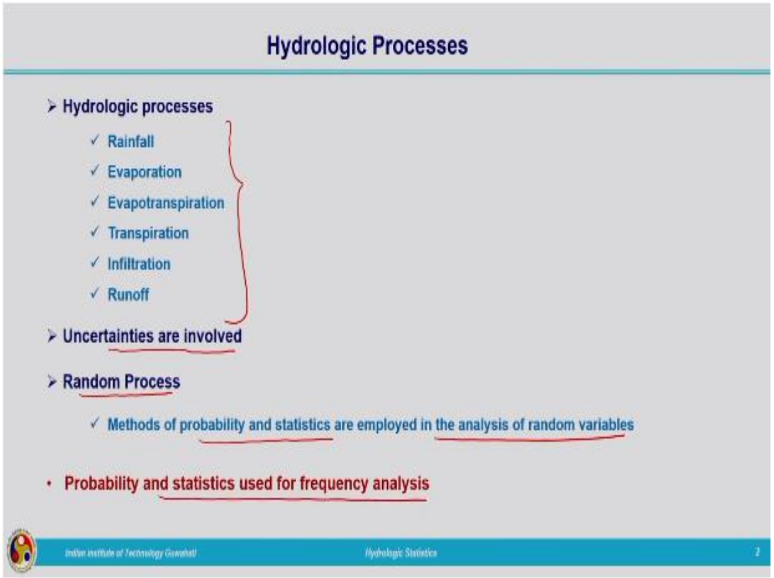
Engineering Hydrology
Dr. Sreeja Pekkat
Department of Civil Engineering
Indian Institute of Technology, Guwahati
Module: 6
Lecture 72: Hydrologic Statistics

Hello all, welcome back, today we are going to start with a new module on Hydrologic Statistics. We have completed already five modules related to different hydrologic processes and also hydrologic analysis. Specifically, if you are talking about hydrologic analysis we were dealing with the hydrograph analysis. Let us start with the module on hydrologic statistics today. I will not be covering in depth related to probability and statistics, in the minimal way which are required for engineering hydrology, only will be included in this module.

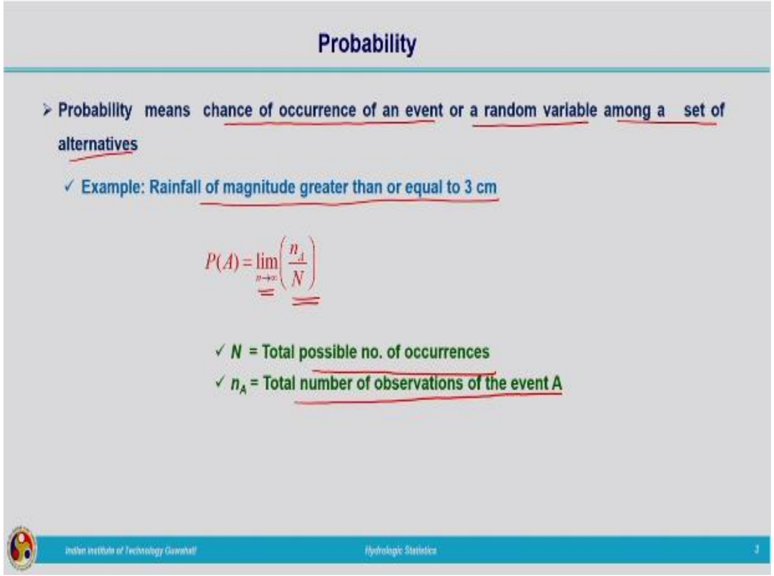
(Refer Slide Time: 1:30)



We will start with the hydrologic processes which we have already covered. Different hydrologic processes are rainfall, evaporation, evapotranspiration, transpiration, infiltration and runoff. Now, when we talk about these hydrologic processes, certain uncertainties are involved with this, these uncertainties may be arising due to the characteristics of the variables which are causing this particular process.

The variables which are involved in the process called rainfall itself are stochastic in nature. That is why these processes can be modeled as random process and hence there is a need of methods of probability and statistics for the analysis of random variables. Probability and statistics are used for frequency analysis, this is very, very important as far as hydrologic analysis is concerned.

