6. Probability

Introduction

In our day-to-day life, the term probability is a measure of one's belief in the occurrence of a future event. This measure can range from 0 to 1, where 0 implying "impossible" and 1 implying "certain" or "sure". This is also expressed as the "chance" of occurrence of the event. In this case the scale is from 0% to 100%.

Consider the following events:

1. {Heavy rain in Colombo tomorrow}
2. {A tsunami occurring tomorrow in some part of Sri Lanka}
3. {no sun-rise for tomorrow}
4. {A death occurring somewhere in Sri Lanka tomorrow}
5. {an accident reported in Colombo tomorrow}

Can you assign **reasonable probabilities** to these events?

Some reasonable answers: (1) **.20** (possible, but less probable)
(2) **.00001** (we now know that it is not impossible, but is highly improbable)
(3) **0** (it is an impossible event)
(4) **.9999** (very highly probable)
(5) **.55** (probable event)

[Note: The answer to question 2 would have been different if it was asked in 2003. We would definitely assign **0** thinking that it is an impossible event]

Probability is always associated with "**random**" experiments, i.e. experiments with several possible outcomes but the outcome on any single trial can not be predicted in advance. In event 4 above, if the "experiment" is to observe the number of deaths reported in Sri Lanka tomorrow, the possible outcomes are: 0, 1, 2, 3, …., and so on. This is not an infinite sequence of numbers. It will be a finite sequence.

How do we assign probabilities to (random) events?
Is there a method (or methods) that we can follow?
Answer is yes, but the method will depend on the way probability is interpreted.

Interpretation of Probability

1. **Relative Frequency Interpretation**:

   Probability of a single outcome of a random experiment is interpreted as the relative frequency of the outcome if the experiment were repeated a "large number of times" under "similar conditions".

   Shortcomings: (i) How large is "large"? (ii) No experiment can be repeated under similar conditions!

<u>Example</u>: Suppose a coin was tossed 10,000 times and Heads (H) appeared 9000 times.
Now, the relative frequency of H is 9000/10000 = 0.9 . Therefore, it is reasonable to assign 0.9 as the **probability of H in any single toss <u>for this coin</u>**.
On the other hand, if Heads appeared say, 5010 times, then the relative frequency of H is 5010/10000 = .501 ≈ .5 .
Here, it is very reasonable to assume that the coin is balanced (and assign .5 for probability of H).

Note: A natural phenomenon is that when the number of trials gets bigger and bigger the relative frequency of an outcome tends to converge to some constant value. This value is defined as the probability of that outcome in a single trial.

2. **Classical Interpretation**:

This is based on the concept of "equally likely" outcomes. Suppose a random experiment has *n* possible outcomes which are equally likely. [For example, a <u>balanced</u> die has 6 equally likely outcomes]. Then, **each possible outcome has probability** $\frac{1}{n}$ **.**

Shortcomings: (i) How to assign probabilities to outcomes that are not equally likely? There is no systematic way! (ii) What if *n* is infinitely many?

3. **Subjective Interpretation**:

According to this, probability assigned to an outcome by a person depends on that person's <u>own belief and information</u> about the likelihood of that outcome. [For example, if a gambler has information that a die is weighted so that "6" is more likely than any other number, he would assign a probability greater than $\frac{1}{6}$ for "6" and try his bets on "6"].

Shortcomings: (i) No objective basis for 2 or more scientists to reach a common agreement about the knowledge of an outcome. (ii) Cannot assign probabilities consistently to infinitely many events.

<u>Example</u>: Suppose neurosurgeon-A has a history of 180 successful cases of a type of brain surgery out of 200 that he carried out. If a new patient consults him and asks about the success rate for this type of brain surgery, he would say 90%. **This is purely based on his information, for him**! On the other hand, neurosurgeon-B may say it is 75%, simply based on his surgical history.

