COURSE OUTCOME 1

THE RANDOM VARIABLE AND MULTIPLE RANDOM VARIABLES:

Review of Probability Theory

Trail: A well-defined single performance for experiment.
Random experiment: An experiment is called a **random experiment** when repeated under the same conditions, it is such way that the outcome cannot be predicated with certainty but all possible outcomes can be determined prior to the performance of the experiment.
Ex: Throwing of a die, tossing of a coin, drawing two playing cards from a pack of cards.

Sample space: The set of all possible outcomes of a Random experiment is called the Sample space and is represented by the symbol **S.**
Ex: 1) When a coin is tossed the sample space is $S = \{H, T\}$.

2) When a six faced die is rolled the sample space is $S = \{1,2,3,4,5,6\}$.

Event: An event is subset of a sample space.
Ex: When a six faced die is rolled $A = \{2,4,6\}$ is a event and represent the occurrence of an even numbers of dots.
Events are denoted by $A, B, C, \ldots$ or $E_1, E_2, E_3, \ldots$
An event may be a subset that includes the entire sample space S called entire event, or a subset of S called the null set and denoted by the symbol ϕ, which contains no elements at all called null event.
For instance, if we let A be the event of detecting a microscopic organism by the naked eye in a biological experiment, then $A = \phi$.
Also, if $B = \{x / x \, is \, an \, even \; factor \, of \; 15\}$, then B must be the null set.

Complement of an event: The complement of an event A with respect to S is the sub set of all elements of S which are not in A. We denote the complement of A by the symbol A^1 or A^c or $\bar{A}$.
Ex: Let A be an event that an even number of dots occurred when a die is rolled then A^1 is an event that an odd number of dots occurred.

Intersection of two events: the intersection of two events A and B denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.
Ex: Let C be the event that a student selected at random is a second year student and M be the event that student is a boy then $C \cap M$ is the event of all second year boys.

Mutually exclusive events: Two events A and B are mutually exclusive, or disjoint if $A \cap B = \phi$, i.e., if A and B have no elements in common.
Ex: In the die tossing experiment, if $A = \{1,3,5\}$ and $B = \{2,4.6\}$ then the events A and B are mutually exclusive.
Union of two events: The union of two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.

Ex: In die tossing experiment, if $A=\{3,6\}$ and $B=\{2,4,6\}$ then $A\cup B=\{2,3,4,6\}$ and it represent the event of getting an even number or a multiple of 3 dots.
The following results can be observed:
1) $A\cap\phi=\phi.$
2) $A\cup\phi=A.$
3) $A\cap A^1=\phi.$
4) $A\cup A^1=S.$
5) $S^1=\phi.$
6) $\phi^1=S.$
7) $\left(A^1\right)^1=A.$
8) $(A\cap B)^1=A^1\cup B^1.$
9) $(A\cup B)^1=A^1\cap B^1.$

Classical definition of Probability:
If there are n outcomes mutually exclusive and equally likely outcomes of a random experiment, out of which, $'s'$ outcomes are favorable for a particular E, then we define the probability of E, as $P(E)=\frac{s}{n}=\frac{Number\ of\ favourable\ outcomes\ of\ the\ experiment}{Number\ of\ total\ outcomes\ of\ the\ experiment}.$
This probability is also known as probability of success of E.
In this experiment $'s'$ results are favorable to E, and hence the remaining n-s results are not favorable to the event E. This set of unfavorable events denoted by E^1 or E^c or $\overline{E}$.
$\therefore$ Then probability of $P(E^c)=\frac{n-s}{n}=1-\frac{s}{n}=1-P(E).$

The relative frequency interpretation of probability or Statistical definition of probability:
Let m be the frequency of occurrence of the event A associated with the n independent trails of the random experiment. Then probability of the event A, denoted by the symbol $P(A)$) is given by
$P(A)=\underset{n\to\infty}{Lt}\frac{m}{n}.$ We may note that $\frac{m}{n}$ is the relative frequency of the event A in n-trails. If n is very large then the relative frequency $\frac{m}{n}$ is very close to actual probability.

Axiomatic definition of probability:
Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties.
If S is the sample space and E is any event in a random experiment,
1. $0\le P(E)\le 1$ for each event e in S.
2. $P(S)=1.$
3. If E_1 and E_2 are any mutually exclusive events in S, then $P(E_1\cup E_2)=P(E_1)+P(E_2).$

Example 1: A box contains 25 parts of which 10 are defective. Two parts are being drawn simultaneously in a random manner from the box. The probability of both parts being good is
A) 7/20 B) 42/125 C) 25/29 D) 5/9
Solution: Number of ways of drawing 2 parts from 25 parts=$25c_2$
Number of ways of drawing 2 good parts from the 15 good parts=$15c_2$
Probability that both parts are good=$\frac{15_{c_2}}{25_{c_2}} = \frac{7}{20}$

Example 2: In a housing society, half of the families have a single child per family, while the remain half have two children per family. The probability that a child picked at random, has a sibling
Solution:
The child picked at random will have a sibling if the family has two children
Probability of this event=1/2.

Example 3: An unbiased coin is tossed an infinite number of times. The probability that the fourth head appears at the 10^{th} toss is
A) 0.067 B) 0.073 C) 0.082 D) 0.091
Solution:
Total number of possibilities for the first ten slips is 2^{10}=1024.
For the fourth head to occur on 10^{th} slip.
We need first 3 heads to occur in the first 9 slips. This is given by
$9c_3$=84.
There is only one way for 4^{th} head occur on 10^{th} slip =84*(1/1024)=21/256=0.082.
(or) for the 4th head to occur at the 10th toss, you have to first get 3 heads and 6 tails in the 1st 9 toss, and then a head at the 10th toss.
So prob. = (9C3)(.5)^3 (1-.5)^6 (0.5)=.082.

Example: 4 A fair dice is tossed 10 times. What is probability that only the first two tosses will yield heads
A) (1/2)^2 B) 10C2 (1/2)^2 C) (1/2)^10 D) 10C2(1/2)^10
Solution: $(1/2)^2(1/2^8=(1/2)^{10}$

Addition Theorem on Probability
If A and B are two events the $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Note:
1) If A and B are mutually exclusive then $P(A \cup B) = P(A) + P(B)$.

2) For three events A, B and c then
$P(A \cup B \cup c) = P(A) + P(B) + P(C) - P(A \cap B). - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$

3) If A, B and C are mutually exclusive then $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

4) If $A_1, A_2, \ldots, A_n$ are n mutually exclusive events then
$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$.

Conditional Probability

The Conditional Probability of B, given A, denoted by P(B/A) is defined by provided P(A) > 0.
The Conditional Probability of A, given B, denoted by P(A/B) is defined by provided P(B) > 0.

Independent Events

If the occurrence of B had no impact on the odds of occurrence of A then A and B are said to be independent.

Ex: If P(A)=0.65, P(B)=0.40 and P(C∩D)=0.24, are the events C and D independent?
Two events A and B are independent if and only if $P(B/A) = P(A)$ *or* $P(A/B) = P(B)$, provided the existence of the conditional probability.

Multiplicative Rule

If in a experiment the events A and B can both occur, then $P(A \cap B) = P(B/A)P(A)$ provided P(A) > 0. We can also write $P(A \cap B) = P(A/B)P(B)$ in other words, it does not matter which event is referred to as A and which event is refered to as B.
Note:

1) Two events A and B are independent if and only if $P(A \cap B) = P(A)P(B)$.

   Eg: :Let A be the event that raw material is available when needed and B be the event that the matching time is less than one had. If P(A)=0.8 and P(B)=0.7. What is P(A∩B).

2) If A, B, C are any three events then the multiplicative rule $P(A \cap B \cap C) = P(B/C)P(A)P(C/A \cap B)$.

P(J/T$\cap I \cap G \cap P$)
P(T$\cap I \cap G \cap P$)=P(T).P(I/T).P(G/ T$\cap I$) P(P/ T$\cap I \cap G$)

3) If A, B, and C are independent events if and only if $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Total Probability

If the events $B_1, B_2, \ldots, B_K$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ *for* $i = 1,2,\ldots,k$, then for any event A of S,

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A/B_i).$$

Baye's Rule:

If the events $B_1, B_2, \ldots, B_K$ constitute a partition of the sample space S such that $P(B_i) \neq 0 \ for \ i = 1,2,\ldots,k,$ then for any event A in S, such that P(A)≠0,

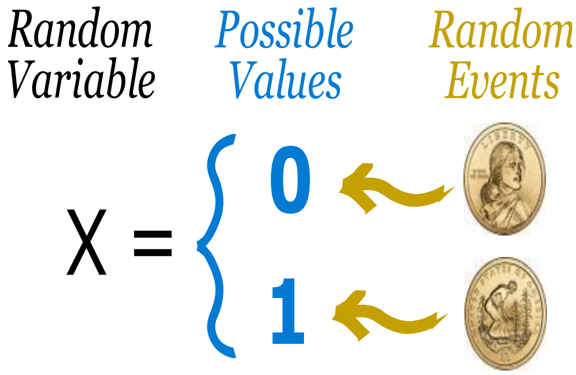
$$P(B_r / A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{K} P(B_i \cap A)} = \frac{P(B_r)}{\sum_{i=1}^{k} P(B_i) P(A / B_i)}, \ for = 1, 2, ..., k$$

RANDOM VARIABLE AND THEIR TYPES

A random variable is a function that associates a real number with each element in the sample space.