(Refer Slide Time: 2:40)



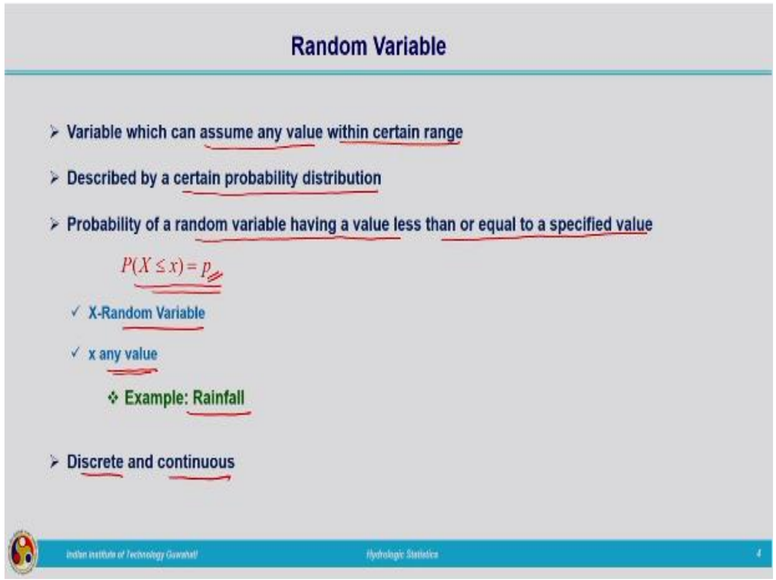
Let us start with probability. Probability means the chance of occurrence of an event or a random variable among a set of alternatives. For example, if you are talking about rainfall, rainfall of magnitude ≥ 3 cm, whether we will be able to state that tomorrow at 10 am there will be a rainfall of exactly 3 cm, it is impossible. So, these type of events or hydrologic variables can be considered as random variable.

So, the probability of occurrence of a rainfall of magnitude ≥ to 3 cm can be mathematically written as

$$P(A)=\lim_{n\to\infty}\left(\frac{n_A}{N}\right)$$

N is the total possible number of occurrences and n_A is the total number of observations of the particular event (that is rainfall magnitude ≥ 3 cm).

(Refer Slide Time: 4:29)



Now, coming to random variable. What is meant by random variable? Just as I have explained with the case of a rainfall, random variable is the variable which can have any value within certain range. We cannot exactly point out at a particular location at a particular time this will be the rainfall which is going to occur tomorrow or any date, it can be any value within certain range.

Those type of variables can be described by a certain probability distribution. What is meant by probability distribution, we will come to it, but these random variables can be expressed by means of certain probability distribution. So, probability of a random variable having a value less than or equal to a specified value, we have seen the expression and that is denoted by

$$P(X \le x) = p$$

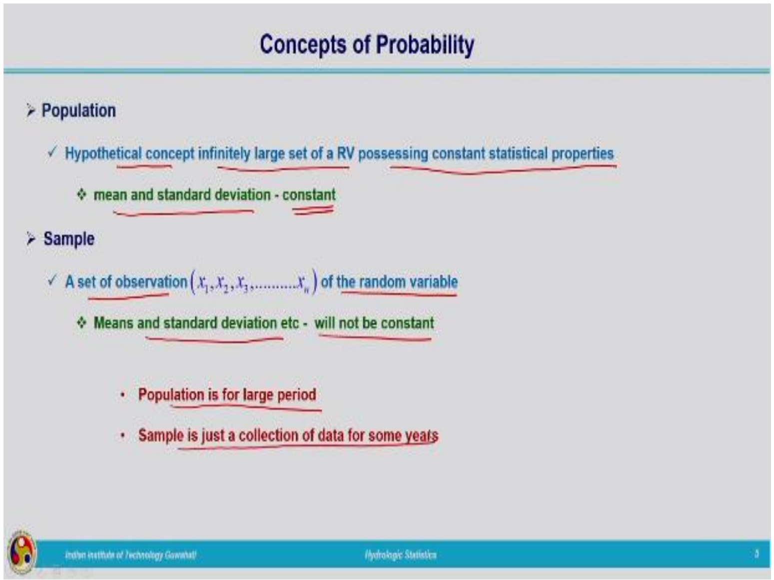
In this expression X is the random variable and x is any value corresponding to this particular random variable. Example, it can be rainfall, it can be stream flow. For example, if you are talking about stream flow, peak flow in a river, if you are considering for last so many years based on that we can express the value corresponding to future year that it will be within certain range, we cannot exactly tell that this will be the peak stream flow or peak flow in a river at a particular location. So, that way stream flow, rainfall all these hydrologic variables can be considered as random variable.

