ω, **A**, P) is called Discrete probability space.

If A = $\bigcup_1^{\infty}\{A_i\}$ $where\ A_i\ is\ the\ partition\ of$ **Ω** any event A is a countable union of $\{A_i\}$, the probability of A can be determined using the probabilities $of\ A_i\ ie., p_i = Pr(A)_i$.

If Ω is countable, A_i would be a singleton $\{w_i\}$ (i:1,2,….). The set $\{p_i;\ i = 1,2,\ldots\}$ is called the probability of Ω and A coincides with the power set.

General probability space

Let A be the system of events; The elements of A be measurable sets. (Ω, A) be the measurable space; P be the probability measure over the measurable space (Ω, A).

The triplet (Ω, A, P) with Ω as non-empty set, A as σ-field over Ω and a mapping
P : A ⟶ R+ with i) P(Ω) = 1 (normalization),
ii) P(A) ≥ 0 (non-negativity),
iii) For any sequence $\{A_n\}$, mutually disjoint sets A we have
$P[\cup_1^n A_n] = \sum_1^n P[A_n]$ (σ-additivity)
is called general probability space; Ω is called basic space.

Finite Probability Spaces

A finite probability space is a finite set Ω ≠ ∅ together with a function P : Ω ⟶ R+ such that
i) ∀ ω∈ Ω, P(ω) > 0

ii) ∀ω∈Ω P(ω) = 1.

The set Ω is the sample space and the function P is the probability function. The elements ω ∈ Ω are called atomic events or elementary events. An event is a subset of Ω. For A ⊆ Ω, we define the probability of A to be P(A)= $\sum_{\omega\in A}$ P(ω). For atomic events

we have P({ω}) = P(ω); and P(∅) = 0, P(Ω) = 1.

The trivial events are those with probability 0 or 1, i. e. ∅ and Ω.

The uniform distribution over the sample space Ω is defined by setting P(ω) = 1/|Ω| for every ω ∈ Ω. With this distribution, in the uniform probability space over Ω, the calculation of probabilities amounts to counting: P(A) = |A|/|Ω|.

Conditional Probability Space

The concept of a conditional probability space Ω to be defined below is identical with the more standard concept of a σ-finite measure space. The term was introduced by Renyi (1970).

Definition:

A conditional probability space is a set Ω equipped with a σ-finite measure P defined on a σ-algebra E of sets in Ω. The members of the σ-algebra are called the events. An event A is an elementary condition if 0 < P(A) < 1. The conditional probability, given an elementary condition, is defined by

P(B | A) = P(A ∩ B) / P(A)

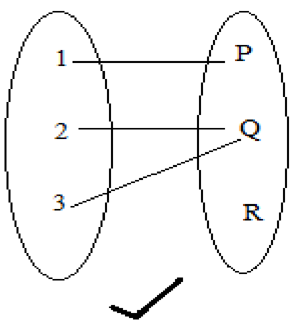
A probability space is a conditional probability space such that P(Ω) = 1. An improper probability is a measure such that P(Ω) = 1.

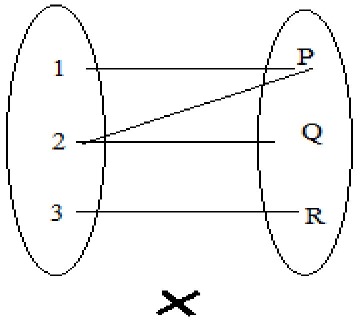
Functions and Inverse Functions

Functions

A function is a relation between a set of inputs and a set of permissible outputs with the property that each input is related to exactly one output.

Let A & B be any two non-empty sets, mapping from A to B will be a function only when every element in set A has one end only one image in set B.





Another definition of functions is that it is a relation "f" in which each element of set "A" is mapped with only one element belonging to set "B". Also in a function, there can't be two pairs with the same first element.

A Condition for a Function:

- Set **A** and Set **B** should be non-empty.
- In a function, a particular input is given to get a particular output. So, A function **f: A->B** denotes that f is a function from **A** to **B**, where **A** is a **domain** and **B** is a **co-domain**.