Mathematically, a random variable is defined as a function that maps the outcomes of a sample space to real numbers.

Formally, let's consider an experiment or random process with a sample space Ω. A random variable, denoted as X, is a function that assigns a real number to each outcome in Ω. In other words, $X: \Omega \rightarrow R$, where R represents the set of real numbers.



Example: The testing of three of electronic components for defectives is a random experiment. The associated sample space is

$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$ where N denotes non defective and D denotes "defective". If X is the number defectives. Then for each point in the sample space, X associates real numbers 0, 1, 2, or 3. **(explain from the definition of trial, experiment, event, sample space, random variable)**

<u>Types of random variables</u>

Random variables are of two types: Discrete random variable and continuous random variable.

Discrete random variable: A random variable is called a discrete random variable if its set of possible outcomes is countable.

•A discrete random variable has a finite number of possible values or an infinite sequence of countable real numbers.
–X: number of hits when trying 20 free throws.
–X: number of customers who arrive at the bank from 8:30 – 9:30AM Mon--Fri.
–E.g. Binomial, Poisson...

Continuous random Variable: When a random variable can take on values on a continuous scale, it is called a continuous random variable.

Example: Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X take on values x for which $x \geq 0$.

Mixed Random Variables: A mixed random variable is a random variable that has both discrete and continuous components. It combines the characteristics of both discrete and continuous random variables. The probability distribution of a mixed random variable can be expressed as a combination of probability masses for discrete values and probability densities for continuous values.

Example:

Suppose we are interested in modeling the amount of time it takes for a customer to complete a transaction at a bank. For some **transactions, such as cash withdrawals or check deposits, the time taken can be measured in discrete units** (e.g., minutes). However, for other **transactions, such as loan applications or account openings, the time taken can be more accurately represented as a continuous variable (e.g., seconds).**

In this case, we can define a **mixed random variable**, let's call it T, to represent the transaction time. T can take on discrete values for certain types of transactions and continuous values for others.

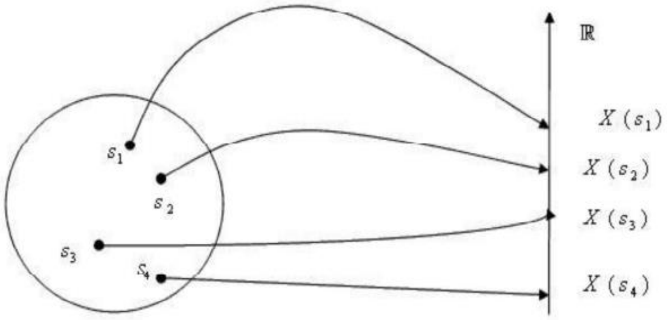
For example, let's say for cash withdrawals and check deposits, the time taken follows a discrete distribution with values {1, 2, 3, ...} representing the number of minutes required. This discrete distribution can be modeled using a probability mass function (PMF).

On the other hand, for loan applications and account openings, the time taken follows a continuous distribution, such as an exponential distribution with a certain rate parameter. This continuous distribution can be modeled using a probability density function (PDF).

By combining both discrete and continuous distributions, we have a mixed random variable T that captures the variability in transaction time for different types of transactions at the bank.

PROBABILITY DISTRIBUTION FUNCTION

Discrete probability distribution function: The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function. More formally, the probability distribution of a discrete random variable X is a function which gives the probability $P(s)$ that the random variable equals s, for each **value** $\boldsymbol{s}$: $\boldsymbol{P(s) = P(S = s)}$, in general we denoted by $\boldsymbol{x}$: $\boldsymbol{P(x) = P(X = x)}$,



Ex: In the case of tossing a coin three times, the variable X, representing the number of heads has the following probability distribution.

X	0	1	2	3
f(x)=P(X=x)	1/8	3/8	3/8	1/8

Probability function: The set of ordered pairs (x, f(x)) is called the probability function or **Probability Mass Function (PMF)** or probability distribution of the discrete random variable X, if for each possible outcome x,

1. $f(x) \geq 0.$
2. $\sum_x f(x) = 1$
3. $P(X = x) = f(x).$

Cumulative distribution: The cumulative distribution function F(x) of a discrete random variable X with probability distribution f(x) is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \ \ for \ -\infty < x \ \infty.$$

Properties:

(i) $F(-\infty) = 0$ (ii) $F(\infty) = 1$ (iii) $P(a \leq X \leq b) = F(b) - F(a)$ (iv) F(x) is a non-decreasing function.

Continuous Probability Distribution: A continuous random variable has a probability of zero of assuming exactly any of its values. Consequently, its probability distribution cannot be given in tabular form.

Density Function: In dealing with continuous variables, f(x) is usually called the probability density function or simply the density function of X.

The function f(x) is a probability density function for the continuous random variable X, defined over the set of real numbers R, if

1) $f(x) \geq 0, \quad for\, all\, x \in R$
2) $\int_{-\infty}^{\infty} f(x)dx = 1$
3) $P(a < x < b) = \int_{a}^{b} f(x)dx.$

Cumulative distribution function: The cumulative distribution function F(x) of a continuous random variable X with density function f(x) is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad for\ -\infty < x < \infty$$

Note: $P(a \leq X \leq b) = F(b) - F(a)$ and $f(x) = \frac{dF(x)}{dx}$, if the derivate exits.

Example: For the probability distribution

X	0	1	2	3	4	5	6	7
$f(x) = P(X = x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2 + k$

Find (i) k (ii) $P(x < 6)$; $P(x \geq 6)$; $P(0 < x < 5)$ (iii) distribution function $F(x)$ (iv) if $P(x \leq c) > \frac{1}{2}$ find minimum value of c (v) find $P(\frac{\{1.5<x<4.5\}}{\{x>2\}})$

Solution:

(i) If $P(x)$ is probability mass function, then $\sum P(x) = 1$.

$$10k\text{^}2 + 9k = 1$$

$$10k^2 + 9k - 1 = 0.$$

By solving, $we\ get\ k = -1, \frac{1}{10}$. since $P(x) \geq 0$, this takes only positive $k = \frac{1}{10}$

X	0	1	2	3	4	5	6	7
$f(x) = P(X = x)$	0	0.1	0.2	0.4	0.3	0.01	0.02	0.17

.

(ii) $P(x < 6) = 1 - P(x \geq 6)$ Since $p + q = 1$

$= 1 - (P(6) + P(7))$

$= 1 - 0.19 = 0.81.$

$$(iii)\quad \mathrm{F(x)} = \begin{cases} 0, & \mathrm{x} \le 0 \\ 0.1, & x \le 1 \\ 0.3, & x \le 2 \\ 0.5, & x \le 3 \\ 0.8, & x \le 4 \\ 0.81, & x \le 5 \\ 0.83, & x \le 6 \\ 1, & \mathrm{x} \le 7 \end{cases}$$

this is required F(x).

(iv) $P(x \le 0) = 0, P(x \le 1) = 0.1, P(x \le 2) = 0.3, P(x \le 3) = 0.5,$ $\boldsymbol{P(x \le 4) = 0.8}$

Among this, , $P(x \le 4) = 0.8 > \frac{1}{2}$, this gives $c = 4$.

(i) $P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$

$$P\left(\frac{1.5 < x < 4.5}{x > 2}\right) = \frac{P((1.5 < x < 4.5) \cap x > 2)}{P(x > 2)}$$

$$= \frac{P((2,3,4)\ \cap (3,4,5,6,7))}{1 - P(x \le 2)}$$

$$= \frac{(P(3) + P(4))}{(1 - (P(0) + P(1) + P(2)))} = 5/7 .$$

MEAN / EXPECTED AND VARIANCE VALUE OF A RANDOM VARIABLE

Mean or Expected Value of Random Variable:

Let X be a random variable with probability distribution f(x). The mean or expected value of X is $\mu = E(X) = \sum_x x f(x),$ if X is discrete and

$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$, If X is continuous.

Note: If X is a random variable, a function of X, g(X) is also a random variable.

The expected value of the random variable g(X) is $\mu_{g(X)} = E(g(X)) = \sum_x g(x) f(x)$, if X is discrete, and $\mu_{g(X)} = \int_{-\infty}^{\infty} g(x) f(x) dx$, if X is continuous.

Variance of random variable:

Let X be a random variable with probability distribution f(x) and mean μ . The variance of X is $\sigma^2 = E\left[(X-\mu)^2\right] = \sum_x (X-\mu)^2 \ f(x),$ if X is discrete, and

$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (X-\mu)^2 \ f(x)$, if X is continuous.

