MATHEMATICAL EXPECTATION :
CONTINUOUS RANDOM VARIABLES

TABLE OF CONTENTS

Content Developer

Chandra Goswami, Associate Professor, Department of Economics

Dyal Singh College, University of Delhi

Reference

Jay L. Devore: *Probability and Statistics for Engineering and the Sciences*, Cengage Learning, 8th edition [Chapter 4]

MATHEMATICAL EXPECTATION: CONTINUOUS RANDOM VARIABLES

***Learning objectives*:**

In this chapter you will learn how to obtain two main characteristics of the probability distribution of a continuous random variable. You will learn how to derive the mean and variance of distributions of continuous random variables. You will also learn how to apply the rules of mathematical expectation to functions of random variables as well as to sums of random variables. The mean, variance, median and mode, and coefficients of skewness and kurtosis will help you to identify the characteristics and shape of the distribution.

Chapter Outline

1. Expected value of a continuous random variable
2. Expectation of a function of a continuous random variable
3. Variance of a continuous random variable
4. Variance of a function of a continuous random variable
5. Rules of mathematical expectation
6. Expectation and variance of sums of continuous random variables
7. Characteristics of the probability density function

1 EXPECTED VALUE OF A CONTINUOUS RANDOM VARIABLE

The mean of a distribution is the point on the number line where the distribution is centered. The mean of the distribution of a continuous random variable (probability density function or pdf) is its expected value. Expected value of a continuous random variable is obtained as a weighted average of the values of the rv where the probability densities are the weights. For discrete random variables method of summation was used. In case of continuous random variables, expected value is obtained by method of integration.

Definition **1**

The ***expected value*** or mean value of a continuous random variable *X* with probability density function *f(x)* is

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

When the pdf *f(x)* specifies a model for the distribution of *X* values in a numerical population, then μ_X is the population mean.

Example 1.1

If a contractor's profits on a construction job can be looked upon as a continuous rv having the pdf

$$f(x) = \begin{cases} \frac{1}{18}(x+1) & -1 < x < 5 \\ 0 & otherwise \end{cases}$$

where the units are \$1000, her expected profit is

$$E(X) = \int_{-\infty}^{\infty} x.\frac{1}{18}(x+1)dx$$

$$= \frac{1}{18}\int_{-1}^{5}(x^2 + x)dx$$

$$= \frac{1}{18}\left[\frac{x^3}{3} + \frac{x^2}{2}\right]_{-1}^{5}$$

$$= \frac{1}{18}\left[\left(\frac{125}{3} + \frac{25}{2}\right) - \left(\frac{-1}{3} + \frac{1}{2}\right)\right]$$

$$= \frac{1}{18}\left[\frac{126}{3} + \frac{24}{2}\right]$$

$$= \frac{42+12}{18} = \frac{54}{18} = 3$$

Therefore, expected profit is \$3,000

Exercise 1

The tread wear (in thousands of kilometers) that car owners get with a certain kind of tyre is a rv *X* whose pdf is given by

$$f(x)=\begin{cases}\frac{1}{30}e^{-\frac{x}{30}} & x>0\\ 0 & x\le 0\end{cases}$$

What tread wear can a car owner expect to get with one of the tyres?

Solution

$$E(X)=\int_{-\infty}^{\infty} x.\frac{1}{30}e^{-\frac{x}{30}}dx$$

$$=\frac{1}{30}\int_{0}^{\infty} x.e^{-\frac{x}{30}}dx$$

Integrating by parts,

$$E(X)=\frac{1}{30}\left[x\int_0^{\infty} e^{-\frac{x}{30}}dx-\int_0^{\infty}\left(1\int_0^{\infty} e^{-\frac{x}{30}}\right)dx\right]$$

$$=\frac{1}{30}\left[x\frac{e^{-\frac{x}{30}}}{-\frac{1}{30}}\Bigg|_0^{\infty}-\int_0^{\infty}\frac{e^{-\frac{x}{30}}}{-\frac{1}{30}}dx\right]$$

$$=\frac{1}{30}[0]+\int_0^{\infty} e^{-\frac{x}{30}}dx \quad =-\frac{e^{-\frac{x}{30}}}{\frac{1}{30}}\Bigg|_0^{\infty}$$

$$=0-\left(-\frac{1}{\frac{1}{30}}\right) \quad =30$$

Therefore, average tread wear a car owner can expect to get is 30,000 km.