Random variables can be of two types, discrete random variable and continuous random variable. If you are talking about the number of rainy days that can be expressed as an integer, those are discrete values, number of rainy days in a particular month it can be represented by means of integer. So, that type of variables are discrete random variables.

Now, if you are talking about the amount of rainfall, amount of peak flow, these things cannot be expressed by means of a single value, it will be within a range. Sometimes it will be having fractional values also, decimal values it will be taking. So, those type of random variables are termed as continuous random variables.

Whether there will be a flood chances are 0 or 1, if there is a flood, the probability will be represented by means of 1 and if there is no flood it can be represented by means of 0. So, these are represented by certain numbers, so those are discrete random variables and which are represented by means of a range of values, that particular value can be within that particular range, then it is termed as continuous random variable.

(Refer Slide Time: 8:16)



Before going to the hydrologic statistics, let us have the preliminary understanding about probability and statistics, then we will deal with the hydrologic variables. So, first we need to have the understanding about population and sample.

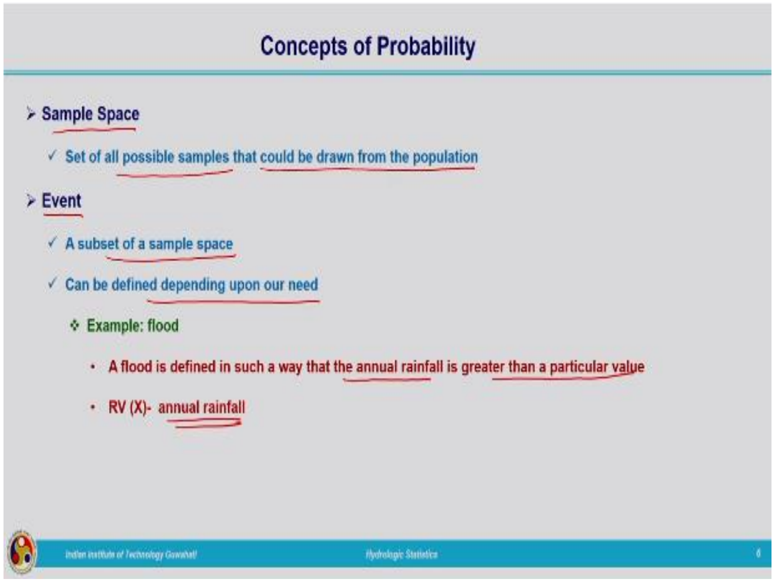
Population is a hypothetical concept which is used for an infinitely large set of a random variable possessing constant statistical properties. For example, if we are carrying out the rainfall analysis related to annual maximum rainfall value, we will be considering certain number of years data, for example, 30 years data. But population is the one which is dealing with large set of data which has occurred in the past and which is going to happen in the future, also for which the statistical properties (mean, standard deviation etc.) are constant, there would not be any change in the statistical values. That is why it is termed as a hypothetical concept.

Sample is the set of observations of random variables. If you are representing set of observations by $\left(x_1, x_2, x_3, \ldots\ldots\ldots. x_n\right)$, for example, if I am considering annual maximum rainfall for the last 30 years, so 30 data points will be there (every year one maximum value) for further analysis, that is why sample is considered as a subset of population.

Population on the other hand, we cannot specify how many years are included in that, all the data points related to annual maximum rainfall which has already occurred in the past and also the values which are going to happen in the future is included in population, that is why the statistical properties of this population is constant, but in the case of sample it is not constant. Means and standard deviation etc. will not be constant.