Note: 1) $\sigma^2 = E(X^2) - \mu^2$

2) $E(aX+b) = aE(X)+b$

3) $V(aX+b) = a^2 V(X)$

4) If X and Y are independent random variables, then $V(aX+bY) = a^2V(X) + b^2V(Y)$

Example:

For the continuous probability function $f(x) = kx^2e^{-x}$ when $x \geq 0$, find

(i) k

(ii) Mean

(iii)Variance

Solution:

(i) We have $\int_{-\infty}^{\infty} f(x) dx = 1$

$$\therefore \int_0^\infty kx^2 e^{-x} dx = 1 \; (\because x \geq 0)$$

$$i.e, k[x^2(-e^{-x}) - 2x(e^{-x}) + 2(-e^{-x})]_0^\infty = 1$$

$$i.e, k[(-e^{-x})(x^2 + 2x + 2]_0^\infty = 1$$

$$k(0+1) = 1 \quad \text{or} \quad \mathrm{k} = \tfrac{1}{2}$$

(ii) mean= $\int_{-\infty}^\infty xf(x)dx = \int_0^\infty kx^3 e^{-x} dx$

$$= k[x^3(-e^{-x}) - 3x^2(e^{-x}) + 6x(-e^{-x}) - 6(e^{-x})]_0^\infty$$

$$= k[(-e^{-x})(x^3 - 3x^2 + 6x - 6)]_0^\infty = k0 + 6] = 6k$$

$$\mu = 6(\tfrac{1}{2}) = 3 \quad (\because k = \tfrac{1}{2})$$

(iii) Variance= $\int_{-\infty}^\infty x^2 f(x)dx - \mu^2 = \int_0^\infty x^2 kx^3 e^{-x} dx - 3^2$

$$= k\int_0^\infty x^4 e^{-x} dx - 9$$

$$= k[x^4(-e^{-x}) - 4x^3(e^{-x}) + 12x^2(-e^{-x}) - 24(-e^{-x}) + 24e^{-x}]_0^\infty - 9$$

$$= k[(-e^{-x})(x^4 + 4x^3 + 12x^2 + 24 - 24]_0^\infty - 9$$

$$= \tfrac{1}{2} 0 + 24] - 9 = 12 - 9 = 3$$

DIFFERENT TYPES OF DISCRETE DISTRIBUTION

BINOMIAL DISTRIBUTION

Bernoulli process: An experiment often consists of repeated trials, each with two possible outcomes that may be labelled **success or failure**. As an example, the testing of items as they come off an assembly line, where each test or trial may indicate a defective or non-defective item. We may choose to define either outcome as a success. The process is referred to as a **Bernoulli process**. Each trial is called a Bernoulli trial.

Properties of Bernoulli Process:

The Bernoulli process must possess the following properties:

1. The experiment consists of n-repeated trials.
2. Each trial results in an outcome that may be classified as a success or a failure.
3. The probability of success, denoted by p, remains constant from trial to trial.
4. The repeated trials are independent.

Consider a sequence of n-independent trials of a Bernoulli process. Let X be the number of successes in n-Bernoulli trials and it is a random variable. The distribution of the random variable X is called Binomial distribution.

Definition: A Bernoulli trial can result in a success with probability p and a failure with probability q =1-p. Then the probability distribution of the Binomial random variable X, the number of successes in n-independent trials, is

$$b\,(x;\, n,\, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n,\ p + q = 1$$

n: number of trials

x: no. of success

p: probability of getting success

q: probability of getting failure

Note

1) We write, X ~ b(n, p) to denote that X follows binomial distribution with parameters n and p

2) The mean of the binomial distribution is ‘np’

3) The variance of the binomial distribution is ‘npq’

4) The standard deviation of binomial distribution is $\sqrt{npq}$

5) In binomial distribution, the mean is always greater than the variance.

Example:

It has been claimed that in 60% of all solar-heat installations the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by at least one-third in

a) Four of five installations
b) at least four of five installations
c) at the most two installations
d) what are the mean and variance of the number of installations

Solution:

Let X be the number of solar installation where the utility bill is reduced by at least one third. Then the distribution of X is binomial with n=5 and P=0.6

$$P(X = x) \ = 5_{c_x}(0.6)x(0.4)5 - x, \ \ x = 0,1,2,3,4,5$$

a) Probability that in four of 5 installations utility bill is reduced by one third is

$= P(X = 4) = 5_{c_4}(0.6)4(0.4)5 - 4 = 0.259$

b) Probability that in at least 4 installations utility bill is reduced by one third is

$= P(X \geq 4) = P(X = 4) + P(X = 5) = 5_{c_4}(0.6)4(0.4)5 - 4 + 5_{c_5}(0.6)5(0.4)5 - 5$

c) Probability that in at most two installations utility bill is reduced by one third is

$= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$= 5_{c_0}(0.6)0(0.4)5 - 0 + 5_{c_1}(0.6)1(0.4)5 - 1 + 5_{c_2}(0.6)2(0.4)5 - 2$

d) Mean $= np = 5(0.6) = 3$

Variance $= np(1 - p) = 5(0.6)(0.4) = 1.2.$

<u>POISSION DISTRIBUTION</u>

When n is large and p is small, binomial probabilities are often approximated by means of the Poisson distribution with the parameter λ equal to the product np i.e., Poisson distribution is used in case of rare events.

Experiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region (Here n is not known and information regarding the number of occurrences of event is known), are called **Poisson experiments**.

The number X of outcomes occurring during a Poisson experiment is called a Poisson random variable and its probability distribution is called the Poisson distribution.

Definition: The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$P(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0,1,2,\dots..$$

$\text{where } \lambda \text{ is the average numebr of outcomes per unit time, distance, area or volume}$ and e=2.71828

<u>Note:</u> Both the mean and variance of the Poisson distribution P(x; λt) are λt.

1) If the time t is one unit then

$$P(x;\lambda) = \frac{e^{-\lambda}(\lambda)^x}{x!}, \quad x = 0,1,2,\dots..$$

2) We write, X ~ P(x, λ) to denote that X follows Poisson distribution with parameter λ

The following are the some of the examples of random variables following Poisson distribution:

- The number of customers arrived during a time period of length t.
- The number of telephone calls per hour received by an office.
- The number of typing errors per page.
- The number of accidents occurred at a junction per day.

Example:

If a bank received on the average 6 bad checks per day, what are the probabilities that it will receive

a) 4 bad checks on any given day?
b) 10 bad checks over any 2 consecutive days
c) No bad check on any given day
d) What are the mean and variance of the number of bad checks per day?
   Mean=average=np=6

Solution:

Let X be the number of bad checks received per day. Then the distribution of X is Poisson with parameter λ=6.

$$P(X = x) = \frac{e^{-6}6^x}{x!}, x = 0,1,2,\dots$$

a) $P(\text{4 bad checks on any given day}) = P(X = 4) = \frac{e^{-6}6^4}{4!} = 0.134$

. b) $P(\text{10 bad checks over any 2 consecutive days}) = P(X = 10) = \frac{e^{-12}12^x}{x!}, \ x = 0,1,2, \ldots.$

(Here $\lambda = 12$)

.c) $P(\text{no bad check on any day}) = P(X = 0) = \frac{e^{-6}6^0}{0!} = e^{-6}$

d) Mean and variance of the number of bad checks per day $= \lambda = 6$.

DIFFERENT TYPES OF CONTINUOUS DISTRIBUTION

Uniform Distribution:

A continuous random variable X is called uniformly distributed over the interval [a, b],

$-\infty < a < b < \infty$, if its probability density function is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 \;, & \text{otherwise} \end{cases}$$

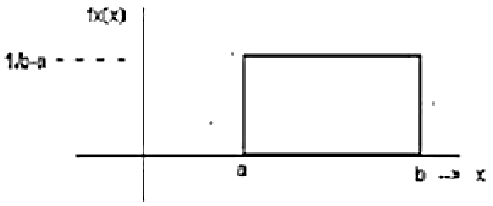


Figure 1

We use the notation $X \sim U(a,b)$ to denote a random variable X uniformly distributed over the interval $[a,b]$

Distribution function $F_X(x)$

For $x < a$

$F_X(x) = 0$

For $a \le x \le b$

$$\int_{-\infty}^{x} f_X(u)du$$

$$= \int_{a}^{x} \frac{du}{b-a}$$

$$= \frac{x-a}{b-a}$$

For $x > b$,

$F_X(x) = 1$

Mean and Variance of a Uniform Random Variable:

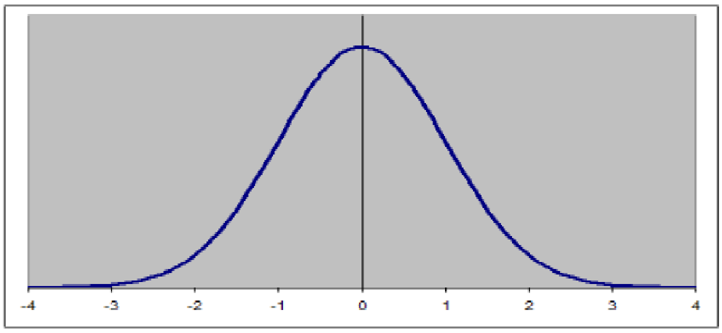
$$\mu_X = EX = \int_{-\infty}^{\infty} x f_X(x)dx = \int_a^b \frac{x}{b-a}dx$$
$$= \frac{a+b}{2}$$
$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_a^b \frac{x^2}{b-a}dx$$
$$= \frac{b^2+ab+a^2}{3}$$
$$\therefore \sigma_X^2 = EX^2 - \mu_X^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4}$$
$$= \frac{(b-a)^2}{12}$$

The characteristic function of the random variable $X \sim U(a,b)$ is given by

$$\phi_X(w) = Ee^{jwx} = \int_a^b \frac{e^{jwx}}{b-a}dx$$
$$= \frac{e^{jwb} - e^{jwa}}{jw(b-a)}$$

Normal / Gaussian Distribution – Its properties and importance

The most important continuous probability distribution in the entire field of statistics is the **normal distribution**. It is graph, called the normal curve, is the bell-shaped curve, which describes approximately many phenomena that occur in nature, industry, and research. Physical measurements in areas such as meteorological experiments, rainfall studies, and measurements of manufactured parts are often more than adequately explained with a normal distribution.



