INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -3

Course Instructor : Prof. P. P. MUJUMDAR
Department of Civil Engg., IISc.

Summary of the previous lecture

- Bivariate distributions, Joint pmf and pdf
- Marginal density functions
- Conditional distributions

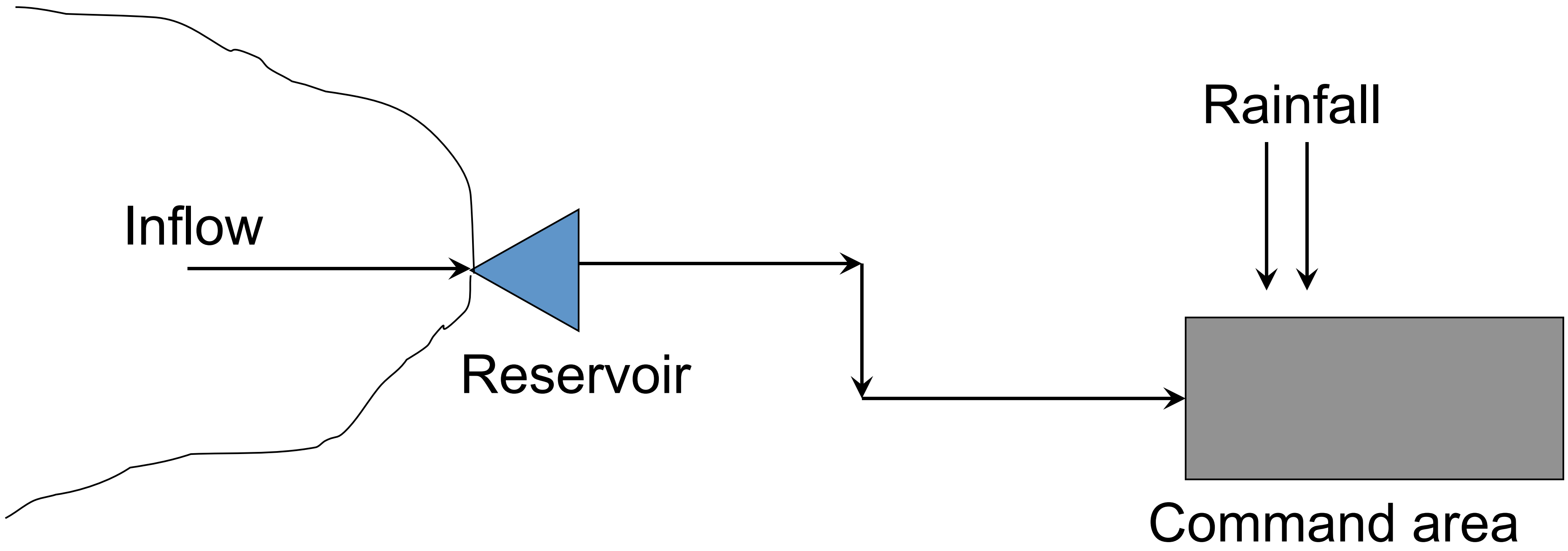
Independent Random Variables

- Intuitively, the rvs X and Y are independent if the distribution of one rv does not in any way influence distribution of the other rv.

- Independence is a useful assumption for hydrologic analysis in many situations. However, there must be a sound physical basis for the assumption.

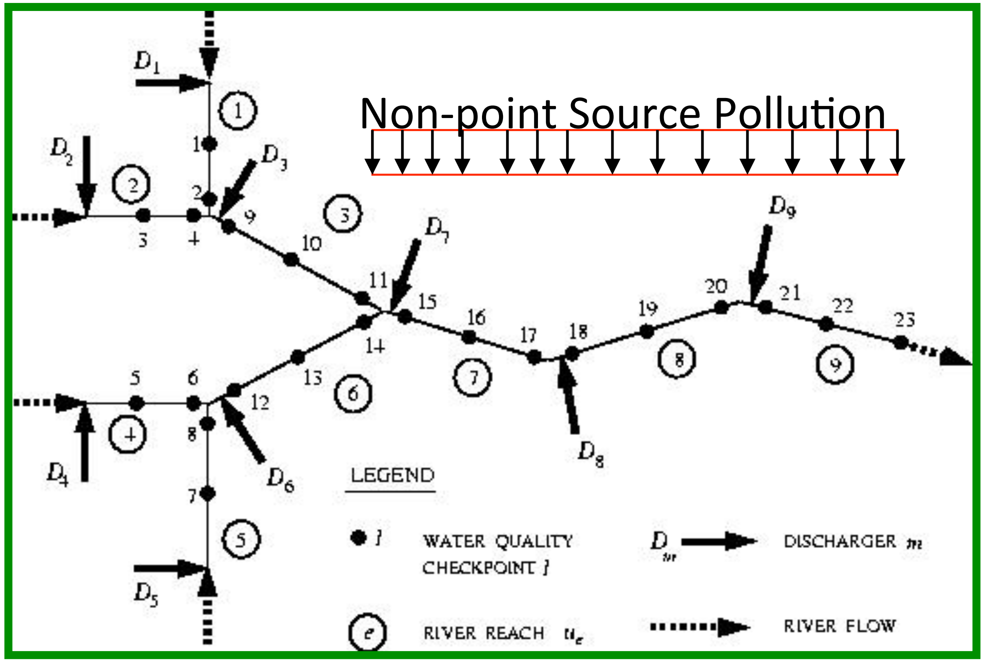
Independent Random Variables

- As an example, inflow to a reservoir (X) and the rainfall in the command area (Y) may be taken as independent, if the command area is far removed from the reservoir.



Independent Random Variables

- In water quality problems, for example, pollutant load (X) and stream flow (Y) may be treated as independent variables. However, stream flow (Y) and water quality indicator, e.g., DO at a location, (Z) are not independent.



