Class No. 8,9,10
Date: September 13,14,15
MA 2302: Introduction to Probability and Statistics

Mean and Variance

Instructor
Prof. Gopal Krishna Panda
Department of Mathematics
NIT Rourkela

Random Variables

Consider the frequency distribution:

x:	x_1	x_2	………..	x_n
f:	f_1	f_2	………..	f_n

Here, X is a variable say mark of a test, which varies from student to student. f_i is the frequency of x_i which means that there are exactly f_i students with mark x_i, and N is the total number of students. The mean and variance of X are given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i\, x_i, \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i\,(x_i - \bar{x})^2.$$

Observe that $p_i = \frac{f_i}{N}$ is the probability that the mark of a randomly chosen student is x_i. Now, $\bar{x}$ and σ^2 can be written as

$$\bar{x} = \sum_{i=1}^{n} x_i\, p_i = \sum_{x} x f(x), \qquad \sigma^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2\, p_i = \sum_{x} (x - \bar{x})^2\, f(x)$$

where $f(x) = Pr\{X = x\}$. The number $\bar{x}$ whose unit is same as that of the variable, is the arithmetic mean (or simply mean) of X and we can write it as $E(X)$. Notice that the variance is nothing but mean of the squared deviation, where the deviations are measured from mean. (If X is a variable, then $X - a$ is called a deviation of X from a and its possible values are $x_1 - a, x_2 - a, \ldots, x_n - a$. Thus, if $\bar{x} = E(X)$, then $\sigma^2 = E(X - \bar{x})^2$. The equivalence of E is $\frac{1}{N}\sum_{i=1}^{n} f_i$.

Mean and Variance

Now let X be a discrete or continuous random variable with pmf or pdf $f(x)$. If X is discrete, the mean and variance of X are defined as

$$\mu = \sum_x x f(x), \qquad \sigma^2 = \sum_x (x-\mu)^2 f(x).$$

If X is continuous, then the mean and variance of X are defined as

$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx, \qquad \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx.$$

σ is known as the standard deviation of X. The mean of X is also known as expectation of X and written as $E(X)$. In general, the mean or expectation of a function of X is defined as

$$E\big(g(X)\big) = \sum_x g(x)\, f(x)$$

or as

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

Similarly, with $\mu^* = E\big(g(X)\big)$, the variance of $g(X)$ can be calculated as

$$Var(g(X)) = E(g(X) - \mu^*)^2 = \sum_x (g(x) - \mu^*)^2 f(x) \quad \text{OR} \quad \int_{-\infty}^{\infty} (g(x) - \mu^*)^2 f(x)\, dx.$$

Mean and Variance

Example 1: Find the mean and variance of X= the number a fair die turns up.

Ans. Observe that the distribution of X is discrete uniform.

x:	1	2	3	4	5	6
$f(x)$:	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
$xf(x)$:	$1/6$	$2/6$	$3/6$	$4/6$	$5/6$	$6/6$
$(x-\mu)^2 f(x)$:	$(1-3.5)^2 \cdot \frac{1}{6}$	$(2-3.5)^2 \cdot \frac{1}{6}$	$(3-3.5)^2 \cdot \frac{1}{6}$	$(4-3.5)^2 \cdot \frac{1}{6}$	$(5-3.5)^2 \cdot \frac{1}{6}$	$(6-3.5)^2 \cdot \frac{1}{6}$

Thus,

$$\mu = \sum_x x f(x) = \frac{1}{6}(1+2+3+4+5+6) = 3.5.$$

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \frac{1}{6}(2.5^2 + 1.5^2 + 0.5^2 + 0.5^2 + 1.5^2 + 2.5^2) = \frac{17.5}{6} = \frac{35}{12} = 2.92\,.$$

Mean and Variance

Example 2: Let X be the number of times an unfair coin is flipped until the first head appears. Find the mean of X.

