Theoretical Distribution

Computer Oriented Numerical and Statistical Methods

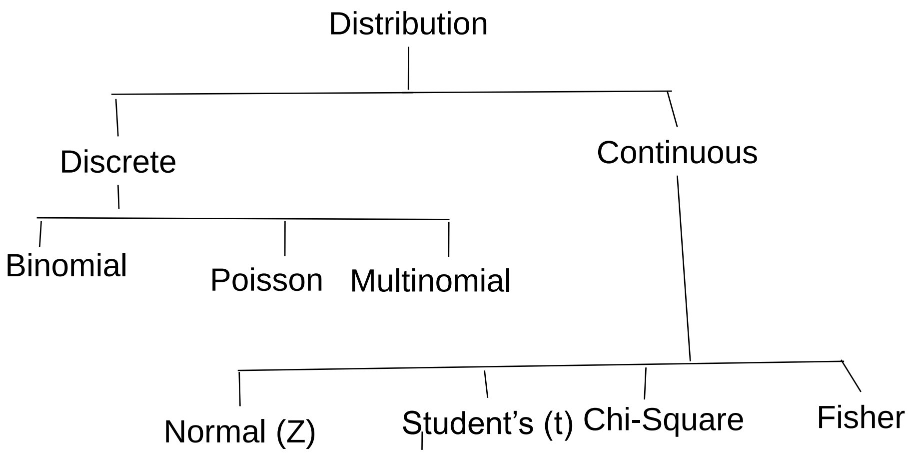
Outline

- Introduction
- Types of Theoretical Distribution

Introduction

- Theoretical Distribution refers to mathematical models of relative frequencies of a finite number of observations of a variable.
- It is a systematic arrangement of probabilities associated with mutually exclusive and collectively exhaustive elementary elements in an experiment.
- Where the relative frequency distributions are based on actual observations, the above distribution are based on mathematical functions.
- Important features of these distributions is that with some known parameters like mean and S.D. or the number of trials and the chances of success, the probabilities of various values of a variate can be found in the form of a complete distribution.

Types of Theoretical Distribution



Discrete Distribution

- The Discrete probability distributions are known as point functions defined over a sample space.
- The random variables in these takes only a finite integer value.
- These are normally represented by line graphs when not grouped and by histograms when grouped, each bar is raised on the mid-value of the class.
- The cumulative probabilities in this case are represented by a staircase type of histogram.

Binomial Distribution

- Binomial is also known as the “Bernoulli Distribution” after the Swiss mathematician James Bernoulli (1654-1705)
- The distribution can be used under the following conditions:
  - The random experiment is performed repeatedly a finite and fixed number of times.
  - The result of any trial can be classified only under two mutually exclusive categories called success (the occurrence of event) and failure (the non-occurrence of event).

Binomial Distribution

- The proportion of outcomes falling in the "success" category are denoted generally by p and the proportion of item falling in the category of "failure" by q = 1 - p
- The probability of success in each trial remains constant and does not change from one trail to another.
- The trails are independent, so that the result of any trial in ineffective by the result of previous trials.

Probability Function Of Binomial Distribution

- If a trial of an experiment can result in success with the probability p and failure with probability q = 1 – p, the probability of exactly x success in n trails is given by $P(X=x) = P(x) = {}^{n}C_{x}p^{x}q^{n-x}$; x = 0, 1, 2,…n
- The quantities n and p are called parameters of the binomial distribution and the notation b(x:n,p) reads "the binomial probability of x given n and p."
- The entire probabilities distribution of x = 0, 1, 2, …, n success with two types of expression, where p stands for success and q for non-success can be written as follows :

Probability Function Of Binomial Distribution

BINOMIAL PROBABILITY DISTRIBUTION			
No. of success x	$(p+q)^n$ Prob. p(x) p(x)	No. of success x	$(p+q)^n$ Prob. p(x) p(x)
n	${}^nC_n p^n q^0$	0	${}^nC_0 p^0 q^n$
n-1	${}^nC_{n-1} p^{n-1} q^1$	1	${}^nC_1 p^1 q^{n-1}$
n-2	${}^nC_{n-2} p^{n-2} q^2$	2	${}^nC_2 p^2 q^{n-2}$
…..	……	…..	……
0	${}^nC_0 p^0 q^n$	n	${}^nC_n p^n q^0$
Total	1		1

Obtaining Coefficient Of The Binomial

- To find the term of the expression of $(q+p)^n$
- The first term is q^n
- The second term is ${}^nC_1 q^{n-1} p$
- In each succeeding term the power of q is reduced by 1 and the power of p is increased by 1.
- The coefficient of any term is found by Pascal's triangle.
- The mean of binomial distribution is n*p
- The S.D. of binomial distribution is $\sqrt{npq}$