Independent Random Variables

- When two rvs are independent, g(x/y)=g(x)
  - Distribution of X given Y is independent of Y and hence the conditional pdf is equalt to the marginal pdf.

$$g\left(x/y\right) = \frac{f(x,y)}{h(y)} \qquad \textbf{h(y) > 0}$$

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

$$f(x,y) = g(x).h(y)$$

Independent Random Variables

- The random variables X and Y are stochastically independent if and only if their joint density is equal to the product of their marginal densities.
- Discrete case: the two r.v.s are independent if and only if

$$p(x_i, y_j) = p(x_i) \cdot p(y_j) \;\forall\; i,j$$

Example-1

Consider the joint pdf

$f(x,y) = x+y \qquad 0 \leq x \leq 1$

$\qquad\qquad\qquad\quad 0 \leq y \leq 1$

$\qquad\quad = 0 \qquad\quad$ elsewhere

For independence of X and Y, the following condition must be satisfied

$$f(x, y) = g(x).h(y)$$

Example-1(contd.)

$$g(x) = \int_0^1 f(x,y)\,dy = \int_0^1 (x+y)\,dy$$

$$= \left[xy + \frac{y^2}{2}\right]_0^1 = x + \frac{1}{2} \qquad 0 \le x \le 1$$

$$h(y) = \int_0^1 f(x,y)\,dx = \int_0^1 (x+y)\,dx$$

$$= \left[\frac{x^2}{2} + xy\right]_0^1 = y + \frac{1}{2} \qquad 0 \le y \le 1$$

Example-1(contd.)

$$g(x)\times h(y)=\left(x+\frac{1}{2}\right)\times\left(y+\frac{1}{2}\right)$$

$$f(x,y)\neq g(x).h(y)$$

Therefore X and Y are not stochastically independent.

Example-2

Consider the joint pdf

$$f(x,y) = e^{-(x+y)} \qquad x > 0,\ y > 0$$

$$= 0 \qquad \text{elsewhere}$$

For independence,

$$f(x,y) = g(x).h(y)$$

Example-2(contd.)

$$g(x) = \int_0^\infty f(x,y)\,dy = \int_0^\infty e^{-(x+y)}\,dy$$

$$= e^{-x}\int_0^\infty e^{-y}\,dy = e^{-x} \qquad x > 0$$

$$h(y) = \int_0^\infty f(x,y)\,dx = \int_0^\infty e^{-(x+y)}\,dx$$

$$= e^{-y}\int_0^\infty e^{-x}\,dx = e^{-y} \qquad y > 0$$

Example-2 (contd.)

$$g(x)\times h(y) = e^{-x}\times e^{-y}$$
$$= e^{-(x+y)}$$

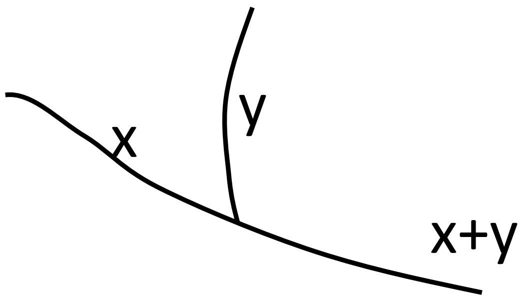
$$f(x,y) = g(x).h(y)$$

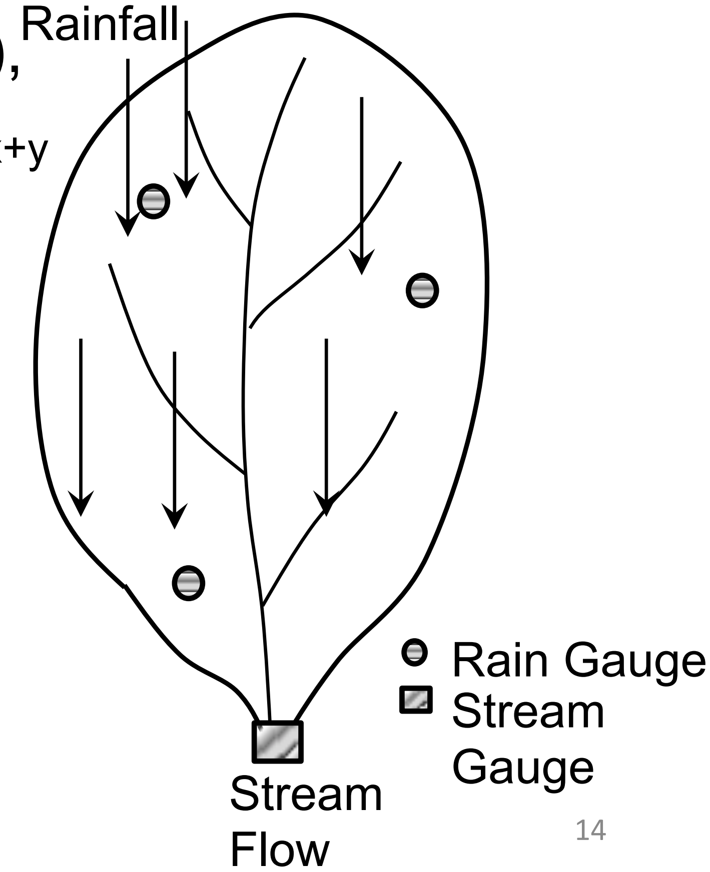
Therefore X and Y are stochastically independent

Functions of Random Variable

- Situations often arise when we will be interested in the distributions of functions of r.v.s. For example,

  Given the joint distribution f(x, y), we will be interested in getting $f_{x+y}(x+y)$





Functions of Random Variable

- X: discrete; Y = H(X), a function of X.
- The pmf of X is known
- Enumerate possible values of Y for the discrete values of X
- Then obtain the probabilities of the possible values of Y from the probabilities of the corresponding values of X.

Example for discrete case

$p(x) = 60/77x$; $x = 2, 3, 4, 5$

discrete values

$y = x^2-7x+5$

x	2	3	4	5
y	-5	-7	-7	-5
p(x)	30/77	20/77	15/77	12/77

Distribution of y:

$p(Y=-7) = p(X=3)+p(X=4) = 20/77+ 15/77 = 35/77 = 5/11$

$p(Y=-5) = p(X=2)+p(X=5) = 30/77+ 12/77 = 42/77 = 6/11$

General procedure for functions of continuous random variables:

X → continuous, Y=H(X) → continuous function of X

We are interested in getting the pdf g(y).

a. Obtain G, the cdf of 'Y', where G(y) = P [Y $\le$ y] by finding the event in the range space of 'X' which is equivalent to the event Y $\le$ y.

$$Y = 2X + 5$$

$$Y \le y$$

$$2X + 5 \le y$$

$$X \le \frac{y-5}{2}$$

$$P\left[Y \le y\right] = P\left[X \le \frac{y-5}{2}\right]$$

Given f(x), you will get P [Y $\le$ y]