Ans. Let p be the probability of a head in a single toss and $q = 1 - p$. The first head will appear in the x-th toss if the previous $x - 1$ tosses result in tails. Hence, for $x = 1,2,3,\ldots$

$$f(x) = \Pr\{X = x\} = P(TTTT \cdots \cdots \cdots TTH) = q^{x-1}p.$$

Hence,

$$\mu = \sum_{x} xf(x) = \sum_{x=1}^{\infty} xq^{x-1}p = p\sum_{x=1}^{\infty} xq^{x-1} = p(1-q)^{-2} = \frac{1}{p}.$$

If the coin is fair then $p = 1/2$ and $\mu = 2$. What is the meaning of this?

If you toss a fair coin, the probability of a head is 1/2. Thus, in every toss, you expect 1/2 of a head. Hence, for a complete head, you need 2 tosses.

Mean and Variance

A simplified formula for the variance:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\,dx$$

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

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - 2\mu \times \mu + \mu^2 \times 1 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2.$$

Thus,

$$\boldsymbol{\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2} = E(X^2) - \mu^2 = E(X^2) - E^2(X).$$

In case of discrete random variables,

$$\boldsymbol{\sigma^2 = \sum_x x^2 f(x) - \mu^2}.$$

Mean and Variance

Example 3: If the diameter X [cm] of certain bolts has the density $f(x) = K(x - 0.9)(1.1 - x)$ for $0.9 < x < 1.1$ and $f(x) = 0$ otherwise, find K, μ and σ^2.

Ans: K: $\int_{-\infty}^{\infty} f(x)\, dx = 1$. Hence, with a change of variable $y = x - 1$,

$$K \int_{0.9}^{1.1} (x - 0.9)(1.1 - x)\, dx = K \int_{-0.1}^{0.1} (0.1 + y)(0.1 - y)\, dy$$

$$= 2K \int_{0}^{0.1} (0.01 - y^2)\, dy = \frac{K}{750} = 1.$$

Hence, $K = 750$. Now,

$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx = 750 \int_{0.9}^{1.1} x(x - 0.9)(1.1 - x)\, dx$$

$$= 750 \int_{0.9}^{1.1} (x - 1)(x - 0.9)(1.1 - x)\, dx + 750 \int_{0.9}^{1.1} (x - 0.9)(1.1 - x)\, dx$$

$= 750 \int_{-0.1}^{0.1} y(0.01 - y^2)\, dy + 1 = 0 + 1 = 1$. (since $y(0.01 - y^2)$ is an odd function of y)

Mean and Variance

Now, $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = 750 \int_{0.9}^{1.1} (x-1)^2 (x-0.9)(1.1-x)\, dx$

$$= 750 \int_{-0.1}^{0.1} y^2 (0.01 - y^2)\, dy = 2 \times 750 \int_{0}^{0.1} y^2 (0.01 - y^2)\, dy$$

$$= 1500 \left(0.01 \times \frac{0.1^3}{3} - \frac{0.1^5}{5} \right) = 3.$$

Example 4: If, in Ex. 3, a defective bolt is one that deviates from 1.00 cm by more than 0.06 cm, what percentage of defectives should we expect?

Ans. $\Pr\{\text{A bolt is nondefective}\} = p = \Pr\{0.94 \le X \le 1.06\} = \int_{0.94}^{1.06} f(x)\, dx$

$$= 750 \int_{0.94}^{1.06} (x-0.9)(1.1-x)\, dx = 750 \int_{-0.06}^{0.06} (0.01 - y^2)\, dy = 1500 \int_{0}^{0.06} (0.01 - y^2)\, dy$$
$$= 0.90.$$

Hence, $\Pr\{\text{A bolt is defective}\} = 1 - p = 0.10$. Thus, we can expect about 10% defective bolts.

Mean and Variance

Let X be a discrete or continuous random variable with mean μ and variance σ^2 and let $X^* = c_1X + c_2$, where c_1 and c_2 are constants. Then, by definition of expectation.

