AAiT

Addis Ababa University
Institute of Technology (AAiT)

Chapter Four
Stochastic Analysis of Stream flow

- Introduction
- Descriptive statistics
- Probability and random variables
- Hydrological statistics and extremes
- Random functions
- Time series analysis
- Geostatistics
- Forward stochastic modelling
- Optimal state prediction and the Kalman filter

AAiT

Addis Ababa University
Institute of Technology (AAiT)

Stochastic

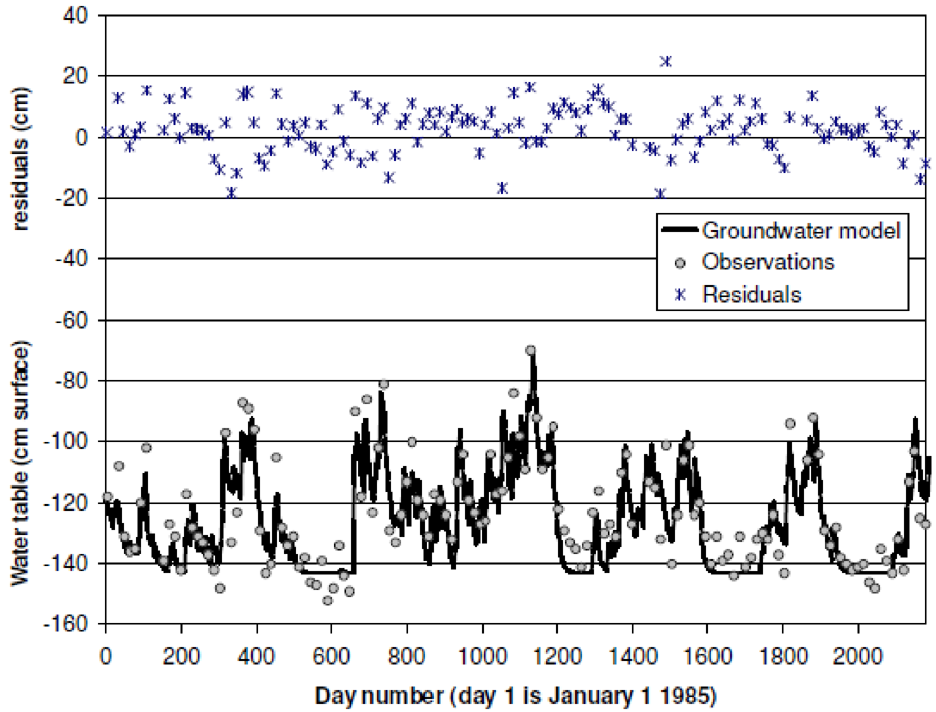
- The term “Stochastic” derives from the Greek word “Stochasticos”, which in turn is derived from “Stochazesthai” means
  - (a) to shoot (an arrow) at a target,
  - (b) to guess or conjecture (the target),
  - (c) to imagine, think deeply, bethink, contemplate, cogitate, meditate

- In the modern sense “stochastic” in stochastic methods refers to the random element incorporated in these methods.

- Stochastic methods thus aim
  - at predicting the value of some variable at non-observed times or at non-observed locations, while also stating how uncertain we are when making these predictions

Addis Ababa University
Institute of Technology (AAiT)

uncertainty associated with our predictions



- observation errors
- errors in boundary conditions, initial conditions and input
- *unknown heterogeneity and parameters*
- scale discrepancy
- model or system errors

Hydrologic Models

- Deterministic (eg. Rainfall runoff analysis)
  - Analysis of hydrological processes using deterministic approaches
  - Hydrological parameters are based on physical relations of the various components of the hydrologic cycle.
  - Do not consider randomness; a given input produces the same output.
- **Stochastic (eg. flood frequency analysis)**
  - Probabilistic description and modeling of hydrologic phenomena
  - Stastical analysis of hydrologic data based on their randomness.

Statistics in Hydrology

- **Mean, median and mode (central tendency)**
- **Dispersion: the spread of the items in a data set around its central value**

Measure of central tendency
- Mode
- Median
- Mean

Measure of dispersion
- Range
- Quartile deviation
- Standard dev.