General procedure for functions of continuous random variables:

b. Differentiate G(y) w.r.t 'y' to get g(y)

c. Since g(y) must be non-negative, determine those values of y over which g(y) $\geq$ 0

   and check,

$$\int_{-\infty}^{\infty} g(y)dy = 1$$

Example-1

The rv X has a pdf

$f(x) = x/2 \qquad 0 \le x \le 2$

$\quad\;\; = 0 \qquad \text{elsewhere}$

Let H(X) = 4X+1

Find the pdf of Y=H(X)

a. Get the CDF of Y

$G(y) = P[Y \le y]$

$\quad = P[4X+1 \le y]$

$$= P\left[X \le \frac{y-1}{4}\right]$$

Example-1(contd.)

$$G(y) = \int_0^{(y-1)/4} f(x)dx$$

$$= \int_0^{(y-1)/4} \frac{x}{2}\, dx$$

$$= \left[\frac{x^2}{4}\right]_0^{(y-1)/4}$$

$$= \frac{(y-1)^2}{64}$$

Example-1(contd.)

b. g(y) = $\frac{dG(y)}{dy} = \frac{d}{dy}\left(\frac{(y-1)^2}{64}\right)$

$$= \frac{2}{64}(y-1) = \frac{y-1}{32}$$

c. From 0 $\leq$ x $\leq$ 1, we get

$$g(y) = \frac{(y-1)}{32} \quad 1 < y < 9$$

y=4x+1

Example-1(contd.)

Check : $\int_1^9 \frac{(y-1)}{32} dy = 1$

$$\int_1^9 \frac{(y-1)}{32} dy = \frac{1}{32}\left[\frac{(y-1)^2}{2}\right]_1^9$$

$$= \frac{1}{64}(8^2 - 0)$$

$$= 1$$

Example-2

The rv X has a pdf

$$f(x) = 3e^{-3x} \qquad 0 \le x \le \infty$$
$$= 0 \qquad \text{elsewhere}$$

Let $H(X) = e^X$

To find the pdf of $Y=H(X)$

a. Get the CDF of Y

$$G(y) = P\,[Y \le y]$$
$$= P\,[e^X \le y]$$
$$= P\,[X \le \ln y]$$

Example-2(contd.)

$$G(y) = \int_0^{\ln y} f(x)dx$$

$$= \int_0^{\ln y} 3e^{-3x} \; dx$$

$$= \left[\frac{3e^{-3x}}{-3} \right]_0^{\ln y}$$

$$= -e^{-3\ln y} - (-1)$$

$$= 1 - e^{\ln y^{-3}}$$

$$= 1 - y^{-3}$$

Example-2(contd.)

b. g(y) = $\dfrac{dG(y)}{dy} = \dfrac{d}{dy}\left(1 - y^{-3}\right)$

$$= 0 - \left(-3y^{-4}\right)$$

$$= 3y^{-4}$$

c. From 0 $\leq$ x $\leq$ ∞, we get

$$g(y) = 3y^{-4} \qquad 1 < y < \infty$$

Example-2 (contd.)

Check $$\int_1^{\infty} 3y^{-4}\,dy = 1$$

$$\int_1^{\infty} 3y^{-4}\,dy = \left[\frac{3(y)^{-3}}{-3}\right]_1^{\infty}$$

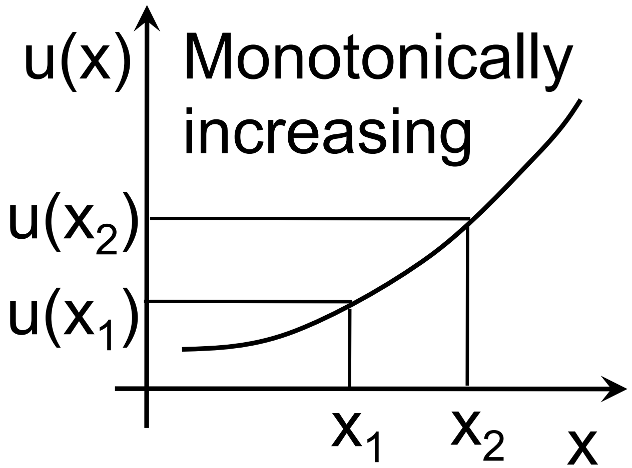
$$= \left(\infty^{-3} - (-1)\right)$$

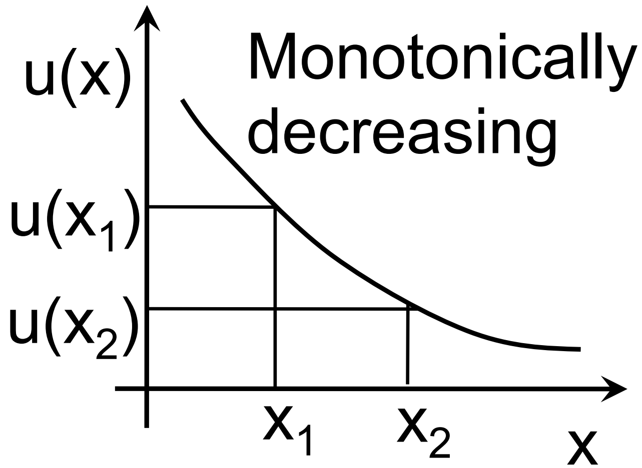
$$= 1 - \frac{1}{\infty}$$

$$= 1 - 0 = 1$$

Generalization for monotonous function

- $u(x)$ is a monotonically increasing function of 'x' if $u(x_2)>u(x_1)\ \forall\ x_2>x_1$ (as 'x' increases, $u(x)$ increases)
- $u(x)$ is a monotonically decreasing function of 'x' if $u(x_2)<u(x_1)\ \forall\ x_2>x_1$ (as 'x' increases, $u(x)$ decreases)





Generalization for monotonous function

- Let ‘X’ be a continuous rv with pdf f(x), where f(x)≥0 for a<x<b.
- Suppose that Y=H(X) is a strictly monotonic (increasing or decreasing) function of ‘X’.
- If this function is differentiable and continuous for all ‘x’, then the rv Y = H(X) has a pdf g(y) given by

$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$

‘x’ and f(x) are expressed in terms of ‘y’

Example-2

Consider the previously solved example-2

$f(x) = 3e^{-3x}$ $\quad 0 \leq x \leq \infty$

$= 0$ $\quad$ elsewhere

Let $H(X) = e^X$

Find the pdf of $Y=H(X)$

'x' is expressed in terms of 'y'

$y = e^X$

$x = \ln y$

$$\frac{dx}{dy} = \frac{1}{y}$$

Example-2(contd.)

$$
\begin{aligned}
f(x) &= 3e^{-3x} \\
&= 3e^{-3\ln y} \\
&= 3e^{\ln y^{-3}} \\
&= 3y^{-3}
\end{aligned}
$$

$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$

$$
\begin{aligned}
g(y) &= 3y^{-3} \times \frac{1}{y} \\
&= 3y^{-4}
\end{aligned}
$$

Example-2(contd.)