The basic from of normal distribution is that of a bell, it has single mode and is symmetric about its central values. The flexibility of using normal distribution is due to the fact that the curve may be centered over any number on the real line and it may be flat or peaked to correspond to the amount of dispersion in the values of random variable.

The Normal Distribution (N.D.) was first discovered by De-Moivre as the limiting form of the binomial model in 1733. The normal distribution is often referred to as the Gaussian distribution, in honour of Karl Friedrich Gauss (1777-1855), who also derived its equation from a study of errors in repeated measurements of the same quantity.

Normal distribution is often used to model the distribution of discrete random variable as well as the distribution of other continuous random variables. Normal distribution provided the basis for which much of the theory of inductive statistics is founded.

Definition

A random variable X is said to follow a Normal Distribution with parameter mean (μ) and variance (σ^2) if its density function is given by the probability law

$$f(x) = n(x; u, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

Symbolically we can represent the distribution of normal variate as $X \sim N(\mu, \sigma^2)$

The Properties of normal probability curve

1. The mode which is point on the horizontal axis where the curve is a maximum, occurs at x=μ. Hence the mean, median and mode of normal distribution are equal.

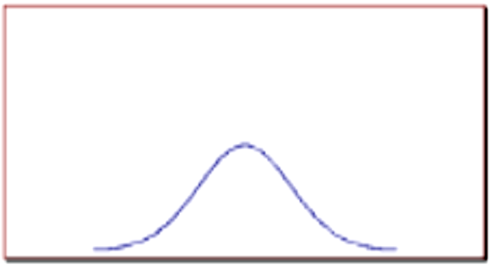
2. The curve is symmetric about a vertical axis through the mean μ.

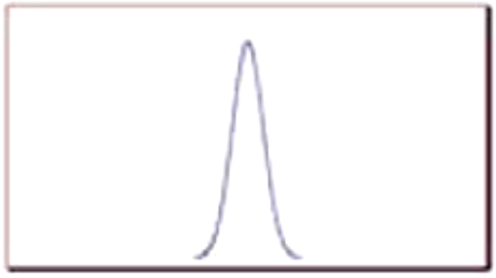
3. The curve has its points of inflexion at $x=\mu\pm\sigma$, it concave downward if $\mu-\sigma<X<\mu+\sigma$, and is concave upward otherwise.

4. The curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.

5. The total area under the curve and above the horizontal axis =1.

6. The graph of the normal distribution depends on two factors – the mean and the standard deviation. The mean of the distribution determines the location of the center of the graph, and the standard deviation determines the height and width of the graph. When the standard deviation is large, the curve is short and wide; when the standard deviation is small, the curve is tall and narrow. All normal distributions look like a symmetric, bell-shaped curve, as shown below.



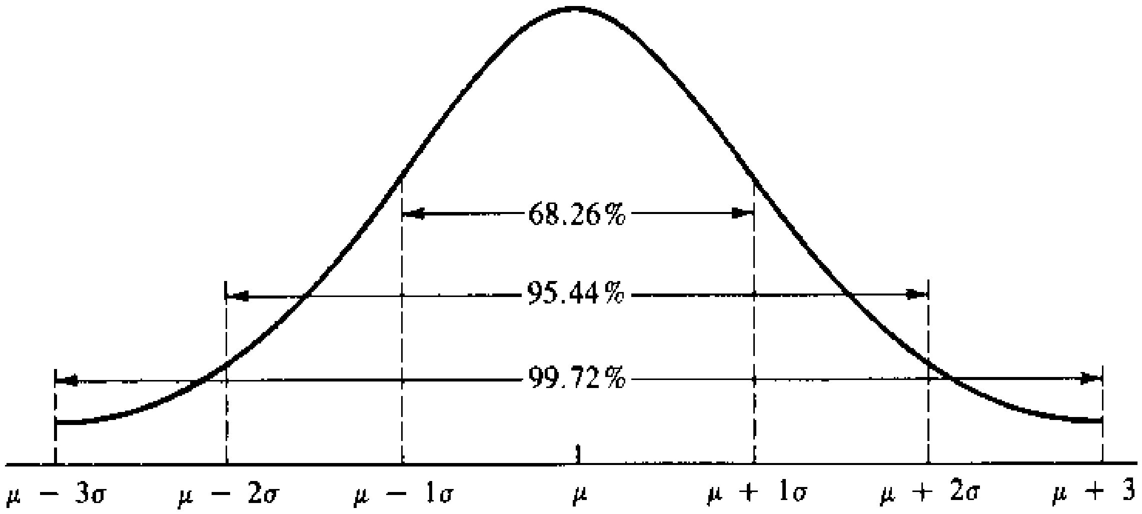


The curve on the left is shorter and wider than the curve on the right, because the curve on the left has a bigger standard deviation.

7. Linear combination of independent normal variates is also a normal variate

8. The total area under the normal curve ($\int_{-\infty}^{+\infty} f(x)dx = 1$) is distributed as follows

- $(\mu - \sigma) < x < (\mu + \sigma)$ covers 68.26% of the area
- $(\mu - 2\sigma) < x < (\mu + 2\sigma)$ covers 95.44% of the area
- $(\mu - 3\sigma) < x < (\mu + 3\sigma)$ covers 99.74% of the area, and it can be represented as follows



Standard Normal Distribution:

The distribution of a random variable with mean ‘0’ and variance ‘1’ is called a standard normal distribution. If Z is a standard normal variate then Z~N(0,1).

The probability density function of the standard normal variate Z is given by the probability law

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < +\infty$$

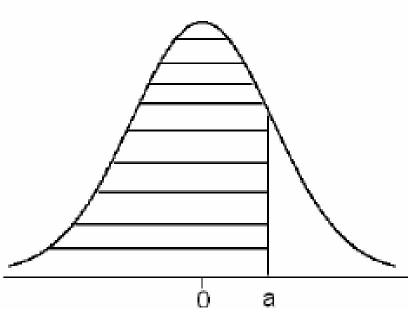
The standard normal distribution, N (0, 1), is very important because probabilities of any normal distribution can be calculated from the probabilities of the standard normal distribution.

Note:

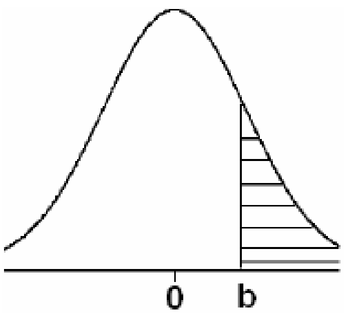
1. If X is a normal random variable with mean μ and standard deviation σ, then $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable and hence $P(x_1 < X < x_2) = P(\frac{x_1-\mu}{\sigma} < Z < \frac{x_2-\mu}{\sigma})$

2. Suppose Z ~ N(0, 1) is standard normal variate then by using the standard normal distribution area tables, we can calculate the various probabilities as explained below:

i) $P(Z<a) \cong P(Z \le a)$. This probability can be read from the table and is described in the following figure

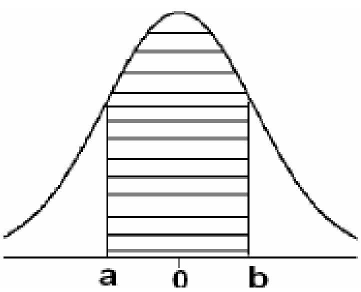


ii) $P(Z>b)$. This probability can be represented by using the following graph of standard normal distribution and it cannot be read directly from the standard normal tables



$\therefore$ $P(Z>b)=1- P(Z\le b)$, where $P(Z\le b)$ available directly from table.

iii) $P(a\le Z\le b)$. This probability can be represented by using the following graph of normal distribution



$\therefore$ $P(a \le Z \le b) = P(Z \le b) - P(Z \le a)$, where $P(Z \le b)$ and $P(Z \le a)$ are available directly from standard normal tables.

The Normal approximation to the Binomial distribution

Given X is a random variable which follows the binomial distribution with parameters n and p, then the limiting form of the distribution is standard normal distribution i.e., $Z = \frac{X-np}{\sqrt{npq}}$ $, where\ q = 1 - p$ provided, if n is large and p is not close to 0 or 1 is a standard normal variate.. If both np and nq are greater than 5, the approximation will be good.

