INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -25

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

Summary of the previous lecture

- Introduction to frequency analysis
  - Extreme events
- Recurrence interval
- Return period
  - Expected value of recurrence interval
- $P[X \geq x_T] = p = 1/T$
- Probability that a T year return period event will occur at least once in N years is

$$1-\left(1-\frac{1}{T}\right)^{N}$$

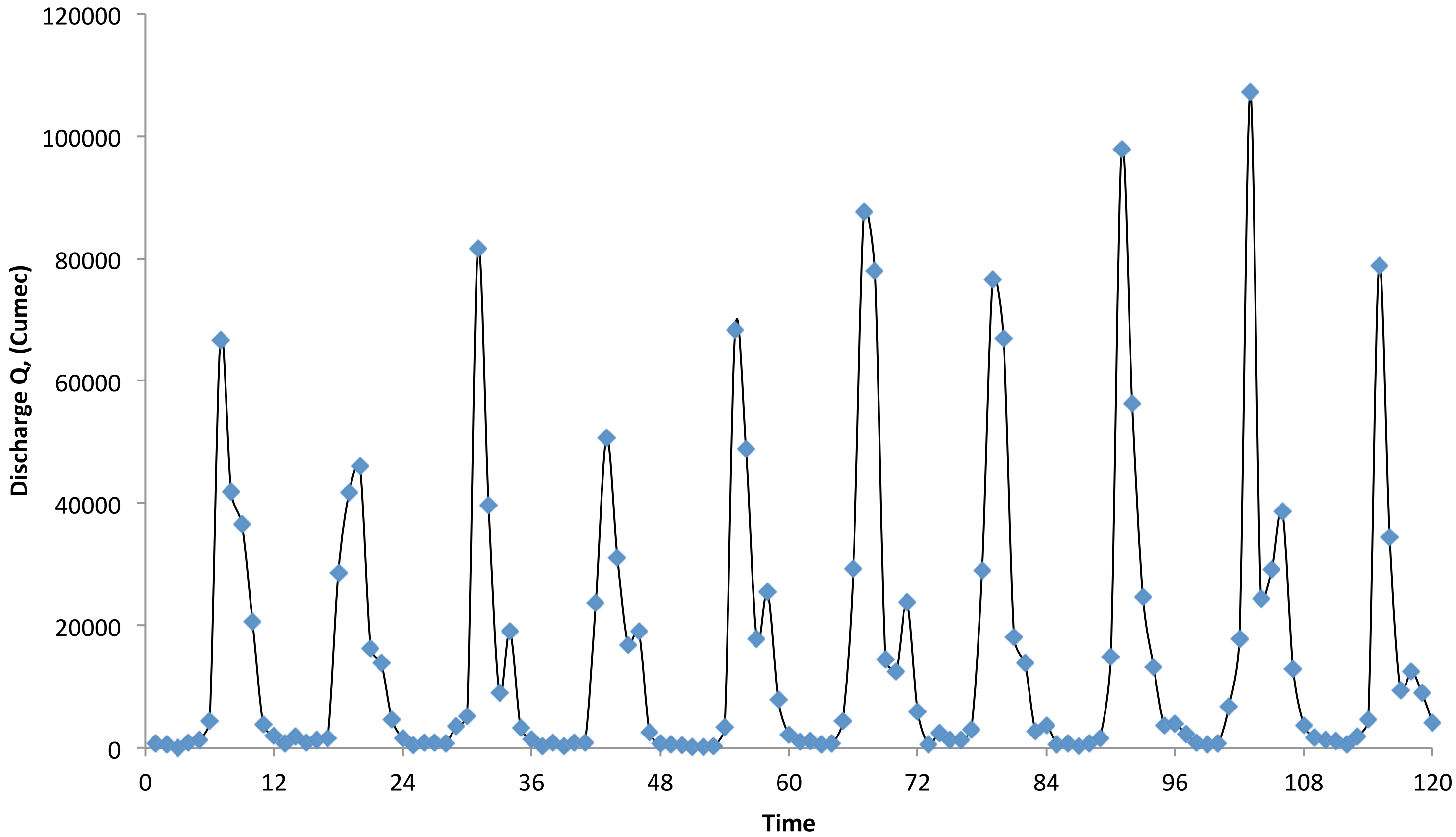
Frequency Analysis

Hydrologic data series:

- Complete duration series: A series containing all the available data.
- Partial duration series: A series of data which are selected so that their magnitude is greater than a predefined base value.
- Annual exceedence series: If the base value is selected so that the number of values is equal to the number of years.
- Extreme value series: Series including the largest or smallest values occurring in each of the equally long time intervals of the record.