(**Mean and variance of linear transformations**)

$$\mu^* = E(X^*) = \int_{-\infty}^{\infty} x^* \, f(x)\, dx = \int_{-\infty}^{\infty} (c_1x + c_2)\, f(x)\, dx$$

$$= c_1 \int_{-\infty}^{\infty} x\, f(x)\, dx + c_2 \int_{-\infty}^{\infty} f(x)\, dx = c_1\mu + c_2.$$

$$\sigma^{*2} = Var(X^*) = E(X^* - \mu^*)^2$$

$$= \int_{-\infty}^{\infty} (x^* - \mu^*)^2 f(x)\, dx = \int_{-\infty}^{\infty} (c_1x - c_1\mu)^2 f(x)\, dx$$

$$= c_1^2 \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx = c_1^2\sigma^2.$$

Mean and Variance

Let X be a discrete or continuous random variable with mean μ and variance σ^2. Then

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

is known as a standardized variable corresponding to X. Putting $c_1 = 1/\sigma$ and $c_2 = -\mu/\sigma$, it is easy to see that the mean of Z is 0 and the variance of Z is equal to 1. That is

$$\mu_z = 0, \sigma_z^2 = 1$$

and, in particular, the standard deviation $\sigma_z = 1$.

Example 5: A small filling station is supplied with gasoline every Saturday afternoon. Assume that its volume X of sales in ten thousands of gallons has the probability density $f(x) = 6x(1 - x)$ if $0 \le x \le 1$ and $f(x) = 0$ otherwise. Determine the mean, the variance, and the standardized variable corresponding to X.

Ans. The mean is given by

$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 x \cdot 6x(1 - x)\, dx = 2 - \frac{3}{2} = \frac{1}{2}$. (means 5000 gallons)

Mean and Variance

Furthermore,

$$E(X^2) = \int_0^1 x^2 \cdot 6x(1-x)\,dx = \frac{3}{2} - \frac{6}{5} = \frac{3}{10}.$$

Thus, the variance is given by

$$\sigma^2 = E(X^2) - \mu^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

The standardized variable corresponding to X is

$$Z = \frac{X-\mu}{\sigma} = \frac{X - \frac{1}{2}}{\sqrt{\frac{1}{20}}} = \sqrt{5}(2X-1).$$

Example 6: What capacity must the tank in Ex. 5 have in order that the probability that the tank will be emptied in a given week be 5% ?

Ans. Let C be the required capacity. Then we have to find C such that $\Pr\{X \geq C\} = 0.05$, i.e.

$\int_0^C 6x(1-x)\,dx = 0.95 \Rightarrow 3C^2 - 2C^3 = 0.95 \Rightarrow C = 0.864649$ Capacity should be 8646.5 gallons. (solution between 0 and 1 and close to 1 is to be found out).

Mean and Variance

Example 7: What is the mean life of a light bulb whose life X [hours] has the density $f(x) = 0.001e^{-0.001x}$ if $x > 0$ and $f(x) = 0$ otherwise.

Ans. The mean is given by

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx = 0.001 \int_{0}^{\infty} x\,e^{-0.001x}\,dx = \frac{1}{0.001} = 1000 \text{ hours.}$$

(use integration by parts)

Example 8: James rolls 2 fair dice, and Harry pays k cents to James, where k is the product of the two faces that shown on the dice. How much should James pay to Harry for each game to make the game fair?

The game will be fair if the expected gain of any player is equal to zero. Calculate how much on the average Harry is paying to James in a single game. That much James must pay to Harry before each game so that the game will be fair. Rest you do.