- For an element, **a**, which belongs to **A**, $a \epsilon A$, a unique element **b**, $b \epsilon B$ is there such that (a,b) ϵ f.

The unique element **b** to which **f** relates **a**, is denoted by **f(a)** and is called **f of a**, or the value of **f at a**, or **the image of a under f.**

- The *range* of **f** (image of **a** under f)
- It is the set of all values of **f(x)** taken together.
- **Range of f** = { y ϵ Y | y = f (x), for some x in X}

A real-valued function has either **P** or any one of its subsets as its range. Further, if its domain is also either **P** or a subset of **P**, it is called a **real function**.

Functions

A function f : X → Y between sets X, Y assigns to each x ∈ X a unique element f(x) ∈ Y . Functions are also called maps, mappings, or transformations. The set X on which f is defined is called the domain of f and the set Y in which it takes its values is called the codomain. We write f : x → f(x) to indicate that f is the function that maps x to f(x).

Example . Let A = {2, 3, 5, 7, 11} , B = {1, 3, 5, 7, 9, 11}. We can define a function **f : A → B** by f(2) = 7, f(3) = 1, f(5) = 11, f(7) = 3, f(11) = 9, and a function g : B → A by g(1) = 3, g(3) = 7, g(5) = 2, g(7) = 2, g(9) = 5, g(11) = 11.

Inverse Functions:

1. Onto function

If the range of a function is equal to the co-domain then the function is called an onto function. Otherwise it is called an into function.

In f:A→B, the range of f or the image set f(A) is equal to the co-domain B i.e. f(A) = B then the function is onto.

Example 7.10

Let A = {1, 2, 3, 4}, B = {5, 6}. The function f is defined as follows: $f(1) = 5$, $f(2) = 5$, $f(3) = 6$, $f(4) = 6$. Show that f is an onto function.

Solution:

$f = \{(1, 5), (2, 5), (3, 6), (4, 6)\}$

The range of f, $f(A) = \{5, 6\}$

co-domain $B = \{5, 6\}$

i.e. $f(A) = B$

⇒ the given function is onto

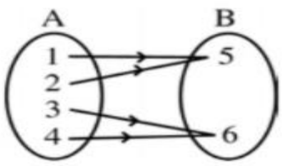


Fig 7. 12

Example 7.11: Let X = {a, b}, Y = {c, d, e} and $f = \{(a, c), (b, d)\}$. Show that f is not an onto function.

Solution:

Draw the diagram

The range of f is {c, d}

The co-domain is {c, d, e}

The range and the co-domain are not equal, and hence the given function is not onto

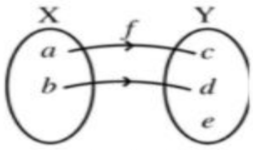


Fig 7. 13

Note :

(1) For an onto function for each element (image) in the co-domain, there must be a corresponding element or elements (pre-image) in the domain.

(2) Another name for onto function is surjective function.

Definition: A function f is onto if to each element b in the co-domain, there is atleast one element a in the domain such that $b = f(a)$

2. One-to-one function:

A function is said to be one-to-one if each element of the range is associated with exactly one element of the domain.

i.e. two different elements in the domain (A) have different images in the co-domain (B).

i.e. $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$ $a_1, a_2 \in A$,

Equivalently $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$

The function defined in 7.11 is one-to-one but the function defined in 7.10 is not one-to-one.

Example 7.12: Let A = {1, 2, 3}, B = {a, b, c}. Prove that the function f defined by f = {(1, a), (2, b), (3,c)} is a one-to-one function.

Solution:

Here 1, 2 and 3 are associated with a, b and c respectively.

The different elements in A have different images in B under the function f. Therefore f is one-to-one.

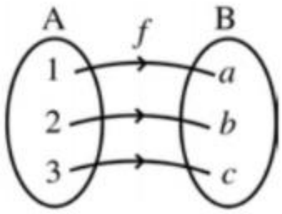


Fig 7. 14

Example 7.13: Show that the function $y = x^2$ is not one-to-one.

