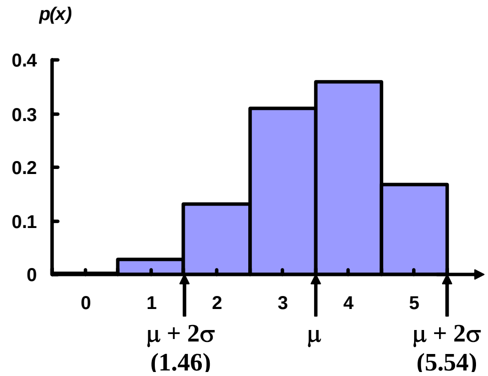
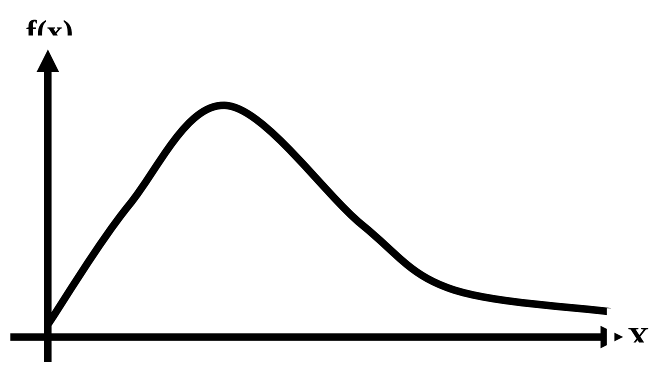


This probability is P(2) + P(3) + P(4) + P(5)
= .132 +.309 + .360 + .168
= .969.
Therefore, 96.9% of the probability distribution lies within 2 standard deviations of the mean. This percentage is *CONSISTENT* with both the Chebychev's rule and the Empirical Rule.

d) Fewer than two successful ventures out of five implies that x = 0 or x = 1. Since both these values of x lie outside the interval $\mu + 2\sigma$, we know from the Empirical Rule that such a result is unlikely (with approximate probability of only .05).The exact probability, $P(x < 1)$, is P(0) + P(1) = .002 + .029 = .031.
Consequently, in a *single* experiment where we invest in five business ventures, we would *not* expect to observe fewer than two successful ones. The *key* question: What is the ***significance*** of the Chebychev's Inequality and the Empirical Rule?
The answer to this question is that both these rules assist us in having a certain *IDEA* regarding amount of data lying between the mean minus a certain number of standard deviations and mean plus that same number of standard deviations. Given any data-set, the moment we compute the mean and standard deviation, we *HAVE* an idea regarding the two points (i.e. mean minus two standard deviations, and mean plus two standard deviations) between which the *BULK* of our data lies. If our data-set is hump-shaped, we obtain this idea through the *Empirical Rule*, and if we don't have any reason to believe that our data-set is hump-shaped, then we obtain this idea through the Chebychev's Rule
We now begin the discussion of CONTINUOUS RANDOM VARIABLES – quantities that are measurable. As stated in the very first lecture, continuous variables result from measurement, and can therefore take any value within a certain range. For example, the height of a normal Pakistani adult male may take any value between 5 feet 4 inches and 6 feet. The temperature at a place, the amount of rainfall, time to failure for an electronic system, etc. are all *examples* of continuous random variable. Formally speaking, a continuous random variable can be defined as follows:

<u>CONTINUOUS RANDOM VARIABLE</u>

A random variable X is defined to be continuous if it can assume every possible value in an interval [a, b], $a < b$, where a and b may be $-\infty$ and $+\infty$ respectively. The function f(x) is called the *probability density function*, abbreviated to *p.d.f.*, or simply density function of the random variable X. A continuous probability distribution looks something like this:



A p.d.f. has the following properties:

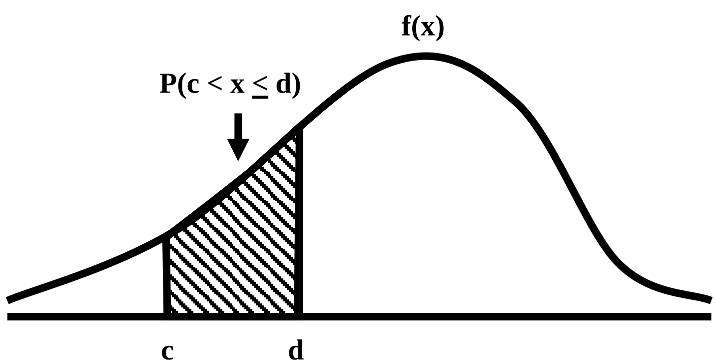
i) **$f(x) \geq 0$, for all x**

ii) $$\int_{-\infty}^{\infty} f(x)\ dx = 1$$

iii) The probability that X takes on a value in the interval [c, d], c < d is given by:

$$P(c < x < d) = \int_{c}^{d} f(x)\,dx$$

which is the area under the curve y = f(x) between X = c and X = d, as shown in the following figure:



The *TOTAL* area under the curve is 1. In other words:

- f(x) a non-negative function,
- The integration takes place over all possible values of the random variable X *between the specified limits*, and
- The probabilities are given by appropriate areas under the curve.

Since

$$P(X = k) = \int_{k}^{k} f(x)\ dx = 0,$$

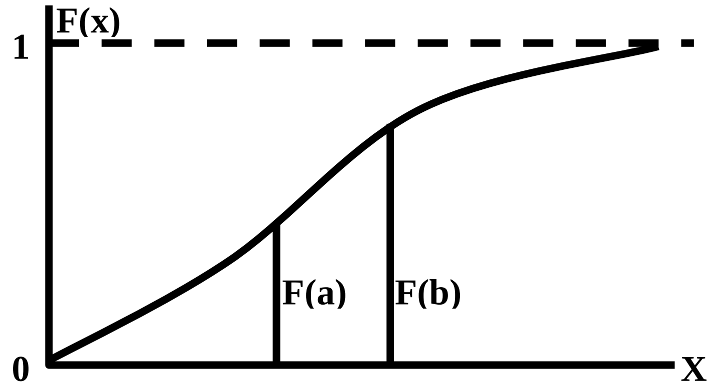
It should therefore be noted that the probability of a continuous random variable X taking any *particular* value k is always *zero*. That is why probability for a continuous random variable is measurable only over a given interval.
Further, since for a continuous random variable X, P(X = x) = 0 for every x, the following four probabilities are regarded as the same:

P(c < X < d), P(c < X < d),
P(c < X < d) and P(c < X d).

They may be different for a discrete random variable. The values (expressed as intervals) of a continuous random variable and their associated probabilities can be expressed by means of a formula.
We now discuss the *distribution function* of a continuous random variable.

CONTINUOUS RANDOM VARIABLE

A random variable X may also be defined as continuous if its *distribution function* F(x) is continuous and is differentiable everywhere except at isolated points in the given range. In contrast with the graph of the distribution function of a discrete variable, the graph of F(x) in the case of a continuous variable has no jumps or steps but is a *continuous* function for all x-values, as shown in the following figure:



Since F(x) is a non-decreasing function of x, we have
i) f(x) > 0,

ii) $F(x) = \int_{-\infty}^{x} f(x)\, dx$, for all x.

The relationship between f(x) and F(x) is as follows: f(x) is obtained by finding the derivative of F(x), i.e.

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

EXAMPLE

a) Find the value of k so that the function f(x) defined as follows, may be a density function

f(x) = kx, 0 < x < 2
= 0, elsewhere

b) Compute P(X = 1).

c) Compute P(X > 1).

d) Compute the distribution function F(x).

e) $P(X < 1/2 \ / \ 1/3 < X < 2/3)$

SOLUTION

a) The function f(x) will be a density function, if
i) f(x) > 0 for every x, and
ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

The first condition is satisfied when k > 0. The second condition will be satisfied, if

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

$$\text{i.e. if} \quad 1 = \int_{-\infty}^{0} f(x)\, dx + \int_{0}^{2} f(x)\, dx + \int_{2}^{\infty} f(x)\, dx$$

$$\text{i.e. if} \quad 1 = \int_{-\infty}^{0} 0\ dx + \int_{0}^{2} kx\ dx + \int_{2}^{\infty} 0\ dx$$

$$\text{i.e. if} \quad 1 == 0 + \left[k\frac{x^2}{2} \right]_0^2 + 0 = 2k$$

We had

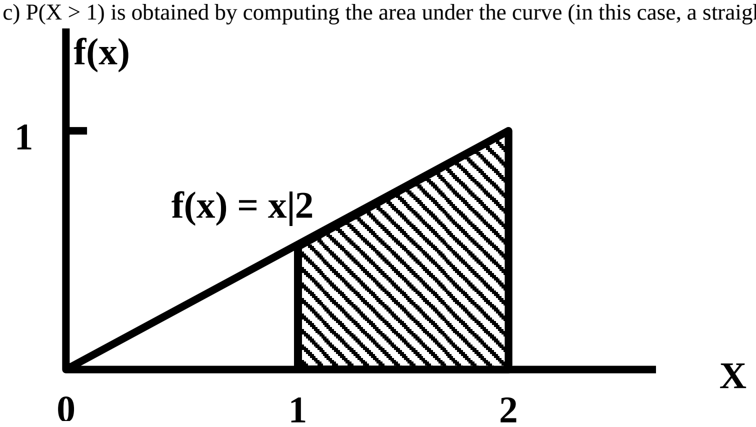
This gives k = 1/2

f(x) = kx, 0 < x < 2
= 0, elsewhere
and since we have obtained
k = 1/2, hence:

$$f(x) = \begin{cases} \frac{x}{2}, & \text{for } 0 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

b) Since f(x) is continuous probability function, therefore (X = 1) = 0.

c) P(X > 1) is obtained by computing the area under the curve (in this case, a straight line) between X=1 and X=2:



This area is obtained as follows:

$$P(X > 1) = \text{area of shaded region}$$
$$= \int_1^2 f(x)\, dx$$
$$= \int_1^2 \frac{x}{2}\, dx = \left[\frac{x^2}{4}\right]_1^2 = \frac{3}{4}$$

d) To compute the distribution function, we need to find:

$$\mathbf{F(x) = P(X < x) =} \int_{-\infty}^{x} f(x)\, dx$$

We do so *step by step*, as shown below:

For any x such that -∞ < x ≤ 0,

$$\mathbf{F(x) =} \int_{-\infty}^{x} 0\, dx = 0,$$

If 0 < x ≤ 2, we have

$$F(x) = \int_{-\infty}^{0} 0\, dx + \int_0^x \left(\frac{x}{2}\right) dx = \left[\frac{x^2}{4}\right]_0^x = \frac{x^2}{4},$$

and, finally, for x > 2 we have

$$F(x) = \int_{-\infty}^{0} 0\, dx + \int_0^2 \frac{x}{2} dx + \int_0^2 0\, dx = 1$$

Hence

F(x) = 0, for x < 0

$= \frac{x^2}{4}$, for $0 \le x \le 2$

=1, for x > 2.

We will discuss the computation of the conditional probability

$P\left(X < 1/2 \;/\; 1/3 < X < 2/3\right)$

LECTURE NO. 25

- Mathematical Expectation, Variance & Moments of a Continuous Probability Distribution
- BIVARIATE Probability Distribution

In the last lecture, we were dealing with an example of a continuous probability distribution in which we were interested in computing a conditional probability. We now discuss this particular concept

EXAMPLE

a) Find the value of k so that the function f(x) defined as follows, may be a density function

f(x) = kx, 0 < x < 2
= 0, elsewhere

b) Compute P(X = 1).
c) Compute P(X > 1).
d) Compute the distribution function F(x).
e) $P\left(X < 1/2 \;\middle/\; 1/3 < X < 2/3\right)$

SOLUTION

We had
f(x) = kx, 0 < x < 2
= 0, elsewhere
and we obtained k = 1/2.

Hence:

$$f(x) = \begin{cases} \frac{x}{2}, & for\ 0 \le x \le 2 \\ 0, & elsewhere \end{cases}$$

e) Applying the definition of conditional probability, we get

$$P\left(X \le \tfrac{1}{2} \mid \tfrac{1}{3} \le X \le \tfrac{2}{3}\right) = \frac{P\left(\frac{1}{3} \le X \le \frac{1}{2}\right)}{P\left(\frac{1}{3} \le X \le \frac{2}{3}\right)} = \frac{\int_{1/3}^{1/2} \frac{x}{2}\,dx}{\int_{1/3}^{2/3} \frac{x}{2}\,dx}$$

$$= \left[\frac{x^2}{4}\right]_{1/3}^{1/2} \div \left[\frac{x^2}{4}\right]_{1/3}^{2/3}$$

The above example was of the simplest case when the graph of our continuous probability distribution is in the form of a straight line.

Let us now consider a slightly more *complicated* situation.

EXAMPLE

A continuous random variable X has the d.f. F(x) as follows:

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ \frac{2x^2}{5}, & \text{for } 0 < x \le 1, \\ -\frac{3}{5} + \frac{2}{5}\left(3x - \frac{x^2}{2}\right), & \text{for } 1 < x \le 2, \\ 1 & \text{for } x > 2. \end{cases}$$

Find the p.d.f. and P(|X| < 1.5).

SOLUTION

By definition, we have $f(x) = \frac{d}{dx} F(x).$

Therefore $f(x) = \frac{4x}{5}$ for $0 < x \le 1$

$= \frac{2}{5}(3 - x)$ for $1 < x \le 2$

$= 0$ elsewhere.