Moments

If X is a discrete or continuous random variable, then

$$\mu_r' = E(X^r), r = 1,2, \ldots$$

is called the r-th moment (about origin) of X. Moreover, $E(X-a)^r$ is known as the r-th moment about the point a. If $a = \mu = E(X)$, then

$$\mu_r = E(X-\mu)^r, r = 1,2, \ldots$$

is known as the r-th central moment of X. If X is discrete, then

$$\mu_r' = \sum_x x^r f(x), \qquad \mu_r = \sum_x (x-\mu)^r f(x).$$

If X is continuous, then

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x)\, dx, \qquad \mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\, dx.$$

Moments

Observe that

$$\mu_1' = \mu,\ \mu_1 = 0 \text{ and } \mu_2 = \sigma^2.$$

The moments about origin have no other role except μ_1'. These moments are used to calculate the central moments easily.

The relation $\sigma^2 = E(X^2) - \mu^2$ can be written as

$$\mu_2 = \mu_2' - {\mu_1'}^2.$$

Similarly, using binomial theorem, one can easily prove that

$$\mu_3 = \int_{-\infty}^{\infty} (x-\mu)^3 f(x)\, dx = \mu_3' - 3\mu_2'\, \mu_1' + 2{\mu_1'}^3$$

and

$$\mu_4 = \int_{-\infty}^{\infty} (x-\mu)^4 f(x)\, dx = \mu_4' - 4\mu_3'\, \mu_1' + 6\mu_2' {\mu_1'}^2 - 3{\mu_1'}^4.$$

Assignment: Write the corresponding formula for μ_5.

Moments

The quantities

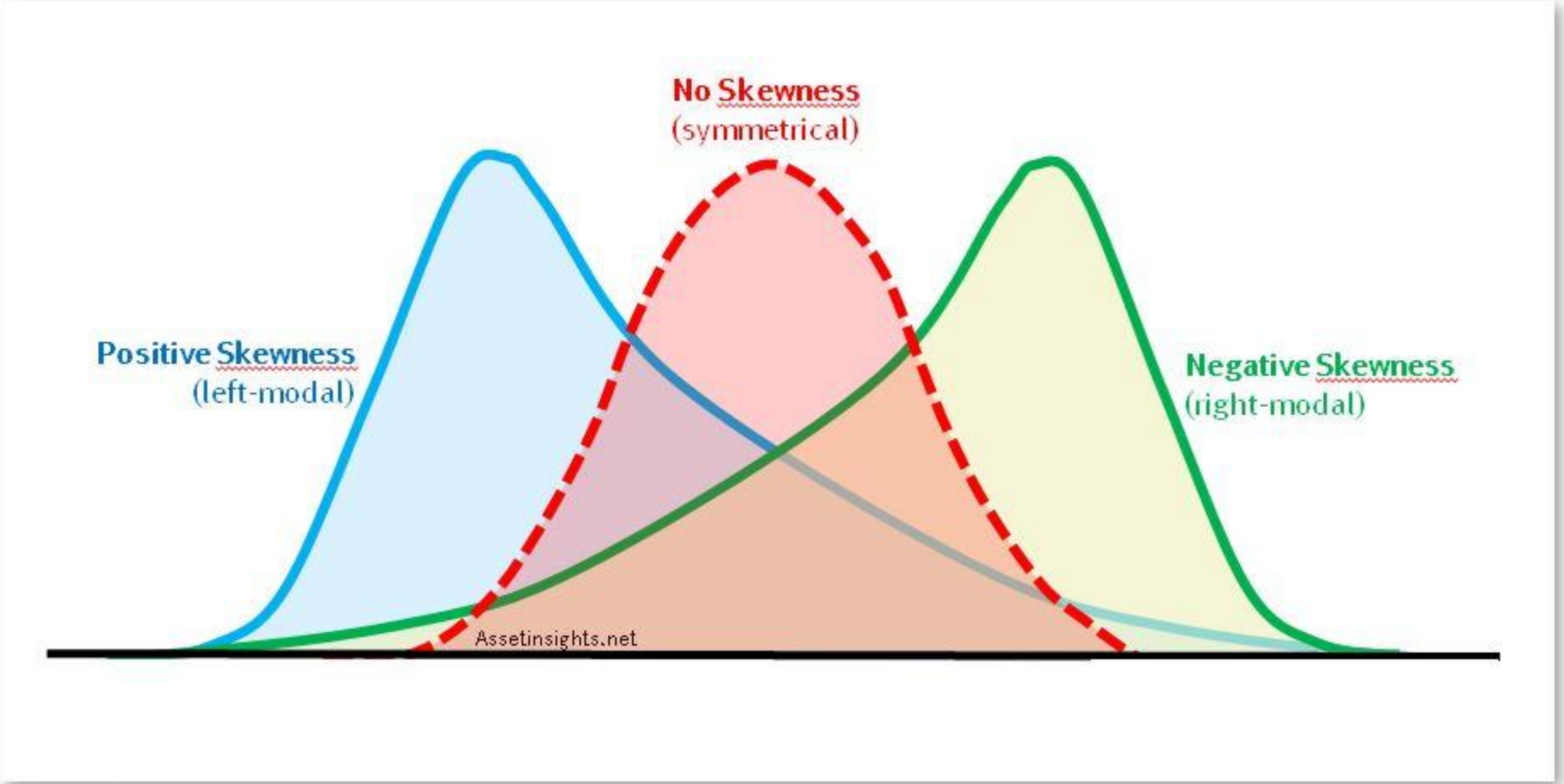
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}$$