Solution:

For the different values of x (say 1, – 1) we have the same value of y. i.e. different elements in the domain have the same element in the co-domain. By definition of one-to-one, it is not one-to-one (OR)

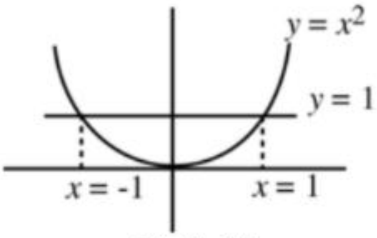


Fig 7. 15

$$y = f(x) = x^2$$
$$f(1) = 1^2 = 1$$
$$f(-1) = (-1)^2 = 1$$
$$\Rightarrow \quad f(1) = f(-1)$$

But $1 \neq -1$. Thus different objects in the domain have the same image.

$\therefore$ The function is not one-to-one.

Note: (1) A function is said to be injective if it is one-to-one.

(2) It is said to be bijective if it is both one-to-one and onto.

(3) The function given in example 7.12 is bijective while the functions given in 7.10, 7.11, 7.13 are not bijective.

3. Identity function:

A function f from a set A to the same set A is said to be an identity function if $f(x) = x$ for all $x \in$ A i.e. f : A $\rightarrow$ A is defined by $f(x) = x$ for all $x \in$ A. Identity function is denoted by I_A or simply I. Therefore $I(x) = x$ always.

Graph of identity function:

The graph of the identity function $f(x) = x$ is the graph of the function $y = x$. It is nothing but the straight line $y = x$ as shown in the fig. (7.16)

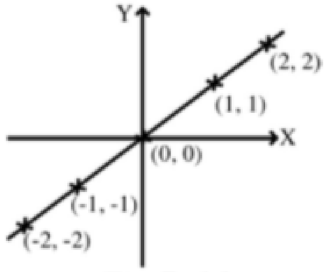


Fig 7. 16

4. Inverse of a function:

To define the inverse of a function f i.e. f^{-1} (read as 'f inverse'), the function f must be one-to-one and onto.

Let A = {1, 2, 3}, B = {a, b, c, d}. Consider a function f = {(1, a), (2, b), (3, c)}. Here the image set or the range is {a, b, c} which is not equal to the co-domain {a, b, c, d}. Therefore, it is not onto.

For the inverse function f^{-1} the co-domain of f becomes domain of f^{-1}.

i.e. If f : A → B then f^{-1} : B → A . According to the definition of domain, each element of the domain must have image in the co-domain. In f^{-1}, the element 'd' has no image in A. Therefore f^{-1} is not a function. This is because the function f is not onto.

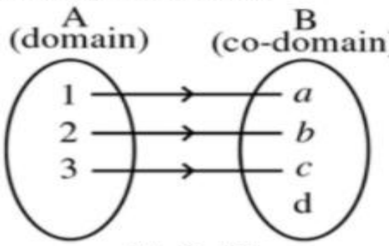


Fig 7. 17 a

$f(1) = a$
$f(2) = b$
$f(3) = c$
All the elements in A have images

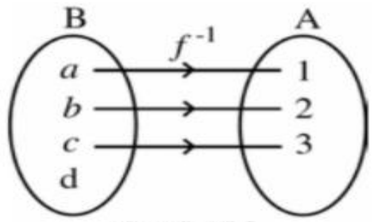


Fig 7.17 b

$f^{-1}(a) = 1$
$f^{-1}(b) = 2$
$f^{-1}(c) = 3$
$f^{-1}(d) = ?$
The element d has no image.

Again consider a function which is not one-to-one. i.e. consider f = {(1, a), (2, a), (3, b)} where A = {1, 2, 3}, B = {a, b}

Here the two different elements '1' and '2' have the same image 'a'. Therefore the function is not one-to-one.

The range = {a, b} = B. ∴ The function is onto.