That is if you are considering a sample, for that a certain value of mean and standard deviation will be there and another 30 years sample you are selecting from the population, for that different standard deviation and mean will be there. We cannot expect different samples having the same standard deviation and mean, but in the case of population it will be having the constant value. Population is for large period but sample is just a collection of data for some years, as I have explained with 30 years data of annual maximum rainfall.

(Refer Slide Time: 11:37)

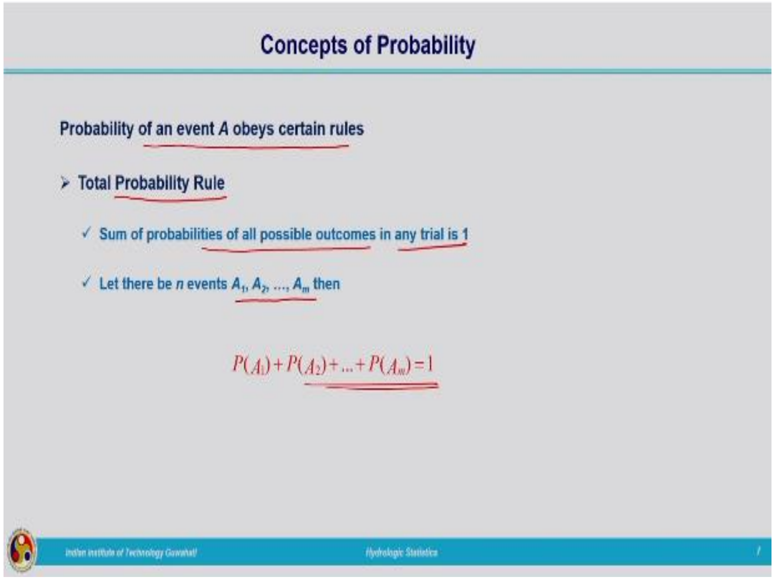


Next term you should understand is sample space. Sample space is the set of all possible samples that could be drawn from the population. For the last 30 years if I am considering annual maximum rainfall data, in the similar way so many samples can be drawn out of the population. These set of all possible samples are termed as sample space.

Now, next term which we commonly use here in this subject is event. Event is a subset of a sample space, it can be defined depending upon our requirement. For example, flood, we need to find out probability of occurrence of a flood in the coming year. So, the event is occurrence of flood, in the similar way it can be occurrence of a drought, it can be occurrence of any other event. So, that way depending on our requirement we will be defining a particular event.

A flood is defined in such a way that the annual rainfall is greater than a particular value. In the similar way if you are defining drought for which the annual rainfall will be less than or equal to particular value, this depends on our data and also our experience. Random variable X is annual rainfall, this annual rainfall which we will be considering should be greater than a particular value in order for a flood to happen. In the similar way it should be less than certain value then that particular year can be considered as drought year. So, these are some examples of events.

(Refer Slide Time: 13:47)

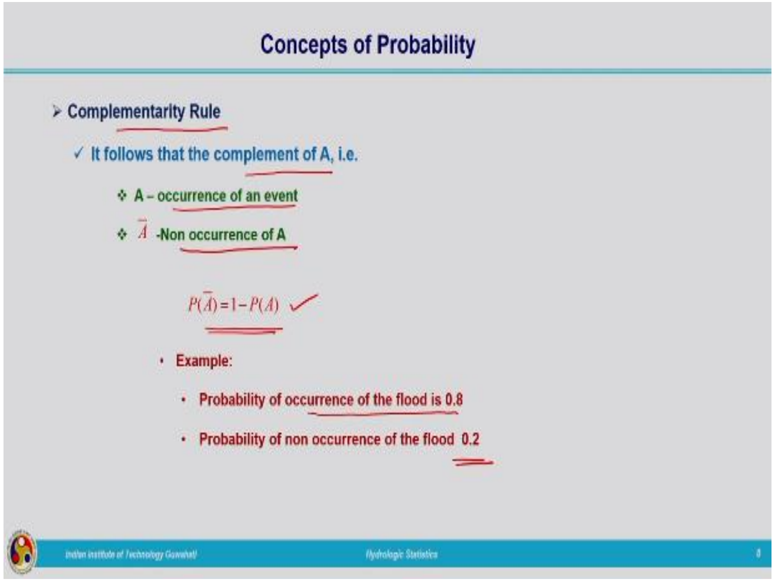


Now, for carrying out the analysis we need to be thorough with the different rules of probability. Probability of an event *A* obeys certain rules. First one is, total probability rule. These rules you have already studied in the topic of probability, so just let me write it down. Total probability rule states that sum of probabilities of all possible outcomes in any trial is 1. Total probability is equal to 1, it cannot be more than 1. Let there be *n* events represented by A_1, A_2, …, A_m, then

$$P(A_1)+P(A_2)+\ldots+P(A_m)=1$$

This is the law of total probability.