6.1 Sampling and Descriptive Statistics

6.1.1 Measures of central tendencies

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. The Mean, Median and Mode are all valid measures of central tendency.

Mean (Arithmetic mean):

$$\text{Mean} = \frac{\sum x_i}{N}$$

where $\sum x_i$ is the sum of all the values in the data set and N is the number of values in the data set. The mean of the population and a sample are denoted by μ and $\bar{x}$ respectively.

Median:
The median is the middle score for a set of data that has been arranged in order of magnitude.

Mode:
The mode is the most frequent score in our data set.

6.1.2 Measures of dispersion

It tells the variation of the data from one another. The Range, Variance, Standard deviation and the Coefficient of Variation are the most commonly used measures of dispersion.

Range:
The range is the difference between two extreme observations of the data set. If X_{max} and X_{min} are the maximum and the minimum values respectively, then

$$\text{Range} = X_{max} - X_{min}$$

Variance:
The variance is the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean.

$$\text{Population Variance } \sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

$$\text{Sample Variance } s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

where x_i's are the observations of the data set, μ and $\bar{x}$ are the mean of the population and the sample respectively, N and n are the size of the population and the sample respectively.

Standard deviation:
A standard deviation is the positive square root of the variance and that of the population and a sample are denoted by σ and s respectively.

Coefficient of Variation:
Coefficient of Variation is the ratio of the standard deviation to the mean.

$$\text{Coefficient of Variation} = \frac{\text{standard deviation}}{mean}$$

6.2 Sample space and events

Definitions:

The ***sample space*** associated with a (random) experiment is the set S of all possible outcomes. Individual outcomes in a sample space are called **sample points.** These are also called **simple events** or **elementary outcomes**.

A ***discrete sample space*** contains either a finite or countably infinite number of sample points.

An ***event*** A is any subset of the sample space S. A **compound event** consists of a collection of sample points whereas a simple event consists of only a single sample point. If S is ***discrete***, then every subset of S is an event. This may not be true for ***nondiscrete*** sample spaces which are not covered in this course.

We can combine events to form new events using various set operations:

(i) $A \cup B$ is the event that occurs if and only if A occurs *or* B occurs (or both).

(ii) $A \cap B$ is the event that occurs if and only if A occurs and B occurs.

(iii) A^c $(or\ \overline{A})$ is the complement of A, that occurs if and only if A does *not* occur.

6.3 Axioms of probability

Let S be the sample space associated with a random experiment. For every event A in S (i.e. for every subset of S) we assign a number, P(A), called the "***probability*** of A" so that the following 3 axioms hold:

1. P(A) $\geq$ 0;
2. P(S) = 1
3. If A_1, A_2, A_3, …. form a sequence of pairwise mutually exclusive events in S (i.e. $A_i \cap A_j = 0$; $i \neq j$) then,

$$P(A_1 \cup A_2 \cup A_3 \,......) = \sum_{i=1}^{\infty} P(A_i).$$

***It can be shown that** axiom 3 implies a similar result for any finite sequence.
That is,

$$P(A_1 \cup A_2 \cup A_3 \cup .... \cup A_n) = \sum_{i=1}^{n} P(A_i).$$

Note: Above definition only states the conditions that must be satisfied by probabilities once they are assigned to events; it does not tell us how to assign probabilities!

Properties of Probability Function, P

Some theorems that follow directly from the above axioms:

Theorem 1. If Φ is the empty set, then P(Φ) = 0.

Proof: Let *A* be any set. Then, $A \cap \Phi = \Phi$ which implies that *A* and Φ are disjoint, and also $A \cup \Phi = A$. Now by axiom 3,
$P(A) = P(A \cup \Phi) = P(A) + P(\Phi)$ which implies that P(Φ) = 0.

Theorem 2. If $\overline{A}$ is the complement of an event *A*, then P($\overline{A}$) = 1 – P(*A*).

