LECTURE 34
05 - APRIL - 2013

FREQUENCY & PROBABILITY FUNCTIONS

Yesterday we discussed on

- → Relative frequency
- → Probability of an event
- → Sample space, etc.

The lecture was concluded with an example.
In the example we introduced you the concept of frequency histogram.

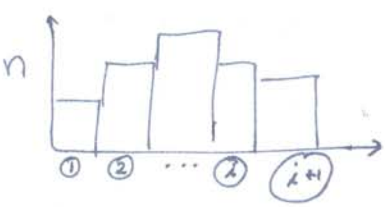
→ For the given rainfall observation data, we ranged the data like 70-80, 80-90, ..., 150-160.

That is the width of each range = 10 cm rainfall.

→ In general therefore we can define ranges based on observation in closed bracket

e.g. $[x_i - \Delta x, \; x_i]$ for the i^{th} range.

where Δx is the width of the range.



The no. of observations in the i^{th} range $= n_i$ (say).

The relative frequency function for this i^{th} range is

$$f_s(x_i) = \frac{n_i}{n}$$

This is almost equivalent to the estimate for probability $P(x_i - \Delta x_0 \le X \le x_i)$

That is the random variable X will lie in this interval $[x_i - \Delta x, x_i]$.

→ We can also estimate the sum of these relative frequencies

$$\sum_{j=1}^{i} f_s(x_j).$$

This is cumulative frequency function $F_s(x)$.

$$F_s(x_i) = \sum_{j=1}^{i} f_s(x_j)$$

It is equivalent to $P(X \le x_i)$ → the cumulative probability

→ As observed : relative frequency, cumulative frequency, etc. are for a sample.
i.e why $f_s(x_i)$, $F_s(x_i)$.

As $n \to \infty$ and $\Delta x \to 0$

$$\lim_{\substack{n \to \infty \\ \Delta x \to 0}} \frac{f_s(x)}{\Delta x}$$ can be defined as a function

$= f(x)$ = probability density function

Again we can define $F(x) = \underset{\substack{n \to \infty \\ \Delta x \to 0}}{Lt} F_s(x)$

This is cumulative probability or also called <u>probability distribution function</u>.

$$F(x) = \underset{\substack{n \to \infty \\ \Delta x \to 0}}{Lt} F_s(x)$$

We can see that $f(x) = \frac{d}{dx} F(x)$

As seen earlier $F(x)$ is $P(X \leq x)$

As $f(x) = \frac{d}{dx} F(x)$

$$\therefore F(x) = \int f(x)\, dx$$

What should be the limits. The range of random variable depends on X.

$\therefore$ In general we can give

$$F(x) = \int_{-\infty}^{x} f(u)\, du \qquad u \to \text{dummy variable}$$

* If you have sample data and you want to check the probability of the random variable.
* First obtain relative frequency $f_s(x)$ and cumulative frequency $F_s(x)$.
* Find the equivalent $F(x)$, of the probability distribution
* From the $F(x)$, evaluate $f(x)$.

$$f_s(x_i) = p(x_i) = P(x_i - \Delta x \leq X \leq x_i)$$

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

$$= F(x_i) - F(x_i - \Delta x)$$

$$= F(x_i) - F(x_{i-1})$$

One of the famous probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Also we can define a standard normal distribution

$$z = \frac{x-\mu}{\sigma}$$

$$\therefore f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2} \quad ; \; -\infty \leq z \leq \infty$$

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, du$$

You can get the values by approximations, tables, etc.

e.g.

$$F(z) = \begin{cases} B & \text{for } z < 0 \\ 1 - B & \text{for } z \geq 0 \end{cases}$$

where

$$B = \frac{1}{2}\left[1 + 0.196854|z| + 0.115194|z|^2 + 0.000344|z|^3 + 0.019527|z|^4\right]^{-4}.$$

e.g.: What is the probability that standard normal variable z will be less than -2.

Soln: ~~$P(z = -2)$~~ $F(-2)$ we need to find.

Now $|z| = 2$.

$$\therefore B = \frac{1}{2}\left[1 + 0.196854 \times 2 + 0.115194 \times 2^2 + 0.000344 \times 2^3 + 0.019527 \times 2^4\right]^{-4}$$

$F(-2) = 0.02256$

Statistical Parameters

Q: What is the objective of statistics.

→ To extract essential information from a set of data

→ To reduce large set of numbers to a small set of numbers.

There are many statistical parameters like

* mean, standard deviation, ~~etc.~~ skewness, etc.

Q: What is a statistical parameter?

It is the expected value E of some function of the random variable in consideration.

e.g. If X is a random variable,

then there can be many expected values to some functions of that random variable X.

→ A very simple expected value of some function of the random variable is the expected value of random variable itself.
It is called <u>mean</u>.

$$E(X) = \mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

It is the product of x and the corresponding probability density $f(x)$.

$$E(X) = \mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

→ First moment about the origin of the random variable (Central tendency).

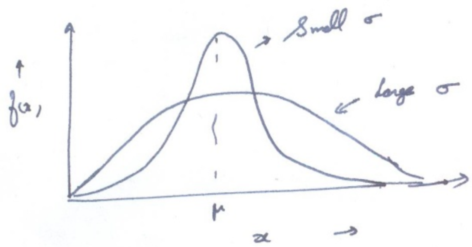
From the sample data we have

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

where n → no: of observations in the sample.

→ From our observations we would like to see where $\bar{x}$ would be in vicinity of μ.

Population Parameter	Sample Statistic
1) Midpoint $\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$	$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
2) Variability. Variance $\sigma^2 = E\left[(x-\mu)^2\right]$; $E\left[(x-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$ (Second moment about mean).	$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
3) Symmetery $\gamma = \frac{E\left[(x-\mu)^3\right]}{\sigma^3}$	$C_s = \frac{n \sum_{i=1}^{n} (x_i - \bar{x})^3}{(n-1)(n-2)\, s^3}$



To fit Probability Distribution function to the hydrological Random Variable

- The probability distribution function → gives the probability of occurrence of a random variable.

→ Therefore, fitting a probability distribution, we can summarise many hydrological data that may be very long and inhuman to handle.