INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -4

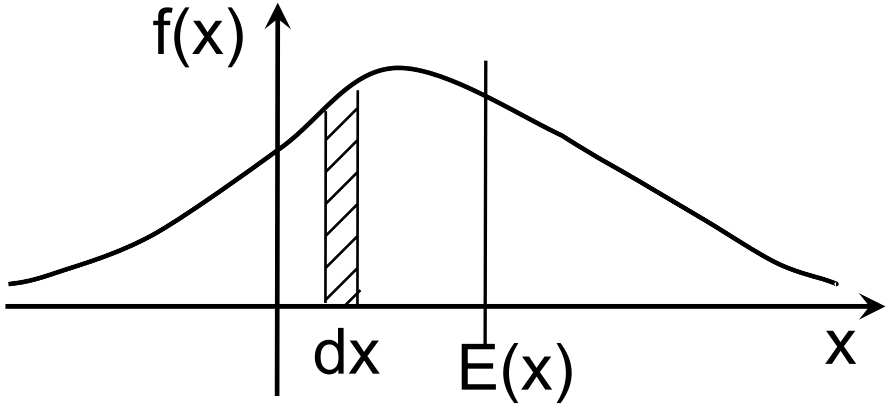
Course Instructor : Prof. P. P. MUJUMDAR

Department of Civil Engg., IISc.

Summary of the previous lecture

- Independent random variables
- Functions of random variables
- Moments of a distribution
- Expected value

Moments of a distribution



n^{th} moment about the origin

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

E(X): Expected value of 'X'

: First moment about the origin

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

n^{th} moment about the expected value

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

Measures of central tendency

Mean:

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

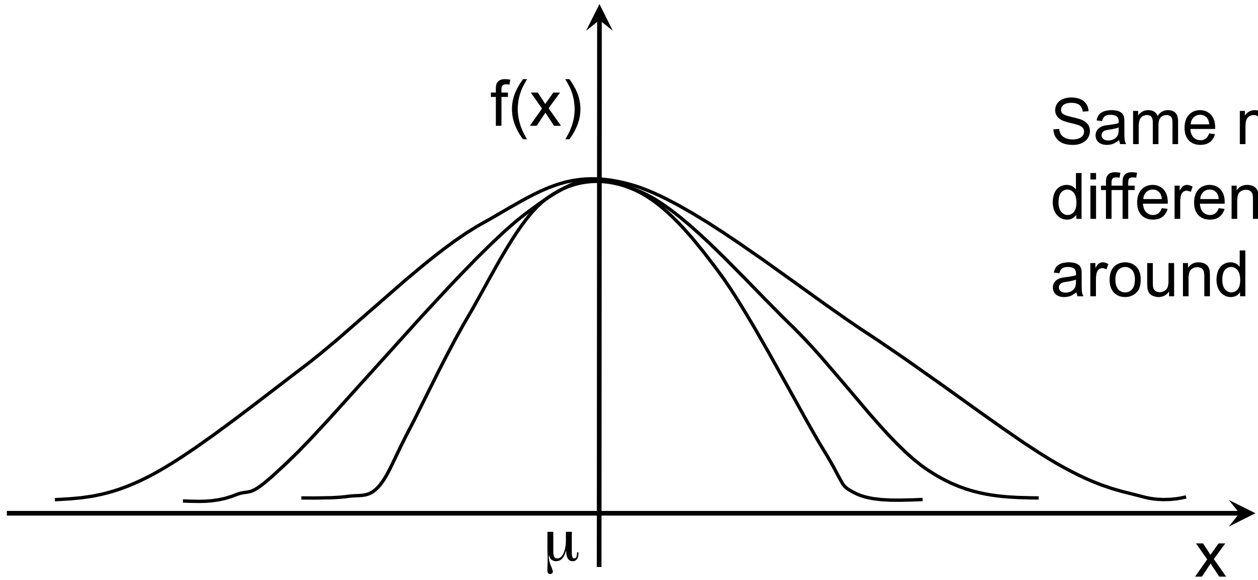
Discrete case: $\mu = \sum_{i=1}^{n} x_i p(x_i)$ n: Sample size

Sample estimate of the mean: $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

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:

$x_{max} - x_{min}$; x_{max} : maximum value of X

x_{min} : minimum value of X

→ 0, 10, 20
9, 10, 11

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 - Variance :

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

n: No. of observations in the sample

Standard deviation:

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

Positive squareroot

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

Coefficient of variation:

$$c_v = \frac{\sigma}{\mu}$$

Population

$$= \frac{s}{\overline{x}}$$

sample space

$$\sigma^2 = E[x - \mu]^2$$

Capital X

$$= E[x^2 - 2x\mu + \mu^2]$$

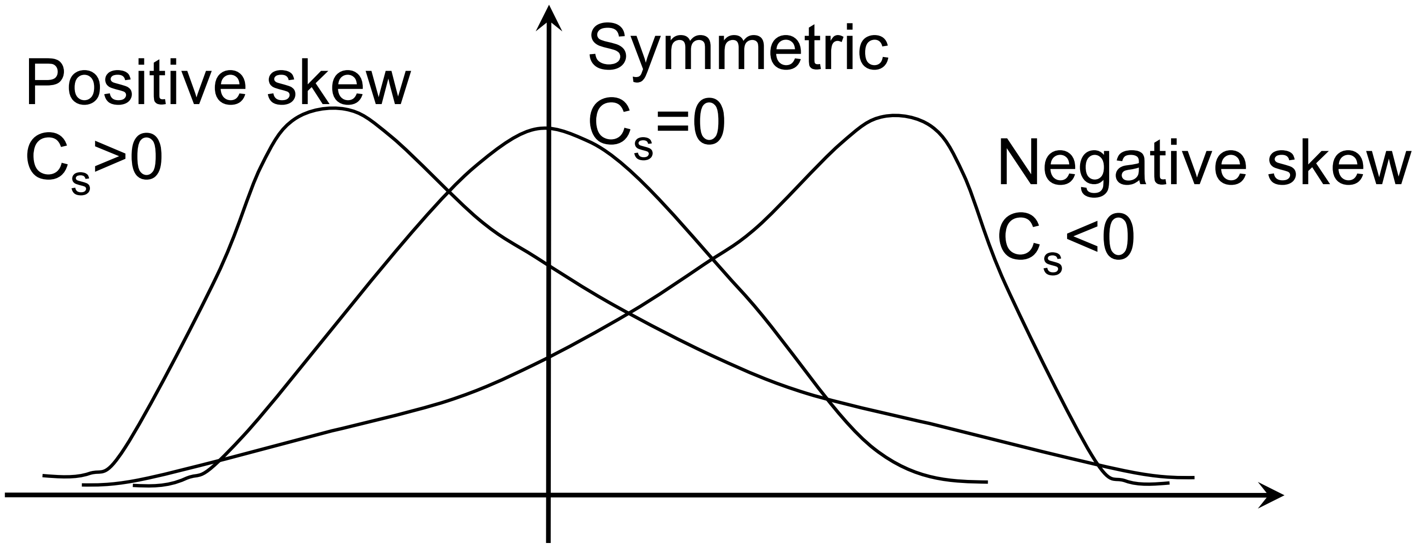
$$= E[x^2] - 2\mu E[x] + E[\mu^2]$$

$$= E[x^2] - 2\mu^2 + \mu^2$$

$$= E[x^2] - \mu^2$$

$$= E[x^2] - \{E[x]\}^2$$

Measures of symmetry



Coefficient of skewness:

Population

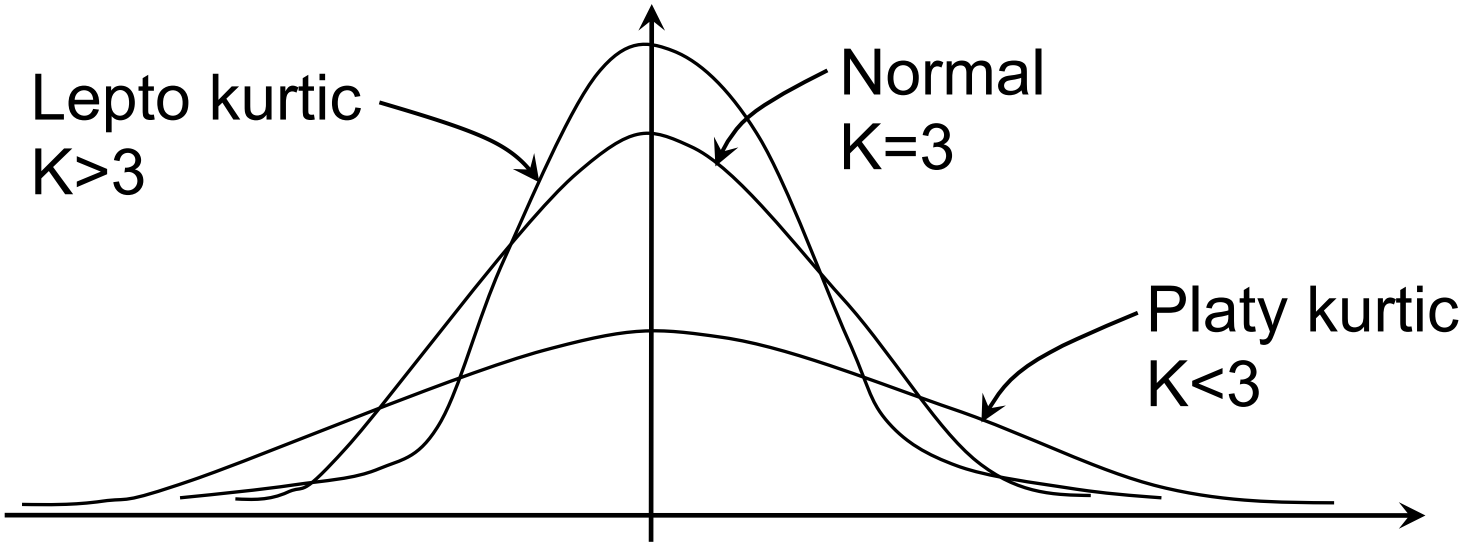
$$\gamma_s = \frac{\mu_3}{\mu_2^{3/2}}$$

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

Sample Estimate

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

Measures of Peakedness



Kurtosis:

Population

$$K = \frac{\mu_4}{\mu_2^2} = \frac{\int_{-\infty}^{\infty} (x-\mu)^4 f(x)dx}{(\sigma^2)^2}$$

Sample Estimate

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

Example-1

Consider the pdf

$$f(x) = 3x^2 \qquad 0 \leq x \leq 1$$
$$= 0 \qquad \text{elsewhere}$$

Obtain

1. E(X)
2. E(3X-2)
3. $E(X^2)$

Example-1(contd.)

$\sigma^2 = E(x^2) - \{E(x)\}^2$

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

2. E(3X-2) $= \int_{-\infty}^{\infty} (3x-2) f(x)\,dx = \int_0^1 (3x-2).3x^2\,dx$

$$= \int_0^1 \left(9x^3 - 6x^2\right)dx = \left[\frac{9x^4}{4} - 2x^3\right]_0^1 = \frac{1}{4}$$

3. E(X²) $= \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2.3x^2\,dx = 3\left[\frac{x^5}{5}\right]_0^1 = \frac{3}{5}$

Example-2

Obtain the sample estimates of mean, standard deviation, coefficient of variation, coefficient of skewness and kurtosis for the following observed data of annual stream flow for 15 years.

Year	1	2	3	4	5	6	7	8	9	10
Avg. yearly stream flow (Mm^3)	150	129	160	152	165	138	149	115	97	154

Year	11	12	13	14	15
Avg. yearly stream flow (Mm^3)	168	110	108	105	125

Example-2 (contd.)

Mean, $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sum_{i=1}^{n} x_i$$ = 150+129+160+152+165+138+149+115+97+154+ 168+110+108+105+125

= 2025

Therefore mean, $\bar{x}$ = 2025/15

= 135 Mm^3

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

Year	Avg. Stream flow $Mm^3 (x_i)$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
1	150	15	225	3375	50625
2	129	-6	36	-216	1296
3	160	25	625	15625	390625
4	152	17	289	4913	83521
5	165	30	900	27000	810000
6	138	3	9	27	81
7	149	14	196	2744	38416
8	115	-20	400	-8000	160000
9	97	-38	1444	-54872	2085136
10	154	19	361	6859	130321
11	168	33	1089	35937	1185921
12	110	-25	625	-15625	390625
13	108	-27	729	-19683	531441
14	105	-30	900	-27000	810000
15	125	-10	100	-1000	10000
Σ	**2025**	**0**	**7928**	**-29916**	**6678008**

Example-2 (contd.)