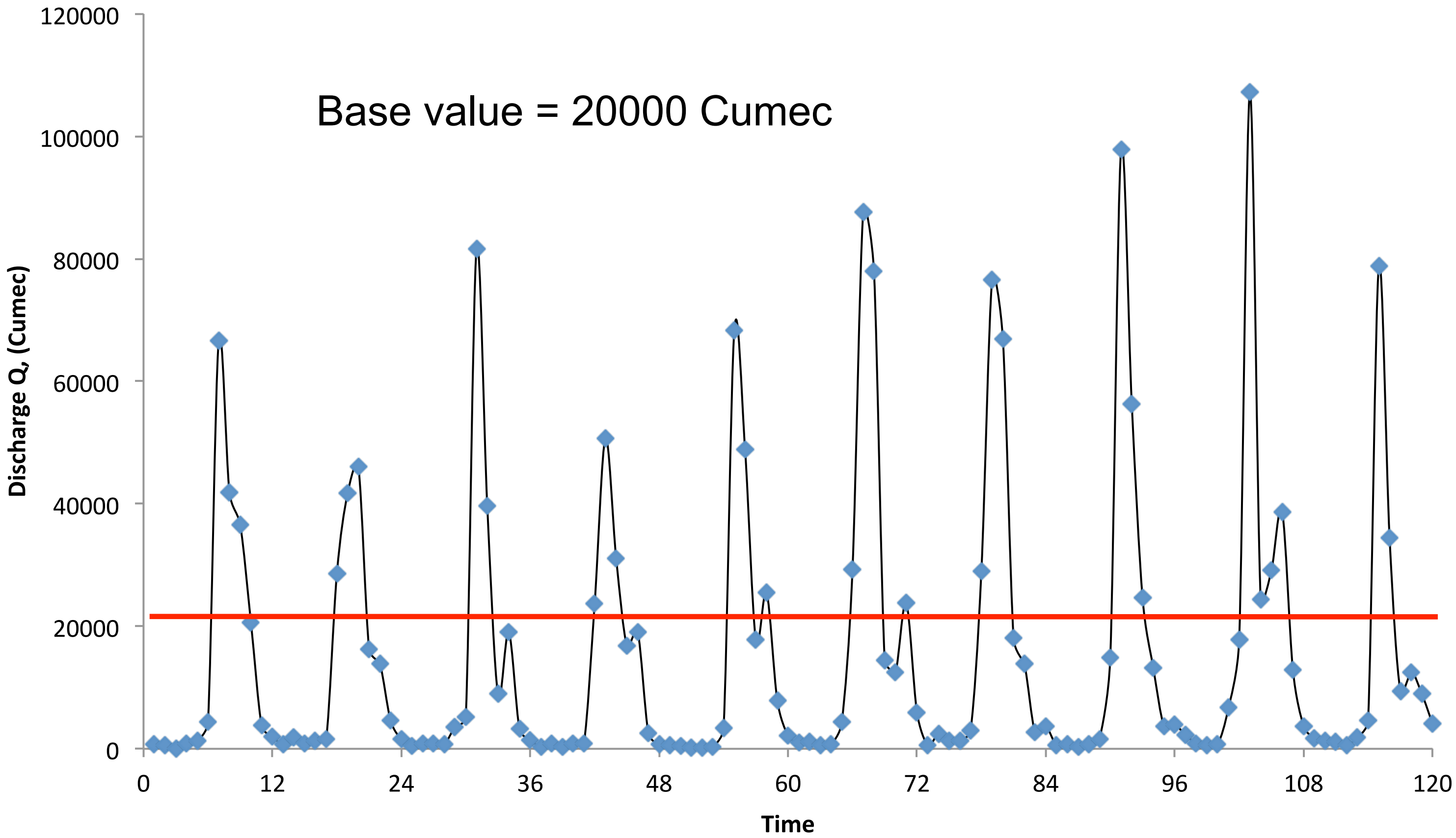
Frequency Analysis

Complete duration series: A series containing all the available data



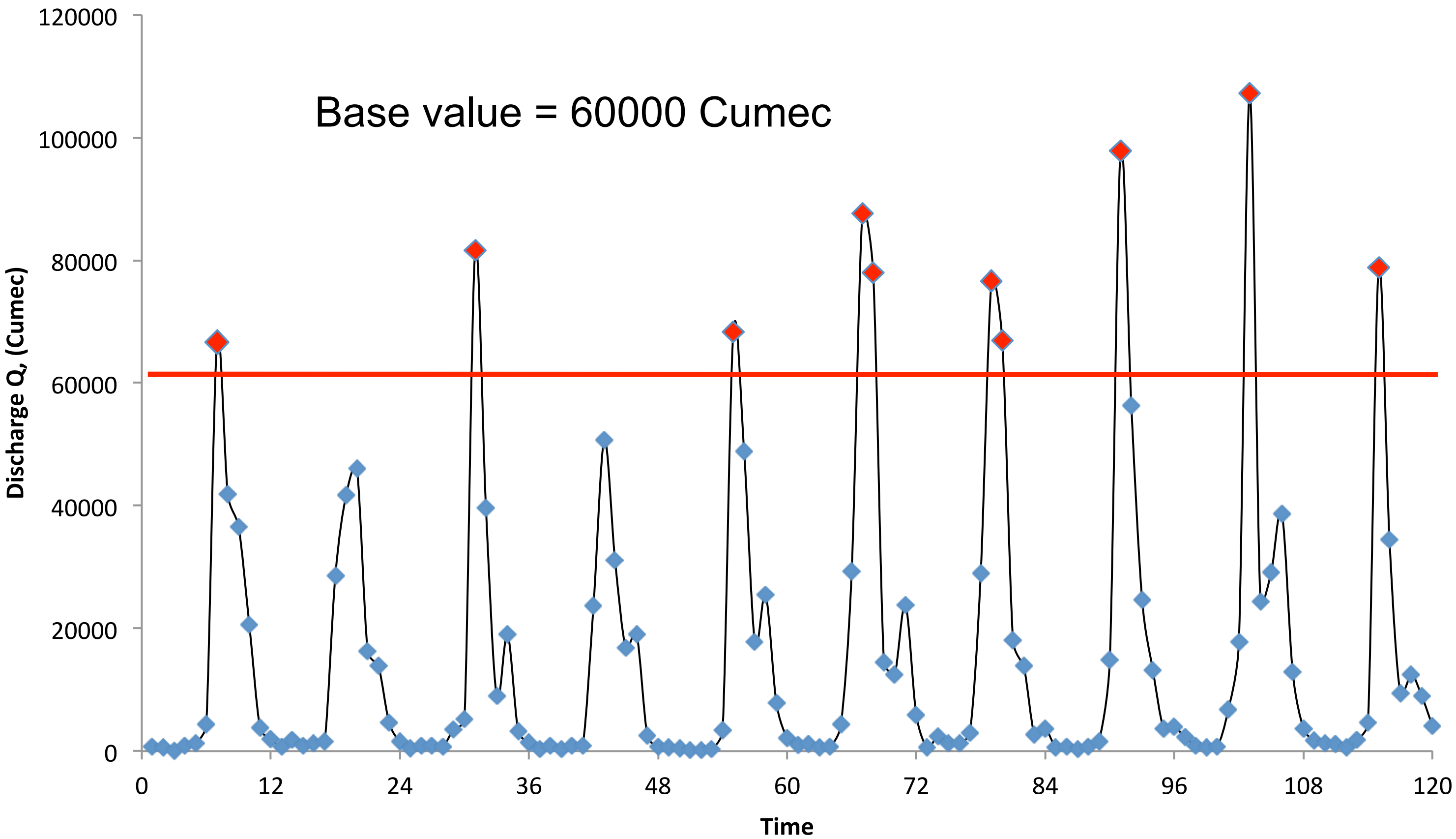
Frequency Analysis

Partial duration series: A series of data which are selected so that their magnitude is greater than a predefined base value.