2 EXPECTED VALUE OF A FUNCTION OF A CONTINUOUS RANDOM VARIABLE

If *X* is a continuous rv with probability density function *f(x)*, then any function of X, *h(X)*, will also have the pdf *f(x)*.

***Definition* 2**

If X is a continuous random variable with pdf f(x) and h(X) is any function of X, then

$$E[h(X)]=\mu_{h(X)}=\int_{-\infty}^{\infty} h(x)f(x)dx \text{ provided that } \int_{-\infty}^{\infty}|h(x)|f(x)dx<\infty$$

***Proposition* 1**

If h(X) is a linear function such as h(X) = aX + b, then E[h(X)] = aE(X) + b

Proof:

E[h(X)] = E[aX + b] = $\int_{-\infty}^{\infty} (ax+b) f(x)dx$

$$= \int_{-\infty}^{\infty} axf(x)dx + \int_{-\infty}^{\infty} bf(x)dx$$

$$= a\int_{-\infty}^{\infty} xf(x)dx + b\int_{-\infty}^{\infty} f(x)dx$$

= aE(X) + b

Example 2.1

If the pdf of *X* is given by

$$f(x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

$$and \quad h(X) = 2X + 1$$

Then,

$$E[h(X)] = \int_{-\infty}^{\infty} (2x+1).2(1-x)dx$$

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

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

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

$$= 2\left[\left(\frac{1}{2} + 1 - \frac{2}{3}\right) - 0\right]$$

$$= 2\left[\frac{3}{2} - \frac{2}{3}\right]$$

$$= \left[3 - \frac{4}{3}\right] \quad = \frac{5}{3} \quad = 1.67$$

Exercise 2

An ecologist wishes to mark off a circular sampling region having radius 10 m. However, the radius of the resulting region is actually a random variable *R* with pdf

$$f(r) = \begin{cases} \frac{3}{4}\left[1-(10-r)^2\right] & 9 \le r \le 11 \\ 0 & otherwise \end{cases}$$

What is the expected area of the resulting circular region?

Solution

Since area of a circle is h(R) = πR^2, therefore expected area of the resulting circle is

E[h(R)] = $E\left[\frac{22}{7}R^2\right] = \frac{22}{7}E\left(R^2\right)$ where $E\left(R^2\right) = \int_9^{11} r^2 f(r)\,dr$

Now,

$$E(R^2) = \frac{3}{4}\int_9^{11} r^2\left[1-100+20r-r^2\right]dr$$

$$= \frac{3}{4}\int_9^{11}\left[-99r^2+20r^3-r^4\right]dr$$

$$= \frac{3}{4}\left[-\frac{99r^3}{3}+\frac{20r^4}{4}-\frac{r^5}{5}\right]_9^{11}$$

$$= \frac{3}{4}\left[33\left(9^3-11^3\right)+5\left(11^4-9^4\right)+\frac{1}{5}\left(9^5-11^5\right)\right]$$

$$= \frac{3}{4}\left[33(729-1331)+5(14641-6561)+\frac{1}{5}(59049-161051)\right]$$

$$= \frac{3}{4}\left[33(-602)+5(8080)+\frac{1}{5}(-102002)\right]$$

$$= \frac{3}{4}\left[-19866+40400-20400.4\right]$$

$$= \frac{3}{4}(133.6) = 100.2$$

Since area of a circle is h(R) = $\pi R^{2,}$ therefore expected area of the resulting circle is

E[h(R)] = $E\left[\frac{22}{7}R^2\right] = \frac{22}{7}E\left(R^2\right) = \frac{22}{7}(100.2) = 314.9143\,sq.m.$

Exercise 3

The weekly demand for propane gas (in 1000s of gallons) from a particular facility is an rv *X* with pdf

$$f(x) = \begin{cases} 2\left(1 - \dfrac{1}{x^2}\right) & 1 \le x \le 2 \\ 0 & otherwise \end{cases}$$

(a) Compute E(X)

(b) If 1.5 thousand gallons are in stock at the beginning of the week, how much of the 1.5 thousand gallons is expected to be left at the end of the week?