Let us now discuss the mathematical expectation of *continuous* random variables through the following example:

EXAMPLE

Find the expected value of the random variable X having the p.d.f
$f(x) = 2(1 - x), \quad 0 < x < 1$
$= 0$, elsewhere

SOLUTION

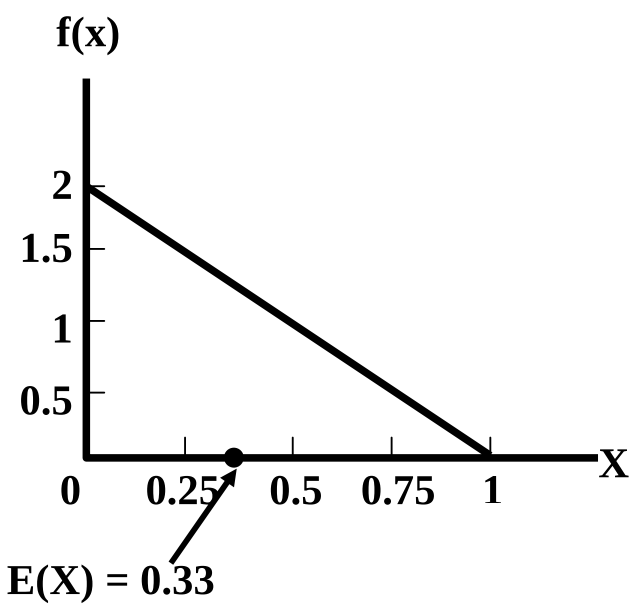
Now

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$= 2\int_{0}^{1} x\,(1 - x)\, dx$$

$$= 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 2\left[\frac{1}{2} - \frac{1}{3}\right] = \frac{1}{3}$$

As indicated earlier, the term 'expected value' implies the *mean* value. The graph of the above probability density function and its *mean* value are presented in the following figure:



Suppose that we are interested in verifying the properties of mathematical expectation that are valid in the case of univariate probability distributions. In the last lecture, we noted that if X is a discrete random variable and if a and b are constants, then

E (aX + b) = a E(X) + b.
This property is equally valid in the case of continuous probability distributions. In this example, suppose that a = 3 and b = 5. Then, we wish to verify that
E(3X + 5) = 3 E(X) + 5.
The right-hand-side of the above equation is:
3 E(X) + 5 = 3(   ) + 5 = 1 + 5 = 6
In order to compute the left-hand-side, we proceed as follows:

$$E(3X+5) = 2\int_0^1 (3x+5)(1-x)dx$$
$$= 2\int_0^1 \left(5-2x-3x^2\right)dx$$
$$= 2\left[5x - x^2 - x^3\right]_0^1$$
$$= 2[5-1-1] = 2(3) = 6.$$

Since the left-hand-side is equal to the right-hand-side, therefore the property is verified.

SPECIAL CASE

We have
E(aX + b) = a E(X) + b.
If b = 0, the above property takes the following simple form:
E(aX) = a E(X).
Next, let us consider the computation of the *moments* and *moment-ratios* in the case of a continuous probability distribution:

EXAMPLE

A continuous random variable X has the p.d.f.

$$f(x) = \frac{3}{4}x(2-x),\ 0 \le x \le 2.$$
$$= 0, \qquad \text{otherwise}$$

Find the first four moments about the mean and the moment-ratios.
We first calculate the moments about origin as

$$\mu'_1 = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$
$$= \frac{3}{4}\int_0^2 x\left(2x - x^2\right)dx = \frac{3}{4}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2$$
$$= \frac{3}{4}\left[\frac{16}{3} - \frac{16}{4}\right] = \frac{3}{4}\left[\frac{16}{12}\right] = 1;$$

$$\mu'_2 = E\left(X^2\right) = \int_{-\infty}^{\infty} x^2\, f(x)\, dx$$
$$= \frac{3}{4}\int_0^2 x^2\left(2x - x^2\right)dx = \frac{3}{4}\left[\frac{2x^4}{4} - \frac{x^5}{5}\right]_0^2$$
$$= \frac{3}{4}\left[8 - \frac{32}{5}\right] = \frac{3}{4}\left[\frac{8}{5}\right] = \frac{6}{5};$$

$$\mu'_3 = E\left(X^3\right) = \int_{-\infty}^{\infty} x^3 f(x)\,dx$$
$$= \frac{3}{4}\int_0^2 x^3\left(2x - x^2\right)dx = \frac{3}{4}\left[\frac{2x^5}{5} - \frac{x^6}{6}\right]_0^2$$
$$= \frac{3}{4}\left[\frac{64}{5} - \frac{64}{6}\right] = \frac{3}{4}\left[\frac{64}{30}\right] = \frac{8}{5};$$

$$\mu'_4 = E\left(X^4\right) = \int_{-\infty}^{\infty} x^4 f(x)\,dx$$
$$= \frac{3}{4}\int_0^2 x^4\left(2x - x^2\right)dx = \frac{3}{4}\left[\frac{2x^6}{6} - \frac{x^7}{7}\right]_0^2$$
$$= \frac{3}{4}\left[\frac{64}{3} - \frac{128}{7}\right] = \frac{3}{4}\left[\frac{64}{21}\right] = \frac{16}{7}.$$

Next, we find the moments about the mean as follows:

$$\mu_1 = 0$$
$$\mu_2 = \mu'_2 - \left(\mu'_1\right)^2 = \frac{6}{5} - (1)^2 = \frac{1}{5}$$

$$\mu_3 = \mu'_3 - 3\mu'_1\,\mu'_2 + 2\left(\mu'_1\right)^3$$
$$= \frac{8}{5} - 3(1)\left(\frac{6}{5}\right) + 2(1)^3 = \frac{8}{5} - \frac{18}{5} + 2 = 0;$$

$$\mu_4 = \mu'_4 - 4\mu'_1\,\mu'_3 + 6\left(\mu'_1\right)^2\mu'_2 - 3\left(\mu'_1\right)^4$$
$$= \frac{16}{7} - 4(1)\left(\frac{8}{5}\right) + 6(1)^2\left(\frac{6}{5}\right) - 3(1)^4$$
$$= \frac{16}{7} - \frac{32}{5} + \frac{36}{5} - 3 = \frac{3}{35}.$$

The first moment-ratio is

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0^2}{\left(\frac{1}{5}\right)^3} = 0.$$

This implies that this particular continuous probability distribution is *absolutely* symmetric
The second moment-ratio is

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\frac{3}{35}}{\left(\frac{1}{5}\right)^2} = 2.14.$$

This implies that this particular continuous probability distribution may be regarded as playkurtic, i.e. flatter than the normal distribution.
The students are encouraged to draw the graph of this distribution in order to develop a visual picture in their minds.

We begin the concept of Bivariate probability distribution by introducing the term 'Joint Distributions':

JOINT DISTRIBUTIONS

The distribution of two or more random variables which are observed simultaneously when an experiment is performed is called their *JOINT* distribution. It is customary to call the distribution of a single random variable as univariate. Likewise, a distribution involving two, three or many r.v.'s simultaneously is referred to as bivariate, trivariate or multivariate. A bivariate distribution may be discrete when the possible values of (X, U) are finite or count ably infinite. It is continuous if (X, Y) can assume all values in some non-countable set of the plane. A bivariate distribution is said mixed when one r.v. is discrete and the other is continuous.

BIVARIATE PROBABILITY FUNCTION

Let X and Y be two discrete r.v.'s defined on the same sample space S, X taking the values x1, x2, …, xm and Y taking the values y1, y2, …, yn. Then the probability that X takes on the value xi and, at the same time, Y takes on the value, denoted by f(xi, yj) or pij, is defined to be the joint probability function or simply the joint distribution of X and Y.
Thus the joint probability function, also called the bivariate probability function f(x, y) is a function whose value at the point (xi, yj) is given byf(xi, yj) = P(X = xi and Y = yj),
i = 1, 2, …, m.
j = 1, 2, …, n.
The joint or bivariate probability distribution consisting of all pairs of values (xi, yj) and their associated probabilities f(xi, yj) i.e. the set of triples [xi, yj, f(xi, yj)] can either be shown in the following two-way table:

Joint Probability Distribution of X and Y

X\Y	y_1	y_2	…	y_j	…	y_n	$P(X = x_i)$
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	…	$f(x_1,y_j)$	…	$f(x_1,y_n)$	$g(x_1)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	…	$f(x_2,y_j)$	…	$f(x_2,y_n)$	$g(x_2)$
⋮	⋮					⋮	⋮
x_i	$f(x_i,y_1)$	$f(x_i,y_2)$	…	$f(x_i,y_j)$	…	$f(x_i,y_n)$	$g(x_i)$
⋮	⋮					⋮	⋮
x_m	$f(x_m,y_1)$	$f(x_m,y_2)$	…	$f(x_m,y_j)$	…	$f(x_m,y_n)$	$g(x_m)$
$P(Y=y_j)$	$h(y_1)$	$h(y_2)$	…	$h(y_j)$	…	$h(y_n)$	1

or be expressed by mean of a formula for f(x, y). The probabilities f(x, y) can be obtained by substituting appropriate values of x and y in the table or formula. A joint probability function has the following properties:

PROPERTIES

i) f(xi, yj)>0,for all(xi, yj),i.e. for i=1,2,…,m; j = 1, 2, …, n.

ii) $$\sum_i \sum_j f(x_i, y_j) = 1$$

MARGINAL PROBABILITY FUNCTIONS

The point to be understood here is that, from the joint probability function for (X, Y), we can obtain the INDIVIDUAL probability function of X and Y. Such individual probability functions are called *MARGINAL* probability functions.

Let f(x, y) be the joint probability function of two discrete r.v.'s X and Y. Then the marginal probability function of X is defined as

$$g(x_i) = \sum_{j=1}^{n} f(x_i, y_j)$$

f(xi, y1) + f(xi, y2) + … + f(xi, yn)
as xi must occur either with y1 or y2 or … or yn
= P(X = xi);

that is, the individual probability function of X is found by adding over the rows of the two-way table. Similarly, the marginal probability function for Y is obtained by adding over the column as

$$h(y_j) = \sum_{i=1}^{m} f(x_i, y_j) = P(Y = y_j)$$

The values of the marginal probabilities are often written in the margins of the joint table as they are the row and column totals in the table. The probabilities in each marginal probability function add to 1.

CONDITIONAL PROBABILITY FUNCTION

Let X and Y be two discrete r.v.'s with joint probability function f(x, y). Then the conditional probability function for X given Y = y, denoted as f(x|y), is defined by

$$\begin{aligned} f(x_i \mid y_j) &= P(X = x_i \mid Y = y_j) \\ &= \frac{P(X = x_i \text{ and } Y = y_j)}{P(Y = y_j)} \\ &= \frac{f(x_i, y_j)}{h(y_j)}, \\ &\text{for } i = 1, 2, \ldots, j = 1, 2, \ldots \end{aligned}$$

Where h(y) is the marginal probability, and h(y) > 0

It gives the probability that X takes on the value xi given that Y has taken on the value yj. The conditional probability f(xi | yj) is non-negative and (for a given fixed yj) adds to 1 on i and hence is a *probability function*. Similarly, the conditional probability function for Y given X = x is

$$\begin{aligned} f(y_j \mid x_i) &= P(Y = y_j \mid X = x_i) \\ &= \frac{P(Y = y_j \text{ and } X = x_i)}{P(X = x_i)} \\ &= \frac{f(x_i, y_j)}{g(x_i)}, \quad \text{where } g(x) > 0. \end{aligned}$$

INDEPENDENCE

Two discrete r.v.'s X and Y are said to be statistically independent, if and only if, for all possible pairs of values (xi, yj) the joint probability function f(x, y) can be expressed as the *product* of the two marginal probability functions.

That is, X and Y are independent, if

$$\begin{aligned} f(x, y) &= P(X = x_i \text{ and } Y = y_j) \\ &= P(X = x_i).\, P(Y = y_j) \\ &\qquad \text{for all } i \text{ and } j. \\ &= g(x)\, h(y). \end{aligned}$$

It should be noted that the joint probability function of X and Y when they are *independent*, can be obtained by *MULTIPLYING* together their marginal probability functions.

EXAMPLE

An urn contains 3 black, 2 red and 3 green balls and 2 balls are selected at random from it. If X is the number of black balls and Y is the number of red balls selected, then find

i) the joint probability function f(x, y);
ii) $P(X + Y \leq 1)$;
iii) the marginal p.d. g(x) and h(y);
iv) the conditional p.d. f(x | 1),
v) P(X = 0 | Y = 1); and
vi) Are x and Y independent?

i) The sample space S for this experiment contains sample points. The possible values of X are 0, 1, and 2, and those for Y are 0, 1, and 2. The values that (X, Y) can take on are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2) and (2, 0). We desire to find f(x, y) for each value (x, y).
The total number of ways in which 2 balls can be drawn out of a total of 8 balls is

$$\binom{8}{2} = \frac{8x7}{2} = 28.$$

Now f(0, 0) = P(X = 0 and Y = 0), where the event (X = 0 and Y = 0) represents that neither black nor red ball is selected, implying that the 2 selected are green balls. This event therefore contains $\binom{3}{0}\binom{2}{0}\binom{3}{2} = 3$ sample points,

and

$$f(0, 0) = P(X = 0 \text{ and } Y = 0) = 3/28$$

Again f(0, 1) = P(X = 0 and Y = 1)

= P (none is black, 1 is red and 1 is green)

$$= \frac{\binom{3}{0}\binom{2}{1}\binom{3}{1}}{28} = \frac{6}{28}$$

Similarly, f(1, 1)

= P(X = 1 and Y = 1)

= P(1 is black 1 is red and none is green)

$$= \frac{\binom{3}{1}\binom{2}{1}\binom{3}{0}}{28} = \frac{6}{28}$$

Similar calculations give the probabilities of other values and the joint probability function of X and Y is given as:

Joint Probability Distribution

X \ Y	0	1	2	$P(X = x_i)$ g(x)
0	3/28	6/28	1/28	10/28
1	9/28	6/28	0	15/28
2	3/28	0	0	3/28
$P(Y = y_j)$ h(y)	15/28	12/28	1/28	1

LECTURE NO. 26

- BIVARIATE Probability Distributions (Discrete and Continuous)
- Properties of Expected Values in the case of Bivariate Probability Distributions

In the last lecture we began the discussion of the example in which we were drawing 2 balls out of an urn containing 3 black, 2 red and 3 green balls, and you will remember that, in this example, we were interested in computing quite a few quantities.

EXAMPLE

An urn contains 3 black, 2 red and 3 green balls and 2 balls are selected at random from it. If X is the number of black balls and Y is the number of red balls selected, then find

i) the joint probability function f(x, y)
ii) $P(X + Y \leq 1)$
iii) the marginal p.d. g(x) and h(y)
iv) the conditional p.d. f(x | 1)
v) P(X = 0 | Y = 1)
vi) Are x and Y independent?

As indicated in the last lecture, using the rule of combinations in conjunction with the classical definition of probability, the probability of the first cell came out to be 3/28. By similar calculations, we obtain all the remaining probabilities, and, as such, we obtain the following bivariate table:

Joint Probability Distribution

X \ Y	0	1	2	$P(X = x_i)$ g(x)
0	3/28	6/28	1/28	10/28
1	9/28	6/28	0	15/28
2	3/28	0	0	3/28
$P(Y = y_j)$ h(y)	15/28	12/28	1/28	1

This joint p.d. of the two r.v.'s (X, Y) can be represented by the formula

$$f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{28} = \begin{matrix} x = 0,1,2 \\ y = 0,1,2 \\ 0 \leq x + y \leq 2. \end{matrix}$$

ii) To compute P(X + Y < 1), we see that x + y < 1 for the cells
(0, 0), (0, 1) and (1, 0).

