LECTURE 33
04-APRIL-2013

STATISTICS IN HYDROLOGY

→ As discussed yesterday, hydrological processes are in most cases not deterministic. The variables describing the hydrological process may be random.

→ We may need to incorporate stochasticity in modeling.
→ For that we require some background in probability.

* As defined yesterday – A random variable X is a variable (here hydrological variable) that is described by probability distribution.

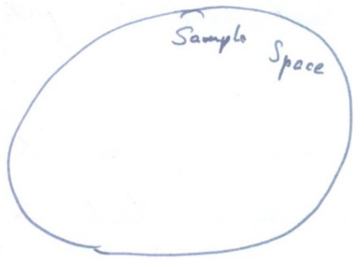
⇒ If X describes annual precipitation of a region Then any observation on annual rainfall (say x_i) should fall within the range specified by the probability distribution of X.

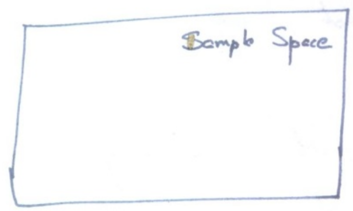
* A set of observations $x_1, x_2, x_3, \ldots, x_n$ of the RV X is called a <u>sample</u>.

→ We assume that samples are drawn from an infinite population that have statistical parameters like mean, variance, skewness, etc.

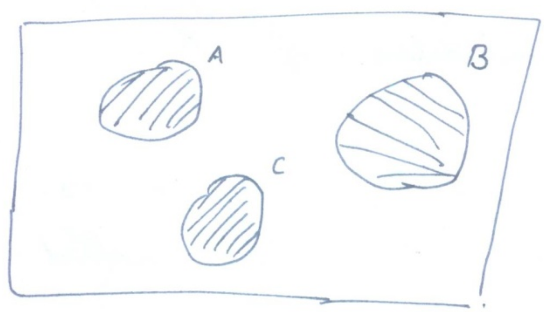
→ The statistical parameters of the sample can vary that from the population.

* If we have a set of all samples that can be drawn from the population, then that is called <u>Sample Space</u>.





* The sample space can enclose many events. (Actually an event is a subset in sample space).

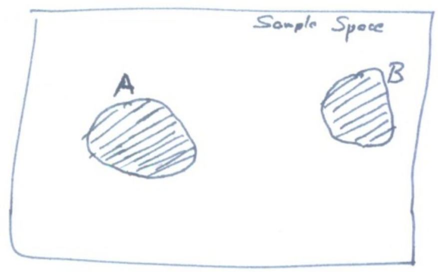


* An event can be anything in the sample space that we want to describe using probability.

If you look into the sample space for annual precipitation - it can range from 0 to $+\infty$.

→ However let us take some realistic values say from 0 to 1500 cm.

→ Then you can draw sample space



For rain
0 to 1500 cm.

* An event in this space can be say A → where rainfall $\leq$ 120 cm

* Another event can be B → where rainfall $>$ 200 cm

Like that we can define many events.

→ Some of the events may overlap.

<u>Probability</u>

Probability of event A in this sample space

→ This means that the chance an annual rainfall observation made at a place falls within this event. P(A).

A sample of 'n' observation of annual rainfall is made. $x_1, x_2, x_3, \ldots, x_n$

→ From these observations, n_A observations adhere to the event A.

Then relative frequency of event A $= \frac{n_A}{n}$

→ If we increase the size n it becomes $n \to \infty$

Then relative frequency becomes better estimate of probability of event A.

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

Probability Principles

(i) If there are 'm' events in a sample space that are not overlapping. (i.e $A_1, A_2, A_3, \ldots, A_m$)

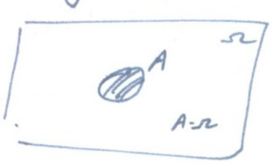
Then $P(A_1) + P(A_2) + \cdots + P(A_m) = P(\Omega) = 1$.

Ω → Sample space symbol.

This is total probability of the sample space.

(ii) Complementary probability :→ $\bar{A} = \Omega - A$

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



(iii) Conditional probability

If there are two events A and B

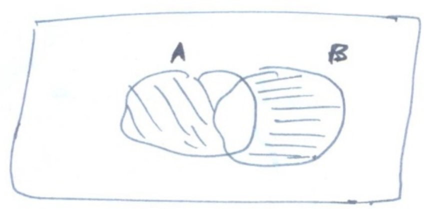
We can define event A as the rainfall in this year ~~2013~~ ≤ 130 cm

Also we can define event B as the rainfall in the following year ≤ 130 cm.

Therefore, in the sample space there may be occurrences of ~~rainfall~~ observations where the ~~rainfall~~ events A and B occurs.

There may be situations where A and B occur a ~~occur~~ overlap.

$\therefore$ $A \cap B$ is an event



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

$P(B/A)$ is called conditional probability.

If occurrence of B does not depend on occurrence of A, then

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

Example

The annual rainfall of a place is observed from 1960-1979 and the data is as given in the table below:

Year	Annual Rainfall (cm)	Year	Annual Rainfall (cm)
1960	115.00	1970	84.75
1961	110.75	1971	79.25
1962	94.50	1972	78.75
1963	74.00	1973	149.00
1964	87.50	1974	126.25
1965	124.25	1975	96.50
1966	91.50	1976	108.50
1967	81.25	1977	71.75
1968	154.25	1978	80.00
1969	118.50	1979	129.50

What is the probability that

i) The annual precipitation R in any given year $\leq$ 87.5 (Event A)

ii) The annual precipitation R is between 87.5 and 112.5 (Event B)

iii) The annual precipitation R is $\geq$ 112.5 (Event C)

What is the probability that there will be two successive years of annual rainfall $\leq$ 100 cm (Event D)

Solution

Number of observations = 20

i) The number of observations for Event A = 7.
Therefore, P(A) = 7/20 = 0.35

ii) The number of observations for Event B = 7
Therefore, P(A) = 0.35

iii) P(C) = 1-P(A)-P(B) = 1- 0.35-0.35 = 0.30

Let D be the event for first year having annual rainfall $\leq$ 100 cm

Let F be the event for the second year having annual rainfall $\leq$ 100 cm

Therefore, $P(D \cap F) = P(F).P(D) = 0.55*0.55=0.3025$

The frequency histogram

Rainfall Range (cm)	No. of Observations
70-80	4
80-90	4
90-100	3
100-110	1
110-120	3
120-130	3
130-140	0
140-150	1
150-160	1