Variance, $s^2 = \frac{7928}{15-1} = 566$ $\quad \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

Standard deviation, S $= +\sqrt{s^2} = 23.8$ Mm³

Coefficient of variation, C_v = S/$\bar{x}$ =23.8/135 =0.176

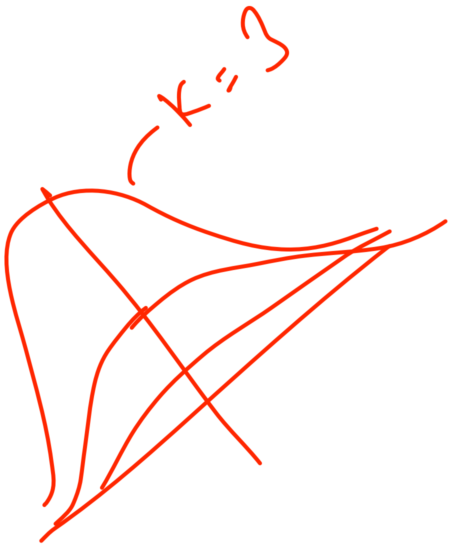
Coefficient of skewness, $C_s = \frac{n\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{(n-1)(n-2)s^3}$

$$= \frac{15 \times (-29916)}{(15-1)(15-2)23.8^3}$$

$= -0.183$ <0, negatively skewed

Example-2 (contd.)

Coefficient of Kurtosis, $k = \dfrac{n^2 \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^4}{(n-1)(n-2)(n-3)s^4}$



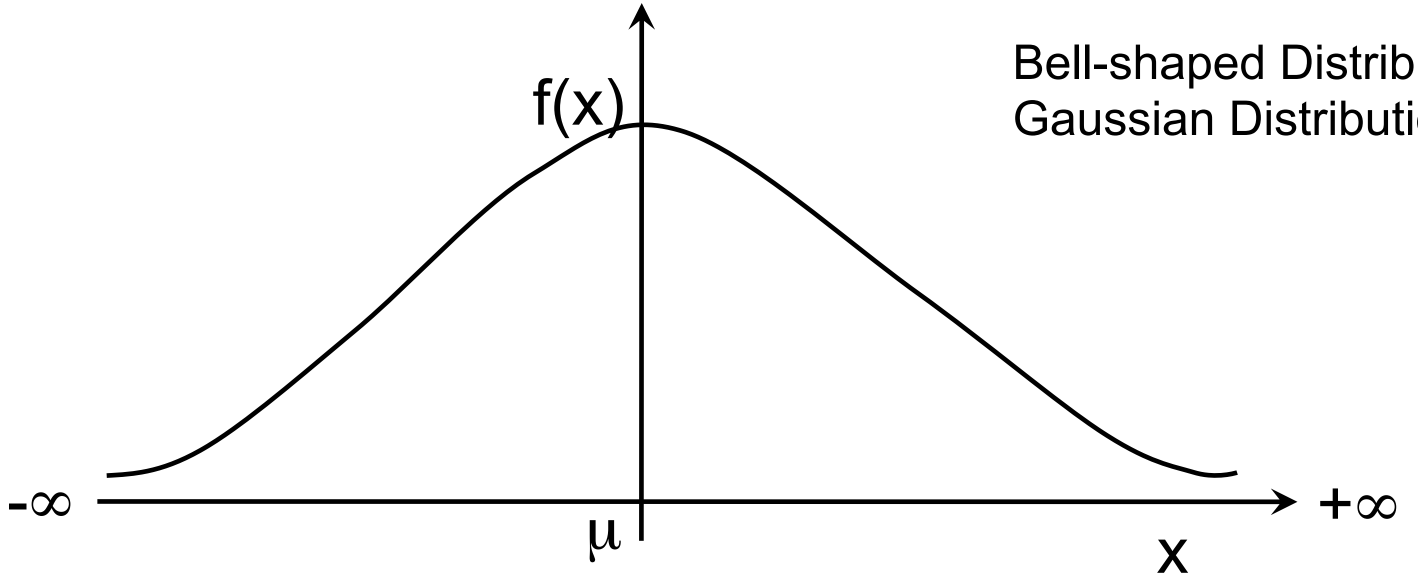
$$= \frac{15^2 \times 6678008}{(15-1)(15-2)(15-3)23.8^4}$$

$$= 2.14$$

< 3, Platy kurtic

COMMONLY USED DISTRIBUTIONS

Normal Distribution



Bell-shaped Distribution;
Gaussian Distribution

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

Symmetric about x = μ

Two parameters, μ & σ

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

f(x) approaches zero as x → $\pm\infty$

Normal Distribution

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

Coefficient of skewness, $\gamma_s = 0$

Kurtosis, K = 3

Y = a + bX – Linear function 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

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

-- Linear function

Y = a + bX

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

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

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

$$: N(0,1)$$

$F(z) = P[Z \le z]$

pdf of z

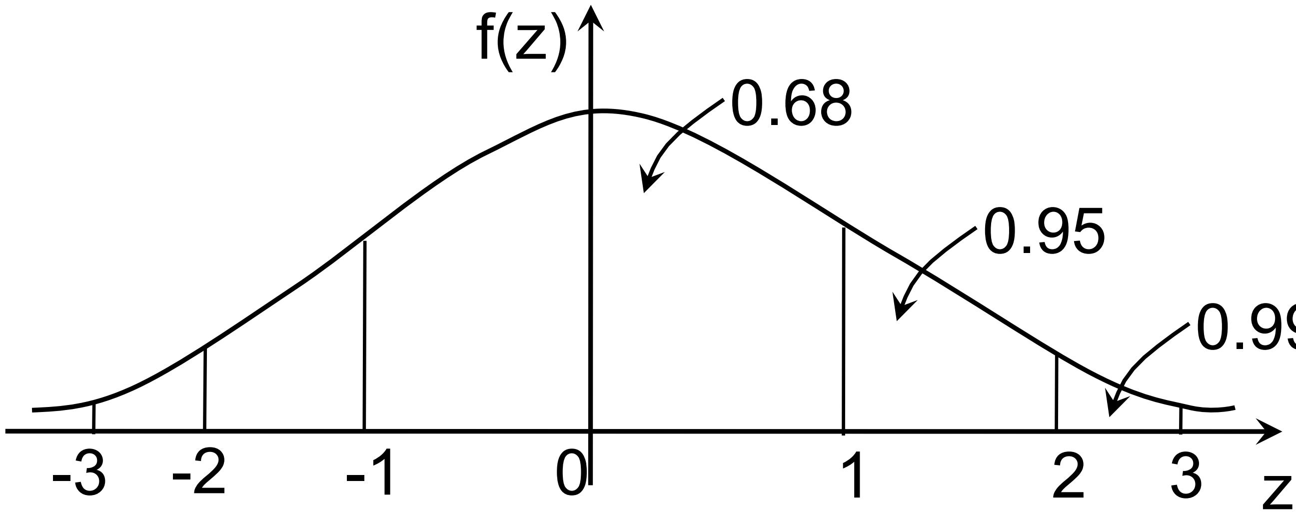
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < +\infty$$