or

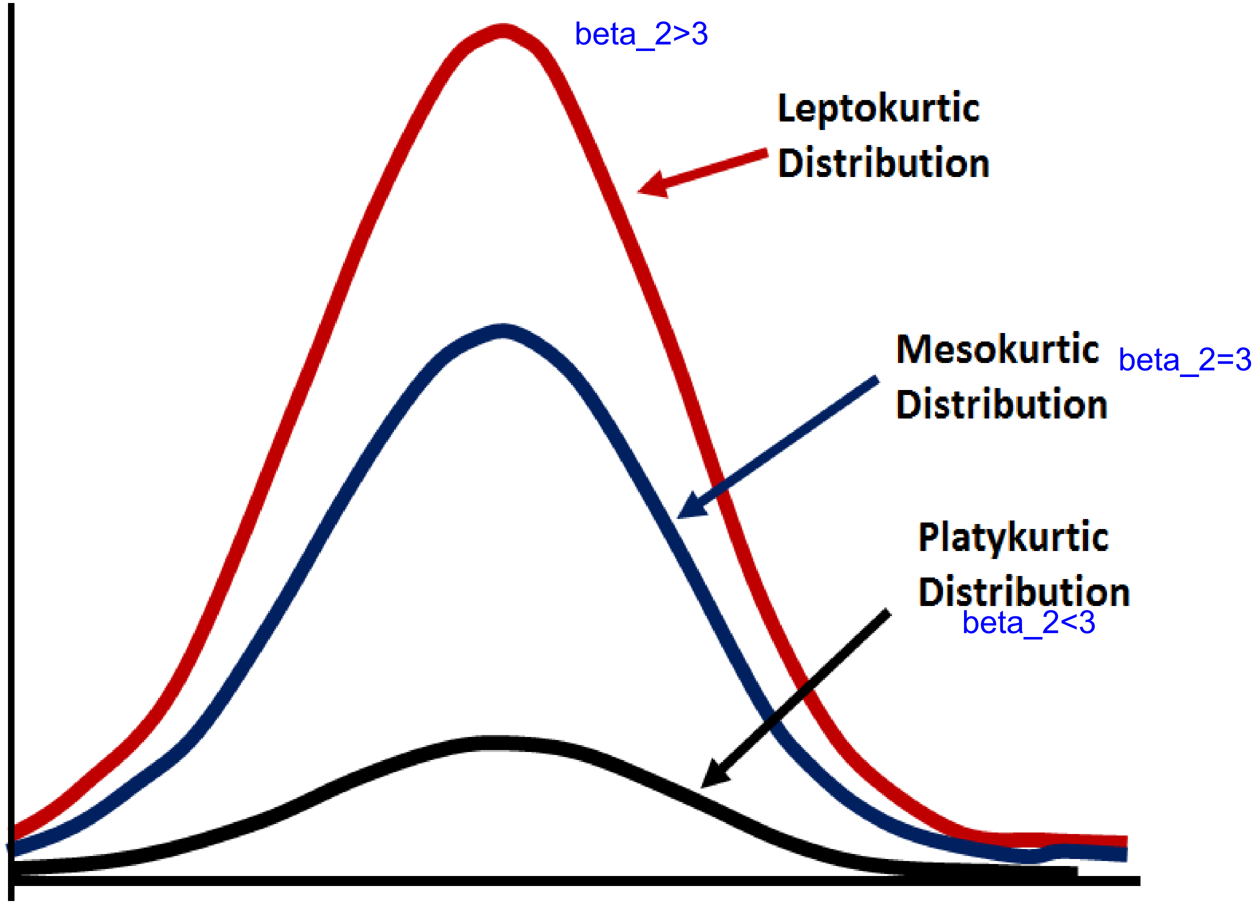
$$\gamma_1 = \sqrt{\beta_1} \text{ and } \gamma_2 = \beta_2 - 3$$