Solution

(a)

$$E(X) = \int_1^2 x.2\left(1 - \frac{1}{x^2}\right)dx$$

$$= 2\int_1^2 \left(x - \frac{1}{x}\right)dx = 2\left[\frac{x^2}{2} - \ln(x)\right]_1^2$$

$$= 2\left[(2 - 0.693) - \left(\frac{1}{2} - 0\right)\right] = 1.614$$

= 1,614 gallons

(b) Amount in stock is 1.5 thousand gallons out of which the demand is a random variable *X* thousand gallons.

Amount left = h(x) = max{(1.5 - x), 0] thousand gallons

$$E[h(x)] = \int_1^2 \max[(1.5 - x), 0] f(x)\,dx$$

$$= \int_1^{1.5} (1.5 - x).2\left(1 - \frac{1}{x^2}\right)dx$$

$$= 2\int_1^{1.5} \left(1.5 - x - \frac{1.5}{x^2} + \frac{1}{x}\right)dx$$

$$= 2\left[1.5x - \frac{x^2}{2} + \frac{1.5}{x} + \ln x\right]_1^{1.5}$$

$$= 2\left[\left\{(1.5)^2 - \frac{(1.5)^2}{2} + \frac{1.5}{1.5} + 0.4055\right\} - \left\{1.5 - \frac{1}{2} + 1.5 + 0\right\}\right]$$

$$= 2[2.5305 - 2.5] \quad = 0.061$$

Therefore, the expected amount left in stock at the end of the week is 61 gallons.

(Since weekly demand for propane can vary between 1000 and 2000 gallons for $1 \leq x \leq 2$, and amount in stock is 1.5 thousand gallons, amount left at the end of the week can vary between the minimum of 0 if demand x = 1.5 or more, and maximum of 500 gallons if demand is x = 1.)

3 VARIANCE OF A CONTINUOUS RANDOM VARIABLE

The variance and standard deviation reflect the spread or dispersion of the probability distribution or the population of x values. Two distributions with the same expected value can be distinguished as two distinct distributions if they have different values of variance.

Definition **3**

The ***variance*** of a continuous random variable *X* with pdf *f(x)* and mean value μ is

$$V(X) = \sigma_X^2 = E[X-\mu]^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

and ***standard deviation*** of X is $\sigma_X = \sqrt{V(X)}$

It can be shown that for continuous random variables, just like in the case of discrete random variables,

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

Proof

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

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

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

$$= E(X^2) - 2\mu E(X) + \mu^2 \qquad \text{since} \int_{-\infty}^{\infty} f(x)dx = 1$$

$$= E(X^2) - \mu^2 \qquad \text{since } E(X) = \mu$$

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

Example 3.1

In *exercise* 3, the pdf of X is given as

$$f(x) = \begin{cases} 2\left(1 - \dfrac{1}{x^2}\right) & 1 \le x \le 2 \\ 0 & otherwise \end{cases}$$

V(X) = E(X^2) – [E(X)]^2 , where E(X) = 1.614

$$E(X^2) = \int^{1} x^2\, 2\left(1 - \frac{1}{x^2}\right) dx$$

$$= 2\int_1^2 (x^2 - 1)\, dx$$

$$= 2\left[\frac{x^3}{3} - x\right]_1^2$$

$$= 2\left[\left(\frac{8}{3} - 2\right) - \left(\frac{1}{3} - 1\right)\right]$$

$$= 2\left[\frac{2}{3} + \frac{2}{3}\right] = \frac{8}{3} = 2.667$$

V(X) = 2.667 – $(1.614)^2$ = 2.667 – 2.605 = 0.062

and σ_x = 0.249

= 249 gallons.

Exercise 4

For what value of *k* does V(Y) =2 when pdf of *Y* is given as

$$f(y) = \begin{cases} \dfrac{2y}{k^2} & 0 \le y \le k \\ 0 & otherwise \end{cases}$$

Solution

$$E(Y) = \int_0^k y \frac{2y}{k^2} dy$$

$$= \frac{2}{k^2}\left[\frac{y^3}{3}\right]_0^k = \frac{2}{3}k$$

$$E(Y^2) = \int_0^k y^2 \frac{2y}{k^2} dy$$

$$= \frac{2}{k^2}\left[\frac{y^4}{4}\right]_0^k = \frac{1}{2}k^2$$

$$V(Y) = \frac{1}{2}k^2 - \frac{4}{9}k^2 = k^2\left[\frac{1}{18}\right]$$

$$Given\, that\, V(Y) = \frac{k^2}{18} = 2,$$

$$k^2 = 36.$$

Therefore, k = 6

The variance is the second moment about the mean, ie the second central moment. It is obtained by taking the difference of the second moment about the origin {ie,E(X^2)} and the square of the first moment about the origin {ie, $[E(X)]^2$}