cdf of z

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

Normal Distribution

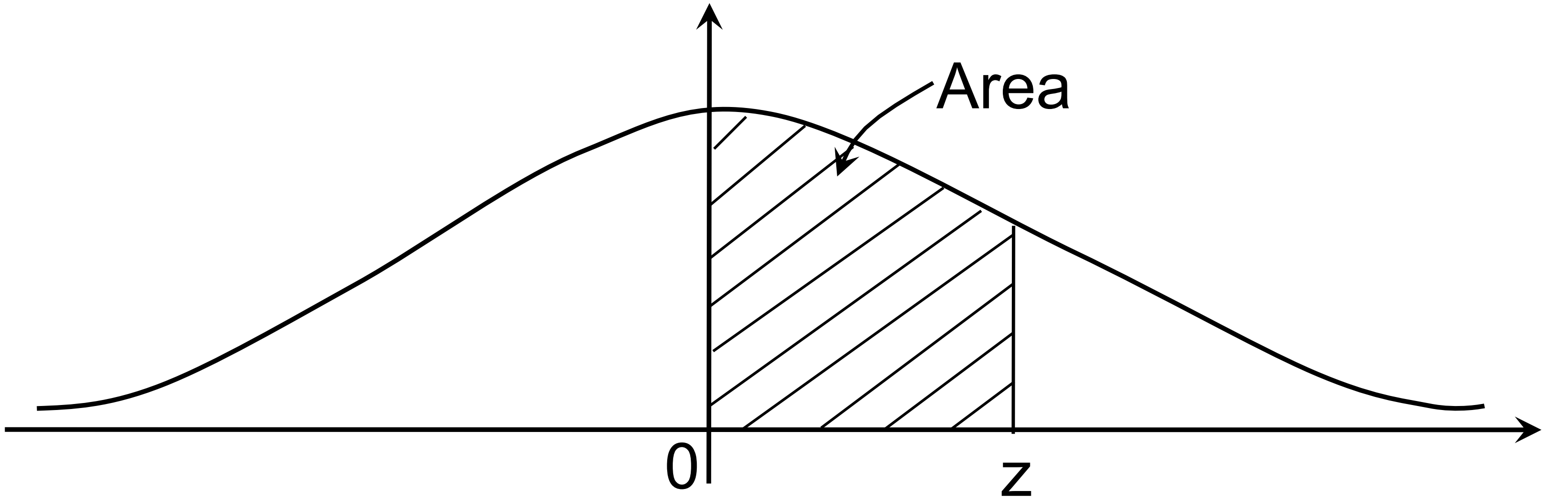
- f(z) is referred as standard normal density function
- The standard normal density curve is as shown
- 99% of area lies between $\underline{+}3\sigma$



- f(z) cannot be integrated analytically by ordinary means
- Methods of numerical integration used
- The values of F(z) are tabulated.

Normal Distribution

Obtaining standard variate 'z' using tables:



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

Normal Distribution Tables

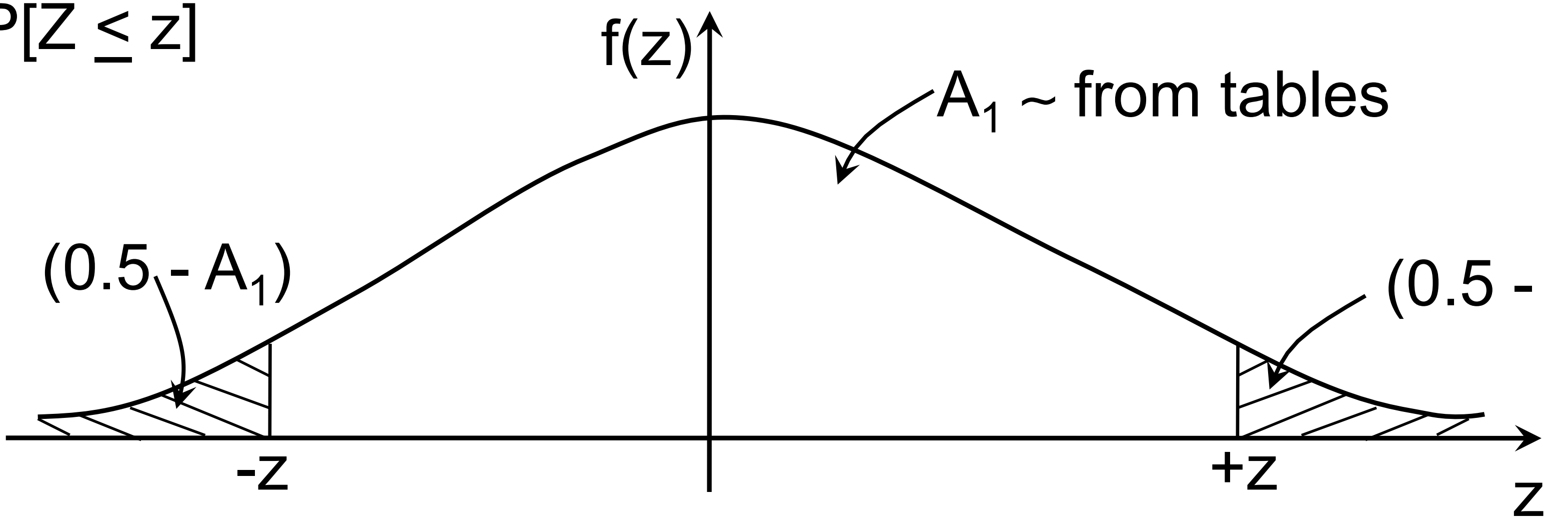
z	0	2	4	6	8
0	0	0.008	0.016	0.0239	0.0319
0.1	0.0398	0.0478	0.0557	0.0636	0.0714
0.2	0.0793	0.0871	0.0948	0.1026	0.1103
0.3	0.1179	0.1255	0.1331	0.1406	0.148
0.4	0.1554	0.1628	0.17	0.1772	0.1844
0.5	0.1915	0.1985	0.2054	0.2123	0.219
0.6	0.2257	0.2324	0.2389	0.2454	0.2517
0.7	0.258	0.2642	0.2704	0.2764	0.2823
0.8	0.2881	0.2939	0.2995	0.3051	0.3106
0.9	0.3159	0.3212	0.3264	0.3315	0.3365
1	0.3413	0.3461	0.3508	0.3554	0.3599