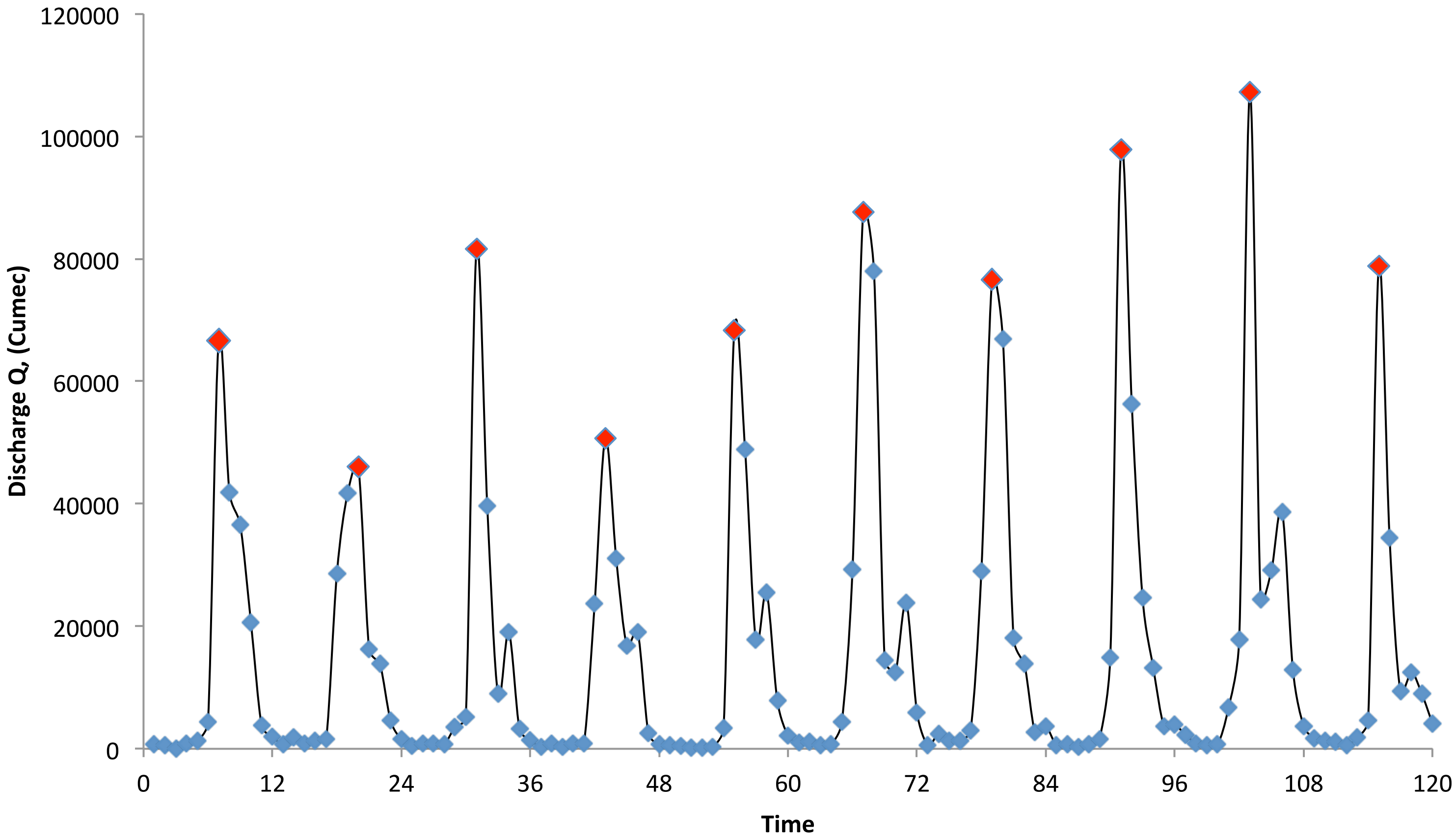
Frequency Analysis

Annual exceedence series: The base value is selected so that the number of values is equal to the number of years



Frequency Analysis

Extreme value series: Series including the largest or smallest values occurring in each of the equally long time intervals of the record



Frequency Analysis

Hydrologic data series:

- Annual maximum (or minimum) series: Series with largest (or smallest) annual values.
- The return period T_e of the event developed from an annual exceedence series is related with the corresponding return period T derived from an annual maximum series by (Chow, 1964)

$$T_e = \left\{ \ln\left(\frac{T}{T-1} \right) \right\}^{-1}$$

Ref: Chow, V.T., Statistical and probability analysis of hydrologic data sec,8.1 in Handbook of applied hydrology, McGraw Hill, New York, 1964

Frequency Analysis

- Annual exceedence series: It may be difficult to verify that all the observations are independent
- The occurrence of a large flood could be related to saturated soil conditions produced during another large flood occurring a short time earlier
- Usually annual maximum series is preferred to use.
- As the return period becomes large, the results from the two approaches become similar as the chance that two such events will occur within any year is very small.

Frequency Analysis

Extreme Value Distributions:

- Extreme events:
  - Peak flood discharge in a stream
  - Maximum rainfall intensity
  - Minimum flow
- The study of extreme hydrologic events involves the selection of largest (or smallest) observations from sets of data
  - e.g., the study of peak flows of a stream uses the largest flow value recorded at gauging station each year.