Therefore

P(X + Y < 1)
= f(0, 0) + f(0, 1) + f(1, 0)
= 3/28 +6/28 + 9/28
= 18/28 = 9/14

iii) The marginal p.d.'s are:

x	0	1	2
g(x)	10/28	15/28	3/28

y	0	1	2
h(x)	15/28	12/28	1/28

iv) By definition, the conditional p.d. f(x | 1) is

f(x | 1) = P(X = x | Y = 1)

$$= \frac{P(X = x \, and \, Y = 1)}{P(Y = 1)} = \frac{f(x,1)}{h(1)}$$

Now

$$h(1) = \sum_{x=0}^{2} f(x,1)$$

$$= \frac{6}{28} + \frac{6}{28} + 0$$

$$= \frac{12}{28} = \frac{3}{7}$$

Therefore

$$f(x \mid 1) = \frac{f(x,1)}{h(1)}$$

That is, $= \frac{3}{7} f(x,1), \qquad x = 0,1,2$

$$f(0 \mid 1) = \frac{7}{3} f(0,1) = \left(\frac{7}{3}\right)\left(\frac{6}{28}\right) = \frac{1}{2}$$

$$f(1 \mid 1) = \frac{7}{3} f(1,1) = \left(\frac{7}{3}\right)\left(\frac{6}{28}\right) = \frac{1}{2}$$

$$f(2 \mid 1) = \frac{7}{3} f(2,1) = \left(\frac{7}{3}\right)(0) = 0$$

Hence the conditional p.d. of X given that Y = 1, is

x	0	1	2
f(x|1)	1/2	1/2	0

vi) We find that f(0, 1) = 6/28,

$$g(0) = \sum_{y=0}^{2} f(0,y)$$

$$= \frac{3}{28} + \frac{6}{28} + \frac{1}{28} = \frac{10}{28}$$

$$h(1) = \sum_{x=0}^{2} f(x,1)$$

$$= \frac{6}{28} + \frac{6}{28} + 0 = \frac{12}{28}$$

v) Finally,

P(X = 0 | Y = 1)
= f(0 | 1) = 1/2

Now $\quad \frac{6}{28} \neq \frac{10}{28} \times \frac{12}{28},$

i.e. $\quad f(0,1) \neq g(0)h(1),$

therefore X and Y are ***NOT***

Satistically independent.

CONTINUOUS BIVARIATE DISTRIBUTIONS

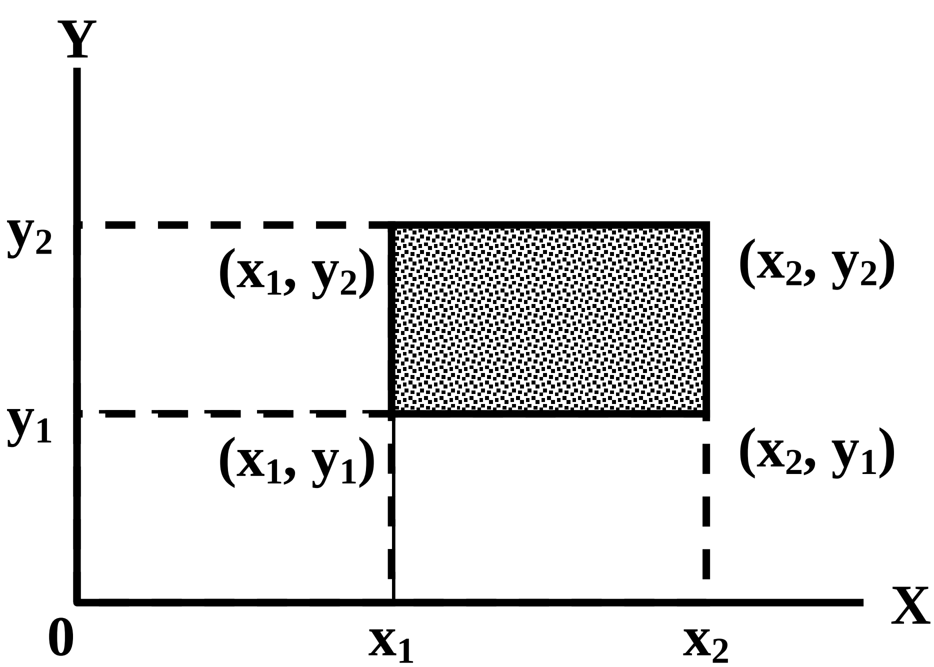
The bivariate probability density function of continuous r.v.'s X and Y is an integral function f(x,y) satisfying the following properties:

i) $f(x,y) \geq 0$ for all (x, y)

ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx \; dy = 1,$ and

iii) $$P(a \leq X \leq b, c \leq Y \leq d) = \int_{a}^{b} \int_{c}^{d} f(x, y) \; dy \; dx.$$

Let us try to understand the graphic picture of a bivariate continuous probability distribution: The region of the XY-plane depicted by the interval (x1 < X < x2; y1 < Y < y2) is shown graphically:



Just as in the case of a continuous univariate situation, the probability function f(x) gives us a curve under which we compute areas in order to find various probabilities, in the case of a continuous bivariate situation, the probability function f(x,y) gives a SURFACE and, when we compute the probability that our random variable X lies between x1 and x2 *AND*, simultaneously, the random variable Y lies between y1 and y2, we will be computing the *VOLUME* under the surface given by our probability function f(x, y) encompassed by this region. The MARGINAL p.d.f. of the continuous r.v. X is

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

and that of the r.v. Y is

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

That is, the marginal p.d.f. of any of the variables is obtained by integrating out the *other* variable from the joint p.d.f. between the limits $-\infty$ and $+\infty$.The CONDITIONAL p.d.f. of the continuous r.v. X given that Y takes the value y, is defined to be

$$f(x \mid y) = \frac{f(x, y)}{h(y)},$$

where f(x,y) and h(y) are respectively the joint p.d.f. of X and Y, and the marginal p.d.f. of Y, and h(y) > 0. Similarly, the conditional p.d.f. of the continuous r.v. Y given that X = x, is

$$f(y \mid x) = \frac{f(x, y)}{g(x)},$$

provided that g(x) > 0

It is worth noting that the conditional p.d.f's satisfy all the requirements for the UNIVARIATE density function.

FINALLY

Two continuous r.v.'s X and Y are said to be Statistically Independent, if and only if their joint density f(x,y) can be factorized in the form f(x,y) = g(x)h(y) for all possible values of X and Y.

EXAMPLE

Given the following joint p.d.f

$$f(x,\ y) = \frac{1}{8}(6 - x - y),\ 0 \le x \le 2;\ 2 \le y \le 4,$$
$$= 0, \quad \text{elsewhere}$$

a) Verify that f(x,y) is a joint density function.

b) Calculate $P\left(X \le \frac{3}{2}, Y \le \frac{5}{2}\right),$

c) Find the marginal p.d.f. g(x) and h(y).

d) Find the conditional p.d.f. f(x | y) and f(y | x).

SOLUTION

a) The joint density f(x,y) will be a p.d.f *if*

(i) f(x,y) > 0 and

(ii)

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

Now f(x,y) is clearly greater than zero for all x and y in the given region, and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\ dx\ dy = \frac{1}{8}\int_{0}^{2}\int_{2}^{4}(6 - x - y)\,dy\ dx$$

$$= \frac{1}{8}\int_{0}^{2}\left[6y - xy - \frac{y^2}{2}\right]_2^4 dx$$

$$= \frac{1}{8}\int_{0}^{2}(6 - 2x)dx = \frac{1}{8}\left[6x - x^2\right]_0^2$$

$$= \frac{1}{8}[12 - 4] = 1$$

Thus f(x,y) *has* the properties of a joint p.d.f.

b) To determine the probability of a value of the r.v. (X, Y) falling in the region X < 3/2, Y < 5/2,

We find $$P\left(X \le \frac{3}{2}, Y \le \frac{5}{2}\right)$$

$$= \int_{x=0}^{\frac{3}{2}} \int_{y=2}^{\frac{5}{2}} \frac{1}{8}(6-x-y)\,dy\ dx$$

$$= \frac{1}{8}\int_0^{\frac{3}{2}} \left[6y - xy - \frac{y^2}{2}\right]_2^{\frac{5}{2}} dx$$

$$= \frac{1}{8}\int_0^{\frac{3}{2}} \left(\frac{15}{8} - \frac{x}{2}\right) dx = \frac{1}{64}\left[15x - 2x^2\right]_0^{\frac{3}{2}} = \frac{9}{32}$$

c) The marginal p.d.f. of X is

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \qquad -\infty < x < \infty$$

$$= \frac{1}{8}\int_2^4 (6-x-y)\,dy, \qquad 0 \le x \le 2$$

$$= \frac{1}{8}\left[6y - xy - \frac{y^2}{2}\right]_2^4 \qquad 0 \le x \le 2$$

$$= \frac{1}{4}(3-x), \qquad 0 \le x \le 2$$

$$= 0, \qquad x < 0 \ \underline{\text{OR}}\ x \ge 2$$

Similarly, the marginal p.d.f. of Y is

$$h(y) = \frac{1}{8}\int_0^2 (6-x-y)\,dx, \quad 2 \le y \le 4$$

$$= \frac{1}{4}(5-y), \qquad 2 \le y \le 4$$

= 0, elsewhere.

d) The conditional p.d.f. of X given
Y = y, is

$$f(x\,|\,y) = \frac{f(x,y)}{h(y)}$$, where h(y) > 0

Hence

$$f(x|y) = \frac{\left(\frac{1}{8}\right)(6-x-y)}{\left(\frac{1}{4}\right)(5-y)} = \frac{6-x-y}{2(5-y)}, \quad 0 \le x \le 2$$

and the conditional p.d.f. of Y given
X = x, is

$$f(y \mid x) = \frac{f(x, y)}{g(x)}, \text{ where } g(x) > 0$$

Hence

$$f(y \mid x) = \frac{\left(\frac{1}{8}\right)(6 - x - y)}{\left(\frac{1}{4}\right)(3 - x)} = \frac{6 - x - y}{2(3 - x)}, \; 2 \le y \le 4$$

Next, we consider *two* important properties of mathematical expectation which are valid in the case of *BIVARIATE* probability distributions:

PROPERTY NO. 1

The expected value of the sum of any two random variables is equal to the *sum* of their expected values, i.e.
E(X + Y) = E(X) + E(Y).
The result also holds for the *difference* of r.v.'s i.e.

E(X – Y) = E(X) – E(Y).

PROPERTY NO. 2

The expected value of the product of two *independent* r.v.'s is equal to the *product* of their expected values, i.e.
E(XY) = E(X) E(Y).
It should be noted that these properties are valid for *continuous* random variable's in which case the summations are replaced by *integrals.*

EXAMPLE

Let X and Y be two discrete r.v.'s with the following joint p.d

y \ x	2	4
1	0.10	0.15
3	0.20	0.30
5	0.10	0.15

Find E(X), E(Y), E(X + Y), and E(XY).

SOLUTION

To determine the expected values of X and Y, we first find the marginal p.d. g(x) and h(y) by adding over the columns and rows of the two-way table as below:

y \ x	2	4	h(y)
1	0.10	0.15	0.25
3	0.20	0.30	0.50
5	0.10	0.15	0.25
g(x)	0.40	0.60	1.00

Now $E(X) = \sum xj\ g(xj)$

$= 2 \times 0.40 + 4 \times 0.60$

$= 0.80 + 2.40 = 3.2$

$E(Y) = \sum yi\ h(yi)$

$= 1 \times 0.25 + 3 \times 0.50 + 5 \times 0.25$

$= 0.25 + 1.50 + 1.25$

$= 3.0$

Hence

$E(X) + E(Y) = 3.2 + 3.0 = 6.2$

In order to compute E(XY) directly, we apply the formula:

$$E(X+Y) = \sum_i \sum_j (x_i + y_j)\, f(x_i, y_j)$$

$$E(XY) = \sum_i \sum_j (x_i y_j)\, f(x_i, y_j)$$

LECTURE NO. 27

- Properties of Expected Values in the case of Bivariate Probability Distributions (*Detailed* discussion)
- Covariance & Correlation
- Some Well-known Discrete Probability Distributions:
  - Discrete Uniform Distribution
  - An Introduction to the Binomial Distribution

EXAMPLE

Let X and Y be two discrete r.v.'s with the following joint p.d.

x \ y	1	3	5
2	0.10	0.20	0.10
4	0.15	0.30	0.15

Find E(X), E(Y), E(X + Y), and E(XY).

SOLUTION

To determine the expected values of X and Y, we first find the marginal p.d. g(x) and h(y) by adding over the columns and rows of the two-way table as below:

x \ y	1	3	5	g(x)
2	0.10	0.20	0.10	0.40
4	0.15	0.30	0.15	0.60
h(y)	0.25	0.50	0.25	1.00

Now $E(X) = \sum x_i\, g(x_i)$
$= 2 \times 0.40 + 4 \times 0.60$
$= 0.80 + 2.40 = 3.2$

$E(Y) = \sum y_j\, h(y_j)$
$= 1 \times 0.25 + 3 \times 0.50 + 5 \times 0.25$
$= 0.25 + 1.50 + 1.25$
$= 3.0$

Hence

$E(X) + E(Y) = 3.2 + 3.0 = 6.2$

$$E(X+Y) = \sum_i \sum_j \left(x_i + y_j\right) f\left(x_i, y_j\right)$$

$= (2 + 1)(0.10) + (2 + 3)(0.20) +$
$(2 + 5)(0.10) + (4 + 1)(0.15) +$
$(4 + 3)(0.30) + (4 + 5)(0.15)$
$= 0.30 + 1.00 + 0.70 + 0.75 +$
$2.10 + 1.35 = 6.20$
$= E(X) + E(Y)$

In order to compute E(XY) directly, we apply the formula:

$$E(XY) = \sum_i \sum_j \left(x_i y_j\right) f\left(x_i, y_j\right)$$

In this example,

$$E(XY) = \sum_i \sum_j \left(x_i y_j\right) f\left(x_i, y_j\right)$$

= (2 × 1) (0.10) + (2 × 3) (0.20) + (2 × 5) (0.10) + (4 × 1) (0.15) + (4 × 3) (0.30) + (4 × 5) (0.15)
= 9.6

Now

E(X) E(Y)
= 3.2 × 3.0
= 9.6

Hence E(XY) = E(X) E(Y) implying that X and Y are independent.