Fig 7. 18

$f(1) = a$
$f(2) = a$
$f(3) = b$

Here all the elements in A has unique image

$f^{-1}(a) = 1$
$f^{-1}(a) = 2$
$f^{-1}(b) = 3$

The element 'a' has two images 1 and 2. It violates the principle of the function that each element has a unique image.

This is because the function is not one-to-one.

Thus, 'f^{-1} exists if and only if f is one-to-one and onto'.

Note:

(1) Since all the function are relations and inverse of a function is also a relation. We conclude that for a function which is not one-to-one and onto, the inverse f^{-1} does not exist

(2) To get the graph of the inverse function, interchange the co-ordinates and plot the points.

To define the mathematical definition of inverse of a function, we need the concept of composition of functions.

5. Composition of functions:

Let A, B and C be any three sets and let $f : A \to B$ and $g : B \to C$ be any two functions. Note that the domain of g is the co-domain of f. Define a new function $(gof) : A \to C$ such that $(gof)(a) = g(f(a))$ for all $a \in A$. Here $f(a)$ is an element of B. $\therefore$ $g(f(a))$ is meaningful. The function gof is called the composition of two functions f and g.

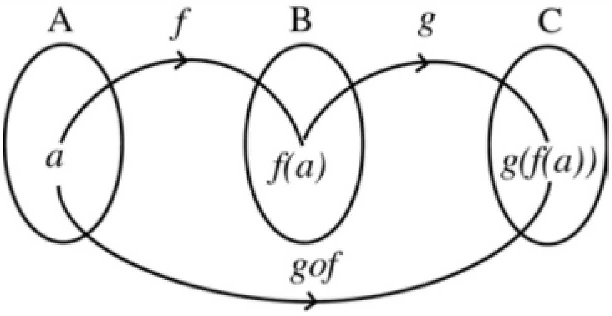


Fig 7. 19

Note:

The small circle o in gof denotes the composition of g and f

Example 7.15: Let A = {1, 2}, B = {3, 4} and C = {5, 6} and $f : A \to B$ and $g : B \to C$ such that $f(1) = 3$, $f(2) = 4$, $g(3) = 5$, $g(4) = 6$. Find gof.

Solution:

gof is a function from $A \to C$.

Identify the images of elements of A under the function gof.

$(gof)(1) = g(f(1)) = g(3) = 5$

$(gof)(2) = g(f(2)) = g(4) = 6$

i.e. image of 1 is 5 and image of 2 is 6 under gof

$\therefore$ $gof = \{(1, 5), (2, 6)\}$

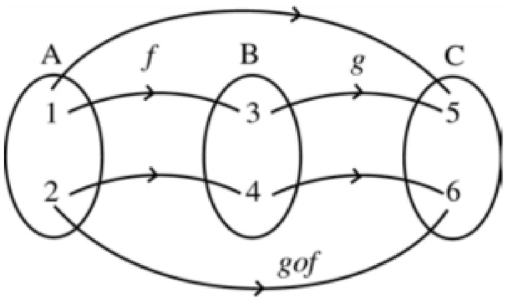


Fig 7. 20

Note:

For the above definition of f and g, we can't find fog. For some functions f and g, we can find both fog and gof. In certain cases fog and gof are equal. In general $fog \neq gof$ i.e. the composition of functions need not be commutative always.

Example 7.16: The two functions $f : R \to R$, $g : R \to R$ are defined by

$f(x) = x^2 + 1$, $g(x) = x - 1$. Find fog and gof and show that $fog \neq gof$.

Solution:

$$(fog)(x) = f(g(x)) = f(x-1) = (x-1)^2 + 1 = x^2 - 2x + 2$$

$$(gof)(x) = g(f(x)) = g(x^2+1) = (x^2+1) - 1 = x^2$$

Thus $(fog)(x) = x^2 - 2x + 2$

$(gof)(x) = x^2$

$\Rightarrow \quad fog \neq gof$

Example 7.17: Let $f, g : R \to R$ be defined by $f(x) = 2x + 1$, and $g(x) = \frac{x-1}{2}$.

Show that $(fog) = (gof)$.