Frequency Analysis

- Three types of Extreme Value distributions are developed based on limited assumptions concerning parent distribution
  - ➢ Extreme value Type-I (EV I) – parent distribution unbounded in direction of the desired extreme and all moments of the distribution exist.
  - ➢ Extreme value Type-II (EV II) – parent distribution unbounded in direction of the desired extreme and all moments of the distribution do not exist.
  - ➢ Extreme value Type-III (EV III) – parent distribution bounded in direction of the desired extreme.

Frequency Analysis

Extreme Value Type-I (EV I) distribution

- The cumulative probability distribution function is

$$F(x) = \exp\left\{-\exp\left(-\frac{x-\beta}{\alpha}\right)\right\} \qquad -\infty \le x \le \infty$$

- Parameters are estimated as

$$\beta = \bar{x} - 0.577\alpha$$

$$\alpha = \frac{\sqrt{6} \times s}{\pi}$$

Frequency Analysis

- Y = (X – β)/ α → transformation

- The cumulative probability distribution function is

$$F(y) = \exp\left\{-\exp\left(-y\right)\right\} \qquad -\infty \le y \le \infty$$

- Solving for y,

$$y = -\ln\left\{-\ln\left(\frac{1}{F\left(y\right)}\right)\right\}$$

Frequency Analysis

$$P\left(X \geq x_T\right) = \frac{1}{T}; \qquad 1 - P\left(X < x_T\right) = \frac{1}{T}$$

$$1 - F\left(x_T\right) = \frac{1}{T}; \qquad F\left(x_T\right) = 1 - \frac{1}{T}$$

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

Therefore

$$y = -\ln\left\{\ln\left(\frac{T}{T-1}\right)\right\}$$

Example – 1

Consider the annual maximum discharge Q (in cumec), of a river for 45 years.

1. Develop a model for annual maximum discharge frequency analysis using Extreme Value Type-I distribution, and
2. Calculate the maximum annual discharge values corresponding to 20-year and 100-year return periods

Example – 1 (Contd.)

Cumec

Data is as follows:

Year	Q	Year	Q	Year	Q	Year	Q
1950	804	1961	507	1972	1651	1983	1254
1951	1090	1962	1303	1973	716	1984	430
1952	1580	1963	197	1974	286	1985	260
1953	487	1964	583	1975	671	1986	276
1954	719	1965	377	1976	3069	1987	1657
1955	140	1966	348	1977	306	1988	937
1956	1583	1967	804	1978	116	1989	714
1957	1642	1968	328	1979	162	1990	855
1958	1586	1969	245	1980	425	1991	399
1959	218	1970	140	1981	1982	1992	1543
1960	623	1971	49	1982	277	1993	360
						1994	348

Example – 1 (Contd.)

Mean of the data, $\bar{x}$ = 756.6 cumec

Standard deviation, s = 639.5 cumec

$$\alpha = \frac{\sqrt{6} \times s}{\pi}$$

$$= \frac{\sqrt{6} \times 639.5}{\pi}$$

$$= 498.6$$

$$\beta = \bar{x} - 0.577\alpha$$

$$= 756.6 - 0.577 \times 498.6$$

$$= 468.8$$

Example – 1 (Contd.)

The probability model is

$$F(x) = P\left[X \le x\right] = \exp\left\{-\exp\left(-\frac{x - 468.8}{498.6}\right)\right\}$$

$$-\infty \leq x \leq \infty$$

To determine the x_T value for a particular return period, the reduced variate y is initially calculated for that particular return period using

$$y = -\ln\left\{\ln\left(\frac{T}{T-1}\right)\right\}$$

Example – 1 (Contd.)

For T = 20 years,

$$y_{20} = -\ln\left\{\ln\left(\frac{20}{20-1}\right)\right\}$$

$$= 2.97$$

$Y = (X - \beta)/\ \alpha$

$x_{20} = \beta + \alpha\ y_{20}$

$= 468.8 + 498.6 * 2.97$

$= 1950$ cumec

Example – 1 (Contd.)

Similarly for T = 100 years,

$$y_{100} = -\ln\left\{\ln\left(\frac{100}{100-1}\right)\right\}$$

$$= 4.6$$

$$x_{100} = \beta + \alpha\, y_{100}$$

$$= 468.8 + 498.6 * 4.6$$

$$= 2762 \text{ cumec}$$

Frequency Analysis

- Extreme value distributions have been widely used in hydrology
- Extreme value Type I distribution (also called as Gumbel's Extreme Value distribution) is most commonly used for modeling storm rainfalls and maximum flows.
- Extreme value Type III distribution (also called as Weibull's distribution) is most commonly used for modeling low flows.