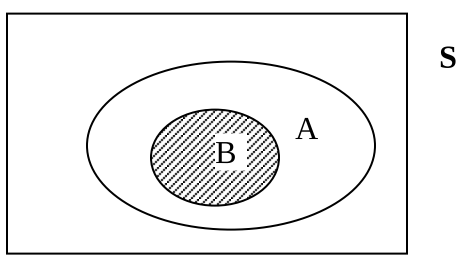
Proof: Note that the sample space S = $A \cup \overline{A}$ where *A* and $\overline{A}$ are mutually exclusive. Then, by axioms 2 and 3,
1= P(S) = P($A \cup \overline{A}$) = P(*A*) + P($\overline{A}$) which implies that P($\overline{A}$) = 1 – P(*A*).

Theorem 3. Let *A* be any event. Then, $P(A) \leq 1$.

Proof: By Theorem 2 we got P($\overline{A}$) = 1 – P(*A*). But, by axiom 1, P($\overline{A}$) ≥ 0 since $\overline{A}$ is any event. Therefore, we get 1 – P(*A*) ≥ 0 which implies $P(A) \leq 1$.

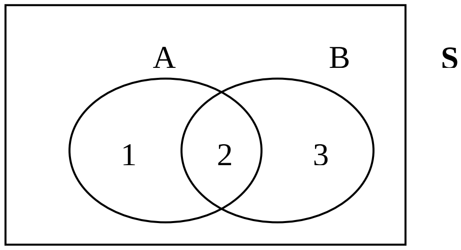
Theorem 4. If $B \subset A$, then $P(B) \leq P(A)$

Proof:



$$A = B \cup (A \cap \overline{B})$$
$\therefore\ P(A) = P(B) + P(A \cap \overline{B})$, since disjoint
$$\therefore\ P(A) - P(B) = P(A \cap \overline{B}) \geq 0$$
$$\therefore\ P(A) \geq P(B)$$

Theorem 5. If A and B are any two events, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Proof:

Note that $A \cup B$ can be written as $A \cup B = A \cup (\bar{A} \cap B)$ which is a union of 2 disjoint events. Now, by axiom 3, $P(A \cup B) = P(A) + P(\bar{A} \cap B) = P(A) + [P(B) - P(A \cap B)]$.
Hence the proof.

6.4 Mutually Exclusive Events

Definition: **Mutually Exclusive** (or **disjoint**) Events

Two sets A and B are said to be ***mutually exclusive*** (or ***disjoint***) if $A \cap B = \Phi$. That is, mutually exclusive sets have no sample points in common.

Example: Consider an experiment of observing the blood type (with Rh factor) of a randomly picked person.
Then, S = {A^+, A^-, B^+, B^-, O^+, O^-, $(AB)^+$, $(AB)^-$}
Define A = {person has blood type A} and
B = {person has blood type B}. Then, A = { A^+, A^-} and B = { B^+, B^-}. Clearly A and B are disjoint events.

6.5 Finite probability spaces

Let S be a finite sample space, say, $S = \{a_1, a_2, ..., a_n\}$. A finite probability space is obtained by assigning to each point $a_i \in S$ a real number p_i called the **probability** of a_i, [written $P(a_i)$] satisfying:

(i) each $p_i \geq 0$, and
(ii) sum of p_i's =1.

The probability $P(A)$ of any event A is the sum of the probabilities of the points of A.

EXAMPLE:

Consider tossing 3 coins and recording the number of heads. The sample space is: $S = \{0,1,2,3\}$. The following assignment of probabilities will give a valid probability space:

$P(0) = 1/12$, $P(1) = 4/12$, $P(2) = 5/12$, $P(3) = 2/12$.

Does it look surprising to you? (this can happen if the coins are not **fair**!).