Y is a monotonically increasing function of X

$Y = e^X$

as 'x' tends to 0, 'y' tends to 1

'x' tends to ∞, 'y' tends to ∞

Therefore $g(y) = 3y^{-4} \qquad 1 \le y \le \infty$

as obtained earlier.

Functions of two dimensional RVs

- In the case of a continuous bivariate r.v., the transformation from f(x,y) to g(u,v), where $U=H_1(X, Y)$ and $V=H_2(X, Y)$ are one-to-one continuously differentiable transformation is given by

$$g(u,v) = f(x,y)\left|J(u,v)\right|$$

J(u, v) is the Jacobian of the transformation, given by

$$J(u,v) = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix}$$

'x', 'y' and f(x, y) are expressed in terms of 'u' and 'v'

Example-3

Consider the joint pdf

$$f(x,y) = \frac{3}{2}\left(x^2 + y^2\right) \quad \begin{matrix} 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \end{matrix}$$

U = X+Y and V=Y/2 what is the joint pdf of (u, v)

'x' and 'y' are expressed in terms of 'u' and 'v'

y = 2v

x = u-2v

Example-3(contd.)

$$f(x,y) = \frac{3}{2}\left(x^2 + y^2\right)$$

$$= \frac{3}{2}\left(\left(u-2v\right)^2 + \left(2v\right)^2\right)$$

$$= \frac{3}{2}\left(u^2 - 4uv + 8v^2\right)$$

$$\frac{1}{J} = \begin{vmatrix} \frac{du}{dx} & \frac{dv}{dx} \\ \frac{du}{dy} & \frac{dv}{dy} \end{vmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1/2 \end{pmatrix} = 1/2$$

Or J = 2

Example-3(contd.)

$$g(u,v) = f(x,y)\left|J(u,v)\right|$$

$$= \frac{3}{2}\left(u^2 - 4uv + 8v^2\right) \times 2$$

$$= 3\left(u^2 - 4uv + 8v^2\right)$$

U = X+Y $0 \leq x \leq 1$

V=Y/2 $0 \leq y \leq 1$

Limits:

y=0, v=0 ; y=1, v=1/2

x = 0, u=2v ; x = 1, u=1+2v

Example-3(contd.)

$$g(u,v) = 3\left(u^2 - 4uv + 8v^2\right) \quad \begin{array}{l} 0 \le v \le 1/2 \\ 2v \le u \le 1+2v \end{array}$$

From this joint distribution, we may obtain the marginal distributions of 'u' and 'v' by integrating over the other variable.

Example-3(contd.)

- In some cases only distribution of U=u(x, y) is desired.
- In such case define a dummy r.v. V=v(x,y),
- find the joint pdf g(u, v) and then integrate over 'v' to get the marginal density of 'u'

Consider the previous joint pdf

$$f(x,y) = \frac{3}{2}\left(x^2 + y^2\right) \quad \begin{array}{l} 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \end{array}$$

U = X+Y and define a dummy variable 'V' as V=Y

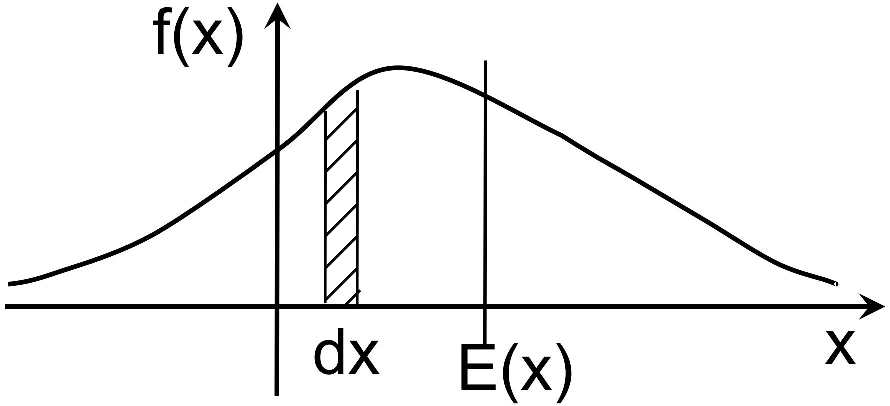
Moments of a distribution

- Population: All possible values of a r.v.
  - E.g., If a r.v. is defined as a page in a book, all pages in the book together constitute the population
- Sample: A subset of population
  - E.g., a chapter in the book
- Realization: A (time)series of the r.v. actually realized
- Observation: A particular value of the r.v. in the realization.

Moments of a distribution

nth moment about the origin

$$\mu_n^0 = \int_{-\infty}^{\infty} x^n f(x)dx$$



E(X): Expected value of ‘X’

: First moment about the origin

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

nth moment about the expected value

$$\mu_n = \int_{-\infty}^{\infty} \left(x-\mu\right)^n f(x)dx$$

Expected value:

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

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

E(cX) = cE(X)

E[c.g(X)] = c. E[g(X)]

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

$E[g_1(X) \pm g_2(X)] = E[g_1(X)] \pm E[g_2(X)]$

Measures of central tendency

Mean:

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

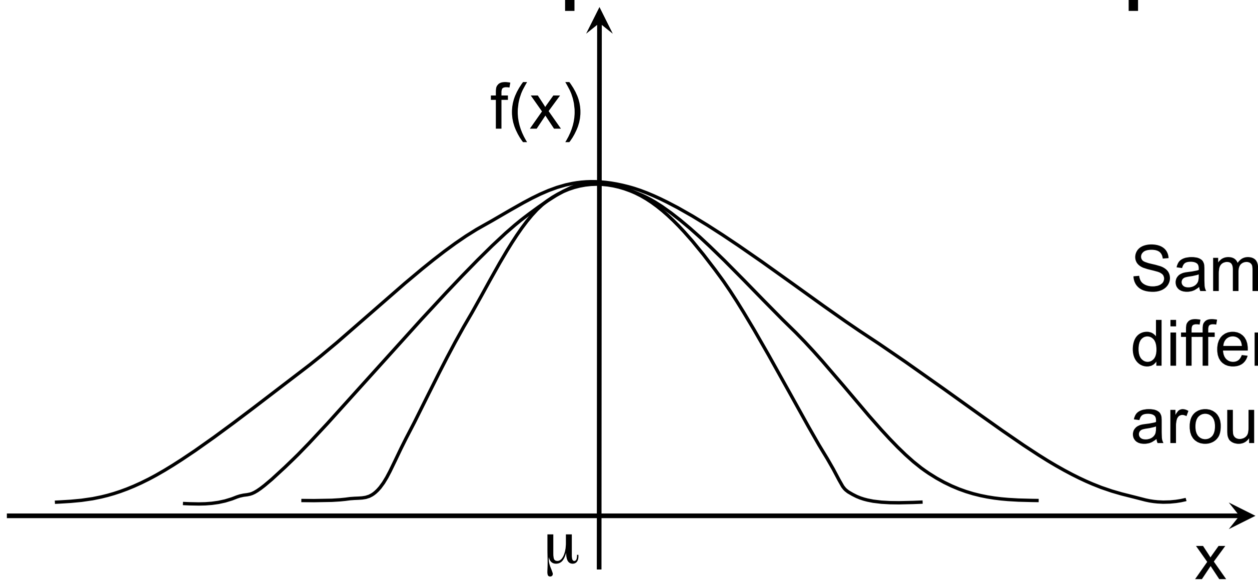
Discrete case: $\mu = \sum_{i=1}^{n} x_i p(x_i)$ n: Sample space

Sample estimate: $\bar{x} = \frac{\sum x_i}{n}$

Mode: Value with highest frequency of occurrence

Median: Value such that 50% of area is on either side

Measures of spread or dispersion



Same mean but different spread around the mean

Range:

$x_{max} - x_{min}$

Variance: Second moment about the mean

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

Measures of spread or dispersion

Sample estimate:

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$$

n: No. of observations in the sample

Standard deviation:

$$\sigma = +\sqrt{\sigma^2}$$

Positive squareroot

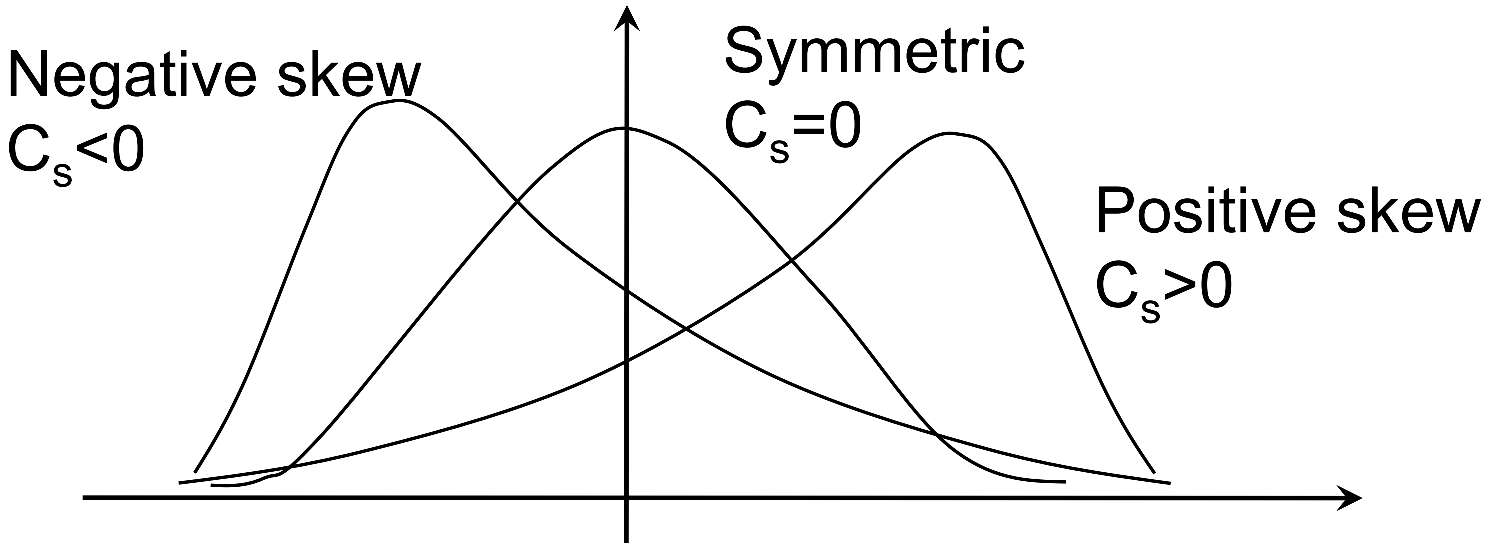
$$s = +\sqrt{s^2}$$

Coefficient of variation:

$$c_r = \frac{\sigma}{\mu}$$ -- Population

$$= \frac{s}{\bar{x}}$$ -- sample space

Measures of symmetry



Coefficient of skewness:

Population

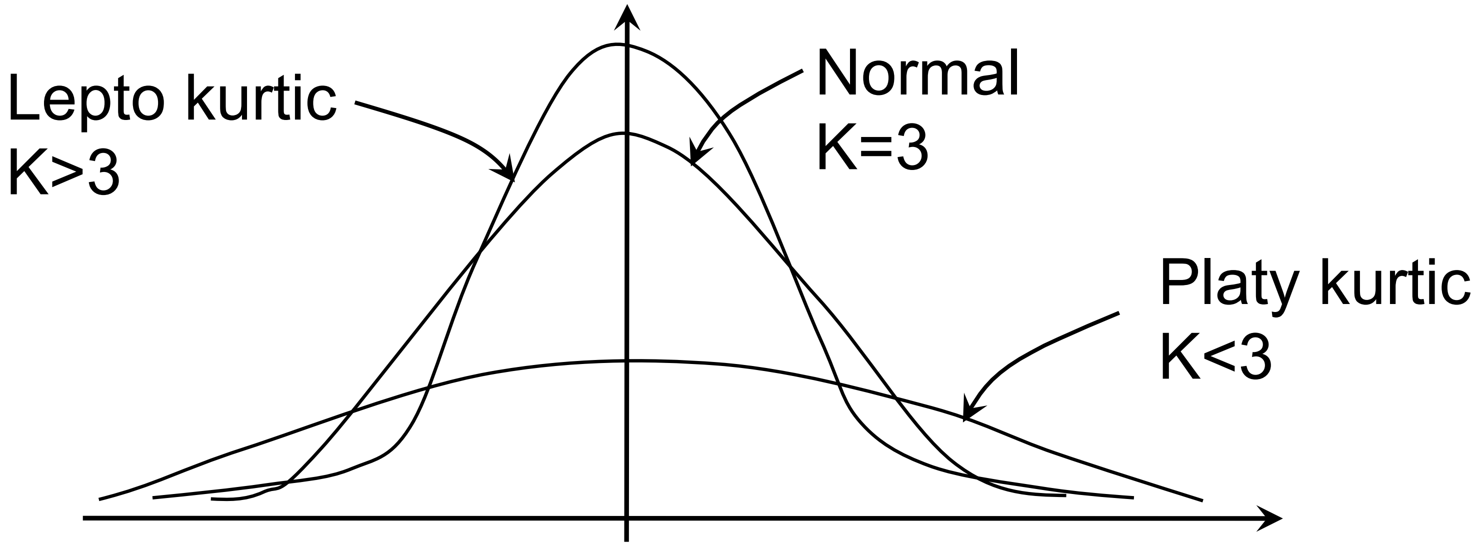
$$\gamma_s = \frac{\mu_3}{\mu_2^{3/2}}$$

$$= \frac{\int_{-\infty}^{\infty} (x-\mu)^3 f(x)dx}{(\sigma^2)^{3/2}}$$

Sample

$$C_s = \frac{n\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{(n-1)(n-2)s^3}$$

Measures of Peakedness



Coefficient of kurtosis:

Population

$$K = \frac{\mu_4}{\mu_2^2}$$

Sample

$$K = \frac{n^2 \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^4}{(n-1)(n-2)(n-3)s^4}$$

Example-1

Consider ‘v’ as wind velocity and pdf is given as

$$f(v) = \frac{1}{5} \qquad 0 \le v \le 5$$

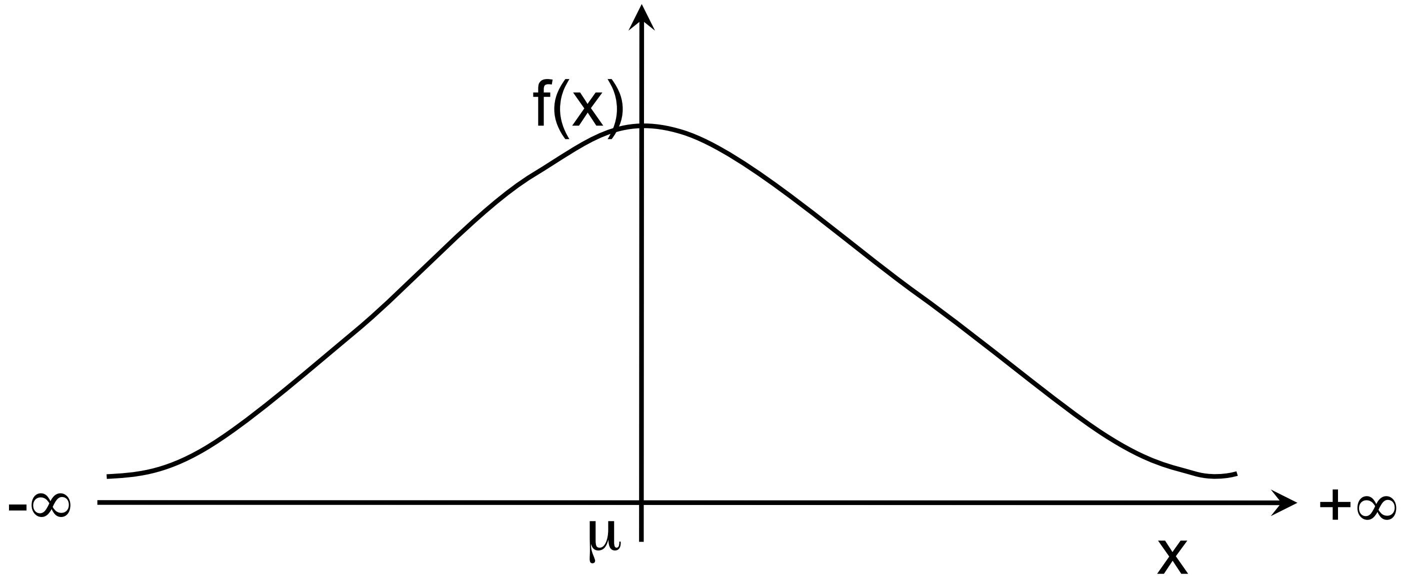
And Pressure w = 0.12v^2, obtain E[w]

E[w] = $\int_0^5 0.12v^2 \times \frac{1}{5} dv$

$$= \frac{0.12}{5} \times \left[\frac{v^3}{3} \right]_0^5 = 1$$

COMMONLY USED DISTRIBUTIONS

Normal Distribution



$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad -\infty < x < +\infty$$

Two parameters, μ & σ

$X \sim N(\mu, \sigma^2)$

F(x) approaches zero as x $\rightarrow \pm\infty$

Normal Distribution

Coefficient of skewness, $\gamma_s = 0$

Kurtosis coefficient, K = 3

y = a + bx - Linear form of 'x'

$y \sim N(a+b\mu,\ b^2\sigma^2)$

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

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \qquad -\infty < x < +\infty$$