4 VARIANCE OF A FUNCTION OF A CONTINUOUS RANDOM VARIABLE

Variance of a function of the random variable, *h(X),* is obtained by substituting *h(X)* in place of *X* in the definitional formula for variance of X. Thus,

$$V[h(X)] = E[h(X) - E\{h(X)\}]^2$$

$$= E[h(X)]^2 - [E\{h(X)\}]^2$$

where

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

and $$E[h(X)]^2 = \int_{-\infty}^{\infty} [h(x)]^2 f(x) dx$$

***Proposition* 2**

If h(X) is a linear function such as h(X) = aX + b, then V[h(X)] = a^2V(X)

Proof:

When h(X) = aX + b, ie, a linear function of the rv,

E[h(X)] = E[aX + b]

= a E(X) + b

= aμ + b

and

$$V[h(X)] = \left[\int_{-\infty}^{\infty} (ax+b)^2 f(x) dx \right] - (a\mu + b)^2$$

$$= \int_{-\infty}^{\infty} \left(a^2x^2 + 2abx + b^2\right) f(x) dx - (a\mu + b)^2$$

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

$$= a^2 E\left(X^2\right) + 2ab\mu + b^2 - a^2\mu^2 - 2ab\mu - b^2$$

$$= a^2 \left[E\left(X^2\right) - \mu^2\right]$$

$$= a^2 \left[E\left(X^2\right) - [E(X)]^2\right]$$

$$= a^2 V(X)$$

Example 4.1

In *example* 1.1 the contractor's profits *X* (in thousand dollars) was a continuous rv with pdf

$$f(x) = \begin{cases} \frac{1}{18}(x+1) & -1 < x < 5 \\ 0 & otherwise \end{cases}$$

Expected profits = E(X) = 3 = $3000

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

$$= \left[\int_{-1}^{5} x^2 . \frac{1}{18}(x+1)\, dx \right] - 9 \quad = \left[\frac{1}{18} \int_{-1}^{5} \left(x^3 + x^2\right) dx \right] - 9$$

$$= \frac{1}{18} \left[\frac{x^4}{4} + \frac{x^3}{3} \right]_{-1}^{5} - 9$$

$$= \frac{1}{18} \left[\left(\frac{625}{4} + \frac{125}{3} \right) - \left(\frac{1}{4} - \frac{1}{3} \right) \right] - 9 = \frac{1}{18}(156 + 42) - 9 \quad = 2$$

$$and\ \sigma_X = \sqrt{2} = 1.41421 = \$1414.21$$

Exercise 5

Let *Y* have the pdf

$$f(y) = \begin{cases} 3(1-y)^2 & 0 < y < 1 \\ 0 & otherwise \end{cases}$$

Find the variance of W, where W = 12 – 5Y

Solution

V(W) = 25V(Y)

$$E(Y) = \int_0^1 y.4(1-y)^2 \, dy$$

$$= 4\int_0^1 \left(y - 2y^2 + y^3\right) dy$$

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

$$= 4\left[\frac{1}{2} - \frac{2}{3} + \frac{1}{4}\right] \quad = \frac{1}{3}$$

$$E(Y^2) = \int_0^1 y^2.4(1-y)^2 \, dy$$

$$= 4\int_0^1 \left(y^2 - 2y^3 + y^4\right) dy$$

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

$$= 4\left[\frac{1}{3} - \frac{2}{4} + \frac{1}{5}\right] \quad = \frac{2}{15}$$

$$V(Y) = \frac{2}{15} - \frac{1}{9} \quad = \frac{1}{45}$$

$$\therefore V(W) = (25)\frac{1}{45} = \frac{5}{9}$$

5 RULES OF MATHEMATICAL EXPECTATION

Let us consider a linear function of the continuous variable X, h(X) = aX + b, where the pdf of *X* is *f(x)* so that the probability distribution of h(X) is f(x)

In *section* 2 it has been proved that E[h(X)] = aE(X) + b.