6.6 Conditional probability and the multiplication rule

Definition: CONDITIONAL PROBABILITY

Let *A* and *B* be two events in a sample space S.
Conditional probability of *A*, <u>given that *B* has occurred</u>, is denoted by $P(A \mid B)$ and is defined as

$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, provided $P(B) > 0$ (otherwise, it is not defined)

[Similarly, $P(B \mid A) = \dfrac{P(B \cap A)}{P(A)} = \dfrac{P(A \cap B)}{P(A)}$, provided $P(A) > 0$.]

$P(A \mid B)$ in a certain sense measures the relative probability of *A* with respect to the **reduced space** ***B***.

<u>Properties of the conditional probability function</u>

Let *A*, *B*, *C*, and *D* be <u>arbitrary</u> events in a sample space S with $P(D) > 0$. Then,

1. $0 \le P(C \mid D) \le 1$
2. $P(S \mid D) = 1$
3. $P(A \cup B \mid D) = P(A \mid D) + P(B \mid D)$, if *A* and *B* are disjoint.
4. $P(A \cup B \mid D) = P(A \mid D) + P(B \mid D) - P(A \cap B \mid D)$, in general.
5. $P(\overline{C} \mid D) = 1 - P(C \mid D)$

EXAMPLE

The following table shows the percentage passing or failing a job competency exam listed according to sex, for a certain population of employees.

	Male (*M*)	Female (*F*)	

Pass (A)	24	36	60
Fail ($\bar{A}$)	16	24	40
	40	60	100

Now, P($M \cap A$) = 24 / 100 ; P(A) = 60 / 100; P(M |A) = $\dfrac{P(M \cap A)}{P(A)}$ = 24 / 60 .

P(A | F) = $\dfrac{P(A \cap F)}{P(F)} = \dfrac{36/100}{60/100}$ = 36 / 60 = "passing rate <u>among</u> females", and so on.

MULTIPLICATION RULE

For <u>any two events</u> A and B in S with P(A) > 0 and P(B) > 0,

P(A$\cap B$) = P(A). P(B | A) = P(B). P(A | B) – follows directly from definition.

<u>Extension to many</u>:

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1).\, P(A_2 \mid A_1).\, P(A_3 \mid A_1 \cap A_2) \ldots P(A_k \mid A_1 \cap A_2 \cap \ldots A_{k-1})$$

EXAMPLE

A bunch of keys has n similar keys, and only one of which will permit access to a room. Suppose a person chooses a key at random and tries it. If it does not work, he removes it from the bunch and randomly picks another one, proceeding in this manner until he finds the correct key.

Define events A_1 = { choose correct key at the 1st try} and

A_2 = {choose correct key at the 2nd try}.

(i) P (open door in the 1st try) = P(A_1) = 1/n.

(ii) P(open door in the 2nd try) = P ($\bar{A}_1 \cap A_2$)

= P($\bar{A}_1$). P (A_2 | $\bar{A}_1$) by the multiplication rule.

Where P($\bar{A}_1$) = $\dfrac{n-1}{n}$, and P (A_2 | $\bar{A}_1$) = $\dfrac{1}{n-1}$

Therefore, P(open door in the 2^{nd} try) = $1/n$.

[Can you guess the answer for opening door in the 5^{th} try, say?]

6.7 Tree diagrams

A **tree diagram** is a convenient tool for describing a sequence of (random) experiments in which each experiment has a finite number of outcomes with given probabilities.

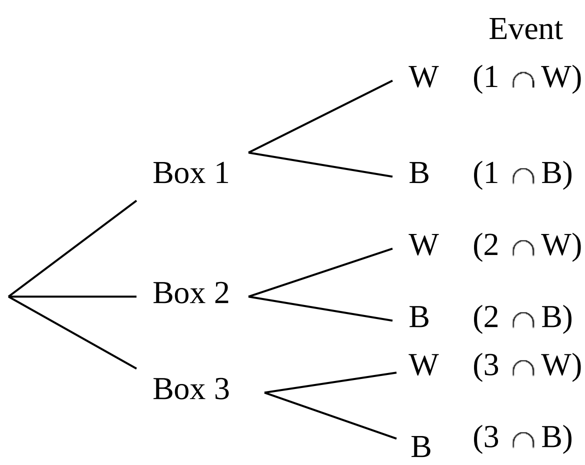
EXAMPLE

Suppose we have 3 boxes as follows:

Box 1 has 5 white balls and 3 black balls;
Box 2 has 10 white balls and 5 black balls; and
Box3 has 6 white balls and 2 black balls.