Normal Distribution

z = $\dfrac{x-\mu}{\sigma}$ -- Linear form

$$a = \frac{-\mu}{\sigma}, b = \frac{1}{\sigma}$$

$$z : N\left[\frac{-\mu}{\sigma} + \frac{\mu}{\sigma}, \frac{1}{\sigma^2} \times \sigma^2\right]$$

$$: N(0,1)$$

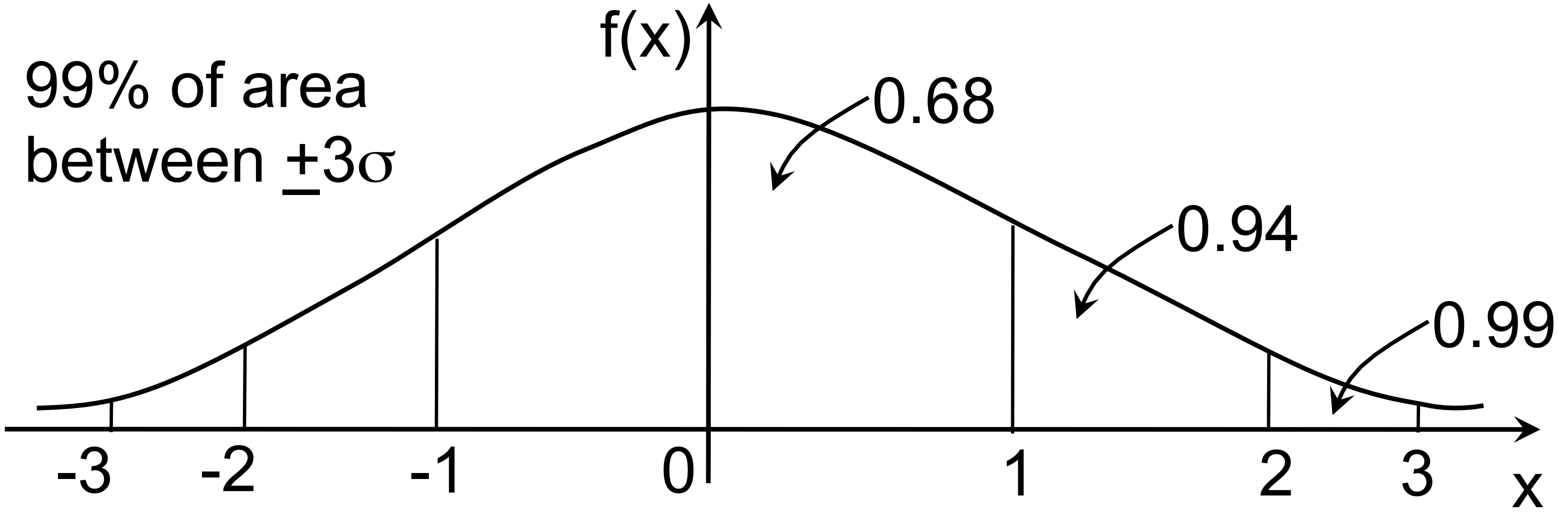
Pdf

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < +\infty$$

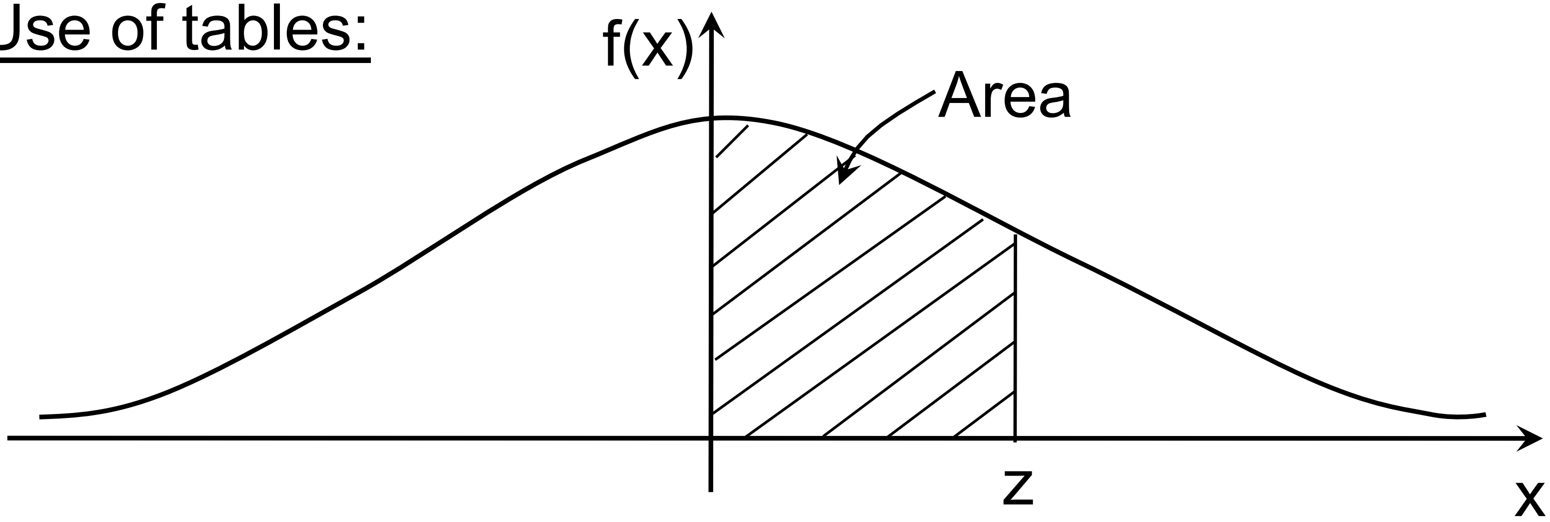
cdf of z

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz$$

Normal Distribution



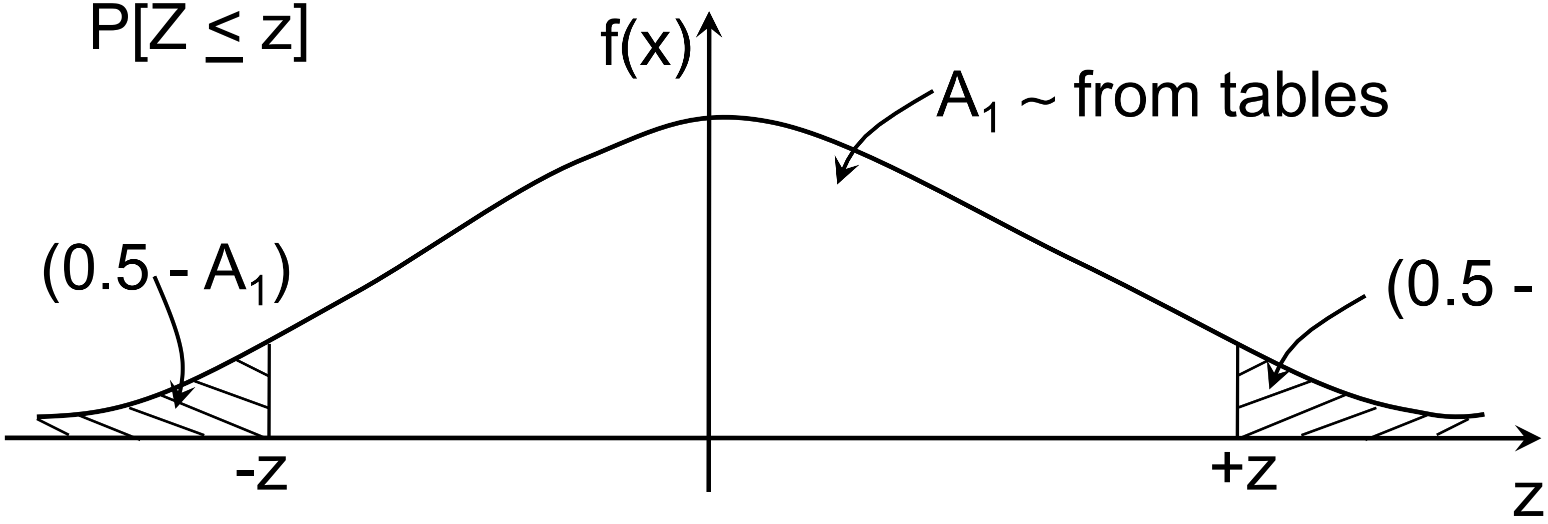
Use of tables:



$P[Z \leq z] = 0.5 + \text{Area from table}$

Normal Distribution

$P[Z \leq z]$



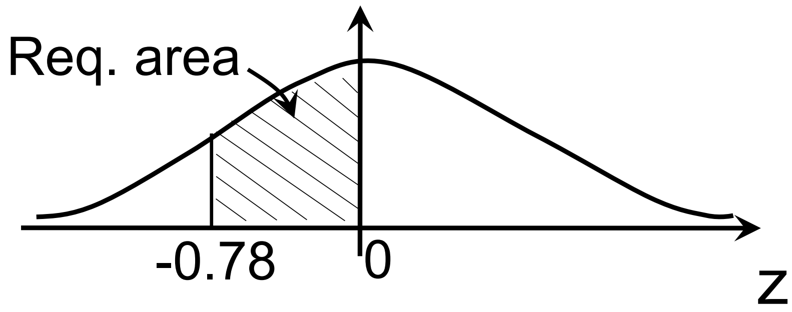
from table

e.g., $P[Z \leq -0.7] = 0.5 - 0.258$

$= 0.242$

Examples on Normal distribution

Obtain the area under the standard normal curve between -0.78 and 0



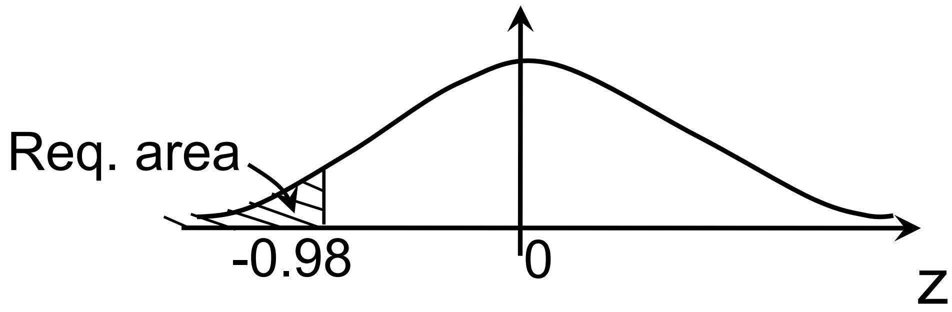
