Pramod Parajuli
Simulation and Modeling, CS-331

Chapter 6
Probability Concepts

Stochastic Variables

-Ordered set of random values

-Stochastic process give rise to stochastic variable

-Stochastic process can be discrete or continuous

Discrete Probability Functions

Probability Mass Function (PMF)

-Distribution of probabilities

-If a variable can take n different values x_n (n = 1, 2, 3, … N), and probability of x_n being take is $p(x_n)$, then the set of numbers comprise of $p(x_n)$ is probability mass function

$$\sum_{n=1}^{N} p(x_i) = 1$$

-In certain cases like dice roll, the PMF values may be known (e.g. 1/6). But most of the time, PMF is counted from input sample

-Example – table 6-1

Discrete Probability Functions

Cumulative Distribution Function (CDF)

-For a continuous variable, probability of any one specific value occurring is considered logically to be zero

$$F(x) = P(-\infty < X \leq x)$$

$$F(x) = P(X \leq x) \qquad F(-\infty) = 0$$

$$gives: 0 \leq F(x) \leq 1 \qquad F(+\infty) = 1$$

-Let f(x) is PDF and the probability that 'x' falls in the range x1 to x2 is given by

$$= \int_{x1}^{x2} f(x)dx \qquad = \int_{-\infty}^{\infty} f(x)dx = 1$$

Discrete Probability Functions

Cumulative Distribution Function

-As we have seen already, value of 'infinity' is implementation dependent

-But, the cumulative distribution function (CDF) will be

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$$gives, 0 \leq F(x) \leq 1$$

-Therefore, probability of x falling in the range x1 to x2 is,

$$F(x_2) - F(x_1)$$

Discrete Probability Functions

Probability Density

-If, $$P(x < X \leq x + \Delta x) = f(x).\Delta x$$

then,

$$f(x) = \frac{P(x < X \leq x + \Delta x)}{\Delta x}$$

is said to be probability density/density function in continuous system

In discrete system, it is called frequency function

Measure of Probability Functions

Mean (Expectation)

$$mean = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad x_i (i = 1,2,3,4,\ldots., N)$$

If observation fall into I groups i.e. x_i element have n_i occurrences

$$m = \frac{1}{N}\sum_{i=1}^{I} n_i . x_i \qquad m = \sum_{i=1}^{I} p(x). x_i$$

For continuous variable;

$$m = \int_{-\infty}^{\infty} x . f(x) dx$$

$$\sum (aX) = a \sum (x)$$

$$\sum (x_1 + x_2) = \sum (x_1) + \sum (x_2)$$

Measure of Probability Functions

Mean (Expectation)

-Most of the times, the selection of expectation has no occurrence in the sample

-Therefore, it is better to take the nearest sample

-Note; if a value is being multiplied in future, then it must not be corrected. Only after the multiplication it should be corrected. (example)

Mode

Peak probability density function

Median

-Half of the random values will fall below this point

-Normally, cumulative function, $F(x) = 0.5$

Measure of Probability Functions

Variance

-Standard deviation

-Measure of the degree to which data are dispersed from the mean value

-Computed as +ve square root of variance

$$s = \left[\frac{1}{(N-1)} \sum_{i=1}^{N} (m - x_i)^2 \right]^{1/2}$$

-If I-groups

$$s = \left[\sum_{i=1}^{I} p(x_i).(m - x_i)^2 \right]^{1/2}$$

Measure of Probability Functions

Variance

-For continuous

$$s = \int_{-\infty}^{\infty} f(x)x^2 dx - m^2$$

-If mean 'm' and data are at same point, standard deviation = 0

Measure of Probability Functions

Coefficient of Variance

$$\frac{s\tan dard - deviation}{mean - value}$$

-Better than variance as it is expressed in relative terms

Auto-correlated

-All the data are supposed to be independent of each other

-But if data at i^{th} step is related to previous data using any kind of relationship they are said to be auto-correlated

-In such case, only the 'mean' holds the meaning, others do not

Numerical Evaluation of probability functions

-Required for random number generation

-For computation, numerical methods required

Representation

-Frequency distribution

-Relative frequency distribution

If $x_i < x \leq x_{i+1}$

then relative frequency distribution= $p_i = \int_{x_i}^{x_{i+1}} f(x)dx$

density function = $\dfrac{p_i}{(x_{i+1} - x_i)}$

For cumulative $F_n = \sum_{i=0}^{n} p_i$