If we pick a box at random, and draw a ball at random, **what is the probability, that the ball is white**? that **the ball is black**?
[Here, if you know the outcome of the first experiment, *i.e*. selecting a box, then the answer is straight forward. But since it's random, the outcome is not known in advance! A tree diagram showing all possible outcomes, can help us find the answer].

In the first experiment, there are 3 outcomes: {box 1, box 2, box 3}.
In the second, box 1 has 2 outcomes: {W, B}; box 2 has 2 outcomes: {W, B}; and box 3 has 2 outcomes: {W, B}.
This process looks like branches of a tree! The following is a **tree diagram** to describe the sequence. The probability of each outcome is shown along the corresponding branch.



Now, to calculate the probability of any event in the last column, simply multiply the probabilities along its path.

For example, probability of selecting Box 1 and then a White ball = $P(1 \cap W)$ = (1/3). (5/8) = 5/24.
Similarly, $P(2 \cap W)$ = (1/3). (2/3) = 2/9.

$P(3 \cap W)$ = (1/3). (3/4) = ¼.

Now since there are 3 mutually exclusive paths which lead to a White ball, the probability that the ball is White is the sum of the probabilities of these 3 paths.

Notationally,

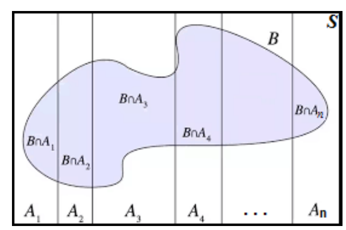
$P(W) = P((1 \cap W) \cup (2 \cap W) \cup (3 \cap W))$ = 5/24 + 2/9 + ¼ = 49/72.

Similarly, P(B) = 23/72.
(This is obtained by using the complement rule, or by the formula.)

6.8 Law of Total Probability

Suppose $A_1, A_2, A_3, \ldots, A_n$ are mutually exclusive and exhaustive events and B is an arbitrary event of the sample space S. Then the total probability of event B for i=1,2,3,…, n is defined as;

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

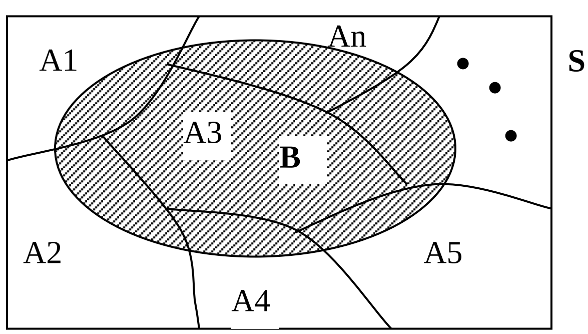


That is

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

6.9 Bayes' theorem

Suppose the events $A_1, A_2, \ldots, A_n$ form a (disjoint) partition of a sample space S; *i.e*. these events are mutually exclusive and their union is S. Let B be any other event in S.
(See the diagram)



Then, $B = S \cap B = (A_1 \cup A_2 \cup \ldots \cup A_n) \cap B$

$= (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B)$

where $A_i \cap B$, for i = 1, 2, …, n, are also mutually exclusive. Therefore,

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B).$$

Now, if one is interested in finding the <u>conditional probability</u> of A_i given B,
we can write

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\,P(B \mid A_i)}{P(B)} \text{ (by the multiplication rule).}$$