Example:

With an eye toward improving performance, industrial engineers studied the ability of scanners to read the bar codes of various food and household products. The maximum reduction in power, just before the scanner cannot read the bar code at a fixed dictionary is called the maximum attenuation. This quantity, measured in decibles, varies from product to product: After collecting the data, the engineers decided to model the variation in maximum attenuation as a normal distribution with mean 10.1 dB and standard deviation 2.7 dB.

a) For the next food product, what is the probability that its maximum attenuation is between 8.5 dB and 13.0 dB?
b) According to the normal model, what proportion of the products has maximum attenuation between 8.5 dB and 13.0 dB?
c) What proportion of the products has maximum attenuation greater than 15.1 dB?

Solution: Let X be the maximum attenuation of the next product, Then X is a normal variable with μ=10.1 and σ=2.7.

$$z = \frac{X - 10.1}{2.7}$$

a) Probability that the maximum attenuation of the next product is between 8.5 dB and 13.0 dB.

$$= P(8.5 \le X \le 13.0) = P(\frac{8.5 - 10.1}{2.7} \le X \le \frac{13.0 - 10.1}{2.7}) = P(-0.59 \le Z \le 1.07) = P(Z \le 1.07) - P(Z \le -0.59)$$

$$= 0.8577 - 0.2776 = 0.5801.$$

b) 0.5801 is the proportion of the product having maximum attenuation between 8.5 and 13.0 dB

c) Proportion of the products having maximum attenuation greater than 15.1 dB

$= P(X > 15.1) = P(Z > (15.1 - 10.1)/(2.7)) = P(Z > 1.85) = 1 - 0.9678 = 0.0322$

Exponential distribution:

Exponential distribution plays an important role in both queuing theory and reliability. Time between arrivals at service facilities, and time to failure of components and electrical systems, often is nicely modeled by the exponential distribution.

Definition: The continuous random variable X has an Exponential distribution, with parameter β (β>0), if its density function is given by

$$f(x, \beta) = \begin{cases} \beta\, e^{-\beta x} & x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Note: 1) The mean and variance of the Exponential distribution are $\mu = \frac{1}{\beta}\, and\, \sigma^2 = \frac{1}{\beta^2}$.

2) Cumulative distribution function F(x) of the Exponential distribution is $F(x) = \beta \int_0^x e^{-\beta t} dt = 1 - e^{-\beta x}$.

$$\beta - \text{failure rate}$$

Memory less property of the exponential distribution:

The types of applications of the exponential distribution in reliability, and component or machine life time problems is influenced by the memory less or lack of memory property of the exponential distribution.

For example, in the case of, say, an electronic component where distribution of life time has an exponential distribution, the probability that the component lasts, say t hours, that is $P(X \geq t)$, is the same as the conditional probability

$$P(X \geq t_0 + t / x \geq t_0).$$

So, if the component "makes it" to t_0 hours, the probability of lasting an additional t hours is the same as the probability of lasting t hours. So, there is no "punishment" through wear that may have ensued for lasting the first t_0 hours. Thus, exponential distribution is more appropriate when the memory less property is justified. But if the failure of a component is a result of gradual or slow wear, then the exponential does not apply and Weibull distribution may be more appropriate.

Note: Cumulative distribution function F(x) for the exponential distribution is

$$F(x) = \beta \int_{0}^{x} e^{-\beta t} dt = 1 - e^{-\beta x}$$

Example: At a receiving dock on an average 3 trucks arrive per hour to be unloaded at a warehouse. What are the probabilities that the time between the arrivals of successive trucks will be

a) less than 5 minutes b) at least 45 minutes c) is between 5 to 30 minutes

Solution: Assuming that the arrivals follow Poisson process, time interval between successive arrivals has an exponential distribution with mean 1/3 hours.

∴ Parameter of the exponential distribution β=3

The probability density is $f(x)=3e^{-3x}$

a) P(inter arrival time is less than 5 minutes)=P(inter arrival time is less than 1/12 hours)

$$= \int_{0}^{1/12} 3e^{-3x} = 1 - e^{-1/4} = 0.221$$

b) P(inter arrival time is at least 45 minutes)=P(inter arrival time is at least ¾ hours)

$$= \int_{3/4}^{\infty} 3e^{-3x} = e^{-9/4} = 0.105$$

c) P(inter arrival time is between 5 to 30 minutes)

$$= \int_{1/12}^{1/2} 3e^{-3x} = \frac{3}{-3} [e^{-3x}]_{1/12}^{1/2} = -[e^{-\frac{3}{2}} - e^{-\frac{3}{12}}].$$

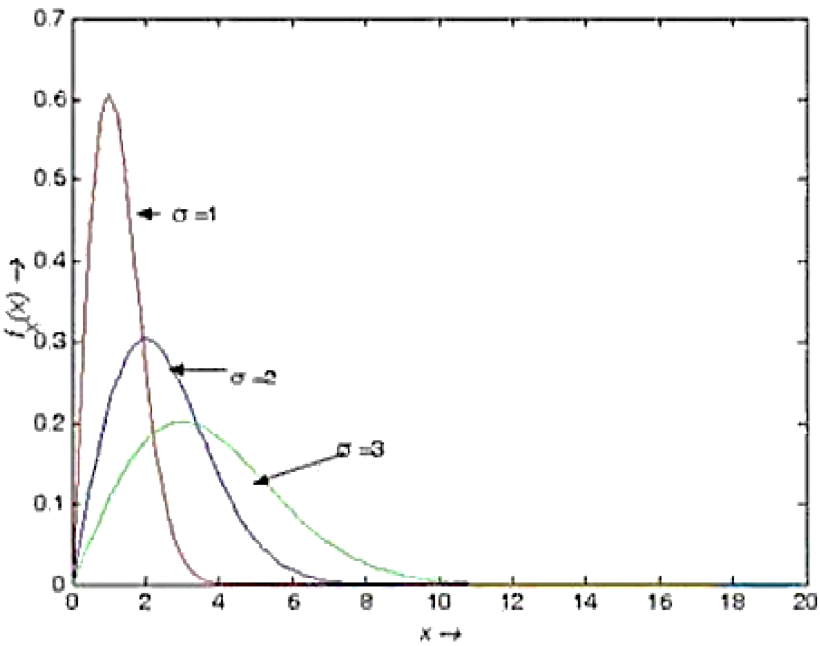
Rayleigh Random Variable

A Rayleigh random variable X is characterized by the PDF

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-x^2/2\sigma^2}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where σ is the parameter of the random variable.

probability density functions for the Rayleigh RVs are illustrated in Figure



Mean and Variance of the Rayleigh Distribution

$$
\begin{aligned}
EX &= \int_{-\infty}^{\infty} x f_X(x)dx \\
&= \int_{0}^{\infty} x \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} dx \\
&= \frac{\sqrt{2\pi}}{\sigma} \int_{0}^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \\
&= \frac{\sqrt{2\pi}}{\sigma} \frac{\sigma^2}{2} \\
&= \sqrt{\frac{\pi}{2}}\sigma
\end{aligned}
$$

similarly

$$
\begin{aligned}
EX^2 &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_{0}^{\infty} x^2 \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} dx \\
&= 2\sigma^2 \int_{0}^{\infty} ue^{-u} du \quad (\textbf{ Substituting } u = \frac{x^2}{2\sigma^2}) \\
&= 2\sigma^2 \qquad (\textbf{ Noting that } \int_{0}^{\infty} ue^{-u} du \textbf{ is the mean of the exponential RV with } \lambda = 1)
\end{aligned}
$$

$$
\begin{aligned}
\therefore \sigma_X^2 &= 2\sigma^2 - \left(\sqrt{\frac{\pi}{2}}\sigma\right)^2 \\
&= (2 - \frac{\pi}{2})\sigma^2
\end{aligned}
$$

Relation between the Rayleigh Distribution and the Gaussian distribution

A Rayleigh RV is related to Gaussian RVs as follow: If $X_1 \sim N(0, \sigma^2)$ and $X_2 \sim N(0, \sigma^2)$ are independent, then the envelope $X = \sqrt{X_1^2 + X_2^2}$ has the Rayleigh distribution with the parameter σ.

We shall prove this result in a later lecture. This important result also suggests the cases where the Rayleigh RV can be used.

Application of the Rayleigh RV

- ✓ Modeling the *root mean square error*-
- ✓ Modeling the envelope of a signal with two *orthogonal components* as in the case of a signal of the following form:

Conditional Distribution and Density functions

We discussed conditional probability in an earlier lecture. For two events A and B with $P(B) \neq 0$, the conditional probability $P(A/B)$ was defined as

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

Clearly, the conditional probability can be defined on events involving a Random Variable X

Conditional distribution function:

Consider the event $\{X \leq x\}$ and any event B involving the random variable X. The conditional distribution function of X given B is defined as

$$\begin{aligned} F_X(x/B) &= P\left[\{X \leq x\}/B\right] \\ &= \frac{P\left[\{X \leq x\} \cap B\right]}{P(B)} \qquad P(B) \neq 0 \end{aligned}$$

We can verify that $F_X(x/B)$ satisfies all the properties of the distribution function. Particularly.