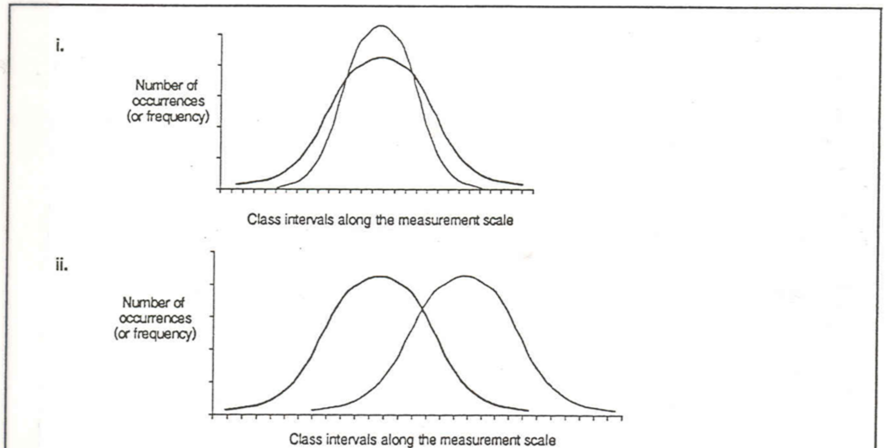
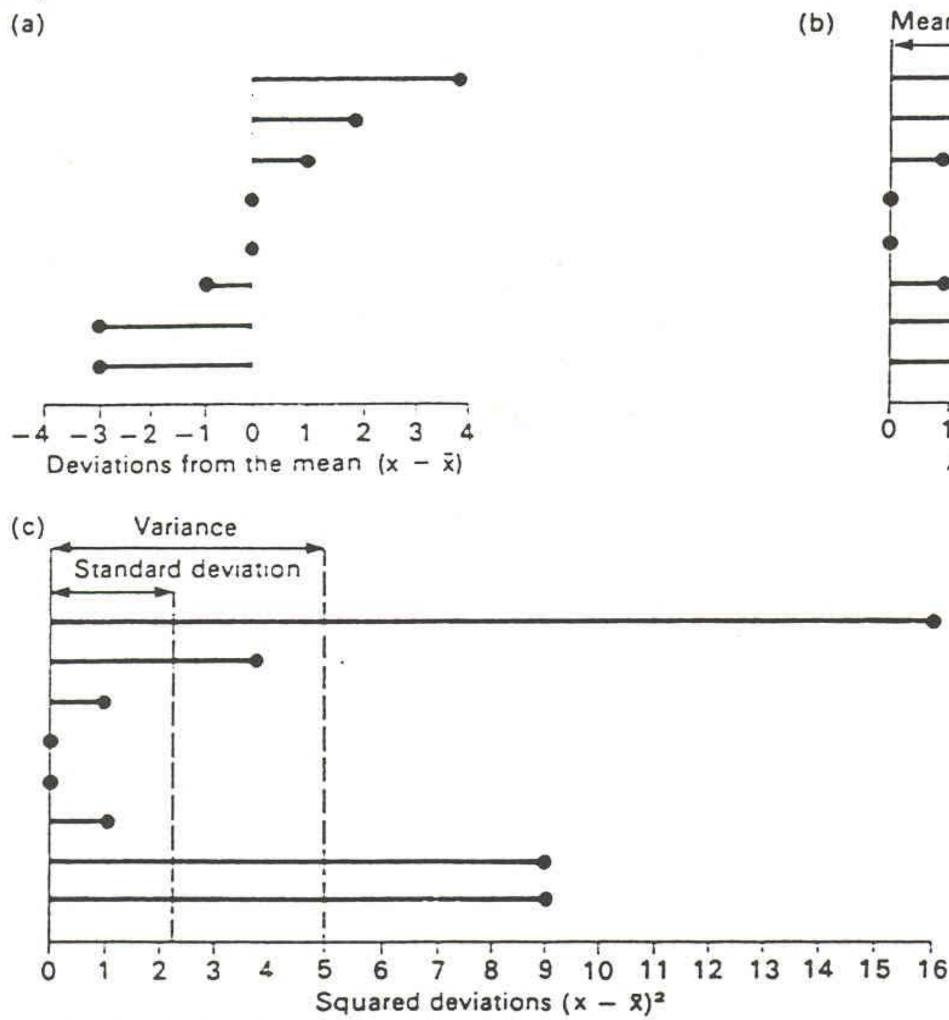


Fig. 3 Central tendency and dispersion.
i. two normal distributions with the same central tendency but different levels of dispersion.
ii. two normal distributions with different central tendency but similar levels of dispersion.

Statistics in Hydrology



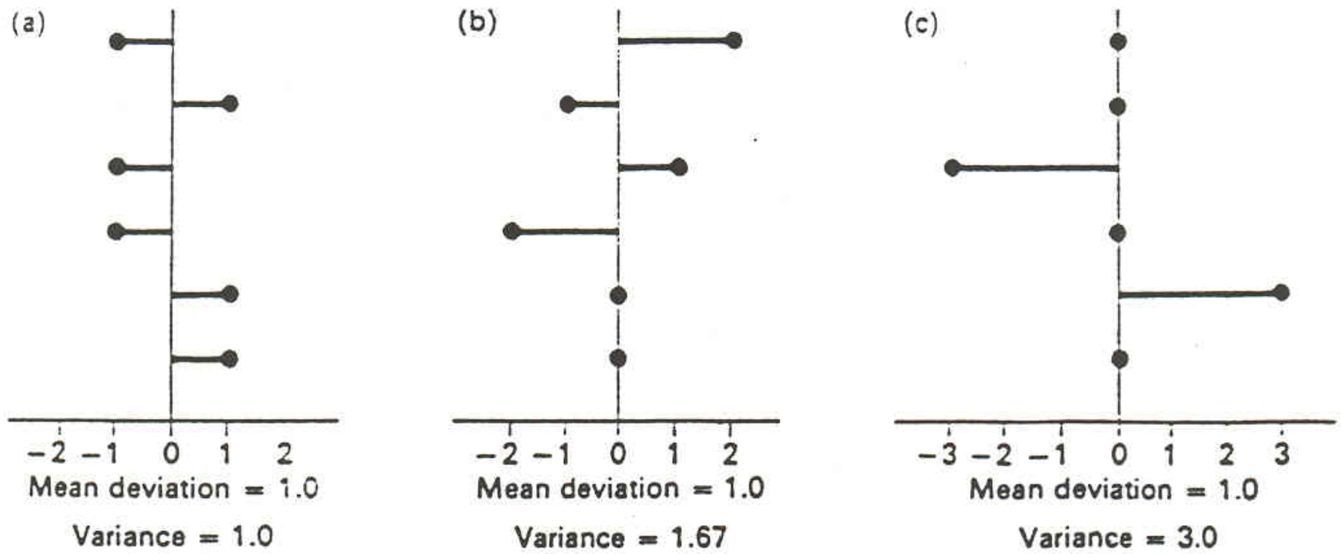
$$\text{Mean deviation} = \frac{\Sigma |x - \bar{x}|}{n}$$

$$\text{Variance} = \frac{\Sigma (x - \bar{x})^2}{n}$$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$$

Statistics in Hydrology

Why do we need to include variance/SD?