Frequency Analysis

Frequency analysis using frequency factors:

- Calculating the magnitudes of extreme events by the method discussed requires that the cumulative probability distribution function to be invertible. That is, given $F(x)$, we must be able to obtain $x = F^{-1}(x)$
- Some probability distribution functions such as the Normal. Lognormal and Pearson type-III distributions are not readily invertible.
- An alternative method of calculating the magnitudes of extreme events is by using frequency factors

Frequency Analysis

- The magnitude of an event x_T is represented as the mean μ plus the deviation Δx_T of the variate from the mean

$$x_T = \mu + \Delta x_T$$

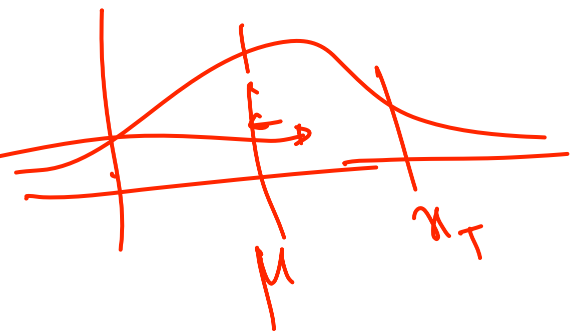
- The deviation is taken as equal to the product of standard deviation σ and a frequency factor K_T

$$\Delta x_T = K_T \sigma$$

$$x_T = \mu + K_T \sigma$$

or

$$x_T = \bar{x} + K_T s$$



Frequency Analysis

- The deviation and the frequency factor are functions of the return period and the type of probability distribution to be used in the analysis
- If in the event, the variable analyzed is y = log x, then the same method is used, with the statistics for the logarithms of the data as

$$y_T = \bar{y} + K_T s_y$$

- For a given distribution, a K – T relationship can be determined
- The relationship can be expressed in terms of mathematical terms or a table.

Frequency Analysis

The procedure for frequency analysis is as follows

- Obtain the statistical parameters required for the probability distribution from the given data
- For a given return period T, the frequency factor is determined from the K – T relationship available for the distribution
- The magnitude of x_T is then computed by

$$x_T = \bar{x} + K_T s$$

Frequency Analysis

Frequency factor for Normal Distribution:

$$x_T = \mu + K_T\sigma$$

$$K_T = \frac{x_T - \mu}{\sigma}$$

- The frequency factor K_T for the normal distribution is equal to standard normal deviate z.
- z corresponding to a given return period T, with p=1/T may be approximated using an intermediate variable, w.

$$w = \left[\ln\left(\frac{1}{p^2}\right)\right]^{1/2} \qquad 0 < p \le 0.5$$

Frequency Analysis

- The frequency factor is expressed in terms of w as

$$K_T = w - \frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}$$

where

$$w = \left[\ln\left(\frac{1}{p^2}\right)\right]^{1/2} \qquad 0 < p \leq 0.5$$

Applied Hydrology by V.T.Chow, D.R.Maidment, L.W.Mays, McGraw-Hill 1998

Frequency Analysis

- When $p > 0.5$, $1 - p$ is substituted and the K_T value computed is given an negative sign
- The error in this formula is less than 0.00045 (Abramowitz and Stegun, 1965)
- The same procedure is applied for the lognormal distribution expect that the logarithm of data is used to calculate mean and standard deviation.

Abramowitz, M., and Stegun, I.A., Handbook of Mathematical Functions, Dover, New York, 1965

Example – 2

Consider the annual maximum discharge Q, of a river for 45 years as in the previous example,

Mean, $\bar{x}$ = 756.6 cumec

Standard deviation, s = 639.5 cumec

Calculate the frequency factor and obtain the maximum annual discharge value corresponding to 20 year return period using Normal distribution

Example – 2 (Contd.)

T = 20

p = 1/20 = 0.05

$$w = \left[\ln\left(\frac{1}{p^2}\right)\right]^{1/2}$$

$$= \left[\ln\left(\frac{1}{0.05^2}\right)\right]^{1/2}$$

$$= 2.45$$

Example – 2 (Contd.)

$$K_{20} = w - \frac{2.515517 + 0.802853w + 0.01032w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}$$

$$= 2.45 - \frac{2.515517 + 0.802853 \times 2.45 + 0.01032 \times 2.45^2}{1 + 1.432788 \times 2.45 + 0.189269 \times 2.45^2 + 0.001308 \times 2.45^3}$$

$$= 1.648$$

$$x_{20} = \overline{x} + K_{20}s$$

$$= 756.6 + 1.648 \times 639.5$$

$= 1810.5$ cumec

Frequency Analysis

Frequency factor for Extreme Value Type I(EV I) Distribution:

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}$$

- To express T in terms of K_T,

$$T = \frac{1}{1 - \exp\left\{-\exp\left[-\left(0.5772 + \frac{\pi K_T}{\sqrt{6}}\right)\right]\right\}}$$

Applied Hydrology by V.T.Chow, D.R.Maidment, L.W.Mays, McGraw-Hill 1998

Frequency Analysis

When $x_T = \mu$ in the equation $K_T = \dfrac{x_T - \mu}{\sigma}$; $K_T = 0$

Substituting $K_T = 0$,

$$T = \frac{1}{1 - \exp\left\{-\exp\left[-\left(0.5772 + \frac{\pi \times 0}{\sqrt{6}}\right)\right]\right\}}$$

$= 2.33$ years

i.e., the return period of mean of a EV I is 2.33 years

Example – 3

Consider the annual maximum discharge of a river for 45years in the previous example,

The mean of the data, $\bar{x}$ = 756.6 cumec

Standard deviation, s = 639.5 cumec

Calculate the frequency factor and obtain the maximum annual discharge value corresponding to 20 year return period using Extreme Value Type I(EV I) distribution

Example – 3 (Contd.)

T = 20 years

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\}$$

$$= -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{20}{20-1}\right)\right]\right\}$$

$$= 1.866$$

$$x_T = \bar{x} + K_T s$$

$$= 756.6 + 1.866 \times 639.5$$

$$= 1949.9 \text{ cumec}$$

Frequency Analysis

Frequency factor for Log Pearson Type III Distribution:

•The data is converted to the logarithmic series by

$y = \log x$.

•Logarithms to base 10 is usually used

•The mean $\bar{y}$, standard deviation s_y, and the coefficient of skewness C_s are calculated for the converted logarithmic data

•The frequency factor depend on the return period and coefficient of skewness

Frequency Analysis

- When $C_s = 0$, the frequency factor is equal to the standard normal deviate z and is calculated as in case of Normal deviation.
- When $C_s \neq 0$, K_T is calculated by (Kite, 1977)

$$K_T = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5$$

Where $k = C_s / 6$

Example – 3

Consider the annual maximum discharge of a river for 45years in the previous example,

Calculate the frequency factor and x_T obtain the maximum annual discharge value corresponding to 20 year return period using Log Person Type III distribution

The logarithmic data series is first obtained

Example – 3 (Contd.)

The mean of the data, $\bar{y}$ = 2.725 cumec

Standard deviation, s = 0.388 cumec

Coefficient of skewness C_s = -0.2664

T = 20 years

$$w = \left[\ln\left(\frac{1}{p^2}\right)\right]^{1/2}$$

$$= \left[\ln\left(\frac{1}{0.05^2}\right)\right]^{1/2}$$

$$= 2.45$$

Example – 3 (Contd.)

$$z = w - \frac{2.515517 + 0.802853w + 0.01032w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}$$

$$= 2.45 - \frac{2.515517 + 0.802853 \times 2.45 + 0.01032 \times 2.45^2}{1 + 1.432788 \times 2.45 + 0.189269 \times 2.45^2 + 0.001308 \times 2.45^3}$$

$$= 1.648$$

$k = C_s / 6$

$= -0.2664/6$

$= -0.0377$

Example – 3 (Contd.)

$$K_T = 1.648 + \left(1.648^2 - 1\right)\left(-0.0377\right) + \frac{1}{3}\left(1.648^3 - 6\times1.648\right)\left(-0.0377\right)^2$$

$$-\left(1.648^2 - 1\right)\left(-0.0377\right)^3 + 1.648\times\left(-0.0377\right)^4 + \frac{1}{3}\left(-0.0377\right)^5$$

= 1.581

$$x_T = \bar{x} + K_T s$$

$$= 756.6 + 1.581\times639.5$$

$$= 1767.6 \text{ cumec}$$

PROBABILITY PLOTTING

Probability Plotting

- To check whether probability distribution fits a set of data or not the data is plotted on specially designed probability paper.
- The cumulative probability of a distribution is represented graphically on probability paper designed for the distribution.
- The plot is prepared with exceedence probability or the return period ‘T’ on abscissa and the magnitude of the event on ordinate.
- The scales of abscissa and ordinate are so designed that the data to be fitted are expected to appear close to straight line

Probability Plotting

- The purpose of using the probability paper is to linearize the probability relationship
- The plot can be used for interpolation, extrapolation and comparison purposes.
- The plot can also be used for estimating magnitudes with other return periods.
- If the plot is used for extrapolation, the effect of various errors is often magnified.