- $F_X(-\infty/B) = 0$ And $F_X(\infty/B) = 1$.
- $0 \leq F_X(x/B) \leq 1$.
- $F_X(x/B)$ Is a non-decreasing function of x.
- $$\begin{aligned} P(\{x_1 < X \leq x_2\}/B) &= P(\{X \leq x_2\}/B) - P(\{X \leq x_1\}/B) \\ &= F_X(x_2/B) - F_X(x_1/B) \end{aligned}$$

Conditional Probability Density Function

In a similar manner, we can define the conditional density function $f_X(x/B)$ of the random variable X given the event B as

$$f_X(x/B) = \frac{d}{dx}F_X(x/B)$$

All the properties of the pdf applies to the conditional pdf and we can easily show that

- $f_X(x/B) \geq 0$
- $\int_{-\infty}^{\infty} f_X(x/B)dx = F_X(\infty/B) = 1$
- $F_X(x/B) = \int_{-\infty}^{x} f_X(u/B)du$

$$P(\{x_1 < X \leq x_2\}/B) = F_X(x_2/B) - F_X(x_1/B)$$
$$= \int_{x_1}^{x_2} f_X(x/B)dx$$

Example 1 Suppose X is a random variable with the distribution function $F_X(x)$. Define $B = \{X \leq b\}$

$$F_X(x/B) = \frac{P(\{X \leq x\} \cap B)}{P(B)}$$
$$= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{P\{X \leq b\}}$$
$$= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{P\{X \leq b\}}$$
$$= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)}$$

Case 1: $x < b$

Then

$$F_X(x/B) = \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)}$$
$$= \frac{P(\{X \leq x\})}{F_X(b)} = \frac{F_X(x)}{F_X(b)}$$

$$f_X(x/B) = \frac{d}{dx}\frac{F_X(x)}{F_X(b)} = \frac{f_X(x)}{F_X(b)}$$

Case 2: $x \geq b$

$$F_X(x/B) = \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)}$$

$$= \frac{P(\{X \leq b\})}{F_X(b)} = \frac{F_X(b)}{F_X(b)} = 1$$

$F_X(x/B)$ and $f_X(x/B)$ are plotted in the following figures.

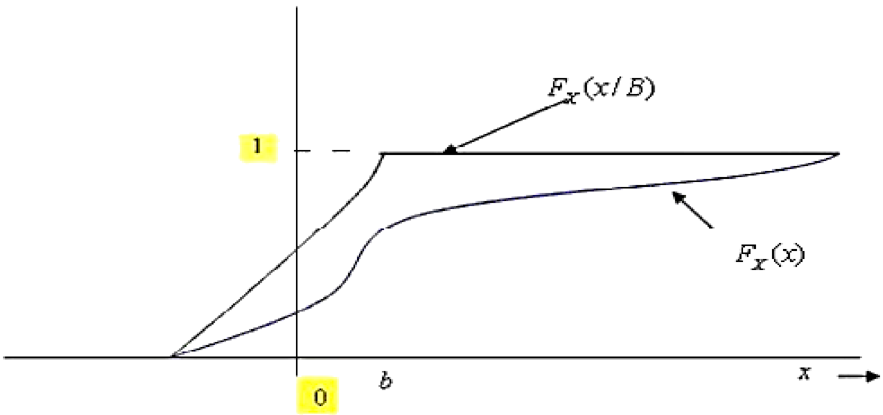
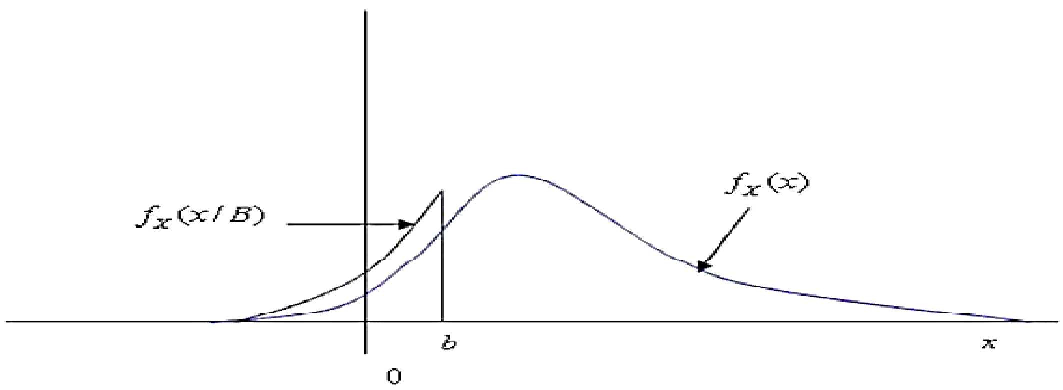


Figure 1



Example2: Suppose X is a random variable with the distribution function $F_X(x)$ and $B = \{X > b\}$.

Then

$$F_X(x/B) = \frac{P(\{X \le x\} \cap B)}{P(B)}$$
$$= \frac{P(\{X \le x\} \cap \{X > b\})}{P\{X > b\}}$$
$$= \frac{P(\{X \le x\} \cap \{X > b\})}{1 - F_X(b)}$$

For $x \le b$, $\{X \le x\} \cap \{X > b\} = \phi$. Therefore,

$$F_X(x/B) = 0 \qquad x \le b$$

For $x > b$, $\{X \le x\} \cap \{X > b\} = \{b < X \le x\}$. Therefore,

$$F_X(x/B) = \frac{P(\{b < X \le x\})}{1 - F_X(b)}$$
$$= \frac{F_X(x) - F_X(b)}{1 - F_X(b)}$$

Thus,

$$F_X(x/B) = \begin{cases} 0 \quad , & x \le b \\ \dfrac{F_X(x) - F_X(b)}{1 - F_X(b)} \quad , & \text{otherwise} \end{cases}$$

the corresponding pdf is given by

$$f_X(x/B) = \begin{cases} 0 \quad , & x \le b \\ \dfrac{f_X(x)}{1 - F_X(b)} & \text{otherwise} \end{cases}$$

Short notes on above all distributions:

Special Distributions

The probability mass functions of some discrete RVs and the probability density functions of some continuous RV*s*, which are of frequent applications, are as follows:

Discrete Distributions

1. If the discrete RV X can take the values 0, 1, 2, ..., n, such that $P(X = i) = nC_i\, p^i\, q^{n-i}$, $i = 0, 1, \ldots, n$, where $p + q = 1$, then X is said to follow a *binomial distribution* with parameters n and p,which is denoted a $B(n, p)$.

2. If the discrete RV X can take the values 0, 1, 2, ..., such that $P(x = i) = \frac{e^{-\lambda} \lambda^i}{i!}$, $i = 0, 1, 2, K$, then X is said to follow a Poisson distribution with parameter λ.

Continuous Distributions

5. If the pdf of a continuous RV X is $f(x) = \frac{1}{b-a}$ (a constant), $a \le x \le b$, then X follows a *uniform distribution* (*or rectangular distribution*).

6. If the pdf of a continuous RV X is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$, $-\infty < x < \infty$, then X is said to follow a *normal distribution* (or *Gaussian distribution*) with parameters μ and σ, which will be hereafter denoted as $N(\mu, \sigma)$.

7. If the pdf of a continuous RV X is $f(x) = \frac{1}{\Gamma(n)} e^{-x} x^{n-1}$, $0 < x < \infty$ and $n > 0$, then X follows a *gamma distribution* with parameter n. Gamma distribution is a particular case of *Erlang distribution*, the pdf of which is $f(x) = \frac{c^n}{\Gamma(n)} x^{n-1} e^{-cx}$, $0 < x < \infty$, $n > 0$, $c > 0$.

8. An Erlang distribution with $n = 1$ [i.e., $f(x) = ce^{-cx}$, $0 < x < \infty$, $c > 0$] is called *an exponential* (or *negative exponential*) *distribution* with parameter c.

9. If the pdf of a continuous RV X is $f(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2}$, $0 < x < \infty$, then X follows a *Rayleigh distribution* with parameter α.

10. If the pdf of a continuous RV X is $f(x) = \frac{\sqrt{2}}{\alpha^3\sqrt{\pi}} x^2 e^{-x^2/2\alpha^2}$, $0 < x < \infty$, then X follows a *Maxwell distribution* with parameter α.

11. If the pdf of a continuous RV X is $f(x) = \frac{1}{2\lambda} e^{-|x-\mu|/\lambda}$, $-\infty < x < \infty$, $\lambda > 0$, then X follows a *Laplace* (or *double exponential*) *distribution* with parameters λ and μ.

Example 1 *From a lot containing 25 items, 5 of which are defective, 4 items are chosen at random. If X is the number of defectives found, obtain the probability distribution of X, when the items are chosen (i) without replacement and (ii) with replacement.*

Solution Since only 4 items are chosen, X can take the values 0, 1, 2, 3 and 4. The lot contains 20 non-defective and 5 defective items.