This gives us the following three ***rules of mathematical expectation***

***Rule* 1**

If *b* = 0, then for any constant *a*, E(aX) = aE(X)

The rule says that expected value in the new units equals the expected value in the old units multiplied by the factor *a*. Multiplication of *X* by the constant *a* changes the unit of measurement

***Rule* 2**

If *a* = 1, then for any constant *b*, E(X + b) = E(X) + b

If a constant is added to each possible value of *X*, there is a change in origin. Then the expected value will be shifted by the same amount *b* to the right or left depending on whether *b* is greater than or less than zero respectively.

***Rule* 3**

If *a* = 0, then for any constant *b*, E(b) = $\int_{-\infty}^{\infty} b.f(x)\,dx = b\int_{-\infty}^{\infty} f(x)\,dx$ = b

That is, the expected value of a constant is just its value.

In *section* 4 we have seen that for the linear function h(X) = aX + b.

V[h(X)] = a^2V(X)

This gives us the following three ***rules for variance***

***Rule* 4**

$V(aX) = \sigma_{aX}^2 = a^2V(X)$ and $\sigma_{aX} = |a|.\sigma_X$ when b= 0 in h(X) = aX + b

A change in the unit of measurement by multiplication or division by a constant *a* impacts the variability. The new standard deviation is a product of the old standard deviation and the absolute value of the conversion factor *a*

***Rule* 5**

$V(X+b) = \sigma_{X+b}^2 = V(X)$ and $\sigma_{X+b} = \sigma_X$ when a= 1 in h(X) = aX + b

A change in origin by adding or subtracting a constant *b* does not affect the variability of the distribution

***Rule* 6**

$V(b) = 0$ and $\sigma_{X+b} = 0$ when a= 0 in h(X) = aX + b

Thus the same rules of mathematical expectation for calculating the mean and variance apply for distributions of both discrete and continuous random variables.

6 EXPECTATION AND VARIANCE OF SUMS OF CONTINUOUS RANDOM VARIABLES

Let us consider the sum of two continuous random variables *X* and *Y*, where f(x) and f(y) are the respective probability density functions. Then, g(X,Y) = X + Y

***Proposition* 3**

Whether or not the random variables *X* and *Y* are independent, E(X+Y) = E(X) + E(Y)

Propositions 1 extends this result to sums of linear functions of random variables.

Let g(X) = aX + b, and h(Y) = cY + d, then

E[g(X) + h(Y)] = E[aX + b + cY + d]

= [aE(X) +b] + [cE(Y) + d]

= E[h(X)] + E[g(Y)]

This result can be extended to more than two linear functions of continuous random variables. For example, if there are linear functions of X, Y and Z, the expected value of the sum of the functions is the sum of the expected value of the functions.

***Proposition* 4**

If the random variables *X* and *Y* are independent, V(X+Y) = V(X) + V(Y)

Proposition 2 extends this result to sums of linear functions of independent random variables.

If g(X) = aX + b, and h(Y) = cY + d, then

V[g(X) + h(Y)] = V[aX + b + cY + d]

= $[a^2V(X) +0] + [c^2V(Y) + 0]$

= $a^2V(X) + c^2V(Y)$

$= V[g(X)] + V[h(Y)]$

and

$\sigma_{g(X)+h(Y)} = \sqrt{V[g(X)+V[h(Y)]}$

This result can be similarly extended to sums of linear functions of three or more than three independent continuous random variables. We see that the methods for obtaining the expectation and variance of sums of linear functions of continuous random variables are the same as that for discrete variables.

Example 6.1

The independent random variables *X* and *Y* have the following density functions:

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

and $$f(y) = \begin{cases} 2(1-y) & 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

Let g(X) = 3X -1 and h(Y) = 2Y + 1.

Then,

E[g(X) + h(Y)] = E[g(X)] + E[h(Y)]

= [3E(X) – 1] + [2E(Y) + 1]

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

$$E[g(X)] = (3)\frac{4}{3} - 1 = 3$$

$$E(Y) = \int_0^1 y.2(1-y)\,dy = 2\int_0^1 (y-y^2)\,dy = 2\left[\frac{y^2}{2}-\frac{y^3}{3}\right]_0^1 = 2\left[\frac{1}{2}-\frac{1}{3}\right] = \frac{1}{3}$$

$$E[h(Y)] = (2)\frac{1}{3} + 1 = \frac{5}{3}$$

$$\therefore E[g(X)+h(Y)] = 3 + \frac{5}{3} = \frac{14}{3} = 4.667$$

Also

V[g(X) + h(Y)] = V[g(X)] + V[h(Y)]