(Refer Slide Time: 14:40)



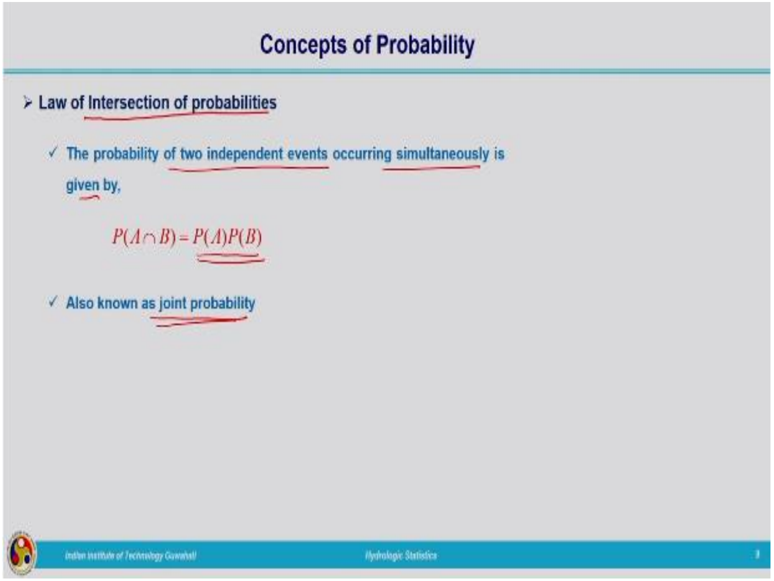
Second rule is complementarity rule. It follows that the complement of A, an event A we are considering corresponding to that event there will be a complement that is we are considering an event called flood, occurrence of flood. For that particular event called flood, there is a case with non-occurrence of flood, that is what is termed as complement of event A. A is occurrence of an event and A compliment is nothing but the non-occurrence of event, occurrence of a flood and non-occurrence of a flood.

If we are having the probability of occurrence of an event ($P(A)$) then we can compute the probability of non-occurrence of an event by making use of the formula

$$P(\bar{A}) = 1 - P(A)$$

That is the sum of occurrence of an event plus non-occurrence of an event is equal to 1, total probability law is that total probability should be equal to 1. This is the rule of complementarity. So, for example, if you are talking about probability of occurrence of the flood is 0.8, we can easily compute the probability of non-occurrence of the flood as 0.2 by making use of the rule of complementarity.

(Refer Slide Time: 16:20)



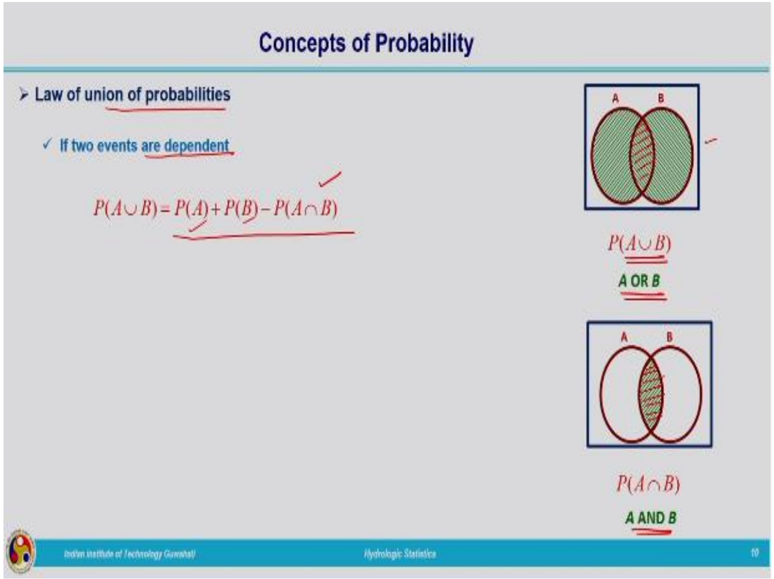
Now, the third rule is law of intersection of probabilities. The probability of two independent events occurring simultaneously is given by the law of intersection of probabilities that is

$$P(A \cap B) = P(A)P(B)$$

That is probability of A and B can be written as $P(A)$ multiplied by $P(B)$.