Req. area = area between 0 and +0.78

= 0.2823

Examples on Normal distribution

Obtain the area under the standard normal curve

$z \leq -0.98$



Req. area = 0.5 – area between 0 and +0.98

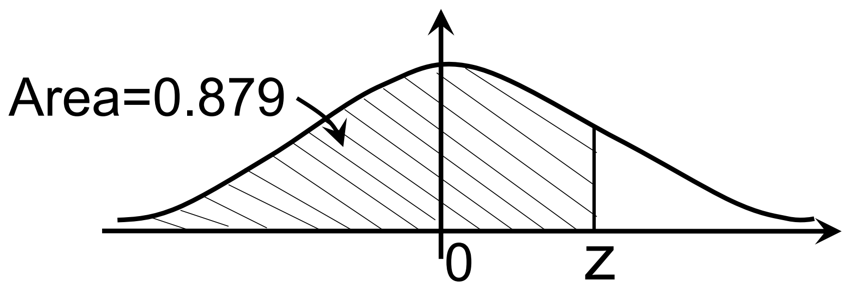
= 0.5 – 0.3365

= 0.1635

Examples on Normal distribution

Obtain 'z' such that P[Z $\leq$ z]=0.879

Since the value is greater than 0.5, 'z' must be +ve



area between 0 to z = 0.5 – 0.879

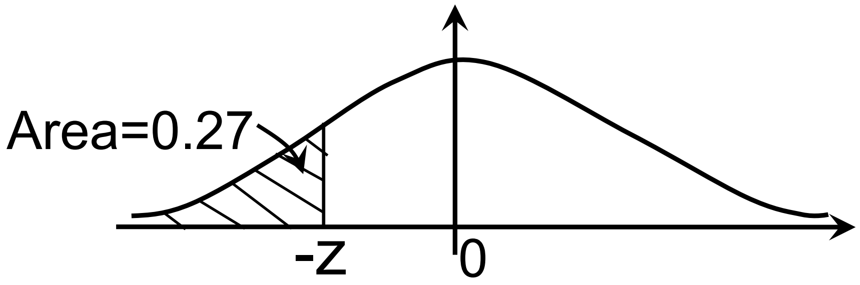
= 0.379

From the table, for the area of 0.379, corresponding z = 1.17

Examples on Normal distribution

Obtain 'x' such that $P[X \geq x]=0.73$ if $\mu=650$; $\sigma = 200$

$P[X \leq x]=0.27$



$P[Z \leq z]=0.27$

area between 0 to -z = 0.5 – 0.27

= 0.23

From the table, z = -0.61

$$z = \frac{x-\mu}{\sigma} \Rightarrow -0.61 = \frac{x-650}{200} \rightarrow x = 528$$