This was the discrete situation; let us now consider an example of the *continuous* situation:

EXAMPLE

Let X and Y be independent r.v.'s with joint p.d.f.

$$f(x, y) = \frac{x\left(1+3y^2\right)}{4}, \quad 0 < x < 2,\ 0 < y < 1$$
$$= 0, \quad \text{elsewhere.}$$

Find E(X), E(Y), E(X + Y) and E(XY). To determine E(X) and E(Y), we first find the marginal p.d.f. g(x) and h(y) as below:

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^1 \frac{x\left(1+3y^2\right)}{4}\,dy$$

$$= \frac{1}{4}\left[xy + xy^3\right]_0^1 = \frac{x}{2}, \quad \text{for } 0 < x < 2.$$

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

$$= \int_0^2 \frac{x\left(1+3y^2\right)}{4}\,dx = \frac{1}{4}\left[\frac{x^2}{2} + 3xy^2\right]_0^2$$

$$= \frac{1}{2}\left(1+3y^2\right), \quad \text{for } 0 < y < 1.$$

Hence

$$E(X) = \int_{-\infty}^{\infty} x\, g(x)\,dx$$

$$= \int_0^2 x\left(\frac{x}{2}\right)dx = \frac{1}{2}\left[\frac{x^3}{3}\right]_0^2 = \frac{4}{3}, \text{and}$$

$$E(Y) = \int_{-\infty}^{\infty} y\ h(y)\,dy = \frac{1}{2}\int_0^1 y\left(1+3y^2\right)dy$$

$$= \frac{1}{2}\left[\frac{y^2}{2} + \frac{3y^4}{4}\right]_0^1 = \frac{1}{2}\left[\frac{1}{2} + \frac{3}{4}\right] = \frac{5}{8},$$

And

$$E(X+Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)\ f(x,y)dx\ dy$$

$$\int_0^2\int_0^1(x+y)\ \frac{x\left(1+3y^2\right)}{4}dy\ dx$$

$$=\int_0^2\int_0^1\ \frac{x^2+3x^2y^2}{4}dx\ dy+\int_0^2\int_0^1\ \frac{xy+3xy^3}{4}dy\ dx$$

$$=\int_0^2\frac{1}{4}\left[x^2y+x^2y^3\right]_0^1dx+\int_0^2\frac{1}{4}\left[\frac{xy^2}{2}+\frac{3xy^4}{4}\right]_0^1dx$$

$$=\int_0^2\frac{1}{4}\left(2x^2\right)dx+\int_0^2\frac{1}{4}\left(\frac{x}{2}+\frac{3x}{4}\right)dx$$

$$=\frac{1}{2}\quad\left[\frac{x^3}{3}\right]_0^2=\frac{1}{4}\left[\frac{x^2}{4}+\frac{3x^2}{8}\right]_0^2$$

$$=\frac{4}{3}+\frac{5}{8}=\frac{47}{24},\ \text{and}$$

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

$$=\int_0^2\int_0^1(xy)\ \frac{x\left(1+3y^2\right)}{4}dy\ dx=\int_0^2\int_0^1\ \frac{x^2y+3x^2y^3}{4}dy\ dx$$

$$=\int_0^2\frac{1}{4}\left[\frac{x^2y^2}{2}+\frac{3x^2y^4}{4}\right]_0^1dx=\int_0^2\frac{1}{4}\left(\frac{5x^2}{4}\right)dx=\frac{1}{4}\quad\left[\frac{5x^3}{12}\right]_0^2=\frac{5}{6}$$

It should be noted that

i) E(X) + E(Y) = 4/3 + 5/8
= 47/24 = E(X + Y), and

ii) E(X) E(Y) = (4/3) (5/8)
= 5/6 = E(XY).

Hence, the two properties of mathematical expectation valid in the case of bivariate probability distributions are verified.

COVARIANCE OF TWO RANDOM VARIABLES

The covariance of two r.v.'s X and Y is a numerical measure of the extent to which their values tend to increase or decrease *together*. It is denoted by σXY or Cov (X, Y), and is defined as the expected value of the product

[X – E(X)] [Y – E(Y)]. That is
Cov (X, Y) = E {[X – E(X)] [Y – E(Y)]}
And the short cut formula is:

Cov (X, Y) = E(XY) – E(X) E(Y).

If X and Y are independent, then
E(XY) = E(X) E(Y), and
Cov (X, Y) = E(XY) – E(X) E(Y)
= 0
It is very important to note that covariance is zero when the r.v.'s X and Y are independent but its converse is not generally true. The covariance of a r.v. with itself is obviously its variance.

CORRELATION CO-EFFICIENT OF TWO RANDOM VARIABLES

Let X and Y be two r.v.'s with non-zero variances σ2X and σ2Y. Then the correlation coefficient which is a measure of linear relationship between X and Y, denoted by ρXY (the Greek letter rho) or Corr(X, Y), is defined as

$$\rho_{XY} = \frac{E[X - E(X)][Y - E(X)]}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sqrt{Var(X)\, Var(Y)}}$$

If X and Y are independent r.v.'s, then ρXY will be zero but zero correlation does not necessarily imply independence.

EXAMPLE

From the following joint p.d. of X and Y, find Var(X), Var(Y), Cov(X,Y) and ρ.

x \ y	0	1	2	3	g(x)
0	0.05	0.05	0.10	0	0.20
1	0.05	0.10	0.25	0.10	0.50
2	0	0.15	0.10	0.05	0.30
h(y)	0.10	0.30	0.45	0.15	1.00

Now
E(X) = Σxig(xi)
= 0 × 0.20 + 1 × 0.50 +2 × 0.30
= 0 + 0.50 + 0.60 = 1.10
E(Y) = Σyjh(yj)
= 0 × 0.10 + 1 × 0.30 + 2 × 0.45 + 3 × 0.15
= 0 + 0.30 + 0.90 + 0.45 = 1.65
E(X2) = Σxi2g(xi)
= 0 × 0.20 + 1 × 0.50 + 4 × 0.30
= 1.70
E(Y2) = Σyj2h(yj)
= 0 × 0.10 + 1 × 0.30 + 4 × 0.45 + 9 × 0.15
= 3.45

Thus
Var(X) = E(X2) – [E(X)]2
= 1.70 – (1.10)2 = 0.49,
and

Var(Y) = E(Y2) – [E(Y)]2
= 3.45 – (1.65)2 = 0.7275

Again:
E(XY) $= \sum_i \sum_j (x_i y_j) f(x_i, y_j)$
= 1 × 0.10 + 2 × 0.15 + 2 × 0.25 + 4 × 0.10 + 3 × 0.10 + 6 × 0.05

$= 0.10 + 0.30 + 0.50 + 0.40 + 0.30 + 0.30$
$= 1.90$

$\therefore \quad Cov(X,Y) = E(XY) – E(X)\, E(Y)$
$= 1.90 – 1.10 \times 1.65 = 0.085$, and

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

$$= \frac{0.085}{\sqrt{(0.49)(0.7275)}} = \frac{0.085}{0.595}$$

$$= 0.14$$

Hence, we can say that there is a weak positive linear correlation between the random variables X and Y.

EXAMPLE

If f(x, y)
= x2 + xy/3 , 0 < x < 1, 0 < y < 2
= 0, elsewhere,

Find
Var(X), Var(Y) and Corr(X,Y)

SOLUTION

The marginal p.d.f.'s are

$$g(x) = \int_0^2 \left(x^2 + \frac{xy}{3}\right) dy = 2x^2 + \frac{3}{2}x,$$

$$0 \le x \le 1$$

and

$$h(y) = \int_0^1 \left(x^2 + \frac{xy}{3}\right) dx = \frac{1}{3} + \frac{y}{6},$$

Now

$$0 \le y \le 2$$

$$E(X) = \int_{-\infty}^{\infty} x g(x)\, dx$$

$$= \int_0^1 x \left(2x^2 + \frac{2x}{3}\right) dx = \frac{13}{18},$$

$$E(Y) = \int_{-\infty}^{\infty} y h(y)\, dy$$

Thus

$$= \int_0^2 y \left(\frac{1}{3} + \frac{y}{6}\right) dy = \frac{10}{9}.$$

$$Var(X) = E[X - E(X)]^2$$

$$= \int_{-\infty}^{\infty} (x + \mu_x)^2 g(x)\, dx$$

$$= \int_0^1 \left(x - \frac{13}{18}\right)^2 \left(2x^2 + \frac{2x}{3}\right) dx = \frac{73}{1620}$$

$$\operatorname{Var}(Y) = E[Y - E(Y)]^2$$
$$= \int_{-\infty}^{\infty} (y - \mu_y)^2 h(y)\, dy$$
$$= \int_0^2 \left(y - \frac{10}{9}\right)^2 \left(\frac{1}{3} + \frac{y}{6}\right) dy = \frac{26}{81}, \text{ and}$$

Cov(X, Y)

$$= E\{[X - E(X)]\,[Y - E(Y)]\}$$
$$= \int_0^1 \int_0^2 \left(x - \frac{13}{18}\right)\left(y - \frac{10}{9}\right)\left(x^2 + \frac{xy}{3}\right) dy\, dx$$
$$= \int_0^1 \left(-\frac{2}{9}x^3 + \frac{25}{81}x^2 - \frac{26}{243}x\right) dx = \frac{-1}{162}.$$

Hence

$$\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$
$$= \frac{-1/162}{\sqrt{(73/1620)(26/81)}}$$
$$= -0.05$$

Hence we can say that there is a *VERY* weak negative linear correlation between X and Y. In other words, X and Y are almost uncorrelated. This brings us to the end of the discussion of the BASIC concepts of discrete and continuous *Univariate* and *bivariate* probable. We now begin the discussion of some probability distributions that are WELL-KNOWN, and are encountered in *real-life* situations.

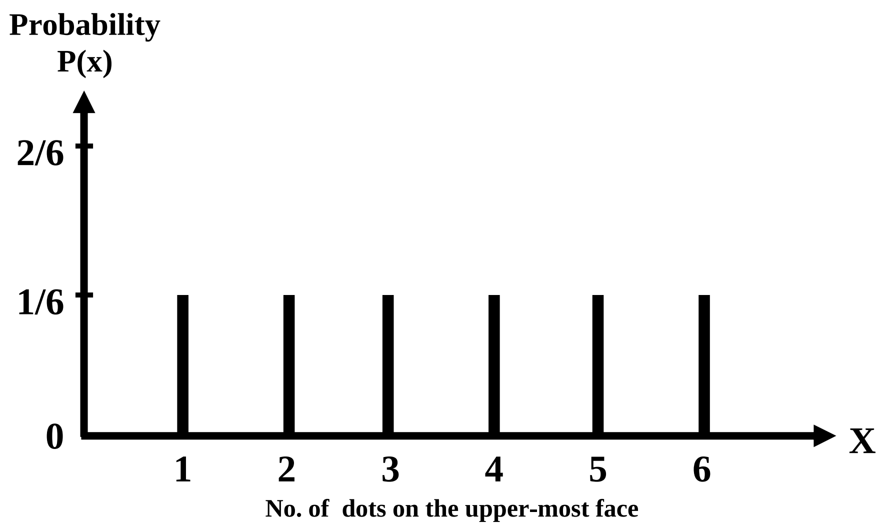
DISCRETE UNIFORM DISTRIBUTION

EXAMPLE

Suppose that we toss a fair die and let X denote the number of dots on the upper-most face. Since the die is *fair*, hence each of the X-values from 1 to 6 is equally likely to occur, and hence the probability distribution of the random variable X is as follows:

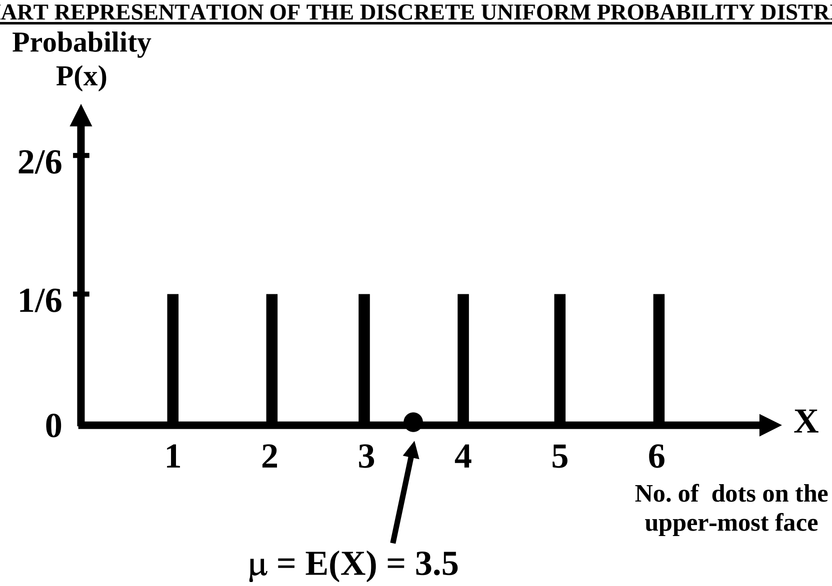
X	P(x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Total	1

If we draw the line chart of this distribution, we obtain Line Chart Representation of the Discrete Uniform Probability Distribution



As all the vertical line segments are of equal height, hence this distribution is called a uniform distribution.
As this distribution is absolutely symmetrical, therefore the mean lies at the *exact centre* of the distribution i.e. the mean is equal to 3.5.

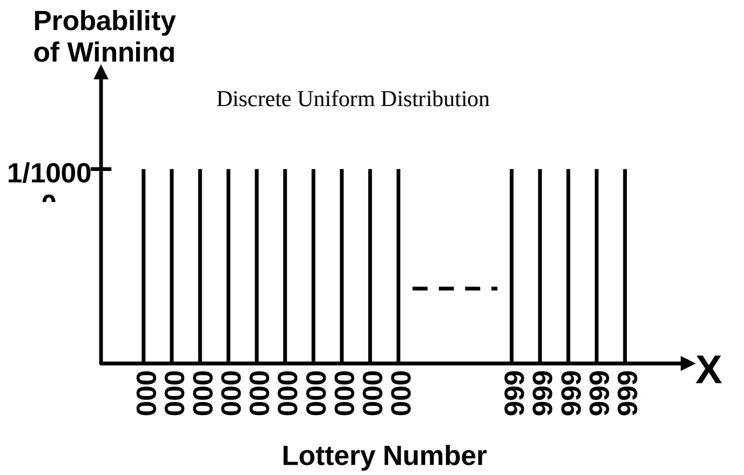
LINE CHART REPRESENTATION OF THE DISCRETE UNIFORM PROBABILITY DISTRIBUTION



What about the spread of this distribution? You are encouraged to compute the standard deviation as well as the coefficient of variation of this distribution on their own. Let us consider another interesting example.

