NATIONAL INSTITUTE OF HYDROLOGY, ROORKEE
WORKSHOP ON
FLOOD FREQUENCY ANALYSIS

LECTURE-1

ROLE OF STATISTICS IN HYDROLOGY

OBJECTIVES

This lecture introduces the objectives of statistical analysis in hydrology and describes the univariate modelling process using frequency analysis approach.

1.1 INTRODUCTION

Flood estimates are required for the design and economic appraisal of a variety of engineering works, including dam spillways, bridges and flood protection works. Flood estimates are also required for the safe operation of flood control structures, for taking emergency measures such as maintenance of flood levees, evacuating the people to safe localities etc. Two main approaches are available for flood estimation viz., deterministic approach and statistical approach. Deterministic approach assumes that input, say, the precipitation is related to the output in a predefined manner and there is no uncertainty involved in arriving at the output, say the discharge. Whereas the statistical approach treats the inter-relationship between processes as governed by theory of statistics. The inter-relationship between processes is established through the measure of correlation. The processes considered may be multivariate or univariate. For example, the rainfall-runoff process may be considered as multivariate while, the cosideration of maximum annual peak series falls under univariate process. Flood frequency analysis deals with univariate process comprising of maximum peak flow values. Before discussing the frequency analysis it is necessary to distinguish between the terminologies of prediction and forecasting from the consideration of their field use.

Estimation of floods arise in hydrology from two different aspects of prognostication. The first which is usually called forecasting, involves the estimation of the values of some hydrological variable at a prescribed future instant or a forecast of the time when a particular value will occur. Forecasting the level of a flood discharge or a water level in a river some days ahead of the event is an obvious example. The second aspect, usually known as prediction is concerned not with the value of a single future event, but with the frequency of occurrence of some prescribed critical condition. For example, how often a particular level will be exceeded. Problems of first category arise most directly in the operation of hydrological controls in the broadest sense including flood warning. Problems of second category are associated with the designs rather than the operation of such controls. Design flood estimation may also be obtained through deterministic approaches say rainfall-runoff models. Therefore essentially predication deals with "how often" problem while the forecasting deals with "when" problem. The former problem is the subject matter of this workshop.

1.2 OBJECTIVES OF STATISTICAL ANALYSIS IN HYDROLOGY

The random variability of such hydrologic variables as streamflow and precipitation has been recognised for centuries. The general field of hydrology was one of the first areas of science and engineering to use statistical concepts in an effort to analyse natural phenomena. The use of statistics in hydrology provides the information about various parameters and distribution of random variables of importance to design and operation of structures. These parameters and distributions are estimated as approximations from the available data because they cannot be determined exactly. This lecture provides an introduction to the application of statistics in hydrology, especially with reference to estimation of design floods using frequency analysis approach.

1.3 CHARACTERISTICS OF STATISTICAL ANALYSIS

Hydrologic data are mostly available as samples of limited sizes. Statistics is the main discipline enabling the extraction of needed information from data and the derivation of conclusions about the characteristics of hydrologic random variables. Statistical estimates are numerical properties of samples. They are necessary in statistical modelling, or for direct use in hydrology. To be effective in application of statistics in hydrology the civil engineer or hydrologist must understand the fundamentals of statistical methods which are employed in existing hydrologic techniques.

1.3.1 Population and Sample

Hydrologic random variables in most cases have infinite populations. The infinite population possesses an infinite number of potential observations. Any survey or any observational attempt cannot exhaust all the values of an infinite population. For large finite population the economy of information procurement forbids an exhaustive approach of obtaining data for the whole finite population.

Assuming that a river basin and its climate stay unchanged forever, the entire population of its storm precipitation, storm runoff etc. cannot be observed. These are infinite populations. However, the population of all river basins of 10 km^2 within a given range of latitude and longitudes, properly defined and without an overlapping of areas within this location is a finite population.

Finite amounts of data collected for a random variable either from an infinite or a finite population constitute the sample. It is conventionally said that the sample is drawn from the respective population. From the practical point of view, samples may be classified as small, medium or large, although there is no generally accepted convention that define these sizes and these words have relative meanings for various problems.

1.3.2 Descriptive and Inferential Statistics

A sample of random variable can be analysed in two basic ways. In the first way, consideration and investigation of the sample data are without any reference to the properties of their population. This approach is usually referred to as the descriptive statistics. In the second way, inferences are made from the sample data about the properties of their population. This approach is referred to as the inferential statistics, which is based on the use of probability theory. Hydrologic analyses use the inferential statistics to estimate the poperties of population of a variable, using the available sample information, for the purpose of estimating the further expected hydrologic characteristics.

1.4 THE UNIVARIATE MODDELLING PROCESS

A model is conceptually a substitution for a system. By system we mean a real world concept which is composed of input elements, a process of conversion, and the generation of output elements. A simple hydrologic system would be rainfall as input, the watershed as a converting device, and the output would be streamflow. This is a real world system. We model this system with a mathematical substitution. The mathematical substitution is a workable simplification of the system. It is used to study the system. We can answer questions about output if we change the input or if we change the form of the process that converts the input to the output.

