$$= \frac{1}{n}\sum_{i=1}^{n}\left\{(x_i-A)^r - \binom{r}{1}(x_i-A)^{r-1}(\bar{x}-A) + \binom{r}{2}(x_i-A)^{r-2}(\bar{x}-A)^2 - \dots + (-1)^r(\bar{x}-A)^r\right\}$$

where, $\binom{n}{k} = {}^nC_k$

$$= m_r'(A) - \binom{r}{1} m_{r-1}'(A)\, m_1'(A) + \binom{r}{2} m_{r-2}'(A)\, m_1'^2(A) - \dots + (-1)^r m_1'^r(A) \quad \dots (6e)$$

In particular, putting $r=1,2,3,4$ in (6e), we get

$$\left.\begin{aligned}
m_1 &= m_1'(A) - m_1'(A) = 0\\
m_2 &= m_2'(A) - m_1'^2(A)\\
m_3 &= m_3'(A) - 3m_2'(A)\,m_1'(A) + 2m_1'^3(A)\\
m_4 &= m_4'(A) - 4m_3'(A)\,m_1'(A) + 6m_2'(A)\,m_1'^2(A) - 3m_1'^4(A)
\end{aligned}\right\} \quad \dots (6f)$$

The values of m_2, m_3 and m_4 are required for obtaining some measures of skewness and Kurtosis.

6.3 Skewness: In relation to a frequency distribution, the term skewness refers to its departure from symmetry. In fact, the frequency distribution of a discrete variable is called symmetrical about the value x_0, if the frequency of x_0-h is same as that of x_0+h, whatever h may be. Again for a continuous variable, the distribution is said to be symmetrical about the value x_0, if the frequency density at x_0-h is same as that at x_0+h, whatever h may be

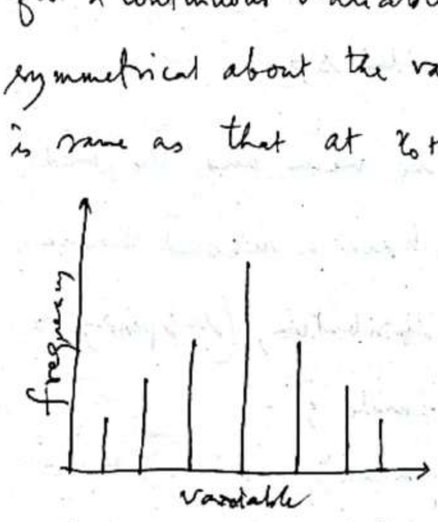

Fig 6.1 A symmetrical distribution (discrete variable)

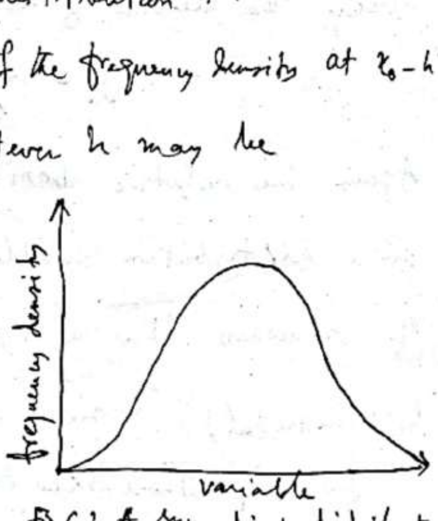

Fig 6.2 A symmetrical distribution (Continuous variable

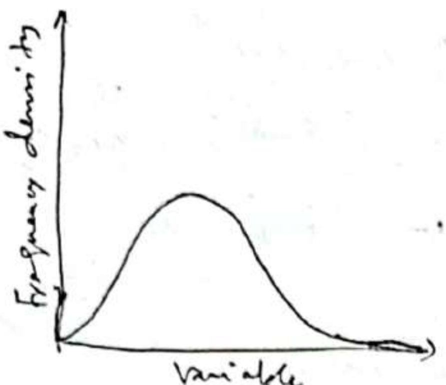

Fig. 6.3 A positively skew distribution

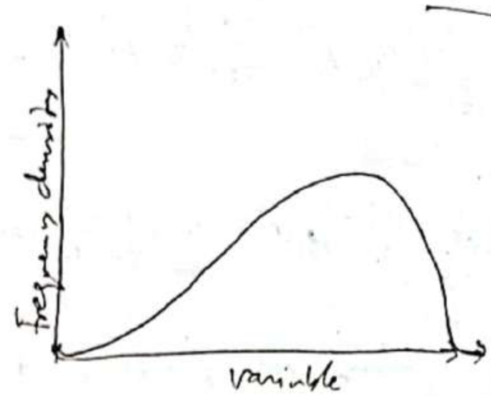

Fig 6.4 A negatively skew distribution

If a distribution is not found to be symmetrical, then it is termed as skew or asymmetrical. The skewness is called positive or negative according as the longer tail of the distribution is towards the higher or lower values of the variable.