Solution:

$$(fog)(x) = f(g(x)) = f\left(\frac{x-1}{2}\right) = 2\left(\frac{x-1}{2}\right) + 1 = x - 1 + 1 = x$$

$$(gof)(x) = g(f(x)) = g(2x+1) = \frac{(2x+1)-1}{2} = x$$

Thus $(fog)(x) = (gof)(x)$

$\Rightarrow \quad fog = gof$

In this example f and g satisfy $(fog)(x) = x$ and $(gof)(x) = x$

Consider the example 7.17. For these f and g, $(fog)(x) = x$ and $(gof)(x) = x$. Thus by the definition of identity function $fog = I$ and $gof = I$ i.e. $fog = gof = I$

Now we can define the inverse of a function f.

Definition:

Let $f : A \to B$ be a function. If there exists a function $g : B \to A$ such that $(fog) = I_B$ and $(gof) = I_A$, then g is called the inverse of f. The inverse of f is denoted by f^{-1}

Note:

(1) The domain and the co-domain of both f and g are same then the above condition can be written as $fog = gof = I$.

(2) If f^{-1} exists then f is said to be invertible.

(3) $f o f^{-1} = f^{-1} o f = I$

Example 7.18: Let $f : R \to R$ be a function defined by $f(x) = 2x + 1$. Find f^{-1}

Solution:

Let $g = f^{-1}$

$(gof)(x) = x \qquad \because gof = I$

$g(f(x)) = x \quad \Rightarrow \quad g(2x+1) = x$

Let $2x + 1 = y \quad \Rightarrow \quad x = \frac{y-1}{2}$

$\therefore \quad g(y) = \frac{y-1}{2}$ or $f^{-1}(y) = \frac{y-1}{2}$

Replace y by x

$$f^{-1}(x) = \frac{x-1}{2}$$

Random Variable

A ***random variable***, usually written *X*, is a variable whose possible values are numerical outcomes of a random phenomenon.

A **random variable** is a measurable function $X : \Omega \to E$ from a set of possible outcomes Ω to a measurable space E. The technical axiomatic definition requires Ω to be a sample space of a probability triple $(\Omega, \mathcal{F}, \mathrm{P})$ (see the measure-theoretic definition). A random variable is often denoted by capital roman letters such as X, Y, Z, T.[3][4]

The probability that X takes on a value in a measurable set $S \subseteq E$ is written as

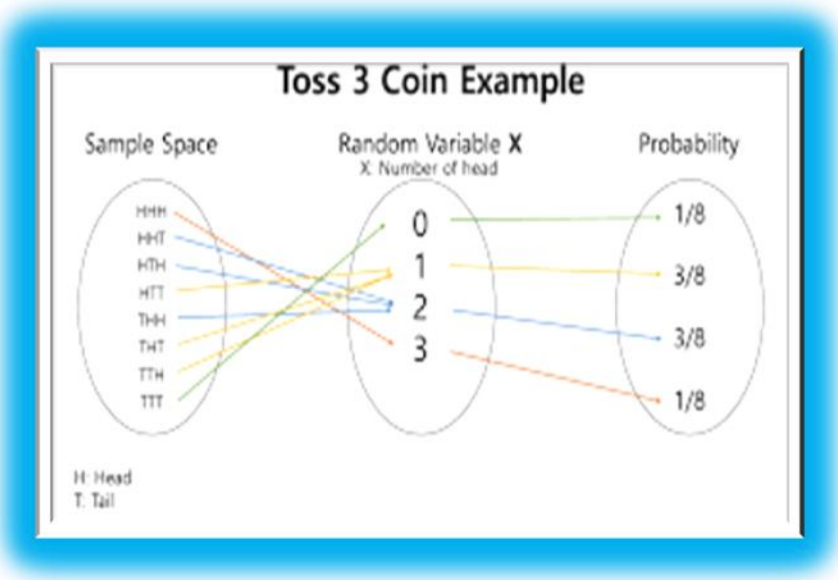
$$\mathrm{P}(X \in S) = \mathrm{P}(\{\omega \in \Omega \mid X(\omega) \in S\})$$[3]

There are two types of random variables, ***discrete*** and ***continuous***.