Events and extreme events

"Man can believe the impossible. But man can never believe the improbable."

Oscar Wilde

"It seems that the rivers know the [extreme value] theory. It only remains to convince the engineers of the validity of this analysis."

E. J. Gumbel

FREQUENCY ANALYSIS

- **Basic Problem:**

  To relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions.

FREQUENCY ANALYSIS

- **Basic Assumptions:**

A. Analyzed Data are to be statistically independent and identically distributed

selection of data (Time dependence, time scale, mechanisms).

B. Change over time due to man-made (eg. urbanization) or natural processes do not alter the frequency relation

temporal trend in data (stationarity)

FREQUENCY ANALYSIS

- **Practical Problems:**
  - Selection of reasonable and simple distribution.
  - Estimation of parameters in distribution.
  - Assessment of risk with reasonable accuracy.

REVIEW OF BASIC CONCEPTS

- **Probabilistic**

  Outcome of a hydrologic event (e.g., rainfall amount & duration; flood peak discharge; wave height, etc.) is random and cannot be predicted with certainty.

- **Population**

  The collection of all possible outcomes relevant to the process of interest. Example:

  (1) Max. 2-hr rainfall depth: all non-negative real numbers;

  (2) No. of storm in June: all non-negative integer numbers.

- ***Sample***

  A measured segment (or subset) of the population.

- Random Variable

  A variable describable by a probability distribution which specifies the chance that the variable will assume a particular value.

REVIEW OF BASIC CONCEPTS

Frequency and Relative Frequency

- For discrete random variables:
  - Frequency is the number of occurrences of a specific event.
  - Relative frequency is resulting from dividing frequency by the total number of events. e.g.

n = no. of years having exactly 50 rainy days;

Let *n*=10 years and *N*=100 years.

Then, the frequency of having exactly 50 rainy days is 10 and the relative frequency of having exactly 50 rainy days in 100 years is $n/N = 0.1$.

- For continuous random variables:
  - Frequency needs to be defined for a class interval.
  - A plot of frequency or relative frequency versus class intervals is called histogram or probability polygon.
  - As the number of sample gets infinitely large and class interval length approaches to zero, the histogram will become a smooth curve, called **probability density function (PDF)**.

REVIEW OF BASIC CONCEPTS

Probability Density Function (PDF) –

- For a continuous random variable, the PDF must satisfy

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

and $f(x) \geq 0$ for all values of x.

- For a discrete random variable, the PDF must satisfy

$$\sum_{all\ x} p(x) = 1$$

and $1 \geq p(x) \geq 0$ for all values of x.

AAiT

Addis Ababa University
Institute of Technology (AAiT)

REVIEW OF BASIC CONCEPTS

Cumulative Distribution Function -

For a continuous random variable,

$$\Pr(X \leq x_o) = \int_{-\infty}^{x_o} f(x)\, dx$$

For a discrete random variable, by

$$\Pr(X \leq x_o) = \sum_{all\ x_i \leq x_o} p(x_i)$$

Statistical Moments

- Descriptors commonly used to show statistical properties of a sample those indicative to population
  - (1) Central tendency;
  - (2) Dispersion;
  - (3) Asymmetry.

- Frequently used descriptors in these three categories are related to statistical moments

- Two types of statistical moments are commonly used in hydrosystem engineering applications:
  - (1) product-moments and
  - (2) L-moments.

Product-Moments

- rth-order <u>product-moment</u> of X about any reference point $X=x_o$ is defined,

for continuous case, as
$$E\left[(X-x_o)^r\right]=\int_{-\infty}^{\infty}(x-x_o)^r f_x(x)\,dx=\int_{-\infty}^{\infty}(x-x_o)^r\,dF_x(x)$$

for discrete case,
$$E\left[(X-x_o)^r\right]=\sum_{k=1}^{K}(x_k-x_o)^r\,p_x(x_k)$$

where $E[\cdot]$ is a <u>statistical expectation operator</u>.

- In practice, the first three moments (r=1, 2, 3) are used to describe the **central tendency, variability, and asymmetry**.

- Two types of product-moments are commonly used:
  - <u>Raw moments:</u> $\mu_r'=E[X^r]$ rth-order moment about the origin;
  - <u>Central moments</u>**:** $\mu_r=E[(X-\mu_x)^r]$ = rth-order central moment

Product-Moments

- Relations between two types of product-moments are:

$$\mu_r = \sum_{i=0}^{r} (-1)^i C_{r,i}\, \mu_x^i \; {}'_{r-i}$$

$$\mu'_r = \sum_{i=0}^{r} C_{r,i}\, \mu_x^i\, \mu_{r-i}$$

where $C_{n,x}$ = binomial coefficient = $n!/(x!(n\text{-}x)!)$

Main disadvantages of the product-moments are:

1. Estimation from sample observations is sensitive to the presence of <u>outliers</u>; and

2. Accuracy of sample product-moments deteriorates rapidly with increase in the order of the moments.

Mean, Mode, Median, and Quantiles

- Expectation (1st-order moment) measures central tendency of random variable *X*

$$E[X] = \mu_x = \int_{-\infty}^{\infty} x f_x(x)\,dx = \int_{-\infty}^{\infty} x\,dF_x(x) = \int_{-\infty}^{\infty} [1 - F_x(x)]\,dx$$

  - Mean (μ) = Expectation = λ_1 = location of the centroid of PDF or PMF.
  - Two operational properties of the expectation are useful:

for dependent $E\left(\sum_{k=1}^{K} a_k X_k\right) = \sum_{k=1}^{K} a_k \mu_k$ in which $\mu_k = E[X^k]$ for k = 1,2, …, K.