To obtain a measure of skewness, we may note that, all odd-order central moments are zero for a symmetrical distribution, positive for a positively skew distribution and negative for a negatively skew distribution. Thus any such moment, excepting m_1 (which is zero for any distribution) may be regarded as a measure of skewness. For the sake of simplicity, m_3 is considered suitable and it is divided by s^3 to make the measure free from units of the variable.

Hence the measure of skewness is

$$g_1 = \frac{m_3}{s^3}, \text{ provided } s \neq 0$$

Again, the relative positions of the mean and the mode in a distribution enable us to have a second measure of skewness. For a symmetrical distribution, (supposing to be unimodal) mean = median = mode ;

for a positively skew distribution mean > median > mode

and for a negatively skew distribution mean < median < mode.

Scanned with CamScanner

Hence (mean-mode) can be taken as a measure of skewness. However, it is divided by s.d ($\neq 0$) to make the measure pure number. So, the second measure of skewness is

$$S_k = \frac{\text{mean} - \text{mode}}{\text{s.d}} = \frac{\bar{x} - M_0}{s} \quad \text{--- (6g)}$$

This measure is called Pearson's coefficient of skewness.

Using the emperical relation between mean, median and mode, we have a third measure as

$$S_k = \frac{3(\text{mean} - \text{median})}{s} = \frac{3(\bar{x} - M_e)}{s} \quad \text{--- (6h)}$$

Also a fourth measure of skewness by Quartiles is

$$S_k = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \quad \text{--- (6i)}$$

This measure, known as Bowley's coefficient of Skewness, lies between -1 and 1.

6.4 Kurtosis : By Kurtosis of a frequency distribution we mean its degree of peakedness or steepness. Two distribution may be identical in respect of central tendency, dispersion and skewness but one may be more peaked than the other. This feature of a frequency distribution is measured by

$$g_2 = \frac{m_4}{s^4} - 3 = \frac{m_4}{m_2^2} - 3, \text{ assuming } s \neq 0. \quad \text{--- (6j)}$$

It is a pure number and its value is zero for a normal distribution. A normal curve is called mesokurtic (i.e. having medium kurtosis). A positive value of g_2 means there is high concentration of values in the neighbourhood of the measure of central tendency and the distribution has high tails compared to a normal distribution with the same standard deviation. The corresponding distribution is called leptokurtic. Again, a negative value of g_2

Scanned with CamScanner

indicates that there is a less concentration of values near the average and the distribution has less tails in comparison with a normal distribution with same standard deviation. Here the distribution is called platykurtic.

It is to be noted that the quantities g_1^2 and g_2+3 are sometimes used as measure of skewness and kurtosis respectively. These are known as b_1 and b_2 coefficients.

Hence

$$b_1 = \frac{m_3^2}{s^6} = \frac{m_3^2}{m_2^3} \quad \text{--- (6k)}$$

$$b_2 = \frac{m_4}{m_2^2} \quad \text{--- (6l)}$$

It is to be noted that while calculating the value of g_1 from b_1, one has to take the sign of m_3 as sign of g_1. Moreover, in terms of b_2, it may be mentioned that

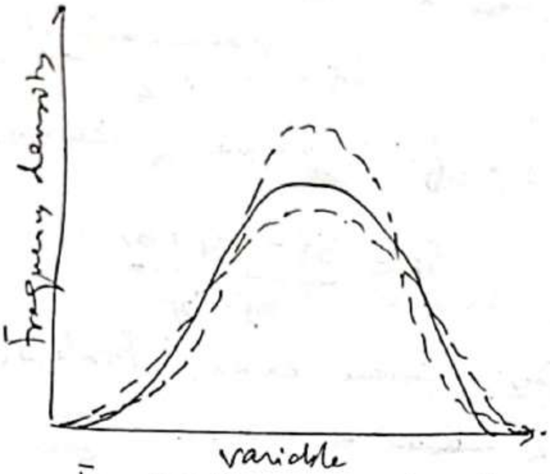

Figure 6.5 Three symmetrical frequency curves with same mean and s.d but with different degree of kurtosis

a distribution is mesokurtic, leptokurtic or platykurtic according as $b_2 = 3$, > 3 or < 3

Example 6.1 : In a certain distribution, the first four moments about the value 4 of the variable are 1, 4, 10 and 45 respectively. Find the moments about mean, b_1 and b_2.

Solution: In our usual notation, we have $A = 4$, $m_1'(4) = 1$, $m_2'(4) = 4$, $m_3'(4) = 10$ and $m_4'(4) = 45$. Now moments about mean may be obtained using the following relations:

$m_1 = 0$ (in all cases)

$m_2 = m_2'(4) - m_1'^2(4) = 4 - 1^2 = 3$

$m_3 = m_3'(4) - 3m_2'(4)m_1'(4) + 2m_1'^3(4) = 10 - 3\times 4\times 1 + 2\times 1^3 = 0$