These are, A and B are two independent events, then the probability of occurrence of these independent events together or simultaneously can be obtained by multiplying the probabilities of individual events, this is also known as joint probability.

(Refer Slide Time: 17:32)

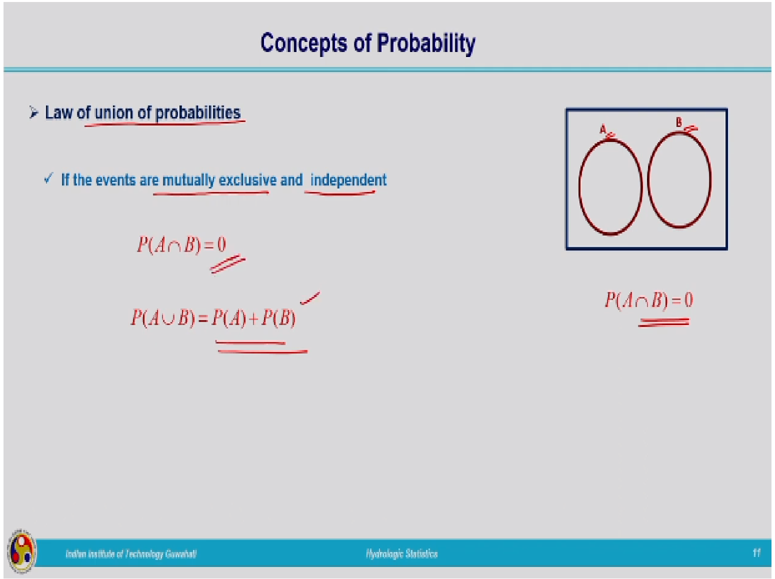


Now, the next Law is related to Union of Probabilities, A union B. If two events are dependent on each other, so that can be represented by means of $P(A \cup B)$. We can make use of Venn diagram for representing that, two events are there, event A and event B. So, the two events are dependent, so, the $P(A \cup B)$ is also termed as probability of A OR B. In the case of $P(A \cap B)$, it is termed as P of A AND B, in the case of union it is A OR B.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

That is probability of event A plus probability of event B minus probability of events A and B. Why we are subtracting this particular term, you look at this Venn diagram, A completely we have taken, B also completely we have taken, twice we are incorporating this central part represented by $P(A \cap B)$, that is why we are subtracting this $P(A \cap B)$ from $P(A) + P(B)$. By means of Venn diagram we can mark $P(A \cap B)$ here like this, this is event A, this is event B, $P(A \cap B)$ can be marked by this shaded region.

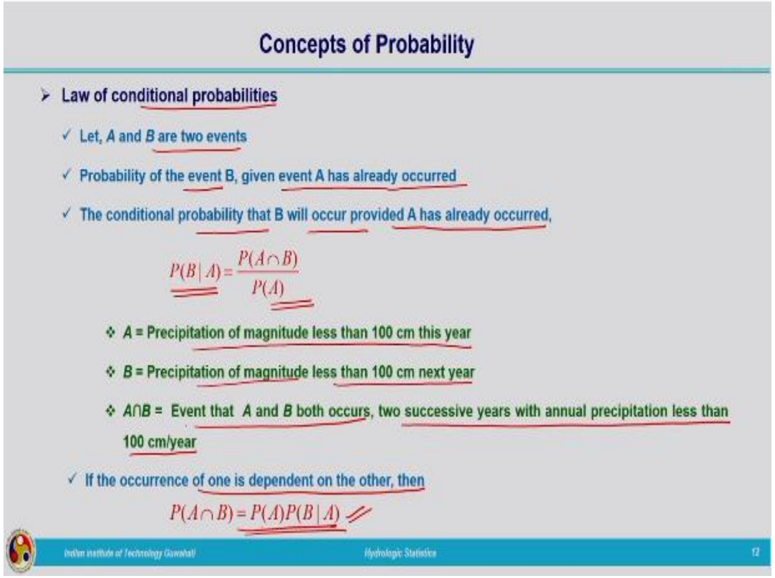
(Refer Slide Time: 19:27)