Now, if we replace $P(B)$ by the result given above, we obtain one of the most important theorems in probability, namely, Bayes' theorem:

BAYES' THEOREM

Suppose $A_1, A_2, \ldots, A_n$ form a partition of S and B is any event in S. Then,

$$P(A_i \mid B) = \frac{P(A_i)\,P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j)\,P(B \mid A_j)}.$$

EXAMPLE

A bowl contains *w* white balls and *b* black balls. One ball is selected at random from the bowl, its color noted, and it is returned to the bowl along with *n* additional balls of the same color. Another single ball is randomly selected from the bowl (now containing $w + b + n$ balls) and it is observed that the ball is black. Show that the (conditional) probability that the first ball selected was white is $\frac{w}{w+b+n}$.

<u>Solution</u>:

Let A_1 = {first ball is white} and A_2 = {first ball is black}. Then,
A_1, A_2 is a (disjoint) partition of S and,
$P(A_1) = \frac{w}{w+b}$ and $P(A_2) = \frac{b}{w+b}$.
Let B = {second ball is black}. We need to find $P(A_1 \mid B)$.

Now, $P(B \mid A_1) = \dfrac{b}{w+b+n}$ and $P(B \mid A_2) = \dfrac{b+n}{w+b+n}$.

Using Bayes' theorem, we have

$$P(A_1 \mid B) = \frac{P(B \mid A_1)\, P(A_1)}{P(B \mid A_1)\, P(A_1) + P(B \mid A_2)\, P(A_2)} = \frac{\left(\frac{b}{w+b+n}\right)\left(\frac{w}{w+b}\right)}{\left(\frac{b}{w+b+n}\right)\left(\frac{w}{w+b}\right) + \left(\frac{b+n}{w+b+n}\right)\left(\frac{b}{w+b}\right)}$$

which is $\dfrac{w}{w+b+n}$.

EXAMPLE

A diagnostic test for a disease is said to be 90% accurate in that if a person has the disease, the test shows positive with probability 0.9 . Also, if a person does not have the disease, the test shows negative with probability 0.9 . Suppose only 1% of the population has the disease in question. If a person is chosen at random from the population and the diagnostic test shows positive, what is the conditional probability that he does, in fact, have the disease? [Are you surprised by the answer? Would you call this diagnostic test reliable?].

<u>Solution</u>:

Define the events, D = {person has the disease}. Then $\overline{D}$ = {does not have}.

Note that D and $\overline{D}$ form a partition of S.
Also $P(D)$ = .01 and P($\overline{D}$) = 1 - .01 = .99 .
Let A = {test shows positive}. Then, $\overline{A}$ = {test shows negative}.
Now, $P(A \mid D)$ = .9 ; $P(\overline{A} \mid \overline{D})$ = .9. Therefore, $P(A \mid \overline{D})$ = 1 - $P(\overline{A} \mid \overline{D})$ = .1
We need, $P(D \mid A)$.

Using Bayes' theorem,

$$P(D \mid A) = \frac{P(A \mid D)\, P(D)}{P(A \mid D)\, P(D) + P(A \mid \overline{D})\, P(\overline{D})} = \frac{(.9)(.01)}{(.9)(.01)+(.1)(.99)} = \tfrac{1}{12}\,.$$

6.10 Finite probability spaces

INDEPENDENT EVENTS

An event B is said to be **independent** of an event A if the probability that B occurs is not influenced by whether A has occurred or not. In other words, B is independent of A if $P(B) = P(B \mid A)$. Similarly, A is independent of B if $P(A) = P(A \mid B)$.

Now, $P(A \cap B) = P(B \mid A).\ P(A) = P(A \mid B).\ P(B)$ (by the multiplication rule)

Consider (i) $P(B) = P(B \mid A)$. This implies $P(A \cap B) = P(A)\ .\ P(B)$.
(ii) $P(A) = P(A \mid B)$. This again implies $P(A \cap B) = P(A)\ .\ P(B)$.
This leads to a formal definition of independence.