For <u>independent</u> random variables, $E\left(\prod_{k=1}^{K} X_k\right) = \prod_{k=1}^{K} \mu_k$

- <u>Mode</u> (x_{mo}) - the value at which its PDF is peaked. The mode, x_{mo}, can be obtained by solving

$$\left[\frac{\partial f_x(x)}{\partial x}\right]_{x=x_{mo}} = 0$$

Mean, Mode, Median, and Quantiles

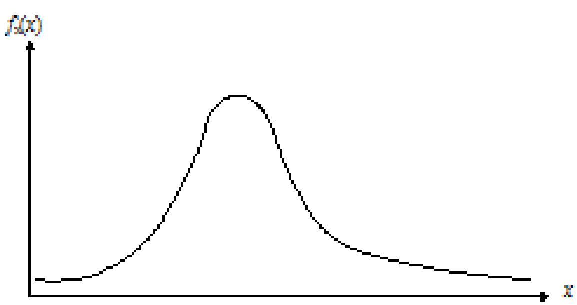
- Median (x_{md}) - value that splits the distribution into two equal halves, i.e,

$$F_x\left(x_{md}\right) = \int_{-\infty}^{x_{md}} f_x\left(x\right) dx = 0.5$$

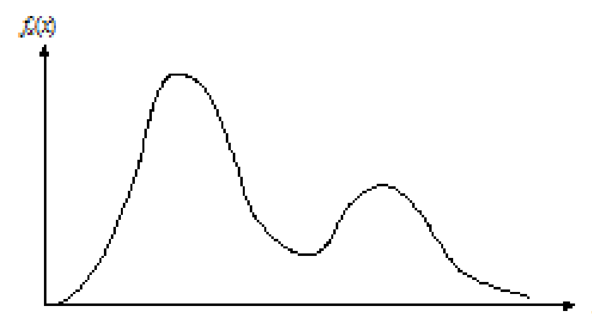
- Quantiles - 100pth quantile of a RV X is a quantity x_p that satisfies

  $P(X \leq x_p) = F_x(x_p) = p$

- A PDF could be uni-modal, bimodal, or multi-modal. Generally, the mean, median, and mode of a random variable are different, unless the PDF is symmetric and uni-modal.



(a) Uni-modal distribution



(b) Bi-modal distribution

Addis Ababa University
Institute of Technology (AAiT)

Variance, Standard Deviation, and Coefficient of Variation

- Variance is the second-order central moment measuring the spreading of a RV over its range,

$$Var[X] = \mu_2 = \sigma_x^2 = E\left[(X-\mu_x)^2\right] = \int_{-\infty}^{\infty} (x-\mu_x)^2 f_x(x)\,dx$$

- Standard deviation (σ_x) is the positive square root of the variance.

- Coefficient of variation, $\Omega_x = \sigma_x / \mu_x$, is a dimensionless measure; useful for comparing the degree of uncertainty of two RVs with different units.

- Three important properties of the variance are:
  - (1) $Var[c] = 0$ when c is a constant.
  - (2) $Var[X] = E[X^2] - E^2[X]$
  - For multiple independent random variables,

$$Var\left(\sum_{k=1}^{K} a_k X_k\right) = \sum_{k=1}^{K} a_k^2 \sigma_k^2$$

where a_k =a constant and σ_k = standard deviation of X_k, k=1,2, ..., K.

Skewness Coefficient

- Measures asymmetry of the PDF of a random variabl

- Skewness coefficient, γ_x, defined as

$$\gamma_x = \frac{\mu_3}{\mu_2^{1.5}} = \frac{E\left[(X - \mu_x)^3\right]}{\sigma_x^3}$$