Discrete Random Variables

A ***discrete random variable*** is one which may take on ***only a countable number of distinct values*** such as 0,1,2,3,4,........ . Discrete random variables are usually (but not necessarily) counts. If a random variable takes only a finite number of distinct values, then it is called discrete random variable.

Example:- Consider an experiment where a coin is tossed three times. If X represents the number of times that the coin comes up heads, then X is a discrete random variable that can only have the values 0, 1, 2, 3 (from no heads in three successive coin tosses to all heads). No other value is possible for X.



The ***probability distribution*** of a discrete random variable is a list of probabilities associated with each of its possible values. It is also called the probability function or the probability mass function.

Suppose a random variable X may take k different values, with the probability that $X = x_i$ defined to be $P(X = x_i) = p_i$. The probabilities p_i must satisfy the following:

i) $0 \leq p_i \leq 1$ *for each i*

ii) $p_1 + p_2 + \ldots + p_k = 1$.

Probability distribution for the random experiment of toss a coin 3 times:

No. of Heads, x	0	1	2	3
Probability, Pr(X=x)=p(x)	Pr(X=0)=p(0) =1/8	3/8	3/8	1/8
Cum. Probability $F_X(x)$=F(x) =F(X<=x)	F(X<=0) =Pr(X<=0) =1/8	Pr(X<=1) =3/8	Pr(X<=2) =7/8	Pr(X<=3) =1

where X is the random variable; p(x)-probability mass function

Examples of discrete random variables:

the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten.

Example

Suppose a variable X can take the values 1, 2, 3, or 4. The probabilities associated with each outcome are:

Outcome	1	2	3	4
Probability	0.1	0.3	0.4	0.2

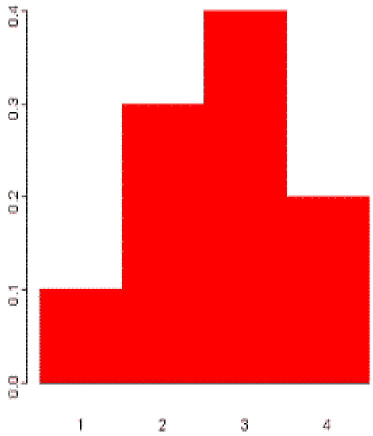
The probability that X is equal to 2 or 3

$= P(X = 2 \text{ or } X = 3)$

$= P(X = 2) + P(X = 3)$

$= 0.3 + 0.4 = 0.7.$

Similarly, $P(X \geq 1) = 1 - P(X = 1) = 1 - 0.1 = 0.9$, by the complement rule.



This distribution may also be described by the ***probability histogram*** shown in the diagram.

Cumulative Distribution Function

All random variables (discrete and continuous) have a ***cumulative distribution function***. It is a function giving the probability that the random variable X is less than or equal to x, for every value x. For a discrete random variable, the cumulative distribution function is found by summing up the probabilities.

Example

The cumulative distribution function for the above probability distribution is:

$P(X \leq 1)=0.1$,

$P(X \leq 2) = 0.1 + 0.3 = 0.4$,

$P(X \leq 3) = 0.1 + 0.3 + 0.4 = 0.8$,

$P(X \leq 4) = 0.1 + 0.3 + 0.4 + 0.2 = 1.0$.

Continuous Random Variables

Continuous random variables can represent any value within a specified range or interval and can take on an infinite number of possible values.

Example:- Consider an experiment that involves measuring the amount of rainfall in a city over a year or the average height of a random group of 25 people.

A ***continuous random variable*** is a variable which takes an infinite number of possible values. Continuous random variables are usually measurements. Example: height, weight, the amount of sugar in an orange, the time required to run a mile.

A continuous random variable is not defined at specific values. Instead, it is defined over an *interval* of values, and is represented by the ***area under a curve*** (this is known as an *integral*). The probability of observing any single value is equal to 0, since the number of values which may be assumed by the random variable is infinite.