DEFINITION

Events *A* and *B* are **independent** if and only if $P(A \cap B) = P(A)\ .\ P(B)$.

(By saying *A* and *B* are independent, we mean that *A* is independent of *B*, and *B* is independent of *A*).

EXTENSION TO THREE EVENTS

Events *A*, *B*, and *C*, are (mutually**) independent** if and only if they are pair-wise independent, and in addition, if $P(A \cap B \cap C) = P(A).P(B).P(C)$.

Note: The definition of pair-wise independence is the one for two events given above.

Note: If *A* and *B* are independent, it can be shown that *A* and $\bar{B}$, $\bar{A}$ and *B*, $\bar{A}$ and $\bar{B}$, are all independent!

EXAMPLE

Let a <u>fair</u> coin be tossed three times. Then the (equiprobable) sample space is

S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}.

Define the events

A = {first toss is H}; *B* = {second toss is H}; and *C* = {at least two H's in a row}. Then,

A = { HHH, HHT, HTH, HTT}; *P*(A) = 4/8 = ½ .

B = { HHH, HHT, THH, THT}; *P*(*B*) = 4/8 = ½ .

C = { HHH, HHT, THH}; *P*(*C*) = 3/8.

$A \cap B$ = {HHH, HHT}; $P(\ A \cap B\ )$ = 2/8 = ¼ .

$A \cap C$ = {HHH, HHT}; $P(A \cap C)$ = 2/8 = ¼ .

$B \cap C$ = {HHH, HHT, THH}; $P(B \cap C)$ = 3/8.

Now, P(A). $P(B)$ = ½ . ½ = ¼ = $P(A \cap B)$. So, A and B are **independent**.

P(A). $P(C)$ = ½ . 3/8 = 3/16 $\neq$ $P(A \cap C)$. So, A and C are **dependent**.

$P(B)$. $P(C)$ = ½ . 3/8 = 3/16 $\neq$ $P(B \cap C)$. So, B and C are **dependent**.

EXAMPLE
Let a <u>fair</u> coin be tossed twice. Then the sample space is

S = {HH, HT, TH, TT}.

Define the events

A = {H on the first toss}; B = {H on the second toss}; and C = {H on exactly one toss}.

Then,
A = {HH, HT}; B = {HH, TH}; and C = {HT, TH}.

It is easy to verify that A, B, and C are **pair-wise independent**, but **not mutually independent**!

INDEPENDENT TRIALS

Consider the above example of tossing a fair coin three times. By saying <u>fair</u> coin, we mean that P(H) = ½ and P(T) = ½ in each toss.

These three trials can be assumed **independent** since the outcome on any one trial has no influence on the outcome on any other trial. Therefore, the probability of any outcome of the experiment can be written as follows:

P(HHH) = P(H). P(H). P(H) = ½. ½ . ½ = 1/8,

P(HHT) = ½ . ½ . ½ = 1/8, and so on. In general, for n independent trials,
$P((s_1, s_2, \ldots, s_n)) = P(s_1).\ P(s_2).\ \ldots\ P(s_n)$.

6.11 One dimensional random variables:

Given a random experiment with sample space S, one dimensional random variable $X(\gamma)$ is a set function that assigns one and only one real number to each element (γ) on S.
Usually, a single letter X is used for $X(\gamma)$ to represents the random variable.