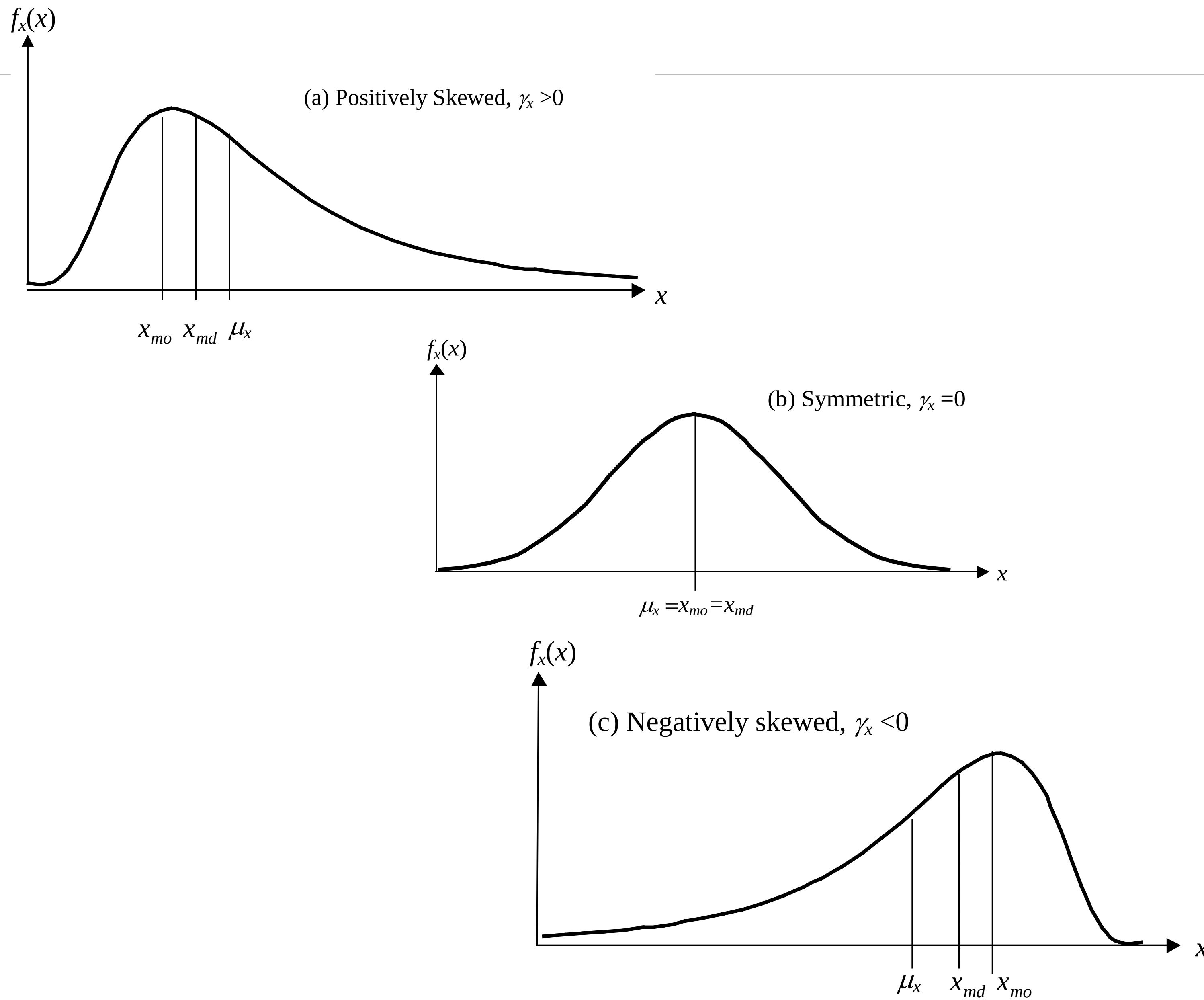
- The sign of the skewness coefficient indicates the degree of symmetry of the probability distribution function.

- Pearson skewness coefficient –

$$\gamma_1 = \frac{\mu_x - x_{mo}}{\sigma_x}$$

- In practice, product-moments higher than 3rd-order are less used because they are unreliable and inaccurate when estimated from a small number of samples

Relative locations of mean, median, and mode for positively skewed, symmetric, and negatively-skewed distributions.



Kurtosis (κ_x)

- Measure of the peakedness of a distribution.

- Related to the 4th central product-moment as

$$\kappa_x = \frac{\mu_4}{\mu_2^2} = \frac{E\left[(X - \mu_x)^4\right]}{\sigma_x^4}$$

- For a normal RV, its kurtosis is equal to 3. Sometimes, <u>coefficient of excess</u>, $\varepsilon_x = \kappa_x - 3$, is used.

- All feasible distribution functions, skewness coefficient and kurtosis must satisfy

$$\gamma_x^2 + 1 \leq \kappa_x$$

Product-moments of random variables

Moment	Measure of	Definition	Continuous Variable	Discrete Variable	Sample Estimator
First	Central Location	Mean, Expected value $E(X)=\mu_x$	$\mu_x = \int_{-\infty}^{\infty} x\, f_x(x)\, dx$	$\mu_x = \sum_{all\ x's} x_k\, p(x_k)$	$\bar{x} = \sum x_i / n$
Second	Dispersion	Variance, $Var(X)=\mu_2=\sigma_x^2$ Standard deviation, σ_x Coefficient of variation, Ω_x	$\sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f_x(x)\, dx$ $\sigma_x = \sqrt{Var(X)}$ $\Omega_x = \sigma_x / \mu_x$	$\sigma_x^2 = \sum_{all\ x's} (x_k - \mu_x)^2 P_x(x_k)$ $\sigma_x = \sqrt{Var(X)}$ $\Omega_x = \sigma_x / \mu_x$	$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$ $s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$ $C_v = s / \bar{x}$
Third	Asymmetry	Skewness Skewness coefficient, γ_x	$\mu_3 = \int_{-\infty}^{\infty} (x-\mu_x)^3 f_x(x)\, dx$ $\gamma_x = \mu_3 / \sigma_x^3$	$\mu_3 = \sum_{all\ x's} (x_k - \mu_x)^3 p_x(x_k)$ $\gamma_x = \mu_3 / \sigma_x^3$	$m_3 = \frac{n}{(n-1)(n-2)}\sum (x_i - \bar{x})^3$ $g = m_3 / s^3$
Fourth	Peakedness	Kurtosis, κ_x Excess coefficient, ε_x	$\mu_4 = \int_{-\infty}^{\infty} (x-\mu_x)^4 f_x(x)\, dx$ $\kappa_x = \mu_4 / \sigma_x^4$ $\varepsilon_x = \kappa_x - 3$	$\mu_4 = \sum_{all\ x's} (x_k - \mu_x)^4 p_x(x_k)$ $\kappa_x = \mu_4 / \sigma_x^4$ $\varepsilon_x = \kappa_x - 3$	$m_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum (x_i - \bar{x})^4$ $k = m_4 / s^4$