Normal Distribution Tables

z	0	2	4	6	8
3.1	0.499	0.4991	0.4992	0.4992	0.4993
3.2	0.4993	0.4994	0.4994	0.4994	0.4995
3.3	0.4995	0.4995	0.4996	0.4996	0.4996
3.4	0.4997	0.4997	0.4997	0.4997	0.4997
3.5	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5	0.5	0.5	0.5	0.5

Normal Distribution

$P[Z \le z]$



from table

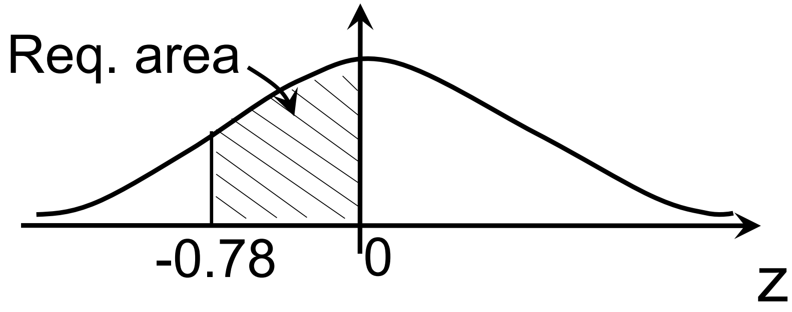
z	0
0.5	0.1915
0.6	0.2257
0.7	0.258

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

$= 0.242$

Example-3

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



P[-0.78 $\leq$ Z $\leq$ 0] = $\frac{1}{\sqrt{2\pi}} \int_{-0.78}^{0} e^{-z^2/2} dz$ = $\frac{1}{\sqrt{2\pi}} \int_{0}^{0.78} e^{-z^2/2} dz$

= 0.2823

From Tables:

Req. area = area betn. 0 and +0.78

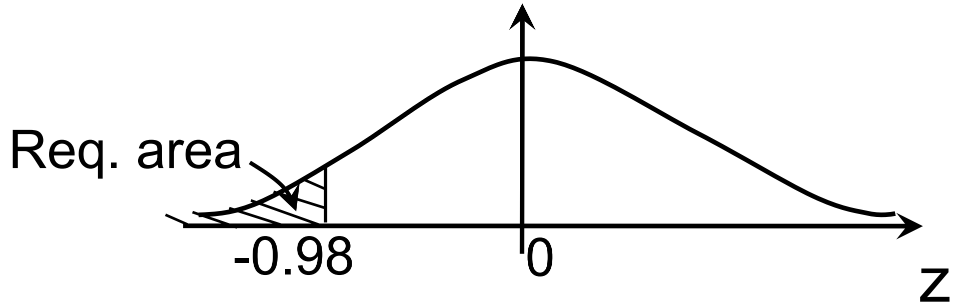
= 0.2823

z	7	8	9
0.6	0.2486	0.2517	0.2549
0.7	0.2794	0.2823	0.2852
0.8	0.3078	0.3106	0.3133

Example-4

Obtain the area under the standard normal curve

$z \leq -0.98$



From tables:

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

= 0.5 – 0.3365

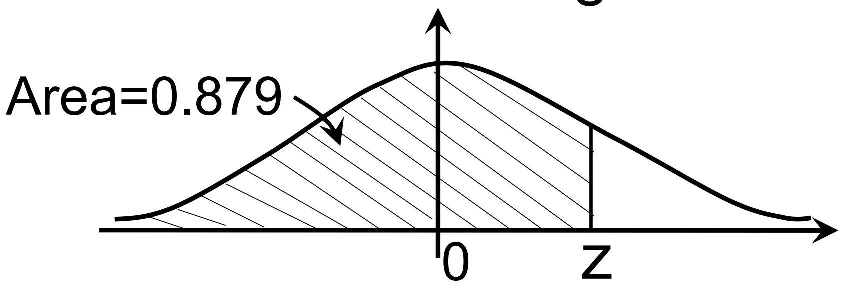
= 0.1635

z	7	8	9
0.8	0.3078	0.3106	0.3133
0.9	0.334	0.3365	0.3389
1	0.3577	0.3599	0.3621
1.1	0.379	0.381	0.383

Example-5

Obtain 'z' such that P[Z ≤ z]=0.879

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



z	6	7	8
1.1	0.377	0.379	0.381
1.2	0.3962	0.398	0.3997

area between 0 to z = 0.879 – 0.5

= 0.379

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

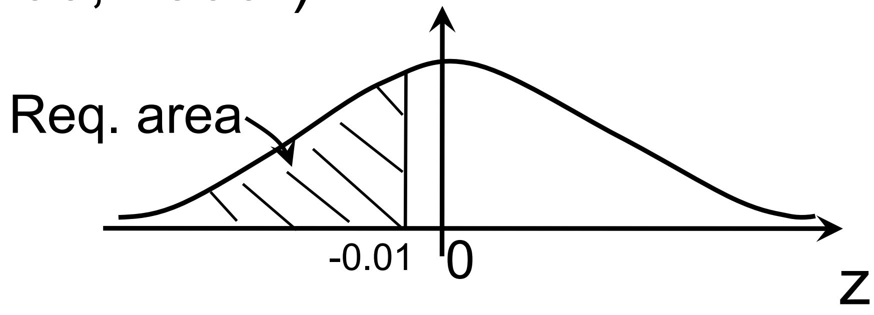
Example-6

Obtain P[X ≤ 75] if N ~ (100, 2500^2)