are known as the coefficients of skewness and kurtosis of X. If X is a symmetric random variable (that if the pmf or pdf of X is a symmetric function), then $\beta_1 = 0$ and the random variable X or its distribution is called a symmetric distribution. If $\mu_3 > 0$, it is called positively skewed and if $\mu_3 < 0$, it is called negatively skewed. Similarly, if $\beta_2 = 3$ or $\gamma_2 = 0$, then the distribution (or curve for $f(x)$) is called *normal* or *mesokurtic*. If $\beta_2 < 3$ or $\gamma_2 < 0$, the curve for $f(x)$ is flatter compared to the normal curve and the distribution or curve is called *platykurtic*. If $\beta_2 > 3$ or $\gamma_2 > 0$, the curve for $f(x)$ is more peaked compared to the normal curve and the distribution or curve is called *lectokurtic*.

Moments


No Skewness
(symmetrical)
Positive Skewness
(left-modal)
Negative Skewness
(right-modal)
Assetinsights.net

Moments


beta_2>3
Leptokurtic
Distribution
Mesokurtic
Distribution
beta_2=3
Platykurtic
Distribution
beta_2<3

Moments

Example 8: Find β_1 and β_2 for the random variable X with pdf $f(x) = 3x^2$ if $0 < x < 1$ and $f(x) = 0$ otherwise.

Ans. The r-th moment of X about origin is

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)\,dx = \int_0^1 x^r \cdot 3x^2\,dx = \frac{3}{r+3}, r = 1,2,\ldots$$

Thus, $\mu_1' = \frac{3}{4}, \mu_2' = \frac{3}{5},\ \mu_3' = \frac{3}{6},\ \mu_4' = \frac{3}{7}$. Rest you do.

Moment generating function:

$$G(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{\infty}\left(1 + tx + \frac{t^2x^2}{2!} + \cdots\right) f(x)\,dx$$

$$= 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} E(X^r) = 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} \mu_r'.$$

Moments

Thus,

$$\mu_r' = \text{Coefficient of } \frac{t^r}{r!} \text{ in the power series expansion of } G(t).$$

In view of the properties of power series,

$$\mu_1' = G'(0), \qquad \mu_2' = G''(0), \quad \mu_3' = G'''(0)$$

and, in general

$$\mu_r' = G^{(r)}(0) = \frac{d^r}{dt^r} G(t)|_{t=0}.$$