It has been noted in Chapter 10 that an important problem in hydrometeorology is to estimate the frequency of maximum rainfall or design storm rainfall of a specified duration that is likely to occur at a selected station or in a selected river catchment for designing hydraulic structures subject to flooding, such as bridges, culverts and dams. In the same chapter it was also mentioned that the statistical method is one of the four methods for estimating design storms. In this method an estimate of the frequency with which a given magnitude of rainfall may be exceeded in the future is based upon the study of the frequency with which it has been exceeded in the past in a probabilistic sense.

In design problems, extreme values are usually used. This is based on the philosophy that if the designed structure can withstand the highest value in a year, it can withstand all other values. Two types of data are generally selected: (1) the annual extreme value series and (2) the partial duration series. This chapter presents some of the applications of statistical methods by which extreme rainfall data series may be quantified and presented in a standard frequency framework. Before proceeding with the application of statistical analysis we define some useful terms.

A hydrometeorological variable \( X \) is one which measures the magnitude of some element of hydrological cycle, say rainfall. The value of \( X \) changes with the individual observation. A variable which can theoretically assume any value between two given values is called a continuous variable, otherwise it is called a discrete variable. A 30-year continuous record of rainfall at a place is a sample of the rainfall history at that place. The set of all records of the same station under a fixed set of conditions is called a population. Statistical theory is based on the concept that the sample is assumed to be representative of the population. In order to draw inferences about a population, the data in the sample must be random, independent and homogenous. When every value in the population has an independent chance of being chosen, the sample is called a random sample. A degree of
independence varies with the nature of the data being considered. Successive daily rainfalls or stream flows are not independent because a stream that is high one day will probably be high the next day. Monthly or extreme daily rainfall data are much more independent of each other. Homogenous data belong to the same population. Homogeneity and randomness of data may be tested through the use of statistical tests described later in this chapter.

11.1 DATA SERIES FOR FREQUENCY ANALYSIS

11.1.1 Annual Extreme Value Series

The highest observed value (daily rainfall) at any rainfall station during a year is called the annual extreme value. The annual extreme value series, therefore, consists of only one highest value for each year of record. Thus, for \( N \) years of data an annual extreme value series will consist of \( N \) values one for each year.

11.1.2 Partial Duration Series

This is also called partial series. The annual extreme series ignores the second and third highest values of a year which may be higher than many of the extreme annual values of other years in the series. This type of situation is overcome by selecting all the values above an arbitrary threshold value. The series obtained in this manner is known as a partial duration series. Therefore, a partial duration series contains the largest values greater than a base value. The lowest value is generally chosen as a base value in the annual extreme data series. In a complete partial duration series if the lowest values are truncated so that the number of values in the series is equal to the number of years of record then this truncated series is called the annual exceedance series.

Both these types of series or selected data are used in frequency analysis. It is important to mention that if the time interval between the two values is less than a year, the seasonal variation may introduce inhomogeneities in the data series. However, if the data are selected only from a particular season or month, the homogeneity may largely be maintained. Since the partial series is not a true series, the values selected may not be independent of each other and as a result such a series cannot be considered to be completely homogeneous and independent. Nevertheless, the use of both types of the selected data for frequency analysis gives almost similar estimates for frequency of more than 10 years. The difference is significant in the lower frequency estimates. Langbein (1949) studied these two types of selected data series and suggested an empirical relationship between their frequencies based on annual and partial duration floodseries:

\[
T_{AM} = \frac{1}{1 - \exp\left(\frac{1}{T_{PD}}\right)}
\]