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{75-100}{2500}$$

$$= -0.01$$



From the table,

Req. area = 0.5 – area between 0 and +0.01

= 0.5 – 0.004

= 0.496

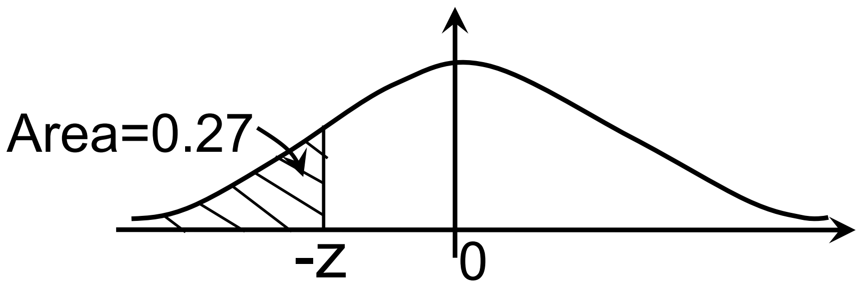
P[X ≤ 75] = 0.496

z	0	1	2
0.0	0	0.004	0.008
0.1	0.0398	0.0438	0.0478

Example-7

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

$P[X \leq x]=0.27$



$P[Z \leq z]=0.27$

z	1	2
0.5	0.195	0.1985
0.6	0.2291	0.2324

area between 0 to -z = 0.5 – 0.27

= 0.23

From the table, z = -0.613

$$z = \frac{x-\mu}{\sigma}; \qquad -0.613 = \frac{x-650}{200} \quad ; \quad x = 527$$

Central limit theorem

- If X_1, X_2,……….. are independent random variables and identically distributed with mean 'μ' and variance 'σ', then the sum

  $S_n = X_1+X_2+\ldots\ldots+X_n$ as $n \rightarrow \infty$

  approaches a normal distribution with mean $n\mu$ and variance $n\sigma^2$.

$$S_n : N\left(n\mu, n\sigma^2\right)$$

iid → independent & identically distributed

Central limit theorem

- For hydrological applications under most general conditions, if X_i 's are all independent with $E[x_i] = \mu_i$ and $var(X_i) = \sigma_i^2$, then the sum

  $S_n = X_1+X_2+\ldots\ldots+X_n$ as $n \rightarrow \infty$

  approaches a normal distribution with

$$E\left[S_n\right] = \sum_{i=1}^{n} \mu_i \;\&$$

$$Var\left[S_n\right] = \sum_{i=1}^{n} \sigma_i^2$$

One condition for this generalised Central Limit Theorem is that each X_i has a negligible effect on the distribution of S_n (Statistical Methods in Hydrology, C.T.Haan, .Affiliated East-West Press Pvt Ltd, 1995, p. 89)

Log-Normal Distribution

- 'X' is said to be log-normally distributed if Y = lnX is normally distributed.
- The probability density function of the log normal distribution is given by

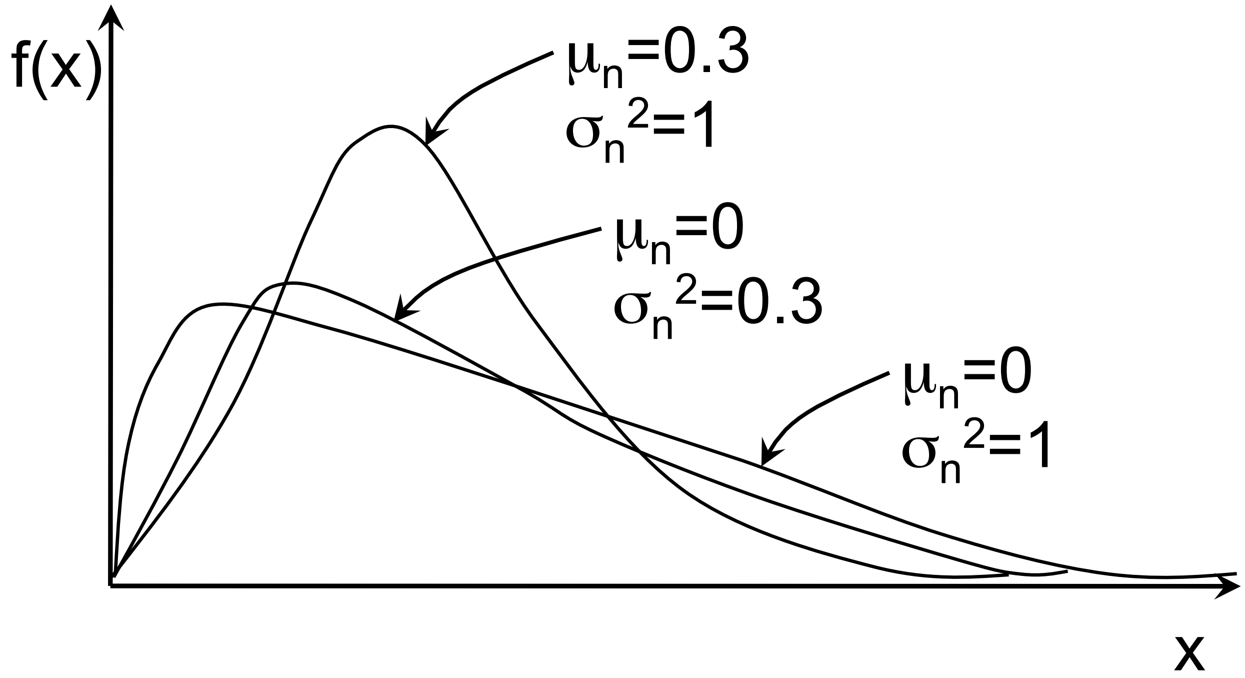
$$f(x) = \frac{1}{\sqrt{2\pi} x\sigma_n} e^{-(\ln x - \mu_n)^2 / 2\sigma_n^2} \qquad 0 < x < \infty, 0 < \mu_n < \infty, \sigma_n > 0$$

- $\gamma_s = 3C_v + C_v^3$

  where C_v is the coefficient of variation of 'X'
- As C_v increases, the skewness, γ_s , increases

$$\mu_y = \frac{1}{2} \ln\left[\frac{\bar{x}^2}{1 + C_v^2}\right], \quad \sigma_y^2 = \ln\left[1 + C_v^2\right] where C_v = \frac{S_x}{\bar{x}}$$

Log-Normal Distribution



- Positively skewed – skewed to the left with long exponential tail on the right.
- Commonly used for monthly streamflow, monthly/ seasonal precipitation, evapotranspiration etc .