Some Commonly Used Distributions

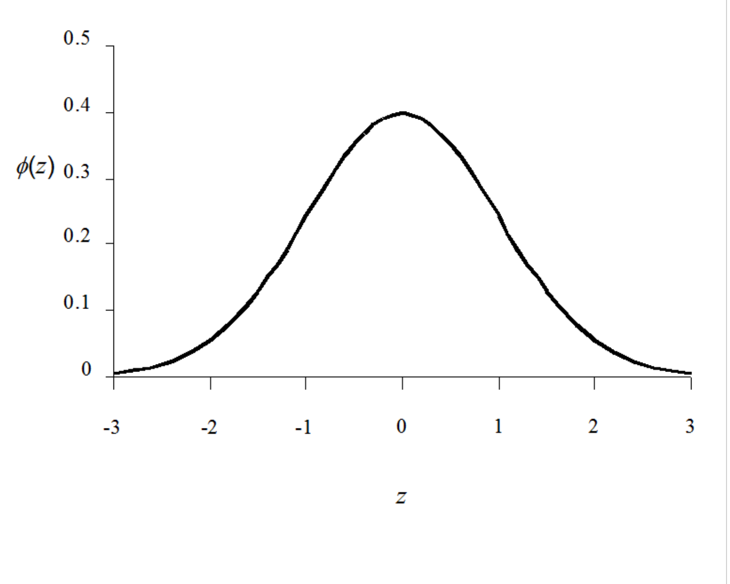
- **NORMAL DISTRIBUTION**

$$f_N\left(x \mid \mu_x, \sigma_x^2\right) = \frac{1}{\sqrt{2\pi}\ \sigma_x} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 \right], \text{ for } -\infty < x < \infty$$

Standardized Variable:

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

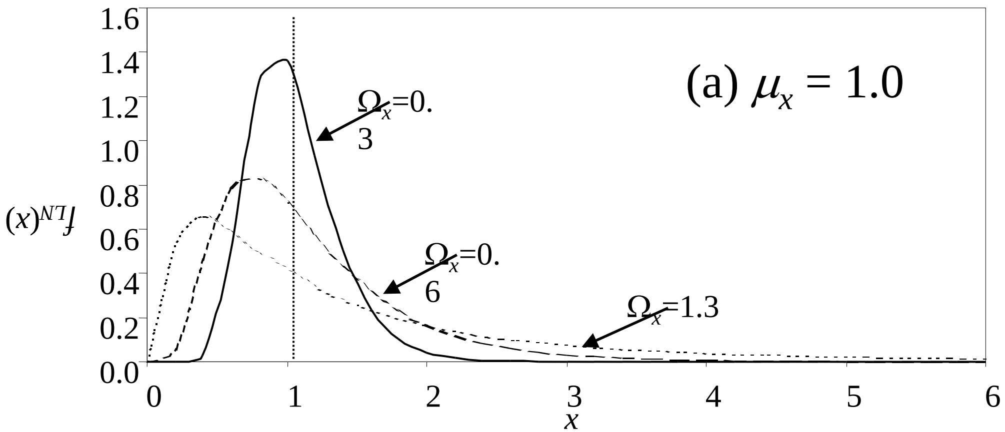
Z has mean 0 and standard deviation 1.

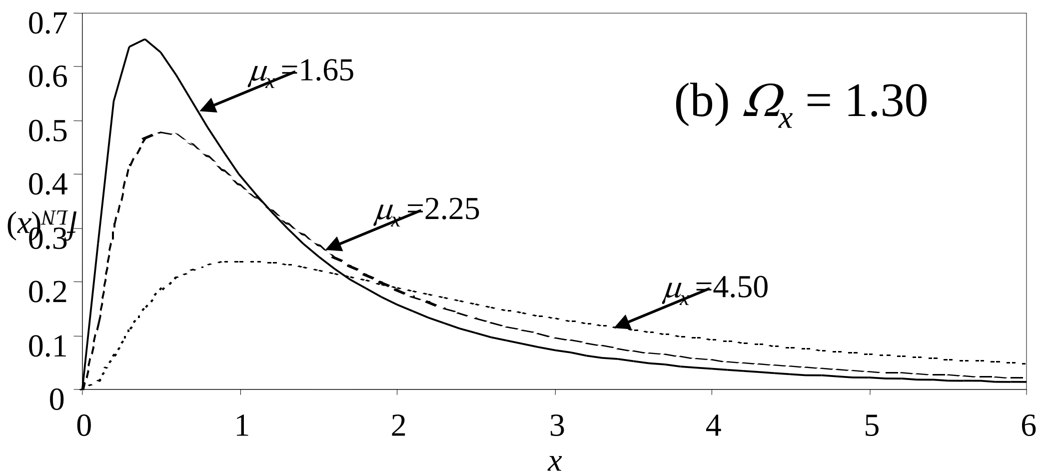


Some Commonly Used Distributions

- **LOG-NORMAL DISTRIBUTION**

$$f_{LN}\left(x \mid \mu_{\ln x}, \sigma^2_{\ln x}\right) = \frac{1}{\sqrt{2\pi}\,\sigma_{\ln x}\, x} \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu_{\ln x}}{\sigma_{\ln x}}\right)^2\right], x > 0$$