Now,

V[g(X)] = 9V(X) = 9[E(X^2) – $(4/3)^2$]

and

V[h(Y)] = 4V(Y) = 4[E(Y^2) – (1/3)^2}

$$E(X^2) = \int_0^2 x^2 . \frac{x}{2} dx = \frac{1}{2}\int_0^2 x^3 dx = \frac{1}{2}\left[\frac{x^4}{4}\right]_0^2 = 2$$

$$\therefore V[g(X)] = 9\left[2 - \frac{16}{9}\right] = 2$$

$$E(Y^2) = \int_0^1 y^2 .2(1-y)\, dy = 2\int_0^1 \left(y^2 - y^3\right) dy = 2\left[\frac{y^3}{3} - \frac{y^4}{4}\right]_0^1 = \frac{1}{6}$$

$$\therefore V[h(Y)] = 4\left[\frac{1}{6} - \frac{1}{9}\right] = \frac{2}{9}$$

$$\therefore V[g(X) + h(Y)] = 2 + \frac{2}{9} = \frac{20}{9} = 2.222$$

7 CHARACTERISTICS OF THE PROBABILITY DENSITY FUNCTION

The probability distribution of a continuous rv is its pdf. Any distribution can be characterized by its mean and variance. The expected value of a rv is called its ***mean*** value. To compute the population average value of *X* we need only the possible values of *X* along with their respective probabilities. The size of the population is immaterial as long as the pdf is given. The mean of the distribution is the point on the number line where the graph of the distribution is centered.

Variance or standard deviation of a distribution measures the spread of the distribution. A smaller value of variance indicates that the distribution is clustered closer to its mean value, whereas a larger value of variance means that the distribution is more spread out. Two distributions having the same mean value but different variances will be two distinct distributions. Similarly, graphs of two distributions with the same variance but different means will have the same spread but centered at different points on the number line.

For a unimodal distribution, by equating the first derivative of the pdf to zero, such that the second derivative is negative, gives us the modal value. The ***mode*** is the value of the rv at which the graph of the distribution reaches its highest point. The ***median*** is obtained by computing the 50th percentile, so that half the distribution is on either side of the median.

Note that the mean, median, mode and standard deviation are all expressed in units of measurement of the rv. If the units are changed by a multiplication factor *a* then the values of all these measures will also be affected accordingly.

A third characteristic of the distribution is the ***skewness*** of the distribution. The distribution may be symmetric or it may be skewed. Just as the first moment E(X) gives us the mean of the distribution and the second moment $E(X^2)$ is used to find the measure of variance, the third moment $E(X^3)$ is used to calculate the measure of skewness. However, a simple inspection of any two of the measures mean, median and mode helps us to identify the shape of the distribution. For a symmetric distribution, mean = median = mode. If the distribution is positively skewed, then mean > median > mode. For a negatively skewed distribution mean < median < mode

Skewness can be measured by the formula $\frac{mean - median}{s\tan dard\ deviation}$. This measure will be zero if the distribution is symmetric, positive for a positively skewed distribution and negative for the negatively skewed distribution.

The moment measure of skewness requires the third moment $E(X^3)$. The moment coefficient of skewness is independent of units of measurement. The formula for the ***moment coefficient of skewness*** is $\frac{E(X-\mu)^3}{\sigma^3}$, where the numerator is the third central moment. This may be expressed in terms of the moments about the origin. $E(X-\mu)^3 = E(X^3) - 3\,E(X^2) + 2[E(X)]^3$. Here too the coefficient will be greater than, less than, or equal to zero if the distribution is positively or negatively skewed or symmetric, respectively. The normal distribution is symmetric.

A fourth characteristic of the distribution is a measure of its peakedness or the degree of flatness near its center. This is measured by the ***coefficient of kurtosis*** and is based on the fourth moment of the distribution. The coefficient of kurtosis,

$$\beta_2 = \frac{E[X-E(X)]^4}{E[X-E(X)]^2} = \frac{E(X^4) - 4E(X^3) + 6E(X^2) - 3[E(X)]^4}{\sigma^4}$$

$$= \frac{4^{th}\ central\ moment}{\left[2^{nd}\ central\ moment\right]^2}$$ [since deviations are taken about E(X) = μ]

The coefficient of kurtosis is 3 for the normal distribution. It is less than 3 for a density function that is flatter than the normal distribution, with short fat tails. If the density function is more peaked than a normal distribution, with long tails, then the coefficient is greater than 3.