Now, based on the principles which we have seen before law of union of probabilities can be written in different cases. If the events are mutually exclusive and independent, that is the events can be represented by means of Venn diagram, event A and even B these are mutually exclusive and independent, then we can write $P(A \cap B) = 0$.

Then $P(A \cup B) = P(A) + P(B)$

(Refer Slide Time: 20:37)



Now, next law is related to law of conditional probability. Law of conditional probability is very important when we carry out hydrologic analysis, this is based on Bayes theorem, so that much we are not going in detail in this particular undergraduate course, but still fundamentally what it is you should have the flavor of that.

So, let A and B are two events, probability of the event B given event A has already occurred. Two events are there represented by A and B, we are going to calculate the probability of occurrence of a particular event but there is a condition that the other event has already occurred. The conditional probability that event B will occur provided A has already occurred is written by

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

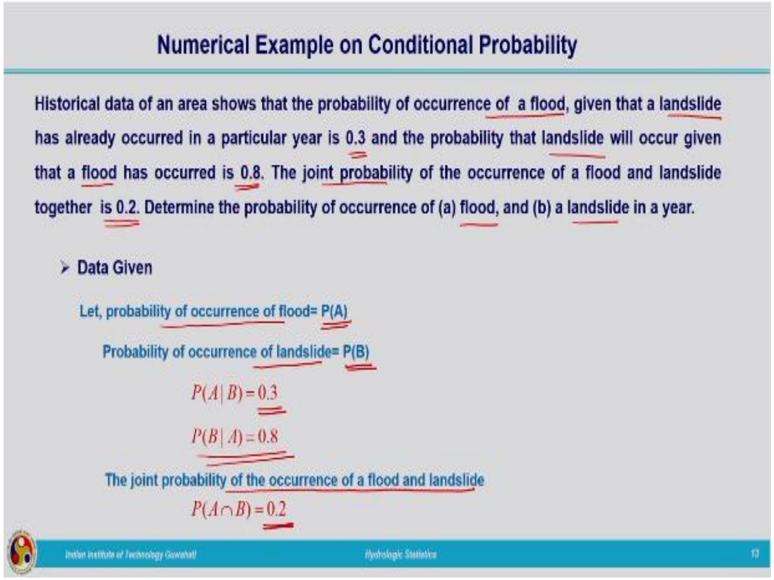
One example we can take, event A is the case with precipitation of magnitude less than 100 cm this year and event B is precipitation of magnitude less than 100 cm next year. $A\cap B$ is event that A and B both occurs, that is two successive years with annual precipitation less than 100 cm per year. $P(B|A)$, that is A has already occurred, the precipitation of magnitude < 100 cm has already occurred this year, what is the probability of occurrence of a precipitation next year which is < 100 cm in the coming year.

So, in such cases we have to go for making use of the formula related to conditional probability.

If the occurrence of one is dependent on the other, then

$$P(A \cap B) = P(A)P(B \mid A)$$

(Refer Slide Time: 24:22)



Now, for understanding the concept of conditional probability we can solve one numerical example.

Q- Historical data of an area shows that probability of occurrence of a flood, given that a landslide has already occurred in a particular year is 0.3 and the probability that landslide will occur given that a flood has occurred already is 0.8. The joint probability of the occurrence of a flood and landslide together is 0.2. Determine the probability of occurrence of a flood and a landslide in a year.