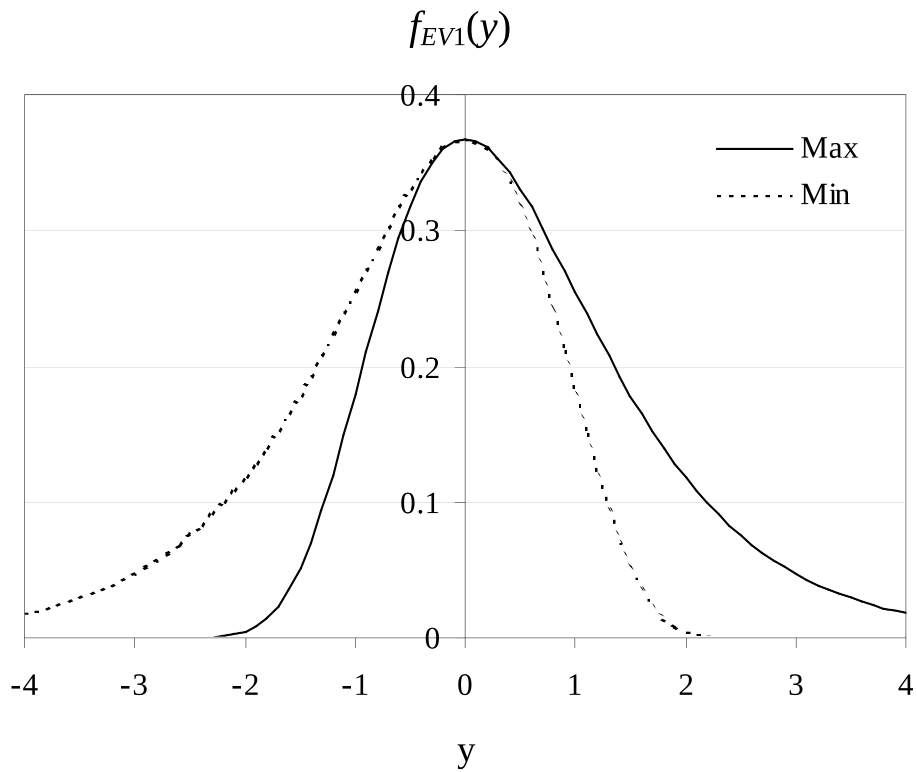


Some Commonly Used Distributions

- **Gumbel (Extreme-Value Type I) Distribution**

$$F_{EV1}(x\,|\,\xi,\beta) = \exp\left\{-\exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]\right\} \quad \text{for maxima}$$
$$= 1-\exp\left\{-\exp\left[+\left(\frac{x-\xi}{\beta}\right)\right]\right\} \quad \text{for minima}$$

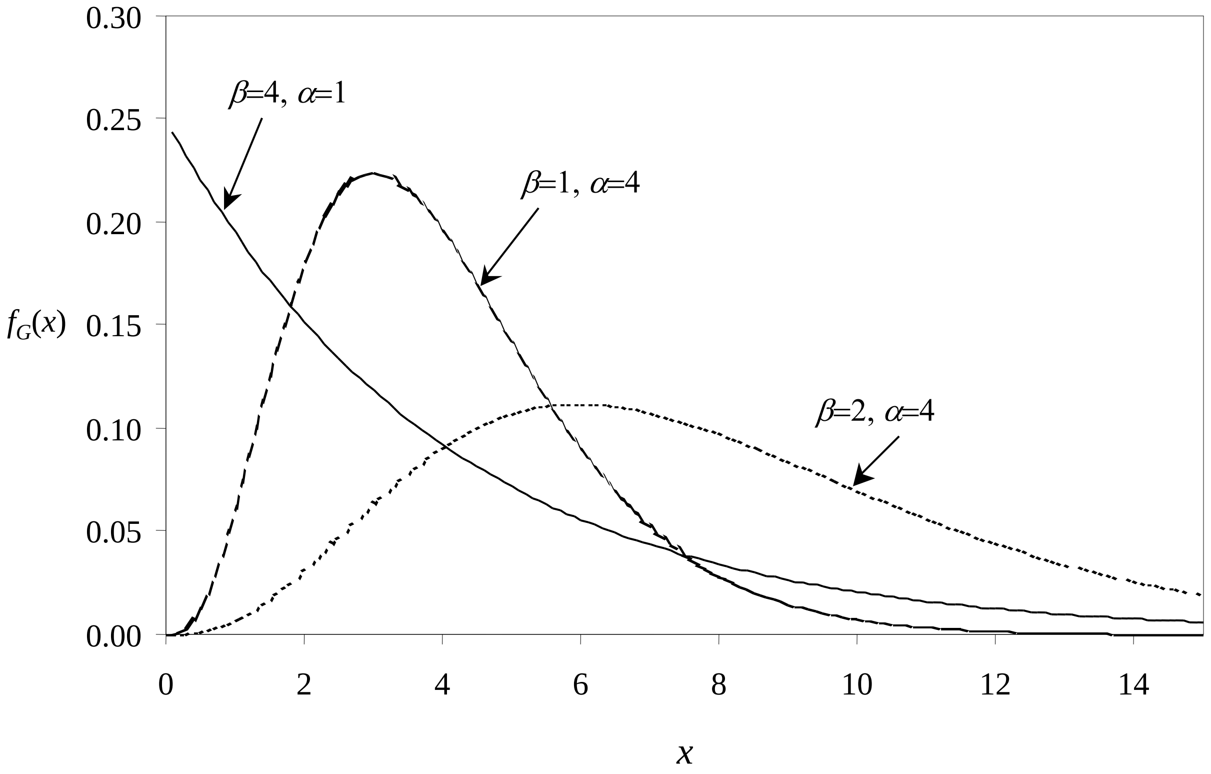
$$f_{EV1}(x\,|\,\xi,\beta) = \frac{1}{\beta}\exp\left\{-\left(\frac{x-\xi}{\beta}\right)-\exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]\right\} \quad \text{for maxima}$$
$$= \frac{1}{\beta}\exp\left\{+\left(\frac{x-\xi}{\beta}\right)-\exp\left[+\left(\frac{x-\xi}{\beta}\right)\right]\right\} \quad \text{for minima}$$