EXAMPLE

The lottery conducted in various countries for purposes of money-making provides a good example of the discrete uniform distribution. Suppose that, in a particular lottery, as many as ten thousand lottery tickets are issued, and the numbering is 0000 to 9999. Since each of these numbers is *equally likely* to occur, hence we have the following situation:



INTERPRETATION

It reflects the fact that winning lottery numbers are selected by a random procedure which makes all numbers equally likely to be selected. The point to be kept in mind is that, whenever we have a situation where the various outcomes are equally likely, and of a form such that we have a random variable X with values 0, 1, 2, … or , as in the above example, 0000, 0001 …, 9999, we will be dealing with the discrete uniform distribution.

BINOMIAL DISTRIBUTION

The binomial distribution is a very important discrete probability distribution. It was discovered by James Bernoulli about the year 1700.We illustrate this distribution with the help of the following example:

EXAMPLE

Suppose that we toss a fair coin 5 times, and we are interested in determining the probability distribution of X, where X represents the number of heads that we obtain.
We note that in tossing a fair coin 5 times:

- every toss results in either a head or a tail,
- the probability of heads (denoted by p) is equal to ½ every time (in other words, the probability of heads remains *constant*),
- every throw is *independent* of every other throw, and
- the total number of tosses i.e. 5 is *fixed* in *advance*.

The above four points represents the *four basic* and vitally important *PROPERTIES* of a binomial experiment

PROPERTIES OF A BINOMIAL EXPERIMENT

- Every trial results in a success or a failure.
- The successive trials are independent.
- The probability of success, p, remains constant from trial to trial.
- The number of trials, n, is fixed in advanced.

LECTURE NO. 28

- Binomial Distribution
- Fitting a Binomial Distribution to Real Data
- An Introduction to the Hyper geometric Distribution

The binomial distribution is a very important discrete probability distribution. We illustrate this distribution with the help of the following example:

EXAMPLE

Suppose that we toss a fair coin 5 times, and we are interested in determining the probability distribution of X, where X represents the number of heads that we obtain. We note that in tossing a fair coin 5 times:

- Every toss results in either a head or a tail,
- The probability of heads (denoted by p) is equal to ½ every time (in other words, the probability of heads remains *constant*),
- Every throw is *independent* of every other throw, and
- The total number of tosses i.e. 5 is *fixed* in *advance*.

The above four points represents the *four basic* and vitally important *PROPERTIES* of binomial experiment. Now, in 5 tosses of the coin, there can be 0, 1, 2, 3, 4 or 5 heads, and the no. of heads is thus a random variable which can take one of these six values. In order to compute the probabilities of these X-values, the formula is:

Binomial Distribution

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Where
n = the total no. of trials
p = probability of success in each trial
q = probability of failure in each trial (i.e. q = 1 - p)
x = no. of successes in n trials.
x = 0, 1, 2, … n

The binomial distribution has two parameters, n and p. In this example, n = 5 since the coin was thrown 5 times, p = ½ since it is a fair coin, q = 1 – p = 1 – ½ = ½ Hence

$$P(X = x) = \binom{5}{x}\left(\tfrac{1}{2}\right)^x \left(\tfrac{1}{2}\right)^{5-x}$$

Putting x = 0

$$\mathrm{P}(\mathrm{X} = 0) = \binom{5}{0}\left(\tfrac{1}{2}\right)^0 \left(\tfrac{1}{2}\right)^{5-0}$$

$$= \frac{5\,!}{0\,!\,5\,!}(1)\left(\tfrac{1}{2}\right)^5$$

$$= 1(1)\left(\tfrac{1}{2}\right)^5 = \frac{1}{32}$$

Putting x = 1

$$\mathrm{P}(\mathrm{X} = 1) = \binom{5}{1}\left(\tfrac{1}{2}\right)^1 \left(\tfrac{1}{2}\right)^{5-1}$$

$$= \frac{5\,!}{1\,!\,4\,!}\left(\tfrac{1}{2}\right)^1 \left(\tfrac{1}{2}\right)^4$$

$$= \frac{(5)}{1}\left(\tfrac{1}{2}\right)$$

$$= 5\left(\tfrac{1}{2}\right)^5 = 5\left(\frac{1}{32}\right) = \frac{5}{32}$$

Similarly, we have:

$$P(X=2)=\binom{5}{2}\left(\tfrac{1}{2}\right)^2\left(\tfrac{1}{2}\right)^{5-2}=\frac{10}{32}$$

$$P(X=3)=\binom{5}{3}\left(\tfrac{1}{2}\right)^3\left(\tfrac{1}{2}\right)^{5-3}=\frac{10}{32}$$

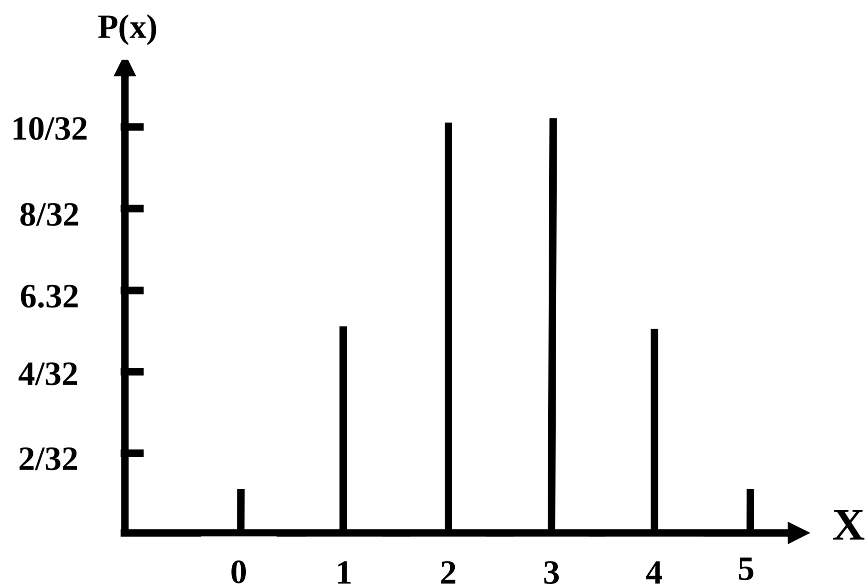
$$P(X=4)=\binom{5}{4}\left(\tfrac{1}{2}\right)^4\left(\tfrac{1}{2}\right)^{5-4}=\frac{5}{32}$$

$$P(X=5)=\binom{5}{5}\left(\tfrac{1}{2}\right)^5\left(\tfrac{1}{2}\right)^{5-5}=\frac{1}{32}$$

Hence, the binomial distribution for this particular example is as follows. Binomial Distribution in the case of tossing a fair coin five times:

Number of Heads X	Probability P(x)
0	1/32
1	5/32
2	10/32
3	10/32
4	5/32
5	1/32
Total	32/32 = 1

Graphical Representation of the above binomial distribution:



The next question is: What about the mean and the standard deviation of this distribution? We can calculate them just as before, using the formulas

Mean of X = E(X) = ∑XP(X)
Var(X) = ∑X2 P(X) – [∑XP(X)]2

but it has been mathematically proved that for a binomial distribution given by

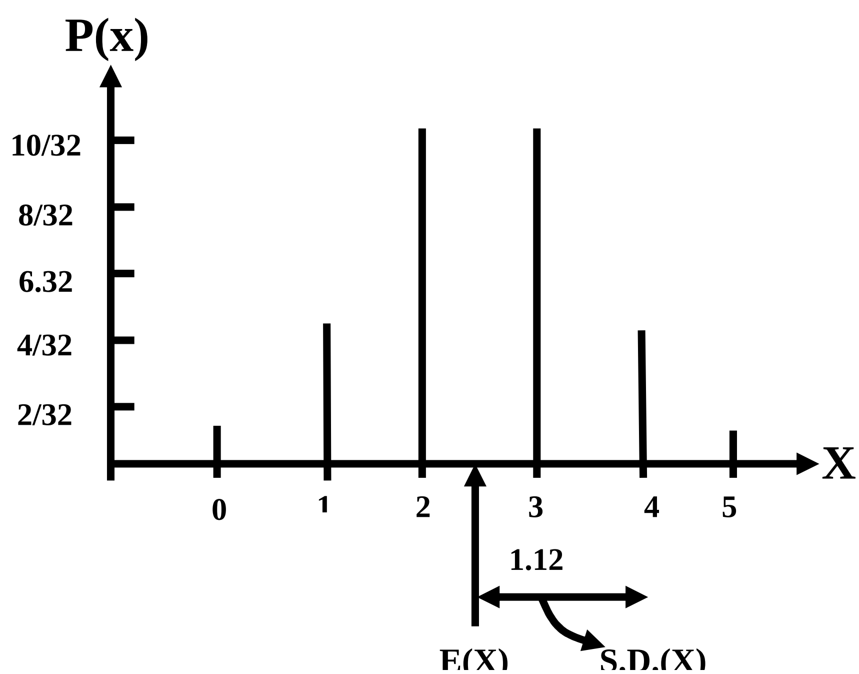
$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

For a binomial distribution
E(X) = np
and Var(X) = npq
so that $S.D.(X) = \sqrt{npq}$

For the above example, n = 5, p = ½ and q = ½
Hence
Mean = E(X) = np = 5(½) = 2.5

and S.D.(X) $= \sqrt{\text{npq}} = \sqrt{5\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{\left(\frac{5}{4}\right)}$ = <u>1.12</u>

We would have got exactly the same answers if we had applied the LENGTHIER procedure.
E(X) = ∑XP(X) and Var X = ∑X2 P(X)-[∑XP(X)]2
Graphical Representation of the Mean and Standard Deviation of the Binomial Distribution (n=5, p=1/2)



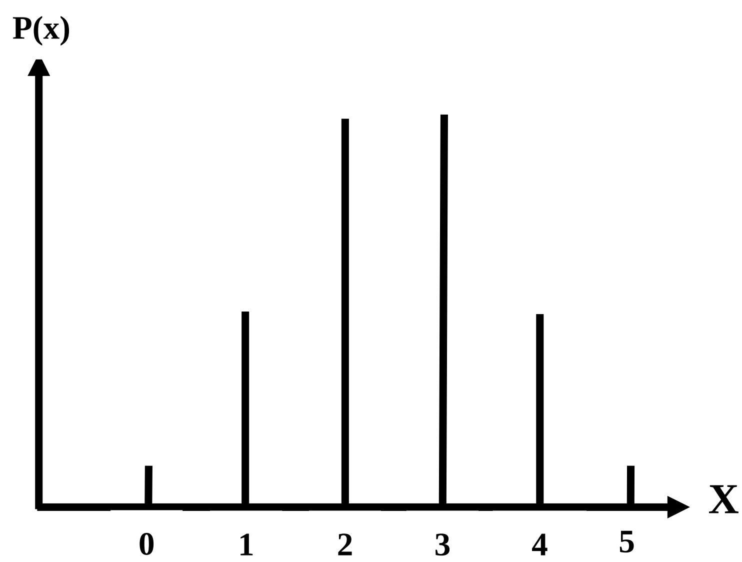
WHAT DOES THIS MEAN?

What this mean is that if 5 *fair* coins are tossed an *INFINITE* no. of times, sometimes we will get no head out of 5, sometimes/head… sometimes all 5 heads. But on the *AVERAGE* we should expect to get 2.5 heads in 5 tosses of the coin, or, a total of 25 heads in 50 tosses of the coin And 1.12 gives a measure of the possible *variability* in the various numbers of heads that can be obtained in 5 tosses. (As you know, in this problem, the number of heads can range from 0 to 5 had the coin been tossed 10 times, the no. of heads possible would vary from 0 to 10 and the standard deviation would probably have been different).

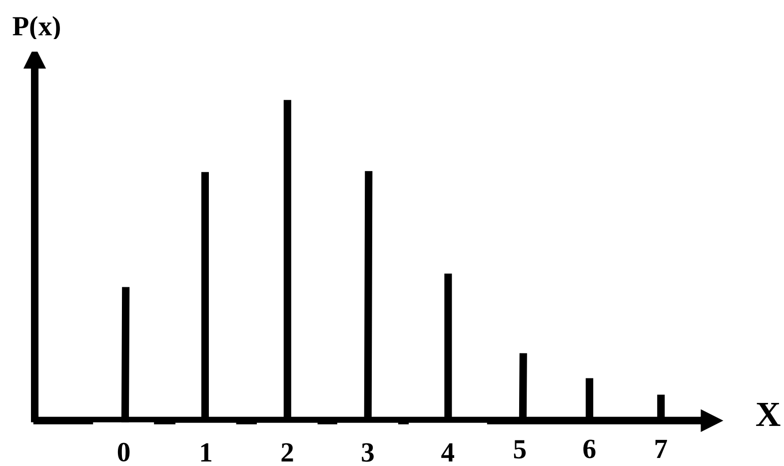
Coefficient of Variation:

$$\text{C.V.} = \frac{\sigma}{\mu} \times 100 = \frac{1.12}{2.5} \times 100 = 44.8\%$$

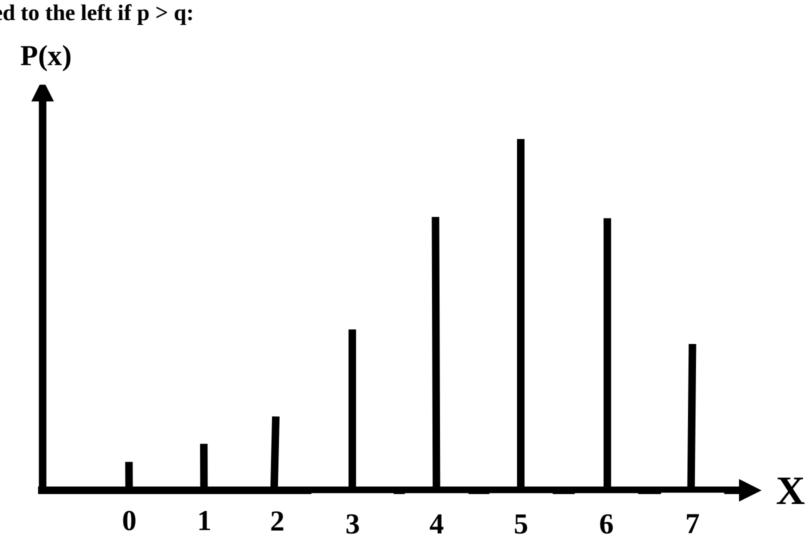
Note that the binomial distribution is not always symmetrical as in the above example. It will be symmetrical *only* when p = q = ½ (as in the above example).