Suppose a random variable X may take all values over an interval of real numbers. Then the probability that X is in the set of outcomes A, $P(A)$, is defined to be the area above A and under a curve. The curve, which represents a function $p(x)$, must satisfy the following:

i) The curve has no negative values ($p(x) \geq 0$ for all x)

ii) The total area under the curve is equal to 1.

A curve meeting these requirements is known as a ***probability density curve***.

Probability Space induced by a random variable

Let (Ω, F, P) be a probability space and $X : \Omega \to R_n$ be a random variable.

Define $\mu_X : B \to [0, 1]$, $B \ni B \to P(X^{-1}(B)) =: \mu_X(B)$. It is possible to verify that μ_X defines a probability measure on (R_n, B). We have thus an "induced" probability space (R_n, B, μ_X) (induced probability space by the random variable X).

- $\mathcal{E}$: given random experiment;
- $(\Omega, \mathcal{P}(\Omega), P)$: probability space associated with $\mathcal{E}$;
- In many situations we may not be directly interested in sample space Ω; rather we may be interested in some numerical aspect of sample space (i.e., we may be interested in a real-valued function defined on sample space Ω).

Example 1:

- $\mathcal{E}$: Tossing a fair can three times independently;
- Sample space $\Omega = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{H, T\}, i = 1, 2, 3\}$; here, in $(\omega_1, \omega_2, \omega_3)$, ω_i $(i = 1, 2, 3)$ indicates the outcome of i^{th} toss. Clearly the sample space has $2^3 = 8$ elements;
- Suppose we are interested in number of heads obtained in three tosses, i.e., we are interested the function $X : \Omega \to \mathbb{R}$, where

$$X(w_1, w_2, w_3) = \begin{cases} 0, & \text{if } (\omega_1, \omega_2, \omega_3) = (T, T, T) \\ 1, & \text{if } (\omega_1, \omega_2, \omega_3) \in \{(H, T, T), (T, H, T), (T, T, H)\} \\ 2, & \text{if } (\omega_1, \omega_2, \omega_3) \in \{(H, H, T), (H, T, H), (T, H, H)\} \\ 3, & \text{if } (\omega_1, \omega_2, \omega_3) = (H, H, H) \end{cases}$$

Definition 1: A real valued function $X : \Omega \to \mathbb{R}$ is called a random variable (r.v.).

Notations:

- $\mathcal{P}(\mathbb{R})$: power set of the real line $\mathbb{R}$;
- For a r.v. X

$$\{X \in A\} \doteq X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}, \quad A \in \mathcal{P}(\mathbb{R}).$$

For example, for a real constant c,

$$\{X = c\} = X^{-1}(\{c\}) = \{\omega \in \Omega : X(\omega) = c\};$$

$$\{X \le c\} = X^{-1}((-\infty, c]) = \{\omega \in \Omega : X(\omega) \le c\};$$

$$\{X > c\} = X^{-1}((c, \infty)) = \{\omega \in \Omega : X(\omega) > c\}.$$

Result 1: Let X be a r.v. Then

(a)
$$X^{-1}\Big(\bigcap_{\alpha \in \Lambda} A_\alpha\Big) = \bigcap_{\alpha \in \Lambda} X^{-1}(A_\alpha);$$

(b)
$$X^{-1}\Big(\bigcup_{\alpha \in \Lambda} A_\alpha\Big) = \bigcup_{\alpha \in \Lambda} X^{-1}(A_\alpha);$$

(c)
$$X^{-1}(A^c) = (X^{-1}(A))^c$$

(d)
$$A \cap B = \phi \quad \Rightarrow \quad X^{-1}(A) \bigcap X^{-1}(B) = \phi.$$

Induced probability space

- X: a given r.v. on probability space $(\Omega, \mathcal{P}(\Omega), P)$;
- Define the set function $P_X : \mathcal{P}(\Omega) \to [0,1]$ as

$$\begin{aligned} P_X(A) &= P(X^{-1}(A)) \\ &\doteq P(\{\omega \in \Omega : X(\omega) \in A\}), \quad A \in \mathcal{P}(\Omega). \end{aligned}$$

Result 2: The set function $P_X(.)$ defined above is a probability function on $\mathcal{P}(\mathbb{R})$, i.e., $(\mathbb{R}, \mathcal{P}(\mathbb{R})), P_X)$ is a probability space.