The derivation of mathematical models of applied statistics is usually connectcd with various prediction problems. In the process of solving practical statistical problems four sequential steps are important :

(1) Selection of the mathematical model,

(2) Estimations of its parameters,

(3) Check on the reasonableness of the model by various tests, and

(4) Derivation of conclusions from the model that affect the solution of the proposed problem.

Univariate models are the simplest possible hydrologic models. A univariate model can be only one of the statistical frequency distributions. Since there is only one variable there is no relationship to another variable. Infact there is no input output relation but that system generates a random output. It is uncontrolled in the sense that we cannot forecast what any particular output element will be at any particular time. The output is controlled, however to that degree which causes it to take a normal, a gamma, or some other distribution. It we know streamflow today, we cannot, with a univariate model, use that streamflow to forecast streamflow tomorrow. A'thougn we do not have relationship form for our model, it is very important to understand that a form exists, and that form is the frequency distribution which we have specified as controlling the probable range of random events under study.

A simple descriptive example may serve to solidify the concept of a univariate model. Suppose that we have taken the maximum temperature each day during the month of May at some meteorological station. The maximum temperature each day is not constant, it varies between certain limits that cluster fairly closely around the mean daily maximum temperature for May. We may plot a graph of our May maximum temperatures for, say, 20 years. Suppose that graph makes us believe that the distribution of these maximum temperatures is essentially normal distribution. By choosing the normal distribution we have specified the form of our model. Now, additionally, a normal distribution is completely quantified when we evaluate its mean and its standard deviation. Suppose we say that we will take the mean daily maximum May temperature as the mean of our model and that we will take the standard deviation of our sample of daily maximum May temperature as the standard deviation of our model. We have then accomplished the second step, which is evaluation or quantification of our model. The third step involves the use of statistical hypothesis test to test the validity of the assumed population, for example the Chi-square, or kolmogorov-Smirnov or D-statistic tests may be used to compare the fitting of the sample data to the assumed population. After the population has been identified and substantiated using the measured data and statistical tests, probability statements about the random variable can be made. The output from a univariate analysis can be either a value of the random variable for some level of probability or an expected probability for a specified value of the random variable. In addition to probability statements, one is usually interested in the accuracy of the probability statement. The accuracy can be assessed using confidence intervals and tolerance limits.

1.5 ELEMENTS OF FLOOD FREQUENCY ANALYSIS

The problem of flood frequency prediction i.e., estimation of the relationship between the magnitude of peak flow and its corresponding return period is a central one in the field of applied hydrology. Attempts to solve this problem are usually based on (a) the analysis of a record of peak flow data at the site in question and (b) use of previously established relationship between the characteristics of other catchments in the region and the parameteric values of the corresponding magnitude return period relationship. The latter approach is known as regional frequency analysis used for estimating floods at sites where there is a very short peak flow record or no record available. This workshop discusses both these approaches of flood frequency analysis.

1.5.1 Assumptions Involved

Three assumptions are fundamental to the flood frequency analysis. They are :

(a) Sample is representative of population,

(b) Peak flows are independent, and

(c) All the peak flows in a given sample belong to a homogeneous population.

(a) Sample is representative of population :

Hydrologic data are mostly available as samples of limited sizes. Using statistical principles we extract the needed information from the available sample data and conclude about the characteristics of tne population. Since any survey or any attempt cannot exhaust all possible events of a variable, we assume that the sample is representative of population.

(b) Independeuce of peak flows :

We assume that that the sample of peak flows available are independent of each other and they are assumed to be evolved from a purely random process.

(c) Homogeneity of peak flows :

When a series of events arranged in time show no systematic variations in time (e,g., a seasonol variation or an increasing or diminishing trend) so that we may say that the probability of an event in a period of t is independent of the location of t, the series is said to be homogeneous The factors which affect the homogeneity of peak flows are the development in the catchment over time such as deforestation, urbanisation, flood control works, earthquakes etc.

1.6 REMARKS

The workshop lectures have been developed considering that the above assumptions have been followed by the given flood peak series. These lectures cover all the univariate modelling steps described earlier. Further there is a lecture devoted to the assessment of risk from the flood frequency analysis consideration. Also there is a lecture devoted to the regional frequency analyis for the purpose of esti~ ating design flood at ungauged sites or at sites with very short records.

References :

Beard, L.R., (1962), 'Statistical Methods in Hydrolgy', Corps of Engineers, Sacramento, CA.

Haan, C.T., (1977), 'Statistical Methods in Hydrology; 'The Iowa State University Press, Ames Iowa.

McCuen, R.H., and W.M. Snyder (1986), 'Hydrologic Modelling-Statistical Methods and Applications', Prentice-Hall, New Jersey.

Nash, J.E., (1980), 'Applied Hydrology : Part 2-Statistical', Lecture notes of the International Post-Graduate Course on Engineering Hydrology, University College, Galway, Ireland.

Natural Environment Research Council (1975), 'Flood Studies Report-Volume I : Hydrological Studies,' London.