If X is a random variable and x is a fixed real value then we can define the event $(X = x)$ as $(X = x) = \{\gamma : X(\gamma) = x\}$

Example: In the experiment of tossing a coin three times the event of obtaining two heads is defined as $(X = 2) = \{\gamma : X(\gamma) = 2\} = \{\text{HHT, HTH, THH}\}$

Similarly, for fixed numbers x, x_1 and x_2 we can define the following events:

$$(X \leq x) = \{\gamma : X(\gamma) \leq x\}$$
$$(X > x) = \{\gamma : X(\gamma) > x\}$$

$$(x_1 < X \leq x_2) = \{\gamma : x_1 < X(\gamma) \leq x_2\}$$

These events have the probabilities hat are denoted by

$$P(X \leq x) = P\{\gamma : X(\gamma) \leq x\}$$
$$P(X \leq x) = P\{\gamma : X(\gamma) \leq x\}$$
$$P(X > x) = P\{\gamma : X(\gamma) > x\}$$

$$P(x_1 < X \leq x_2) = P\{\gamma : x_1 < X(\gamma) \leq x_2\}$$

Example: $P(X = 2) = P\{\text{HHT, HTH, THH}\} = 3/8$

6.11.1 Cumulative distribution function

The cumulative distribution function (cdf) of a random variable X is the function defined by

$$F_X(x) = P(X \leq x) \quad for \quad -\infty < x < \infty$$

6.11.2 Discrete random variables and Continuous random variables

<u>Discrete random variables :</u>

Let X be a random variable with cdf $F_X(x)$. If $F_X(x)$ changes values only in jumps at each value of X and is constant between jumps (that is $FX(x)$ is a staircase or step function) then X is called a discrete random variable

X is a discrete random variable only if it's output contains a finite or countably infinite number of distinct values such as 0,1,2,3,4,…. Discrete random variables are usually (but not necessarily) counts.

<u>Continuous random variables:</u>

Let X be a random variable with cumulative distribution function $F_X(x)$. If $F_X(x)$ is continuous and also has a derivative $dF_X(x)/_{dx}$ which exists everywhere, then X is called a continuous random variable

A continuous random variable is one which takes an infinite number of possible values. Continuous random variables are usually measurements.

6.11.3 Probability (mass) function and probability density function

Probability mass function (pmf):
The probability that a discrete random variable X takes on a particular value x, that is, $P(X = x)$, is frequently denoted $f(x)$. The function $f(x)$ is typically called the probability mass function.

The probability mass function, $P(X = x) = f(x)$ satisfies the following properties:

- $P(X = x) = f(x) > 0,$ if $x \in S$
- $\sum_{x \in S} P(X = x) = 1$
- $F(X = a) = P(X \leq a) = \sum_{x \leq a} P(X = x)$

Probability density function(pdf):
The probability density function of a continuous random variable is an integrable function satisfying;

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Here $P(a < X \leq b) = \int_a^b f_X(x)dx$ which represents the area under the curve $f_X(x)$ for the boundaries a and b of X

The probability density function, $f_X(x)$ satisfies the following properties:

- $f_X(x) > 0$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- $F(X = a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx$
- $P(X = a) = 0$ since area under the curve of a single point a gets 0
- $P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$

6.11.4 Expected value and Variance of a random variable

Expected value:
The mathematical expectation (mean or expected value) of a random variable X denoted by $\mu_X (or\ \mu)$ or $E(X)$ is defined as

$$\mu_X = E(X) = \begin{cases} \sum_{all\ x} xP(X = x) & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} x\, f_X(x)dx & if\ X\ is\ continuous \end{cases}$$

Variance:
The variance of a random variable X denoted by $\sigma_X^2 (or\ \sigma^2)$ or $Var(X)$ is defined as

$$\sigma_X^2 = Var(X) = \begin{cases} \sum_{all\ x} (x - \mu_X)^2 P(X = x) & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} (x - \mu_X)^2\, f_X(x) dx & if\ X\ is\ continuous \end{cases}$$

Thus;

$$\sigma_X^2 = Var(X) = E\{[X - \mu_X]^2\} = E(X^2) - {\mu_X}^2 = E(X^2) - [E(X)]^2$$

$Var(X) \geq 0$

Standard deviation σ_X is the positive square root of $Var(X)$