It is skewed to the right if p < q:



It is skewed to the left if p > q:



But the degree of Skewness (or asymmetry) *decreases* as n increases. Next, we consider the *Fitting* of a Binomial Distribution to *Real* Data. We illustrate this concept with the help of the following example:

EXAMPLE

The following data has been obtained by tossing a *LOADED* die 5 times, and noting the number of times that we obtained a *six*. Fit a binomial distribution to this data.

No. of Sixes	0	1	2	3	4	5	Total
Frequency	12	56	74	39	18	1	200

SOLUTION

To fit a binomial distribution, we need to find n and p.

Here n = 5, the largest x-value.

To find p, we use the relationship $\bar{x} = np$.

The rationale of this step is that, as indicated in the last lecture, the mean of a binomial *probability* distribution is equal to np, i.e.

$\mu = np$

But, here, we are not dealing with a *probability* distribution i.e. the entire *population* of all possible sets of throws of a loaded die --- we only have a *sample* of throws at our disposal.

As such, μ is not available to us, and all we can do is to replace it by its estimate $\bar{X}$.

Hence, our equation becomes $\bar{X} = np$.

Now, we have:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{0+56+148+117+72+5}{200}$$

$$= \frac{398}{200} = 1.99$$

Using the relationship $\bar{x} = np$, we get 5p = 1.99 or p = 0.398.This value of p seems to indicate *clearly* that the die is not fair at all! (Had it been a fair die, the probability of getting a six would have been 1/6 i.e. 0.167; a value of p = 0.398 is *very* different from 0.167.) Letting the random variable X represent the number of sixes, the above calculations yield the fitted binomial distribution as

$$b(x;5,\ 0.398) = \binom{5}{x}(0.398)^x(0.602)^{5-x}$$

Hence the *probabilities* and *expected frequencies* are calculated as below:

No. of Sixes (x)	Probability f(x)		Expected frequency
0	$\binom{5}{0}q^5 = (0.602)^5$	= 0.07907	15.8
1	$\binom{5}{1}q^5 p = 5.(0.602)^4(0.398)$	= 0.26136	52.5
2	$\binom{5}{2}q^3 p^2 = 10.(0.602)^3(0.398)^2$	= 0.34559	69.1
3	$\binom{5}{3}q^2 p^3 = 10.(0.602)(0.398)^3$	= 0.22847	45.7
4	$\binom{5}{4}qp^4 = (0.602)(0.398)^4$	= 0.07553	15.1
5	$\binom{5}{5}p^5 = (0.398)^5$	= 0.00998	2.0
Total		= 1.00000	200.0

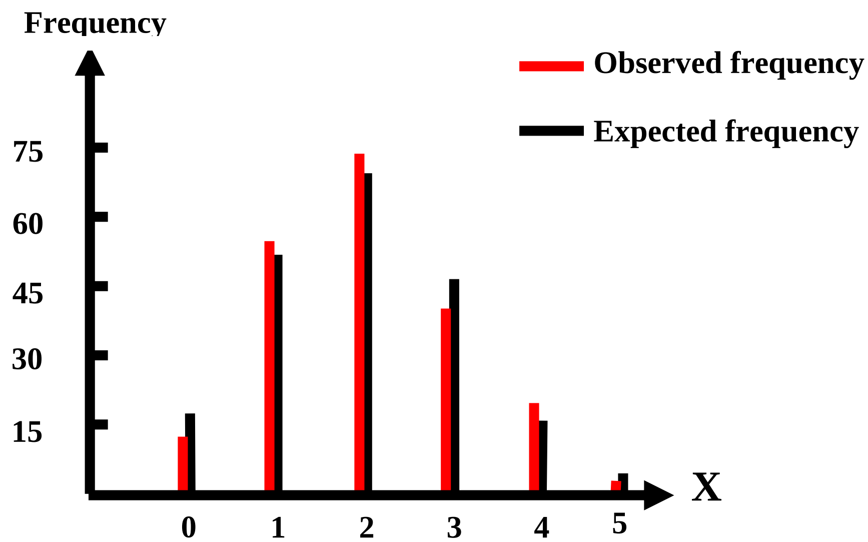
In the above table, the expected frequencies are obtained by multiplying each of the probabilities by 200.

In the entire above procedure, we are assuming that the given frequency distribution has the characteristics of the fitted theoretical binomial distribution, comparing the observed frequencies with the expected frequencies, we obtain:

No. of Sixes x	Observed Frequency f_0	Expected Frequency f_e
0	12	15.8
1	56	52.5
2	74	69.1
3	39	45.7
4	18	15.1
5	1	2.0
Total	200	200.0

The graphical representation of the observed frequencies as well as the expected frequencies is as follows:

Graphical Representation of the Observed and Expected Frequencies:



The above graph quite clearly indicates that there is not much discrepancy between the observed and the expected frequencies. Hence, we can say that it is a reasonably good fit.

There is a procedure known as the Chi-Square Test of Goodness of Fit which enables us to determine in a formal, mathematical manner whether or not the theoretical distribution fits the observed distribution reasonably well. This test comes under the realm of Inferential Statistics --- that area which we will deal with during the last 15 lectures of this course. Let us consider a *real-life* application of the binomial distribution:

AN EXAMPLE FROM INDUSTRY

Suppose that the past record indicates that the proportion of defective articles produced by this factory is 7%.And suppose that a law *NEWLY* instituted in this particular country states that there should not be more than 5% defective.

Suppose that the factory-owner makes the statement that his machinery has been *overhauled* so that the number of defectives has *DECREASED*.

In order to examine this claim, the relevant government department decides to send an inspector to examine a sample of 20 items.

What is the probability that the inspector will find 2 or more defective items in his sample (so that a fine will be imposed on the factory)?

SOLUTION

The first step is to identify the NATURE of the situation, If we study this problem closely, we realize that we are dealing with a binomial experiment because of the fact that all four properties of a binomial experiment are being fulfilled:

PROPERTIES OF A BINOMIAL EXPERIMENT

- Every item selected will either be defective (i.e. *success*) or not defective (i.e. *failure*)
- Every item drawn is independent of every other item
- The probability of obtaining a defective item i.e. 7% is the same (constant) for all items. (This probability figure is according to relative frequency definition of probability.
- The number of items drawn is fixed in advance i.e. 20 hence; we are in a position to apply the binomial formula

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

$$P(X = x) = \binom{20}{x} 0.07^x 0.93^{20-x}$$

Substituting n = 20 and p = 0.07, we obtain:

Now

P(X > 2) = 1 - P(X < 2)
= 1- [P(X = 0) + P(X =1)]

$$= 1 - \binom{20}{0} 0.07^0 0.93^{20-0} - \binom{20}{1} 0.07^1 0.93^{20-1}]$$
$$= 1 - 1 \times 1 \times 0.93^{20} - 20 \times 0.07 \times 0.93^{19}$$
$$= 1 - 0.234 - 0.353$$
$$= 0.413$$
$$= 41.3\%$$

Hence the probability is SUBSTANTIAL i.e. more than 40% that the inspector will find two or more defective articles among the 20 that he will inspect. In other words, there is CONSIDERABLE chance that the factory will be fined.
The point to be realized is that, generally speaking, whenever we are dealing with a 'success / failure' situation, we are dealing with what can be a binomial experiment. (For EXAMPLE, if we are interested in determining any of the following proportions, we are dealing with a BINOMIAL situation:

- Proportion of smokers in a city smoker → success, non-smokers → failure.
- Proportion of literates in a community → literacy rate, literate → success, illiterate → failure.
- Proportion of males in a city → *sex ratio*).

HYPERGEOMETRIC PROBABILITY DISTRIBUTION

There are many experiments in which the condition of independence is violated and the probability of success does not remain constant for all trials. Such experiments are called hyper geometric experiments.
In other words, a hyper geometric experiment has the following properties:

PROPERTIES OF HYPERGEOMETRIC EXPERIMENT

- The outcomes of each trial may be classified into one of two categories, success and failure.
- The probability of success changes on each trial.
- The successive trials are not independent.
- The experiment is repeated a fixed number of times.

The number of success, X in a hyper geometric experiment is called a hyper geometric random variable and its probability distribution is called the hyper geometric distribution. When the hyper geometric random variable X assumes a value x, the hyper geometric probability distribution is given by the formula

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

Where
N = number of units in the population,
n = number of units in the sample, and
k = number of successes in the population.
The hyper geometric probability distribution has three parameters N, n and k.
The hyper geometric probability distribution is appropriate when

- a random sample of size n is drawn *WITHOUT REPLACEMENT* from a *finite* population of N units;
- k of the units are of one kind (classified as success) and the remaining N – k of another kind (classified as failure).

LECTURE NO. 29

- Hyper geometric Distribution
  (in some detail)
- Poisson Distribution
- Limiting Approximation to the Binomial
- Poisson Process
- Continuous Uniform Distribution

In the last lecture, we began the discussion of the HYPERGEOMETRIC PROBABILITY DISTRIBUTION. We now consider this distribution in some detail. As indicated in the last lecture, there are many experiments in which the condition of independence is violated and the probability of success does not remain constant for all trials. Such experiments are called hyper geometric experiments. In other words, a hyper geometric experiment has the following properties:

PROPERTIES OF HYPERGEOMETRIC EXPERIMENT

- The outcomes of each trial may be classified into one of two categories, success and failure.
- The probability of success changes on each trial.
- The successive trials are not independent.
- The experiment is repeated a fixed number of times.

The number of success, X in a hyper geometric experiment is called a hyper geometric random variable and its probability distribution is called the hyper geometric distribution. When the hyper geometric random variable X assumes a value x, the hyper geometric probability distribution is given by the formula

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

where
N = number of units in the population,
n = number of units in the sample,
and
k = number of successes in the population.
The hyper geometric probability distribution has three parameters N, n and k.

- The hyper geometric probability distribution is appropriate when
- a random sample of size n is drawn *WITHOUT REPLACEMENT* from a *finite* population of N units;
- k of the units are of one kind (classified as success) and the remaining N – k of another kind (classified as failure).

EXAMPLE

The names of 5 men and 5 women are written on slips of paper and placed in a hat. Four names are drawn. What is the probability that 2 are men and 2 are women? Let us regard ‘men’ as success. Then X will denote the number of men. We have N = 5 + 5 = 10 names to be drawn from; Also, n = 4, (since we are drawing a sample of size 4 out of a ‘population’ of size 10) In addition, k = 5 (since there are 5 men in the population of 10). In this problem, the possible values of X are 0, 1, 2, 3, 4, i.e. n): The hyper geometric distribution is given by

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

Since N = 10, k = 5 and n = 4, hence, in this problem, the hyper geometric distribution is given by

$$P(X = x) = \frac{\binom{5}{x}\binom{5}{4-x}}{\binom{10}{4}}$$

and the required probability,
i.e P(X = 2) is

$$P(X=2)=\frac{\binom{5}{2}\binom{5}{4-2}}{\binom{10}{4}}$$
$$=\frac{\binom{5}{2}\binom{5}{2}}{\binom{10}{4}}$$
$$=\frac{10\times 10}{210}$$
$$=\frac{10}{21}$$

In other words, the probability is a little less than 50% that two of the four names drawn will be those of MEN. In the above example, just as we have computed the probability of X = 2, we could also have computed the probabilities of X = 0, X = 1, X = 3 and X = 4 (i.e. the probabilities of having zero, one, three *OR* four men among the four names drawn).The students are encouraged to compute these probabilities on their own, to check that the sum of these probabilities is 1, and to draw the line chart of this distribution.
Additionally, the students are encouraged to think about the *centre*, *spread* and *shape* of the distribution. Next, we consider some important *PROPERTIES* of the Hyper
geometric Distribution:

<u>PROPERTIES OF THE HYPERGEOMETRIC DISTRIBUTION</u>

- The mean and the hyper geometric probability distribution are
  $$\mu = n\frac{k}{N} \text{ and } \sigma^2 = n\frac{k}{N}\frac{N-k}{N}\frac{N-n}{N-1},$$
- If N becomes *indefinitely* large, the hyper geometric probability distribution tends to the *BINOMIAL* probability distribution.

The above property will be best understood with reference to the following important points:

- There are two ways of drawing a sample from a population, sampling with replacement, and sampling without replacement.
- Also, a sample can be drawn from either a finite population or an infinite population.

This leads to the following bivariate table: With reference to sampling, the various possible situations are:

Sampling \ Population	Finite	Infinite
With replacement		
Without replacement		

The point to be understood is that, whenever we are sampling with replacement, the population remains undisturbed (because any element that is drawn at any one draw, is re-placed into the population before the next draw).Hence, we can say that the various trials (i.e. draws) are independent, and hence we can use the binomial formula. On the other hand, when we are sampling without replacement from a finite population, the constitution of the population changes at every draw (because any element that is drawn at any one draw is not re-placed into the population before the next draw). Hence, we cannot say that the various trials are independent, and hence the formula that is appropriate in this particular situation is the hyper geometric formula. *But*, if the population size is much larger than the sample size (so that we can regard it as an ‘infinite’ population), then we note that, although we are not re-placing any element that has been drawn back into the population, the population remains almost undisturbed. As such, we can assume that the various trials (i.e. draws) are independent, and, once again, we can apply the binomial formula.
In this regard, the generally accepted rule is that the *binomial* formula *can* be applied when we are drawing a sample from a finite population *without replacement* and the sample size n is not more than 5 percent of the population size N, or, to put it in another way, when $n < 0.05$ N.
When n is greater than 5 percent of N, the *hyper geometric* formula should be used.

Next, we discuss the Poisson Distribution.

POISSON DISTRIBUTION

The Poisson distribution is named after the French mathematician Sime'on Denis Poisson (1781-1840) who published its derivation in the year 1837.THE POISSON DISTRIBUTION ARISES IN THE FOLLOWING TWO SITUATIONS:

- It is a limiting approximation to the binomial distribution, when p, the probability of success is very small but n, the number of trials is so large that the product np = μ is of a moderate size;
- a distribution in its *own* right by considering a *POISSON PROCESS* where events occur *randomly* over a specified interval of *time* or *space* or *length*.

Such random events might be the number of typing errors per page in a book, the number of traffic accidents in a particular city in a 24-hour period, etc.
With regard to the *first* situation, if we assume that n goes to infinity and p approaches zero in such a way that μ = np remains constant, then the limiting form of the binomial probability distribution is

$$\lim_{\substack{n\to\infty \\ p\to 0}} b(x;n,p)=\frac{e^{-\mu}\mu^x}{x!}, \quad x=0,1,2,\ldots,\infty$$

where **e = 2.71828.**
The Poisson distribution has only one parameter $\mu > 0$.
The parameter μ may be interpreted as the mean of the distribution.
Although the theoretical requirement is that n should tend to infinity, and p should tend to zero, but in *PRACTICE*, generally, most statisticians use the Poisson approximation to the binomial when
p is 0.05 or less,
& n is 20 or more,
but *in fact*, the *LARGER* n is and the *SMALLER* p is, the *better* will be the approximation. We illustrate *this* particular application of the Poisson distribution with the help of the following example:

EXAMPLE

Two hundred passengers have made reservations for an airplane flight. If the probability that a passenger who has a reservation will not show up is 0.01, what is the probability that exactly three will not show up?