The following figure illustrates the three types of distributions. For the purpose of comparison of kurtosis, only symmetric distributions are shown.

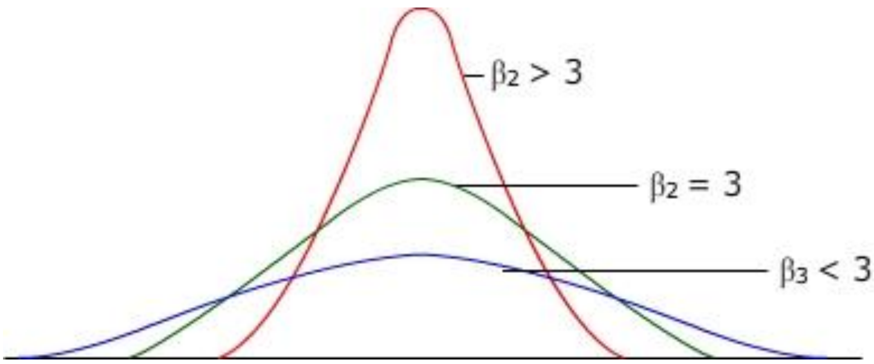

Figure 1: Peakedness of a probability density function: Kurtosis

PRACTICE QUESTIONS

1. Scores earned by students in an economics test (X) is a rv with the pdf

$$f(x) = \begin{cases} \frac{1}{5000}(100 - x) & 0 \le x \le 100 \\ 0 & otherwise \end{cases}$$

The professor announces that to improve the results he will replace each student's marks, X, with new scores, $Y = \sqrt{X}$. Has the professor's strategy been successful in raising the class average above 60?

2. A box is to be constructed so that its height is 5 inches and its base is *Y* inches by *Y* inches, where *Y* is a random variable described by the pdf given below. Find the expected volume of the box.

$$f(y) = \begin{cases} 6y(1-y) & 0 < y < 1 \\ 0 & otherwise \end{cases}$$

3. A tool manufacturing company makes steel gauges. Their annual profit, *Q*, in hundreds of thousands of rupees, can be expressed as a function of product demand, *y*:

$Q(y) = 2(1- e^{-2y})$

Suppose that the demand (in thousands) for their product follows an exponential pdf

$$f(y) = \begin{cases} 6e^{-6y} & y > 0 \\ 0 & otherwise \end{cases}$$

Find the company's expected profit

4. Suppose that *X* is an exponential rv whose pdf is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda y} & y \geq 0 \\ 0 & otherwise \end{cases}$$

Show that the variance of *X* is $1/\lambda^2$

5. The density function of *X* is given by

$$f(x) = \begin{cases} a + bx^2 & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}.$$

If $E(X) = \frac{3}{5}$, find *a* and *b*

6. The median of the distribution of the rv *X*, described by the following pdf, is 3.123

$$f(x) = \begin{cases} \frac{1}{8}(x+1) & 2 < x < 4 \\ 0 & otherwise \end{cases}$$

Is the distribution symmetric? Give reasons for your answer.

7. Suppose the amount of paint, *Y*, in a can of spray paint is a rv with pdf

$$f(y) = \begin{cases} 3y^2 & 0 < y < 1 \\ 0 & otherwise \end{cases}$$

Experience has shown that the largest surface area that can be painted by a can having *Y* amount of paint is twenty times the area that can be generated by a radius of *Y* ft. Can a randomly selected can of spray paint be expected to cover a wall of dimensions 5’ x 8’?

8. A rv *X* is described by the pdf

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

What is the standard deviation of 3X + 2?

9. If the pdf of the rv *Y* is

$$f(y) = \begin{cases} \frac{2y}{k^2} & 0 \le y \le k \\ 0 & otherwise \end{cases}$$

for what value of *k* does V(Y) = 2?

10. The *coefficient of variation* (σ/μ) is a measure of the spread of the distribution that is independent of the unit of measurement of the rv. If the rv *Y* is described by the following pdf,

$$f(y) = \begin{cases} 3(1-y)^2 & 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

what is the coefficient of variation of *X*?