Case (i): When the items are chosen without replacement, we can assume that all the 4 items are chosen simultaneously.

$$\therefore \quad P(X = r) = P(\text{choosing exactly } r \text{ defective items})$$

$$= P(\text{choosing } r \text{ defective and } (4 - r) \text{ good items})$$

$$= \frac{5C_r \times 20C_{4-r}}{25C_4} \quad (r = 0, 1, \ldots, 4)$$

Case (ii): When the items are chosen with replacement, we note that the probability of an item being defective remains the same in each draw.

i.e., $$p = \frac{5}{25} = \frac{1}{5},\ q = \frac{4}{5} \text{ and } n = 4$$

The problem is one of performing 4 Bernoulli's trials and finding the probability of exactly r successes.

$$\therefore \quad P(X = r) = 4C_r \left(\frac{1}{5}\right)^r \left(\frac{4}{5}\right)^{4-r} \quad (r = 0, 1, \ldots, 4)$$

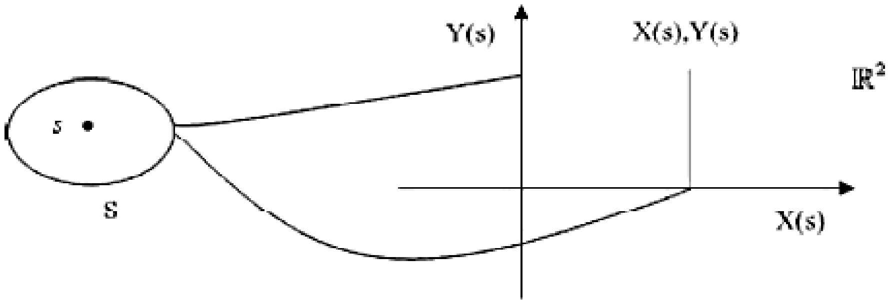
Multiple Random Variables

In many applications we have to deal with more than two random variables. For example, in the navigation problem, the position of a space craft is represented by three random variables denoting the x, y and z coordinates. The noise affecting the R, G, B channels of colour video may be represented by three random variables. In such situations, it is convenient to define the vector-valued random variables where each component of the vector is a random variable.

In this lecture, we extend the concepts of joint random variables to the case of multiple random variables. A generalized analysis will be presented for n random variables defined on the same sample space.

Jointly Distributed Random Variables

We may define two or more random variables on the same sample space. Let X and Y be two real random variables defined on the same probability space $(S,\mathbb{F},P)$. The mapping $S\to\mathbb{R}^2$ such that for $s\in S$, $(X(s),Y(s))\in\mathbb{R}^2$ is called a joint random variable.



Joint Probability distributions:

If X and Y are two discrete random variables, the probability distribution for their simultaneous occurrence can be represented by a function with values f(x,y) = P(X=x, Y=y); the function f(x, y) is a joint Probability distribution or Probability mass function of the random variables X and Y if

1. $f(x,y)\geq 0 \quad for\, all\ (x,y)$
2. $\sum_x\sum_y f(x,y)=1$
3. $P(X=x,Y=y)=f(x,y).$

Joint density function: The function f(x, y) is a joint density function of the continuous random variables X and Y if

1. $f(x,y)\geq 0 \quad for\, all\ (x,y)$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1$
3. $P[(X,Y)\in A]=\iint_A f(x,y).dxdy.$

Cumulative Distribution Function

If (X, Y) is a two-dimensional RV (discrete or continuous), then $F(x, y) = P\{X \le x \text{ and } Y \le y\}$ is called *the cdf of* (X, Y).
In the discrete case,

$$F(x, y) = \sum_{\substack{j \\ y_j \le y}} \sum_{\substack{i \\ x_i \le x}} p_{ij}$$

In the continuous case,

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

Properties of F(x, y)

(i) $F(-\infty, y) = 0 = F(x, -\infty)$ and $F(\infty, \infty) = 1$

(ii) $P\{a < X < b, Y \le y\} = F(b, y) - F(a, y)$

(iii) $P\{X \le x, c < Y < d\} = F(x, d) - F(x, c)$

(iv) $P\{a < X < b, c < Y < d\} = F(b, d) - F(a, d) - F(b, c) + F(a, c)$

(v) At points of continuity of $f(x, y)$

$$\frac{\partial^2 F}{\partial x\, \partial y} = f(x, y)$$

Marginal Distribution:

The Marginal Distribution of X alone and of Y alone are

$g(x) = \sum_{y} f(x, y)$ and $h(y) = \sum_{x} f(x, y)$ for the discrete case and $g(x) = \int_{-\infty}^{\infty} f(x, y)dy$ and $h(y) = \int_{-\infty}^{\infty} f(x, y)dx$

Conditional distribution: Let X and Y be two random variables, discrete or continuous. The conditional distribution of the random variable Y given that X =x is

$$f(y/x) = \frac{f(x,y)}{g(x)},\ g(x) > 0$$

Similarly the conditional distribution of the random variable X given that Y=y is

$$f(x/y) = \frac{f(x,y)}{h(y)},\ h(y) > 0.$$

Statistical Independence: Two random variables X and Y, discrete or continuous with joint probability distribution of f(x, y) and marginal distributions g(x) and h(y), respectively. The random variables X and Y are said to be statistically independent if and only if f(x,y)=g(x)h(y) for all (x,y) within their range.

Example: Given the joint density function

$$f(x, y) = \begin{cases} \dfrac{x(1 + 3y^2)}{4}, & 0 < x < 2, 0 < y < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

Find g(x), h(y), f(x/y), and evaluate $P(\frac{1}{4} < X < \frac{1}{2} / Y = \frac{1}{3})$

Solution: By definition

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{0}^{1} \frac{x(1+3y^2)}{4} dy$$

$$= \left(\frac{xy}{4} + \frac{xy^3}{4}\right)\Big|_{y=0}^{y=1} = \frac{x}{2}, \quad 0 < x < 2$$

and

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{0}^{2} \frac{x(1+3y^2)}{4} dx$$

$$= \left(\frac{x^2}{8} + \frac{3x^2y^2}{8}\right)\Big|_{x=0}^{x=2} = \frac{1+3y^2}{2}, \quad 0 < y < 1$$

Therefore,

$$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{x(1+3y^2)/4}{(1+3y^2)/2} = \frac{x}{2}, \qquad 0 < x < 2$$

and

$$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{x(1+3y^2)/4}{(1+3y^2)/2} = \frac{x}{2}, \qquad 0 < x < 2$$

$$P\left(\frac{1}{4} < X < \frac{1}{2} \middle| Y = \frac{1}{3}\right) = \int_{1/4}^{1/2} \frac{x}{2} dx = \frac{3}{64}.$$

<u>Mean or expected value of a random variable:</u>

Let X be a random variable with probability distribution f(x). The mean or expected value of X is $\mu = E(X) = \sum_x xf(x)$, if X is discrete and

$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$, If X is continuous.

Note: If X is a random variable, a function of X, g(X) is also a random variable.

E(x)=0*P(X=0)+1*P(X=1)+2*P(X=2)+3*P(X=3)=

The expected value of the random variable g(X) is $\mu_{g(X)} = E(g(X)) = \sum_x g(x)f(x)$, if X is discrete, and $\mu_{g(X)} = \int_{-\infty}^{\infty} g(x)f(x)dx$, if X is continuous.

<u>Mean of a joint probability Distribution:</u>

Let X and Y be random variables with joint probability Distribution f(x, y). The mean or expected value of the random variable g(X, Y) is

$\mu_{g(X,Y)} = E\big(g(X,Y)\big) = \sum_x \sum_y g(x,y)f(x,y)$, if X and y are discrete and,

$\mu_{g(X,Y)} = E\big(g(X,Y)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$, X and Y are continuous.

Variance of random variable:

Le X be a random variable with probability distribution f(x) and mean μ . The variance of X is

$\sigma^2 = E\left[(X-\mu)^2\right] = \sum_x (X-\mu)^2\ f(x),$ if X is discrete, and

$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (X-\mu)^2\ f(x)$, if X is continuous.

Note: 1) $\sigma^2 = E(X^2) - \mu^2$

2) $E(aX+b) = aE(X)+b$

3) $V(aX+b) = a^2V(X)$

4) If X and Y are independent random variables, then $V(aX+bY) = a^2V(X)+b^2V(Y)$

Random Vector

Definitions: A vector X: $[X_1, X_2, \ldots, X_n]$ whose components X_i are RVs is called a *random vector*. $(X_1, X_2, \ldots, X_n)$ can assume all values in some region R_n of the n-dimensional space. R_n is called the *range space*.

The joint distribution function of $(X_1, X_2, \ldots \ldots, X_n)$ is defined as $F(x_1, x_2, \ldots, x_n) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n]$

The joint pdf of $(X_1, X_2, \ldots, X_n)$ is defined as $f(x_1, x_2, \ldots, x_n)$

$$= \frac{\partial^n F(x_1, x_2 \cdots x_n)}{\partial x_1 \times \partial x_2 \cdots \partial x_n}$$ and satisfies the following conditions.