Example-8

Consider the annual peak runoff in a river - modeled by a lognormal distribution

μ_n= 5.00 and σ_n = 0.683

Obtain the probability that annual runoff exceeds 300m^3/s

$$
\begin{aligned}
P[X > 300] &= P[Z > (\ln 300-5.00)/0.683] \\
&= P[Z > 1.03] \\
&= 1 - P[Z \leq 1.03] \\
&= 1 - 0.3485 \\
&= 0.6515
\end{aligned}
$$

z	2	3	4
0.9	0.3212	0.3238	0.3264
1	0.3461	0.3485	0.3508
1.1	0.3686	0.3708	0.3729

Example-9

Consider the earlier example,

$\bar{x}$ = 135 Mm^3 ,S = 23.8 Mm^3 and C_v = 0.176

If X follows lognormal distribution

Obtain the P[X $\geq$ 150]

$$\bar{Y} = \frac{1}{2}\ln\left[\frac{\bar{X}^2}{C_v^2+1}\right]$$

$$= \frac{1}{2}\ln\left[\frac{135^2}{0.176^2+1}\right] = 4.89$$

$S_y^2 = \ln(C_v^2+1) = \ln(0.176^2+1) = 0.0305$

$S_y = 0.1747$

Example-9 (contd.)

Y = ln X follows log normal distribution

P[X ≥ 150] = P [Y ≥ ln150]; ln150 = 5.011

$$z = \frac{y - \overline{y}}{S_y}$$

$$= \frac{5.011 - 4.89}{0.1747}$$

$$= 0.693$$

P[Y ≥ ln150] = 1– P[Y ≤ ln150]

= 1– P[Z ≤ 0.693]

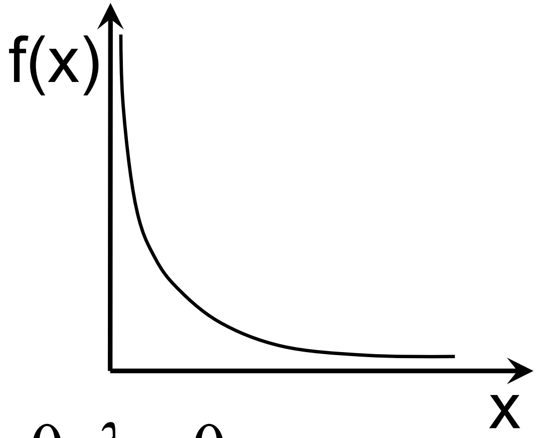
= 1 – 0.25583

= 0.7442

Exponential Distribution

- The probability density function is given by

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



- E[X] = 1/λ
- λ = 1/μ
- Var(X) = $1/\lambda^2$

$$F(x) = \int_0^x f(x)dx = 1 - \lambda e^{-\lambda x} \qquad x > 0, \lambda > 0$$

- $\gamma_s > 0$, therefore positively skewed
- Used for expected time between two critical events (such as floods of a given magnitude), time to failure in hydrologic/water resources systems components

Example-10

The mean time between high intensity rainfall (rainfall intensity above a specified threshold) events occurring during a rainy season is 4 days. Assuming that the mean time follows an exponential distribution.

Obtain the probability of a high intensity rainfall repeating

1. within next 3 to 5 days.
2. within next 2 days

Mean time (μ) = 4

$\lambda = 1/\mu = 1/4$

Example-10 (contd.)

1. P[3 ≤ X ≤ 5] = F(5) - F(3)

   F(5) = $1 - \frac{1}{4}e^{-5/4}$

   = 0.7135

   F(3) = $1 - \frac{1}{4}e^{-3/4}$

   = 0.5276

   P[3 ≤ X ≤ 5] = 0.7135 – 0.5276 = 0.1859

2. P[X ≤ 2] = $1 - \frac{1}{4}e^{-2/4}$ = 0.3935

Gamma Distribution

- The probability density function is given by

$$f(x) = \frac{\lambda^n x^{\eta-1} e^{-\lambda x}}{\Gamma(\eta)} \qquad x, \lambda, \eta > 0$$

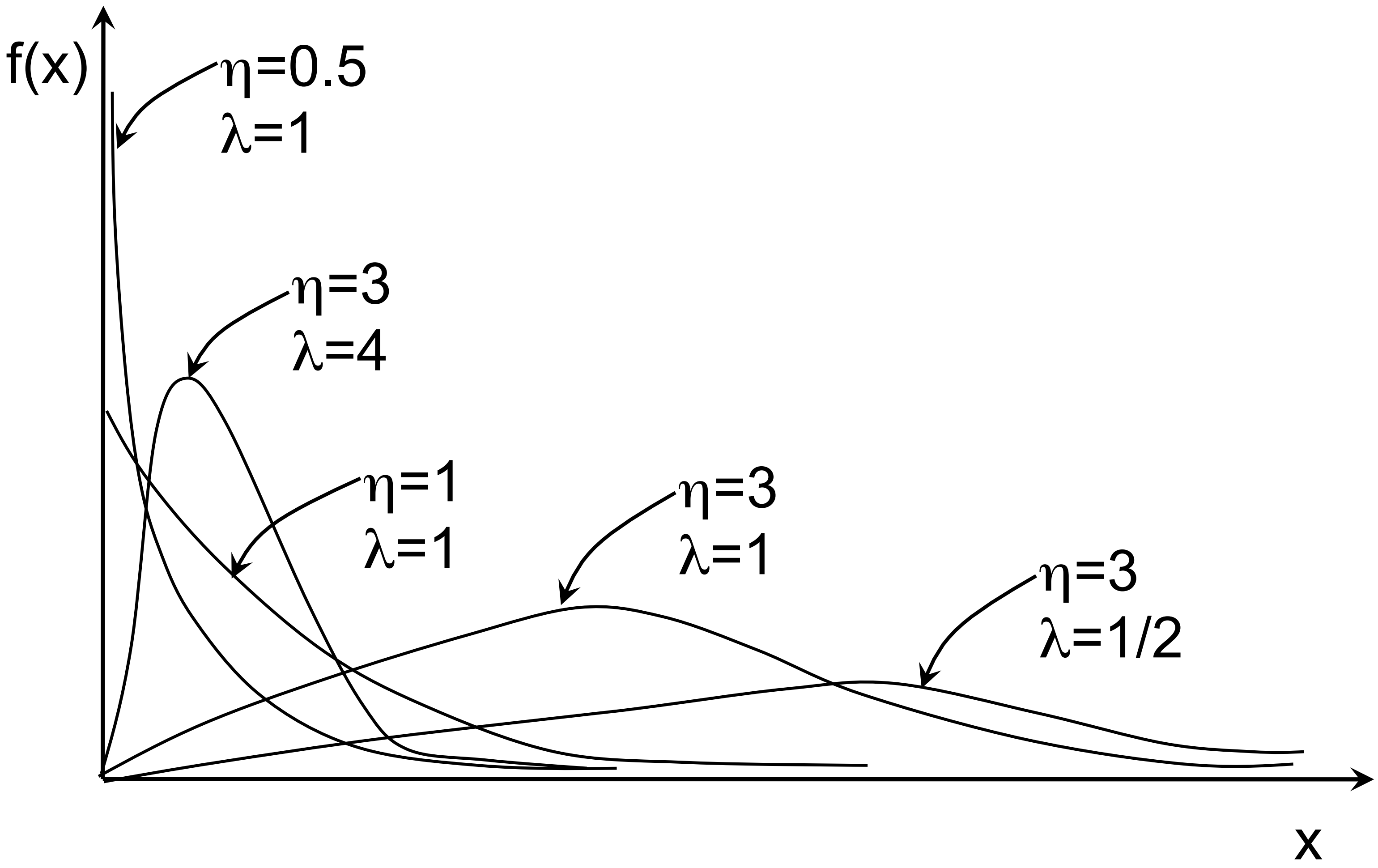
- $\Gamma(\eta)$ is a gamma function
- $\Gamma(\eta) = (\eta-1)!$, $\eta = 1,2,\ldots$
- $\Gamma(\eta+1) = \eta\sqrt{\eta}$ $\quad \eta > 0$
- $\Gamma(\eta) = \int_0^\infty t^{\eta-1} e^{-t} dt$ $\quad \eta > 0$
- Gamma distribution is called as family of distribution