Some Commonly Used Distributions

Log-Pearson Type 3 Distribution

$$f_{P3}(x \mid \xi, \alpha, \beta) = \frac{1}{|\beta|\,\Gamma(\alpha)}\left(\frac{x-\xi}{\beta}\right)^{\alpha-1} e^{-(x-\xi)/\beta}$$



AAiT

Addis Ababa University
Institute of Technology (AAiT)

Concept of Risk

RISK = HAZARD * VULNERABILITY * AMOUNT

Hazard = **Probability of event with a certain magnitude**

Hazard Characteristic	Definition
Magnitude	Only those occurrences that exceed some common level of magnitude are extreme.
Frequency	How often an event of a given magnitude may be expected to occur in the long-run average.
Duration	The length of time over which a hazardous event persists, the onset to peak period.
Areal Extent	The space covered by the hazardous event.
Speed of Onset	The length of time between the first appearance of an event and its peak.
Spatial Dispersion	The pattern of distribution over the space in which its impacts can occur.
Temporal Spacing	The sequencing of events, ranging along a continuum from random to periodic.

AAiT

Addis Ababa University
Institute of Technology (AAiT)

Flow duration curve

- A cumulative frequency curve that shows the percentage of time that specified discharges are equaled or exceeded.

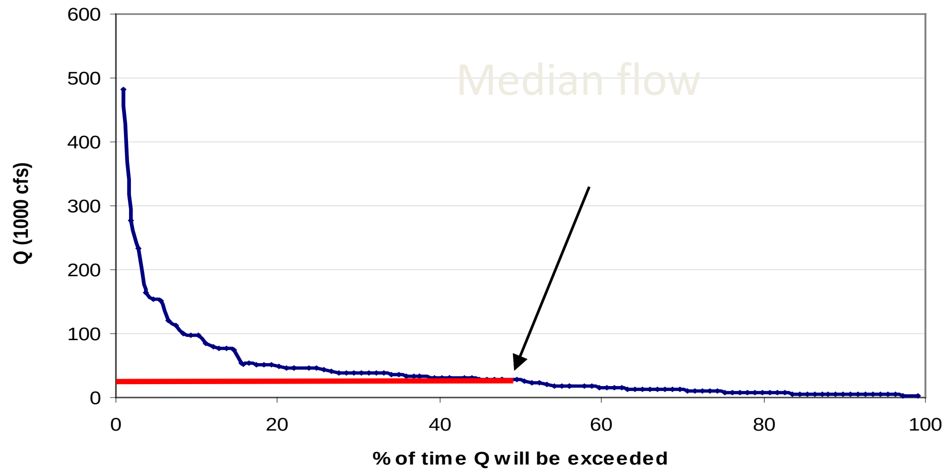
- **Steps**
  - **Arrange flows in chronological order**
  - **Find the number of records (N)**
  - **Sort the data from highest to lowest**
  - **Rank the data (m=1 for the highest value and m=N for the lowest value)**
  - **Compute exceedance probability for each value using the following formula**

$$p = 100 \times \frac{m}{N+1}$$

  - **Plot p on x axis and Q (sorted) on y axis**

Flow duration curve in Excel

	A	B	C	D	E
1	Year	Q	Q_{sorted}	Rank	p
2	1905	30.2	481	1	0.92
3	1905	113	276	2	1.83
4	1900	151	234	3	2.75
5	1901	28.7	164	4	3.67
6	1902	35.9	154	5	4.59
7	1903	33.7	151	6	5.50
8	1904	31.5	120	7	6.42
9	1905	52.9	113	8	7.34
10	1906	78.5	100	9	8.26
11	1907	28.1	98.2	10	9.17
12	1908	100	97.6	11	10.09
13	1909	29.7	84	12	11.01
14	1910	27.4	78.5	13	11.93



DESIGN FLOOD

- Flood adopted for the design of a structure
  - Spillway Design Flood (SDF): the flood specifically compute for the design of a spillway of a storage structure
  - Standard Project Flood (SPF): the flood that would result from a severe combination of meteorological and hydrological factors that reasonably applicable to the region
  - Probable maximum Flood (PMF): the extreme flood that physically possible in a region as a result of severe most combinations of meteorological and hydrological factors
- The criteria used for selecting the design flood for various hydraulics structures vary from one country to another

AAiT

Addis Ababa University
Institute of Technology (AAiT)

ERA Standard for design frequency

Table 2-1 Design Storm Frequency (Yrs) by Geometric Design Criteria

Structure Type	Geometric Design Standard			
	DS1/DS2	DS3/DS4	DS5/6/7	DS8/9/10
Gutters and Inlets*	10/5	2	2	-
Side Ditches	10	10	5	5
Ford/Low-Water Bridge	-	-	-	5
Culvert, pipe (see Note) Span<2m	25	10	5	5
Culvert, 2m<span <6m	50	25	10	10
Short Span Bridges 6m<span<15m	50	50	25	25
Medium Span Bridges 15m<span<50m	100	50	50	50
Long Span Bridges spans>50m	100	100	100	100
Check/Review Flood	200	200	100	100