(i) $f(x_1, x_2, \ldots, x_n) \ge 0$, for all $(x_1, x_2, \ldots, x_n)$

(ii) $\iint\limits_{R_n}\int \cdots \int f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n = 1$

(iii) $P[(X_1, X_2, \ldots, X_n) \in D] = \iint\limits_{D} \ldots \int f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$ where D is a subset of the range space R_n.

The marginal pdf of any subset of the n RVs $X_1, X_2, \ldots, X_n$ is obtained by "integrating out" the variables not in the subset. For example, if $n = 3$, then

$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3) dx_2 dx_3$ is the marginal pdf of the one-dimensional RV X_1 and $f_{X_1, X_2}(x_1, x_2) = \int_{-\infty}^{\infty} f(x_1\ x_2\ x_3)\, dx_3$ is the marginal joint pdf of the two-dimensional RV (X_1, X_2). The concept of independent RVs is also extended in a natural way. The RVs $(X_1, X_2, \ldots, X_n)$ are said to be independent, if

$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n)$

The conditional density functions are defined as in the following examples.

If $n = 3$,

$$f(x_1, x_2 / x_3) = \frac{f(x_1, x_2, x_3)}{f_{X_3}(x_3)}$$ and

$$f(x_1 / x_2, x_3) = \frac{f(x_1, x_2, x_3)}{f_{X_2, X_3}(x_2, x_3)}.$$

Sum of two random variables

Mathematically, the **sum of two random variables** X and Y is denoted as $Z = X + Y$. The probability distribution of Z can be obtained by convolving the probability distributions of X and Y.

If X and Y are discrete random variables, their sum Z will also be a discrete random variable. The probability mass function (PMF) of Z is given by:

$$P(Z = z) = \Sigma\, P(X = x, Y = z - x)$$

where the sum is taken over all possible values of x that satisfy the equation z = x + y.

If X and Y are continuous random variables, their sum Z will be a continuous random variable. The probability density function (PDF) of Z is given by the convolution of the PDFs of X and Y:

$$fZ(z) = \int fX(x)\, fY(z - x)\, dx$$

where the integral is taken over the entire range of possible values of x.

In summary, the sum of two random variables involves combining their probability distributions to obtain the probability distribution of their sum. The specific form of the probability distribution depends on whether the random variables are discrete or continuous.

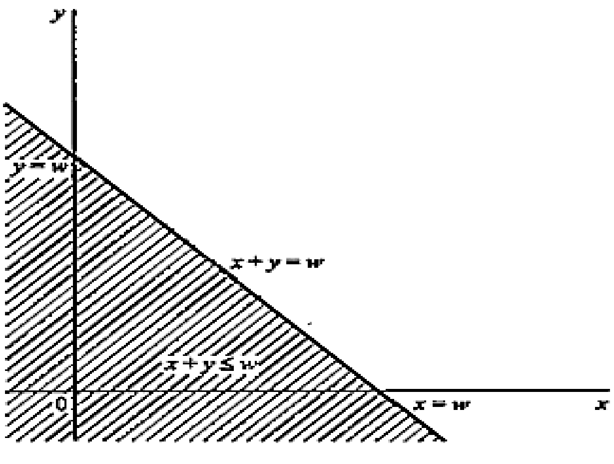


FIGURE 1
Region in xy plane where $x + y \leq w$.

Sum of several random variables:

Mathematically, the sum of **several random variables** $X_1, X_2, ..., X_n$ is denoted as $S = X_1 + X_2 + ... + X_n$. The probability distribution of S can be obtained by convolving the probability distributions of the individual random variables.

If the random variables $X_1, X_2, ..., X_n$ are discrete, their sum S will also be a discrete random variable. The probability mass function (PMF) of S is given by:

$$P(S = s) = \Sigma\, P(X_1 = x_1, X_2 = x_2, ..., X_n = s - x_1 - x_2 - ... - x_n)$$

where the sum is taken over all possible combinations of values $x_1, x_2, ..., x_n$ that satisfy the equation $s = x_1 + x_2 + ... + x_n$.

If the random variables $X_1, X_2, ..., X_n$ are continuous, their sum S will be a continuous random variable. The probability density function (PDF) of S is given by the convolution of the PDFs of the individual random variables:

$$fS(s) = \int \int ... \int\, fX_1(x_1)\, fX_2(x_2) ... fX_n(s - x_1 - x_2 - ... - x_n)\, dx_1\, dx_2 ... dx_n$$

where the integrals are taken over the entire range of possible values for each individual random variable.

Central Limit Theorem: Unequal Distribution, Equal Distributions

Statement of the Central Limit Theorem: Let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distributed random variables with a common mean μ and standard deviation σ. As n approaches infinity, the distribution of the sample mean (or the sum) of the random variables approaches a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

Unequal Distribution: When referring to the Central Limit Theorem with unequal distributions, it means that the random variables $X_1, X_2, ..., X_n$ are not necessarily drawn from the same distribution, but they are still independent and have the same mean and standard deviation. Despite the differences in their individual distributions, the sum or average of these variables still follows an approximately normal distribution as n becomes large.

Equal Distributions: In the context of the Central Limit Theorem, equal distributions imply that the random variables $X_1, X_2, ..., X_n$ are not only independent but also identically distributed. This means that they are drawn from the same distribution with the same parameters. In this case, the CLT guarantees that the distribution of the sample mean or sum converges to a normal distribution as the sample size increases.

In summary, the Central Limit Theorem states that, under certain conditions, the sum or average of a large number of independent and identically distributed random variables will have an approximately normal distribution. This result holds even when the original random variables have different distributions (unequal distribution) or when they are drawn from the same distribution (equal distributions). The CLT is a powerful tool in statistics, allowing for the approximation of probabilities and the estimation of parameters in various fields of study.

Examples:

The heights of a population of adult males follow a normal distribution with a mean of 175 cm and a standard deviation of 6 cm. Suppose a random sample of 100 adult males is selected. Find the probability that the average height of this sample is between 174 cm and 177 cm.

Solution:

To find the probability that the average height of a random sample of 100 adult males is between 174 cm and 177 cm, we can use the Central Limit Theorem and the properties of the normal distribution.

Given:
Population mean $(\mu) = 175\ cm$
Population standard deviation $(\sigma) = 6\ cm$
Sample size $(n) = 100$

According to the Central Limit Theorem, when the sample size is large enough, the distribution of the sample means approaches a normal distribution, regardless of the shape of the population distribution.

The mean of the sample means (also known as the sampling distribution mean) is equal to the population mean, which is 175 cm. The standard deviation of the sample means (also known as the standard error) is equal to the population standard deviation divided by the square root of the sample size, which is $6\ cm / \sqrt{100} = 0.6\ cm$.

To calculate the probability that the average height of the sample falls between 174 cm and 177 cm, we can convert these values to z-scores using the formula:

$$z = (x - \mu) / \sigma$$

where x is the value of interest, μ is the population mean, and σ is the population standard deviation.

For 174 cm:

$$z_1 = (174 - 175) / 0.6 = -1.67$$

For 177 cm:

$$z_2 = (177 - 175) / 0.6 = 3.33$$

We can then use the z-table or a statistical software to find the probabilities associated with these z-scores. The probability that the average height of the sample is between 174 cm and 177 cm can be calculated as the difference between the cumulative probabilities:

$$P(174\, cm \le x \le 177\, cm) = P(z_1 \le Z \le z_2)$$

Using the z-table or a statistical software, we can find the corresponding probabilities and calculate the final result.

Note: The z-table provides the cumulative probabilities for standard normal distribution, so we need to use the calculated z-scores to find the probabilities in the table.

Example:
A manufacturing process produces bolts with a diameter that follows a normal distribution with a mean of 12 mm and a standard deviation of 0.5 mm. A random sample of 50 bolts is selected. Find the probability that the average diameter of this sample is greater than 12.2 mm.

Solution:
To find the probability that the average diameter of a random sample of 50 bolts is greater than 12.2 mm, we can use the Central Limit Theorem and the properties of the normal distribution.

Given:
Population mean $(\mu) = 12\, mm$
Population standard deviation $(\sigma) = 0.5\, mm$
Sample size $(n) = 50$

According to the Central Limit Theorem, when the sample size is large enough, the distribution of the sample means approaches a normal distribution, regardless of the shape of the population distribution.

The mean of the sample means (also known as the sampling distribution mean) is equal to the population mean, which is 12 mm. The standard deviation of the sample means (also known as the standard error) is equal to the population standard deviation divided by the square root of the sample size, which is $0.5\, mm / \sqrt{50} \approx 0.0707\, mm$.

To calculate the probability that the average diameter of the sample is greater than 12.2 mm, we can convert this value to a z-score using the formula:

$$z = (x - \mu) / \sigma$$

where x is the value of interest, μ is the population mean, and σ is the population standard deviation.

For 12.2 mm:

$$z = (12.2 - 12) / 0.0707 \approx 2.828$$

We can then use the z-table or a statistical software to find the cumulative probability associated with this z-score. The probability that the average diameter of the sample is greater than 12.2 mm can be calculated as:

$$P(x > 12.2\, mm) = 1 - P(z \leq 2.828)$$

Note: The z-table provides the cumulative probabilities for standard normal distribution, so we need to use the calculated z-score to find the probability in the table.