Gamma Distribution

- Exponential distribution is a special case of gamma distribution with η=1
- $\lambda \rightarrow$ Scale parameter
- $\eta \rightarrow$ Shape parameter
- Mean = η/λ
- Variance = η/λ^2 $\rightarrow$ $\sigma = \sqrt{\eta}/\lambda$
- Skewness coefficient $\gamma = 2/\sqrt{\eta}$
- As γ decreases, η increases
- Cdf is given by

$$F(x) = 1 - e^{-\lambda x} \sum_{j=0}^{\eta-1} (\lambda x)^j / j! \qquad x, \lambda, \eta > 0$$

Gamma Distribution



Gamma Distribution

- If 'X' and 'Y' are two independent gamma rvs having parameters η_1, λ and η_2, λ respectively then U=X+Y is a gamma rv with parameters $\eta=\eta_1+\eta_2$ and λ
- This property can be extended to sum of 'n' number of independent gamma rvs.
- Gamma distribution is generally used for daily/ monthly/annual rainfall data
- Also used for annual runoff data

Example-11

During the month 1, the mean and standard deviation of the monthly rainfall are 7.5 and 4.33 cm resp. Assume monthly rainfall data can be approximated by Gamma distribution

1.Obtain the probability of receiving more than 3cm rain during month 1.

Given, μ= 7.5, σ= 4.33

Initially the parameters λ, η are obtained.

$\mu = \eta/\lambda \rightarrow \quad 7.5 = \eta/\lambda$

$\lambda = \eta/7.5$

Example-11 (contd.)

$\sigma = \sqrt{\eta}/\lambda \quad \rightarrow \quad 4.33 = \sqrt{\eta}/\lambda$

$\sqrt{\eta}/\eta = 4.33/7.5$

$\eta = 3$

$\lambda = 4$

$$f(x) = \frac{\lambda^n x^{\eta-1} e^{-\lambda x}}{\Gamma(\eta)} \qquad x, \lambda, \eta > 0$$

$$= \frac{4^3 x^{3-1} e^{-4x}}{\Gamma(3)} \qquad \Gamma(3) = (3-1)! = 2!$$

$$= 32x^2 e^{-4x}$$

Example-11 (contd.)

P[X $\geq$ 3] = 1 – P[X $\leq$ 3]

$$= 1 - \int_0^3 32x^2 e^{-4x} dx$$

$$= 1 - \left(1 - \frac{85}{e^{12}}\right)$$

$$= 0.0005$$

Example-11 (contd.)

During the month 2, the mean and standard deviation of the monthly rainfall are 30 and 8.6 cm respectively.

1. Obtain the probability of receiving more than 3cm rain during month 2.
2. Obtain the probability of receiving more than 3cm rain during the two month period assuming that rainfall during both months are independent.

Given, μ= 30, σ= 8.66

Initially the parameters λ, η are obtained.

$\mu = \eta/\lambda \rightarrow \quad 30 = \eta/\lambda$

$\lambda = \eta/30$

Example-11 (contd.)

σ = $\sqrt{\eta}/\lambda$ → 8.66 = $\sqrt{\eta}/\lambda$

$$\sqrt{\eta}/\eta = 8.66/30$$

η = 12

λ = 4

$$f(x) = \frac{\lambda^n x^{\eta-1} e^{-\lambda x}}{\Gamma(\eta)} \qquad x, \lambda, \eta > 0$$

$$= \frac{4^{12} x^{12-1} e^{-4x}}{\Gamma(12)}$$

Γ(12)= (12-1)!=11!

$$= 0.42 x^{11} e^{-4x}$$

Example-11 (contd.)

1. P[X $\geq$ 3] = 1 – P[X $\leq$ 3]

$$= 1 - \int_0^3 0.42 x^{11} e^{-4x} dx$$

$$= 1 - \left(0.993 - \frac{75073}{e^{12}} \right)$$

$$= 0.4683$$

2. Probability of receiving more than 3cm rain during the two month period:

Since λ value is same for both the months and the rainfall during the both months are independent,

Example-11 (contd.)

Then the combined distribution will have the parameters η , λ as

η = 3+12 = 15

λ = 4

Therefore $f(x) = \dfrac{\lambda^n x^{\eta-1} e^{-\lambda x}}{\Gamma(\eta)}$ $\qquad x, \lambda, \eta > 0$

$$= \frac{4^{15} x^{15-1} e^{-4x}}{\Gamma(15)}$$

$$= 0.0123 x^{14} e^{-4x}$$

Example-11 (contd.)

$P[X \geq 3] = 1 - P[X \leq 3]$

$$= 1 - \int_0^3 0.0123 x^{14} e^{-4x} dx$$

$$= 1 - \left(0.99865 - \frac{125481}{e^{12}} \right)$$

$$= 0.7723$$

The values of cumulative gamma distribution can be evaluated using tables with $\chi^2 = 2\lambda x$ and $\nu = 2\eta$