SOLUTION

Let us regard a "no show" as success. Then this is essentially a *binomial* experiment with n = 200 and p = 0.01. Since p is very small and n is considerably large, we shall apply the Poisson distribution, using
μ= np = (200) (0.01) = 2.
Therefore, if X represents the number of successes (not showing up), we have

$$P(X=3)=\frac{e^{-2}(2)^3}{3!}$$
$$=\frac{(0.1353)(8)}{3\times 2\times 1}=0.1804$$
$$\left(\because e^{-2}=\frac{1}{(2.71828)^2}=0.1353\right)$$

POISSON PROCESS

may be defined as a *physical* process governed at least in *part* by some *random* mechanism.
Stated differently a poisson process represents a situation where events occur *randomly* over a specified interval of *time* or *space* or *length*. Such random events might be the number of taxicab arrivals at an intersection per day; the number of traffic deaths per month in a city; the number of radioactive particles emitted in a given period; the number of flaws per unit length of some material; the number of typing errors per page in a book; etc.
The formula valid in the case of a Poisson process is:

$$P(X=x)=\frac{e^{-\lambda t}(\lambda t)^x}{x!},$$

where

λ = average number of occurrences of the outcome of interest per unit of time,

t = number of time-units under consideration, and

x = number of occurrences of the outcome of interest in t units of time.

We illustrate this concept with the help of the following example:

EXAMPLE

Telephone calls are being placed through a certain exchange at random times on the average of four per minute. Assuming a Poisson Process, determine the probability that in a 15-second interval, there are 3 or more calls.

SOLUTION

Step-1: Identify the *unit* of time:
In this problem we take a minute as the unit of time.

Step-2: Identifyλ, the *average* number of occurrences of the outcome of interest per unit of time,
In this problem we have the information that, on the average, 4 calls are received per minute, hence:

$\lambda = 4$

Step-3: Identify t, the number of time-units under consideration. In this problem, we are interested in a 15-second interval, and since 15 seconds are equal to 15/60 = ¼ minutes i.e. 1/4 units of time, therefore t = 1/4

Step-4: Compute λt: In this problem,

$\lambda = 4$, &

t = 1/4,

Hence:

$\lambda t = 4 \times ¼ = 1$

Step-5: Apply the Poisson formula

$$P(X = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!},$$

In this problem, since λt = 1, therefore and since we are interested in 3 or more calls in a 15-second interval, therefore

P(X > 3) = 1 - P(X < 3)
= 1 - [P(X=0)+P(X=1)+P(X=2)]

$$= 1 - \sum_{x=0}^{2} \frac{e^{-}(1)^x}{x!}$$

$$= 1 - \sum_{x=0}^{2} \frac{(0.3679)(1)^x}{x!} \quad (\because e^{-1} = 0.3679)$$

$$= 1 - (0.91975) = 0.08025$$

Hence the probability is only 8% (i.e. a very low probability) that in a 15-second interval, the telephone exchange receives 3 or more calls.

PROPERTIES OF THE POISSON DISTRIBUTION

Some of the main properties of the Poisson distribution are given below:

- If the random variable X has a Poisson distribution with parameter μ, then its mean and variance are given by E(X) = μ and Var(X) = μ.
- (In other words, we can say that the mean of the Poisson distribution is *equal* to its variance.)
- The shape of the Poisson distribution is *positively skewed*. The distribution tends to be symmetrical as μ becomes *larger and larger*.

Comparing the Poisson distribution with the binomial, we note that, whereas the binomial distribution can be symmetric, positively skewed, *or* negatively skewed (depending on whether p = 1/2, p < 1/2, *or* p > 1/2), the Poisson distribution can never be negatively skewed.

FITTING OF A POISSON DISTRIBUTION TO REAL DATA

Just as we discussed the fitting of the binomial distribution to real data in the last lecture, the Poisson distribution can *also* be fitted to real-life data. The procedure is very similar to the one described in the case of the fitting of the binomial distribution: The population mean μ is replaced by the sample mean $\bar{X}$, and the probabilities of the various values of X are computed using the Poisson formula. The *chi-square test of goodness of fit* enables us to determine whether or not it is a good fit i.e. whether or not the discrepancy between the expected frequencies and the observed frequencies is small. Next, we discuss some important mathematical points regarding Poisson distribution.

- **1)** The Poisson approximation to the binomial formula works well when
  $n > 20$ and $p < 0.05$.
- **2)** Suppose that the Poisson is used to approximate the binomial which, in *turn*, is being used to approximate the hyper geometric. Then the *Poisson* is being used to approximate the hyper geometric Putting the two approximation conditions *together*, the rule of *thumb* is that the Poisson distribution can be used to approximate the hyper geometric distribution when $n < 0.05N$, $n > 20$, and $p < 0.05$

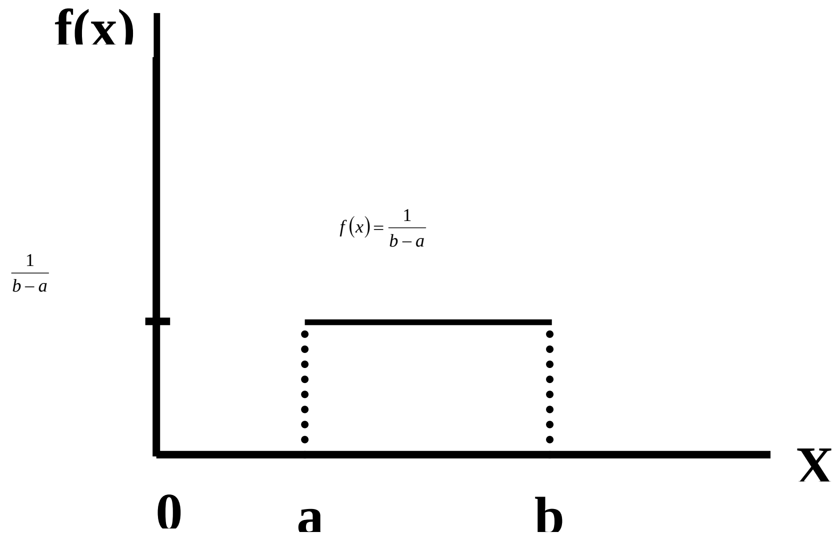
This brings to the *end* of the discussion of some of the most important and well-known Univariate discrete probability distributions. We now begin the discussion some of the well-known Univariate continuous probability distribution.
There are different types of continuous distributions e.g. the *uniform* distribution, the *normal* distribution, the *geometric* distribution, and the *exponential* distribution. Each one has its *own* shape and its *own* mathematical properties. In this course, we will discuss the uniform distribution and the normal distribution.
We begin with the continuous UNIFORM DISTRIBUTION (also known as the RECTANGULAR DISTRIBUTION).

UNIFORM DISTRIBUTION

A random variable X is said to be uniformly distributed if its density function is defined as

$$f(x) = \frac{1}{b-a}, \qquad a \le x \le b$$

The graph of this distribution is as follows



The above function is a *proper* probability density function because of the fact that:
i) Since $a < b$, therefore $f(x) > 0$

ii) $$\int_{-\infty}^{\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx = \frac{1}{b-a}[x]_a^b = \frac{b-a}{b-a} = 1$$

Since the shape of the distribution is like that of a rectangle, therefore the total area of this distribution can *also* be obtained from the simple formula:
Area of rectangle

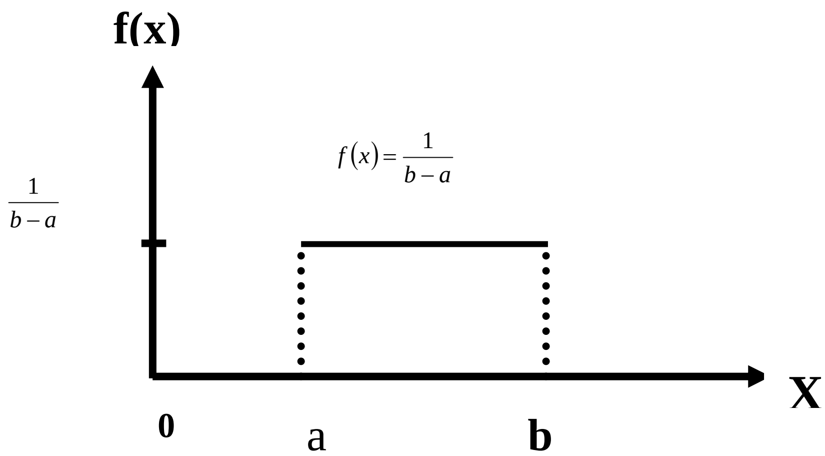
= (Base) × (Height)

$$= (\mathrm{b-a}) \times \left(\frac{1}{\mathrm{b-a}}\right) = 1$$

Area under the Uniform Distribution

= Area of the rectangle
= (Base) × (Height)

$$= (b-a) \times \left(\frac{1}{b-a} \right) = 1$$



The distribution derives its *name* from the fact that its density is constant or *uniform* over the interval [a, b] and is 0 elsewhere. It is also called the rectangular distribution because its total probability is confined to a rectangular region with base equal to (b – a) and height equal to 1/(b – a). The *parameters* of this distribution are a and b with

$$\mu = \frac{a+b}{2} \text{ and variance is } \sigma^2 = \frac{(b-a)^2}{12}$$

PROPERTIES OF THE UNIFORM DISTRIBUTION

Let X has the uniform distribution over [a, b]. Then its mean is

The uniform probability distribution provides a *model* for continuous random variables that are *evenly distributed over a certain interval*. That is, a uniform random variable is one that is just *as* likely to assume a value in *one* interval as it is to assume a value in any *other* interval *of equal size*. There is *no clustering* of values around *any* value. Instead, there is an *even spread* over the *entire* region of possible values. As far as the *real-life application* of the uniform distribution is concerned, the point to be noted is that, for *continuous* random variables there is an *infinite* number of values in the sample space, but in *some* cases, *the values may appear to be equally likely*.

EXAMPLE-1

If a short exists in a 5 meter stretch of electrical wire, it may have an equal probability of being in any particular 1 centimeter segment along the line.

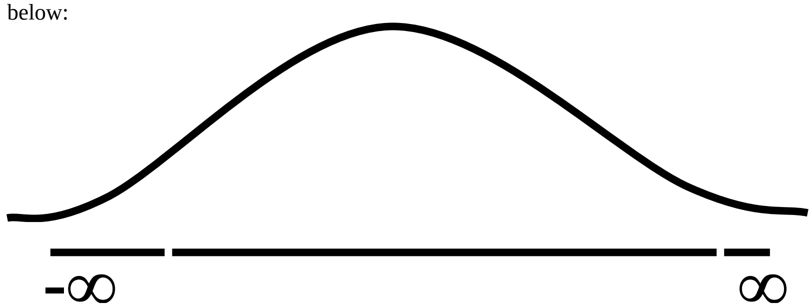
EXAMPLE-2

If a safety inspector plans to choose a time at random during the 4 afternoon work-hours to pay a surprise visit to a certain area of a plant, then each 1 minute time-interval in this 4 work-hour period will have an *equally likely* chance to being selected for the visit. Also, the uniform distribution arises in *the study of rounding off errors*, etc.

LECTURE NO. 30

- Normal Distribution.
  - Mathematical Definition
  - Important Properties
- The Standard Normal Distribution
  - Direct Use of the Area Table
  - Inverse Use of the Area Table
- Normal Approximation to the Binomial Distribution

The normal distribution was discovered in 1733. The normal distribution has a bell-shaped curve of the type shown below:



Let us begin its detailed discussion by considering its formal MATHEMATICAL DEFINITION, and its main PROPERTIES.

NORMAL DISTRIBUTION

A continuous random variable is said to be normally distributed with mean μ and standard deviation σ if its probability density function is given by

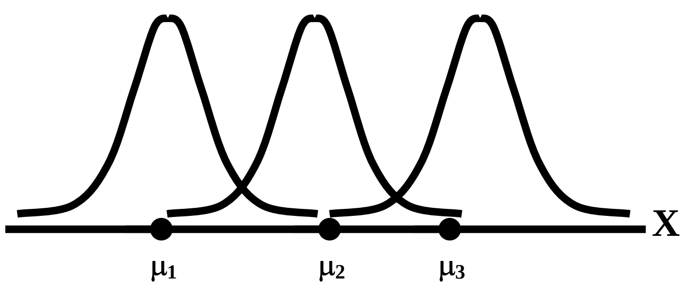
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}, \quad -\infty < x < \infty \qquad \begin{pmatrix}\text{where}\\ \pi = 3.1416 \simeq 22/7,\\ e \simeq 2.71828\end{pmatrix}$$

For any particular value of μ and any particular value of σ, giving different values to x and we obtain a set of ordered pairs (x, f(x)) that yield the bell-shaped curve given above. The formula of the normal distribution defines a *FAMILY* of distributions depending on the values of the two *parameters* μ and σ (as these are the two values that determine the shape of the distribution).