This is based on conditional probability, we need to solve the problem. We are having two events, that is one is flood and the other one is landslide. Let us represent two events, let probability of occurrence of flood be represented by $P(A)$ and the probability of occurrence of landslide be represented by $P(B)$,

Data given includes

$$P(A \mid B) = 0.3$$

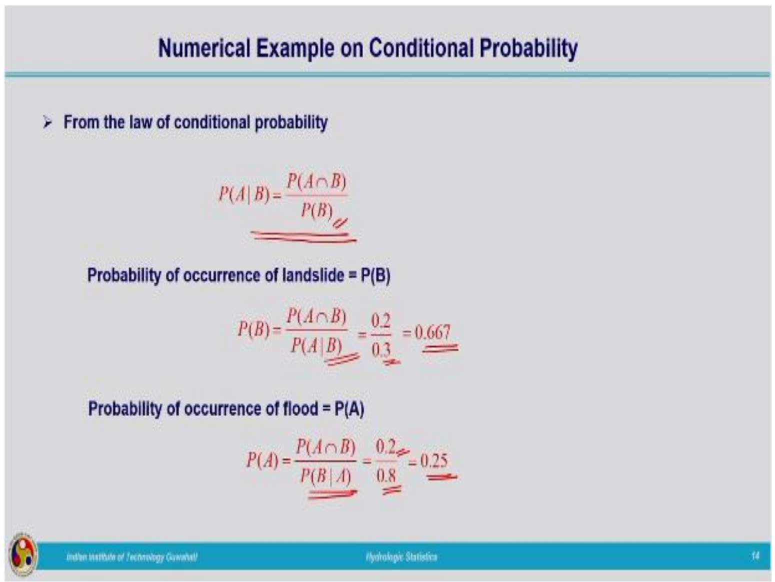
$P(B \mid A) = 0.8$

The joint probability is also given,

$P(A \cap B) = 0.2$

Now, by making use of the principle of conditional probability, we can calculate the probability of occurrence of flood and also probability of occurrence of landslide.

(Refer Slide Time: 26:43)



From the law of conditional probability we are going to make use of the mathematical expression

The joint probability is also given,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

And we can calculate the probability of occurrence of event B that is landslide given by

$$P(B) = \frac{P(A \cap B)}{P(A \mid B)} = \frac{0.2}{0.3} = 0.667$$

So, the probability of occurrence of landslide can be calculated as 0.667.

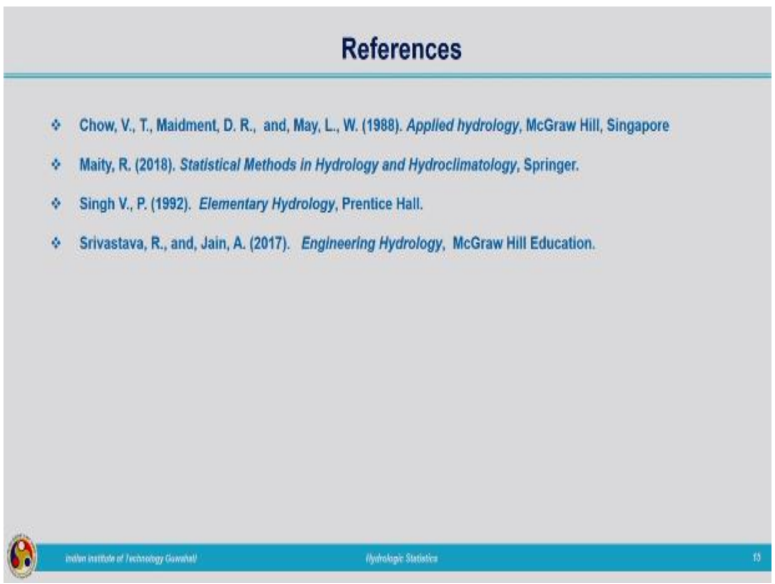
Next is probability of occurrence of a flood P(A).

$$P(A) = \frac{P(A \cap B)}{P(B \mid A)} = \frac{0.2}{0.8} = 0.25$$

So, with the given data we can conclude that the chance of occurrence of landslide is 66.7% and occurrence of flood is only 25%.

So, this is the simple example representing the practical application of conditional probability. In this example we have made use of the concepts of conditional probability. So, here I am winding up this lecture on preliminary concepts related to probability and statistics.

(Refer Slide Time: 28:29)



The references related to this, any of the probability statistics will be including all these concepts, related to hydrology, you can go through these references. Here I am winding up this lecture, thank you.