Plotting Position

- Plotting position is a simple empirical technique
- Relation between the magnitude of an event verses its probability of exceedence.
- Plotting position refers to the probability value assigned to each of the data to be plotted
- Several empirical methods to determine the plotting positions.
- Arrange the given series of data in descending order
- Assign a order number to each of the data (termed as rank of the data)

Plotting Position

- First entry as 1, second as 2 etc.
- Let ‘n’ is the total no. of values to be plotted and ‘m’ is the rank of a value, the exceedence probability (p) of the m^{th} largest value is obtained by various formulae.
- The return period (T) of the event is calculated by $T = 1/p$
- Compute T for all the events
- Plot T verses the magnitude of event on semi log or log log paper

Plotting Position

Formulae for exceedence probability:

California Method:

$$P\left(X \geq x_m\right) = \frac{m}{n}$$

Limitations

– Produces a probability of 100% for m = n

Plotting Position

Modification to California Method:

$$P\left(X \geq x_m\right) = \frac{m-1}{n}$$

Limitations

– Formula does not produce 100% probability

– If m = 1, probability is zero

Plotting Position

Hazen's formula:

$$P\left(X \geq x_m\right) = \frac{m - 0.5}{n}$$

Chegodayev's formula:

$$P\left(X \geq x_m\right) = \frac{m - 0.3}{n + 0.4}$$

Widely used in U.S.S.R and Eastern European countries

Plotting Position

Weibull's formula:

– Most commonly used method
– If 'n' values are distributed uniformly between 0 and 100 percent probability, then there must be n+1 intervals, n–1 between the data points and 2 at the ends.

$$P\left(X \geq x_m\right) = \frac{m}{n+1}$$

– Indicates a return period T one year longer than the period of record for the largest value

Plotting Position

Most plotting position formulae are represented by:

$$P\left(X \geq x_m\right) = \frac{m-b}{n+1-2b}$$

Where b is a parameter

- E.g., b = 0.5 for Hazen's formula, b = 0.5 for Chegodayev's formula, b = 0 for Weibull's formula
- b = 3/8 0.5 for Blom's formula
- b = 1/3 0.5 for Tukey's formula
- b = 0.44 0.5 for Gringorten's formula

Plotting Position

- Cunnane (1978) studied the various available plotting position methods based on unbiasedness and minimum variance criteria.
- If large number of equally sized samples are plotted, the average of the plotted points foe each value of m lie on the theoretical distribution line.
- Minimum variance plotting minimizes the variance of the plotted points about the theoretical line.
- Cunnane concluded that the Weibull's formula is biased and plots the largest values of a sample at too small a return period.

Plotting Position

- For normally distributed data, the best formula is Blom's plotting position formula (b = 3/8).
- For Extreme Value Type I distribution, the Gringorten formula (b = 0.44) is the best.

Example – 3

Consider the annual maximum discharge of a river for 45years, plot the data using Weibull's formula

Year	Data	Year	Data	Year	Data	Year	Data
1950	804	1961	507	1972	1651	1983	1254
1951	1090	1962	1303	1973	716	1984	430
1952	1580	1963	197	1974	286	1985	260
1953	487	1964	583	1975	671	1986	276
1954	719	1965	377	1976	3069	1987	1657
1955	140	1966	348	1977	306	1988	937
1956	1583	1967	804	1978	116	1989	714
1957	1642	1968	328	1979	162	1990	855
1958	1586	1969	245	1980	425	1991	399
1959	218	1970	140	1981	1982	1992	1543
1960	623	1971	49	1982	277	1993	360
						1994	348

Example – 3 (Contd.)

- The data is arranged in descending order
- Rank is assigned to the arranged data
- The probability is obtained using

$$P\left(X \geq x_m\right) = \frac{m}{n+1}$$

- Return period is calculated
- The maximum annual discharge verses the return period is plotted

Example – 3 (Contd.)

Year	Annual Max. Q	Arranged data	Rank (m)	$P(X \geq x_m)$	T
1950	804	3069	1	0.021739	46
1951	1090	1982	2	0.043478	23
1952	1580	1657	3	0.065217	15.33333
1953	487	1651	4	0.086957	11.5
1954	719	1642	5	0.108696	9.2
1955	140	1586	6	0.130435	7.666667
1956	1583	1583	7	0.152174	6.571429
1957	1642	1580	8	0.173913	5.75
1958	1586	1543	9	0.195652	5.111111
1959	218	1303	10	0.217391	4.6
1960	623	1254	11	0.23913	4.181818

Example – 3 (Contd.)