PROPERTIES OF THE NORMAL DISTRIBUTION

Property No. 1

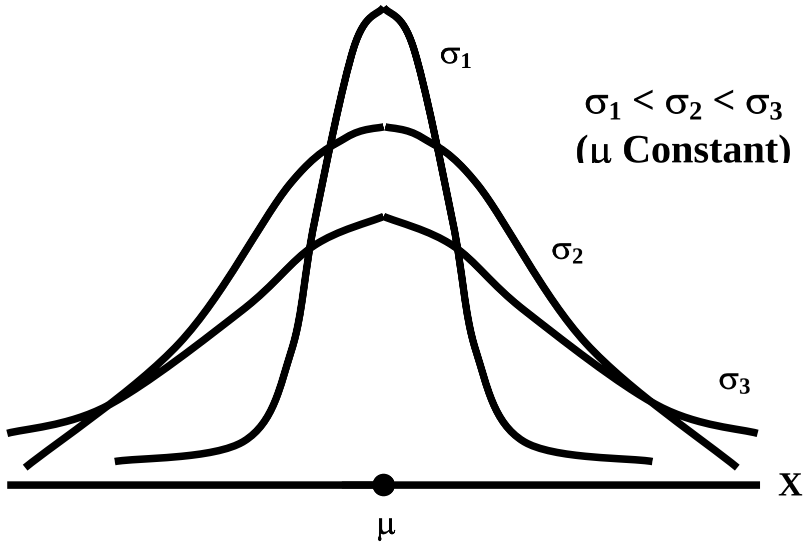
It can be mathematically proved that, for the normal distribution N(μ,σ2), μ represents the *mean*, and σ represents the *standard deviation* of the normal distribution. A change in the mean μ *shifts* the distribution to the left or to the right along the x-axis:



$\mu_1 < \mu_2 < \mu_3$

(σ Constant)

The different values of the standard deviationσ, (which is a measure of *dispersion*), determine the *flatness* or *peakedness* of the normal curve. In other words, achange in the standard deviation on σ *flattens* it or *compresses* it while leaving its centre in the same position:



Property No. 2

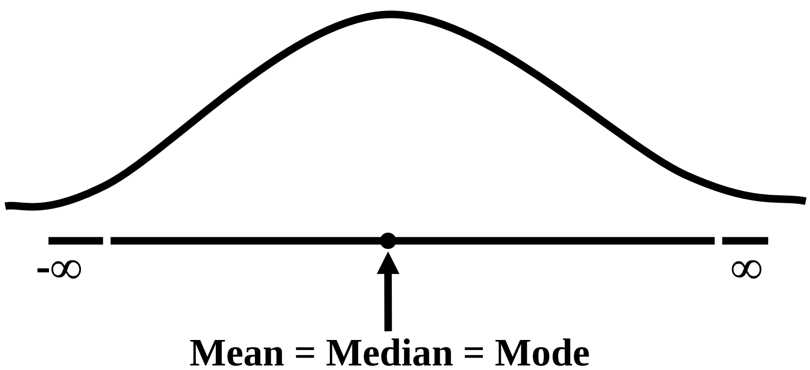
The normal curve is asymptotic to the x-axis as $x \to \pm \infty$.

Property No. 3

Because of the *symmetry* of the normal curve, 50% of the area is to the right of a vertical line erected at the mean, and 50% is to the left.(Since the total area under the normal curve from $-\infty$ to $+\infty$ is unity, therefore the area to the left of μ is 0.5 and the area to the right of μ is also 0.5.)

Property No. 4

The density function attains its maximum value at $x = \mu$ and falls off symmetrically on each side of μ. This is why the mean, median and mode of the normal distribution are all equal to μ.



Property No. 5

Since the normal distribution is absolutely symmetrical, hence μ_3 , the third moment about the mean is zero.

Property No. 6

For the normal distribution, it can be mathematically proved that $\mu_4 = 3\,\sigma^4$

Property No. 7

The moment ratios of the normal distribution come out to be 0 and 3 respectively:

Moment Ratios:

$$\beta_1 = \frac{\mu^2{}_3}{\mu_2{}^3} = \frac{0^2}{\left(\sigma^2\right)^3} = 0,$$

$$\beta_2 = \frac{\mu^4}{{\mu_2}^2} = \frac{3\sigma^4}{\left(\sigma^2\right)^2} = 3$$

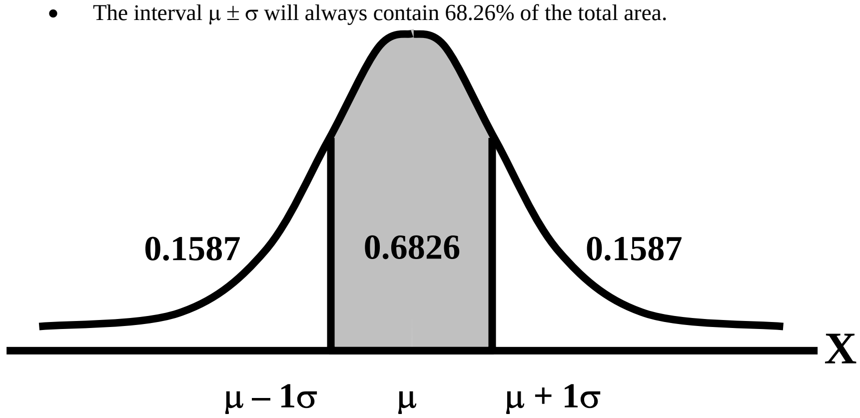
NOTE

Because of the fact that, for the normal distribution, β2 comes out to be 3, this is why this value has been taken as a criterion for measuring the kurtosis of any distribution: The amount of peakedness of the *normal* curve has been taken as a *standard*, and we say that this particular distribution is masochistic. Any distribution for which β2 is greater than 3 is more peaked than the normal curve, and is called leptokurtic; Any distribution for which β2 is less than 3 is less peaked than the normal curve, and is called platykurtic.

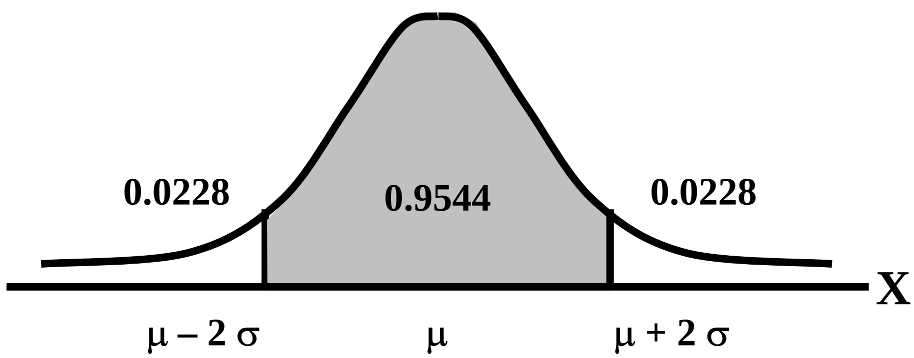
Property No. 8

No matter what the values of μ and σ are, areas under the normal curve remain in certain *fixed* proportions within a *specified* number of standard deviations on either side of μ.
For the normal distribution:

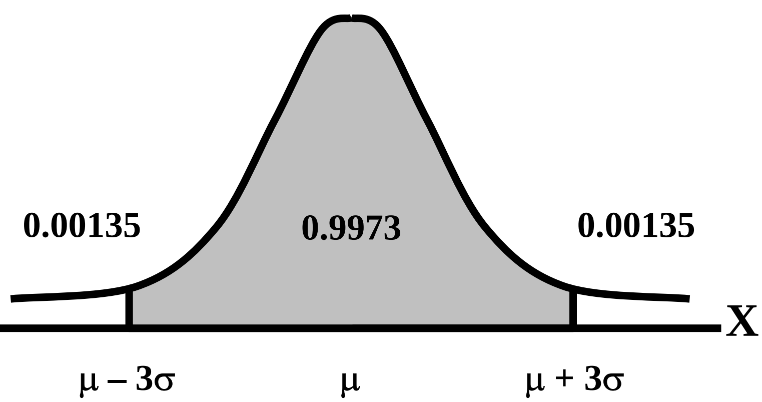
- The interval μ ± σ will always contain 68.26% of the total area.



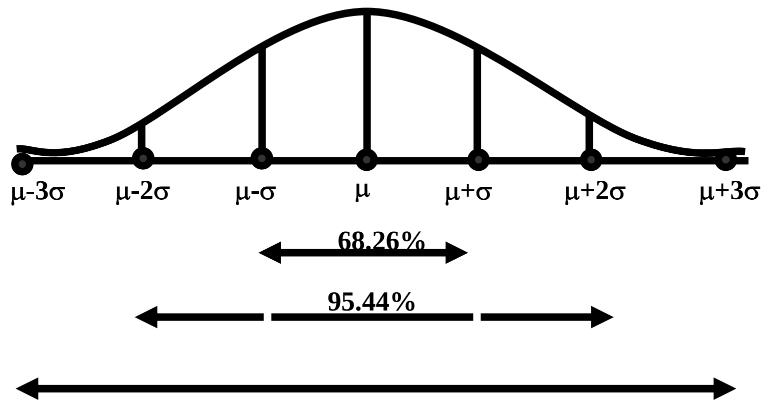
- The interval μ + 2σ will always contain 95.44% of the total area.



- The interval μ ± 3σ will always contain 99.73% of the total area.



Combining the above three results, we have:



At this point, the student are reminded of the Empirical Rule that was discussed during the first part of this course --- that on descriptive statistics. You will recall that, in the case of any approximately symmetric hump-shaped frequency distribution, approximately 68% of the data-values lie between $\bar{X}$ + S, approximately 95% between the $\bar{X}$ + 2S, and approximately 100% between $\bar{X}$ + 3S.You can now recognize the similarity between the empirical rule and the property given above. (In case a distribution is absolutely normal, the areas in the above-mentioned ranges are 68.26%, 95.44% and 99.73%; in case a distribution approximately normal, the areas in these ranges will be *approximately* equal to these percentages.)

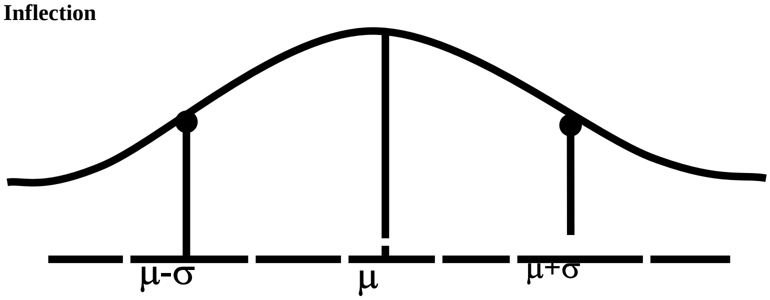
Property No. 9

The normal curve contains points of inflection (where the direction of concavity changes) which are equidistant from the mean. Their coordinates on the XY-plane are

$$\left(\mu-\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right) and \left(\mu+\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$$

respectively.

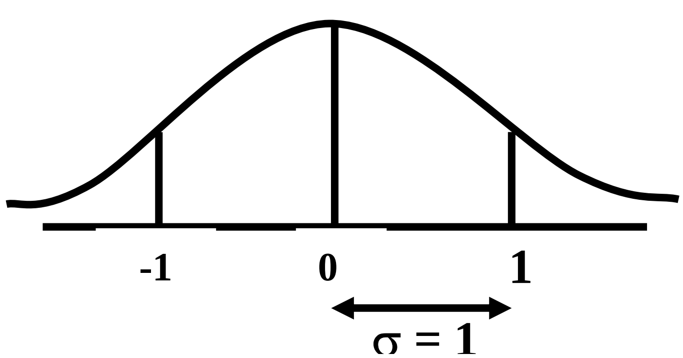
Points of Inflection



Next, we consider the concept of the Standard Normal Distribution:

THE STANDARD NORMAL DISTRIBUTION

A normal distribution whose mean is zero and whose standard deviation is 1 is known as the standard normal distribution.



This distribution has a very important role in *computing areas* under the normal curve. The *reason* is that the mathematical equation of the normal distribution is so complicated that it is not possible to find areas under the normal curve by ordinary integration. Areas under the normal curve have to be found by the more advanced method of *numerical integration*. The point to be noted is that areas under the normal curve have been computed for *that* particular normal distribution whose mean is zero and whose standard deviation is equal to 1, i.e. the standard normal distribution.

Areas under the Standard Normal Curve

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0159	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2083	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2380	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3880
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3990	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4430	0.4441
1.6	0.4452	0.4463	0.4474	0.4485	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4690	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4758	0.4762	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4865	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4980	0.4980	0.4981
2.9	0.4981	0.4982	0.4983	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.49865	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.49903	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993

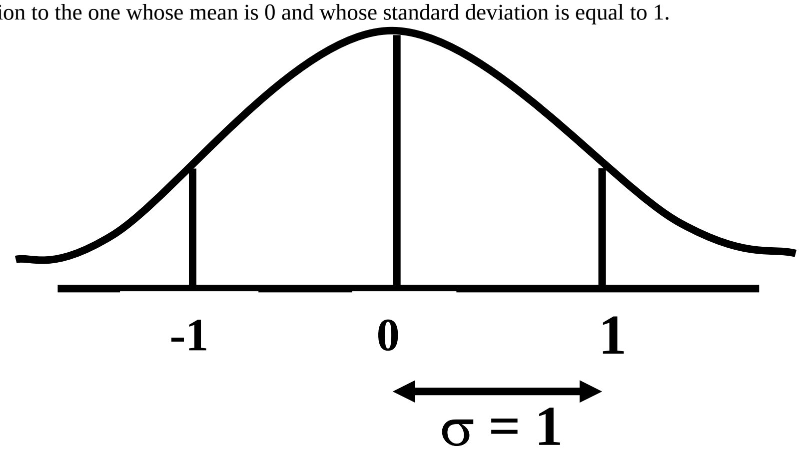
In any problem involving the normal distribution, the generally established procedure is that the normal distribution under consideration is *converted to* the standard normal distribution. This process is called *standardization*. The formula for converting N (μ, σ) to N (0, 1) is:

THE PROCESS OF STANDARDIZATION

The standardization formula is:

$$Z = \frac{X - \mu}{\sigma}$$

If X is N (μ, σ), then Z is N (0, 1). In other words, the standardization formula given above converts our normal distribution to the one whose mean is 0 and whose standard deviation is equal to 1.



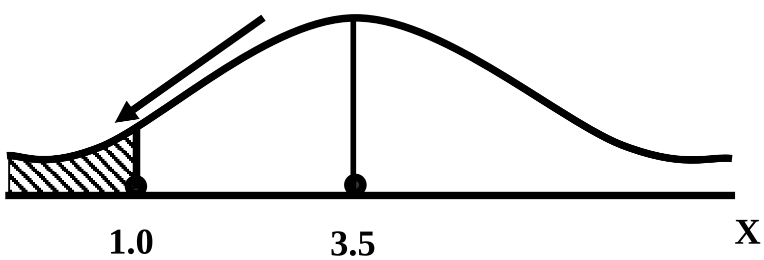
We illustrate this concept with the help of an interesting example:

EXAMPLE

The length of life for an automatic dishwasher is approximately normally distributed with a mean life of 3.5 years and a standard deviation of 1.0 years. If this type of dishwasher is guaranteed for 12 months, what fraction of the sales will require replacement?

SOLUTION

Since 12 months equal one year, hence we need to compute the fraction or *proportion* of dishwashers that will cease to function before a time-span of one year. In other words, we need to find the *probability* that a dishwasher fails before one year.



In order to find this area we nee to standardize normal distribution i.e. to convert N(3.5, 1) to N(0, 1):

The method is

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3.5}{1.0}$$

The X-value representing the warranty
period is 1.0 so

$$Z = \frac{1.0 - 3.5}{1.0} = \frac{-2.5}{1} = -2.5$$