Proof: Since, $P(\cdot)$ is a probability function

$$P_X(A) = P(X^{-1}(A)) \geq 0, \quad \forall\ A \in \mathcal{P}(\mathbb{R}).$$

Let $\{A_i : i \in S\}$ be a countable collection of disjoint events in $\mathcal{P}(\mathbb{R})$. Then

$$\begin{aligned} P_X(\bigcup_{i\in S} A_i) &= P(X^{-1}(\bigcup_{i\in S} A_i)) \\ &= P(\bigcup_{i\in S} X^{-1}(A_i)) \quad \text{(using Result 1 (b))} \\ &= \sum_{i\in S} P(X^{-1}(A_i)) \quad (X^{-1}(A_i)\text{s are disjoint}) \\ &= \sum_{i\in S} P_X(A_i). \end{aligned}$$

Also

$$\begin{aligned} P_X(\mathbb{R}) &= P(X^{-1}(\mathbb{R})) \\ &= P(\Omega) \\ &= 1. \end{aligned}$$

Remark 1:

(a) The probability space $(\mathbb{R}, \mathcal{P}(\mathbb{R}), P_X)$ is called the probability space induced by r.v. X and the probability function $P_X(\cdot)$ is called the probability function induced by r.v. X.

(b) Given a r.v. X, we are generally no longer interested in the original probability space $(\Omega, \mathcal{P}(\Omega), P)$; rather we are then, interested in induced probability space $(\mathbb{R}, \mathcal{P}(\mathbb{R}), P_X)$. We have

$$X : (\Omega, \mathcal{P}(\Omega), P) \quad \to \quad (\mathbb{R}, \mathcal{P}(\mathbb{R}), P_X),$$

where

$$\begin{aligned} P_X(A) &= P(X^{-1}(A)) \\ &= P(\{\omega \in \Omega : \quad X(\omega) \in A\}), \quad A \in \mathcal{P}(\mathbb{R}). \end{aligned}$$

Example 2:

- $\mathcal{E}$: a fair coin is tossed three times independently;
- Sample space $\Omega = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{H, T\}, \quad i = 1, 2, 3\}$;
- Suppose we are interested in number of heads in three tosses of coin, i.e., we are interested in r.v. $X : \Omega \to \mathbb{R}$, defined by

$$X(\omega) = \begin{cases} 0, & \text{if } \omega = (T, T, T) \\ 1, & \text{if } \omega \in \{(H, T, T), (T, H, T), (T, T, H)\} \\ 2, & \text{if } \omega \in \{(H, H, T), (H, T, H), (T, H, H)\} \\ 3, & \text{if } \omega = (H, H, H) \end{cases}.$$

We have

$$\begin{aligned} P_X(\{0\}) &= P(X^{-1}(\{0\})) \\ &= P(\{(T, T, T)\}) \\ &= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}; \\ P_X(\{1\}) &= P(X^{-1}(\{1\})) \\ &= P(\{(H, T, T), (T, H, T), (T, T, H)\}) \\ &= \frac{3}{8}; \\ P_X(\{2\}) &= P(X^{-1}(\{2\})) \\ &= P(\{(H, H, T), (H, T, H), (T, H, H)\}) \\ &= \frac{3}{8}; \\ P_X(\{3\}) &= P(X^{-1}(\{3\})) \\ &= P(\{(H, H, H)\}) \\ &= \frac{1}{8} \end{aligned}$$

For $A \subseteq \mathcal{P}(\mathbb{R})$

$$\begin{aligned} P_X(A) &= P(X^{-1}(A)) \\ &= \sum_{\omega \in A \cap \{0,1,2,3\}} P_X(\{\omega\}) \end{aligned}$$

End of Unit-I