Using Binomial Tables

n = 10									
x	…	p=.20	p=.25	p=.30	**p=.35**	p=.40	p=.45	p=.50	
0	…	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010	10
1	…	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098	9
2	…	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439	8
3	…	0.2013	0.2503	0.2668	**0.2522**	0.2150	0.1665	0.1172	7
4	…	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051	6
5	…	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461	5
6	…	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051	4
7	…	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172	3
8	…	0.0001	**0.0004**	0.0014	0.0043	0.0106	0.0229	0.0439	**2**
9	…	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098	1
10	…	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0
	…	p=.80	**p=.75**	p=.70	p=.65	p=.60	p=.55	p=.50	**x**

Examples:

n = 10, p = 0.35, x = 3: P(x = 3|n =10, p = 0.35) = 0.2522

n = 10, p = 0.75, x = 2: P(x = 2|n =10, p = 0.75) = 0.0004

Poisson Distribution

- In Binomial distribution it was found that there is a sample of a definite size so that it is possible to count the number of times an event is observed; in other words, n is precisely known.
- There are certain situations where this may not be possible.
- The basic reason for is that the events is the rare and causal.
- Successful events in the total events space are few e.g. the events like accidents on a road, defects in a product, goals scared at a football match, etc.

Poisson Distribution

- In these, we known the number of times an event occurs but not how many times it does not occur.
- The total number of trials in regard to a given experiment are not precisely known.
- The Poisson distribution is very suitable in case of such rare events.
- Only the average chance of occurrence based on past experience or a small sample extracted for the purpose will enable us to construct the whole distribution.

Poisson Distribution

- The binomial distribution requires only two parameters (p and n) this distribution requires only the value of m which is the mean of the occurrence of an event (np) based on existing knowledge on the matter.
- Poisson distribution was derived in 1837 by a French-Mathematician Simeon D. Poisson.

- Poisson distribution may be obtained as a limiting case of Binomial probability distribution under the following conditions.
  - n, the number of trails is indefinitely large i.e. $n \to \infty$

Poisson Distribution

- p, the constant probability of success for each trial is indefinitely small i.e. $p \to 0$
- np = m (say) is finite.

- The probability function of random variable X following Poisson Distribution is $P(X=x) = p(x) = (m^x / x!) * e^{-m}$ x = 0, 1, 2,…
  - where X = the number of successes (occurrence of the event)
  - e = 2.71828 [The base of the system of natural logarithm] and x! = x(x-1)(x-2)…3.2.1

Poisson Distribution

Values of variables (x)	0	1	2	3	….	total
Prob. P(X=x)= p(x)	e^{-m}	$e^{-m} * m$	$(e^{-m} * m^2) / 2!$	$(e^{-m} * m^3) / 3!$		1

Constants Of Poisson Distribution

$$\text{Mean } \mu = \lambda$$

$$\text{Variance } \sigma^2 = \lambda$$

$$\text{S.D.} = \sqrt{\lambda}$$

$$p(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0,1,2,.....$$

Minal Shah

17

Using Poisson Tables

x	λ								
	0.10	0.20	0.30	0.40	**0.50**	0.60	0.70	0.80	0.90
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	**0.0758**	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Example: Find P(X = 2) if $\lambda = 0.50$

$$P(X=2) = \frac{e^{-\lambda}\lambda^X}{X!} = \frac{e^{-0.50}(0.50)^2}{2!} = 0.0758$$

Minal Shah

Multinomial Distribution

- The binomial distribution is generalized as follows. Suppose the sample space of an experiment is partitioned into say mutually exclusive events A_1, A_2, A_2, ….., A_s with respective probabilities P_1, P_2, P_3, …., P_s (hence $P_1 + P_2 + P_3 + \ldots + P_s = 1$).
- Theorem : In n repeated trials, the probability that A_1 occurs K_1 times, A_2 occurs K_2 times and A_s occurs K_s times is equal to

$$\frac{n!}{k_1!k_2!k_3!\ldots k_s!} p_1^{k_1} p_2^{k_2} p_3^{k_3} \ldots\ldots p_s^{k_s}$$

  where $k_1 + k_2 + \ldots\ldots + k_s = n$
- The above numbers from the so-called multinomial distribution since they are precisely the terms in the expansion of $(p_1+p_2+\ldots+p_s)$

Continuous Distribution

- These distribution are associated with continuous variables.
- A continuous variable defined over a given range, may take any of the intermediate values. It is always written as an approximate values. For ex. Weight, height etc.
- The continuous variables are generally represented by a smooth curve.
- The cumulative distribution of a continuous variables is also a smooth curve.
- Normal distribution is one of the continuous distribution.

Normal Distribution

- The most important continuous probability distribution used in the entire statistics is the normal distribution.
- Its graph, called the normal curve, is a bell shaped curve that extends indefinitely in both directions, coming closer and closer to the horizontal axis without even reaching it.
- The mathematical equation of normal curve was developed by De-Moivre in 1733.
- The normal distribution is often referred to as the Gaussian distribution in honour of Karl F. Gauss (1777-1855) who also derived the equation from the study of errors in repeated measurements of the same quantity.

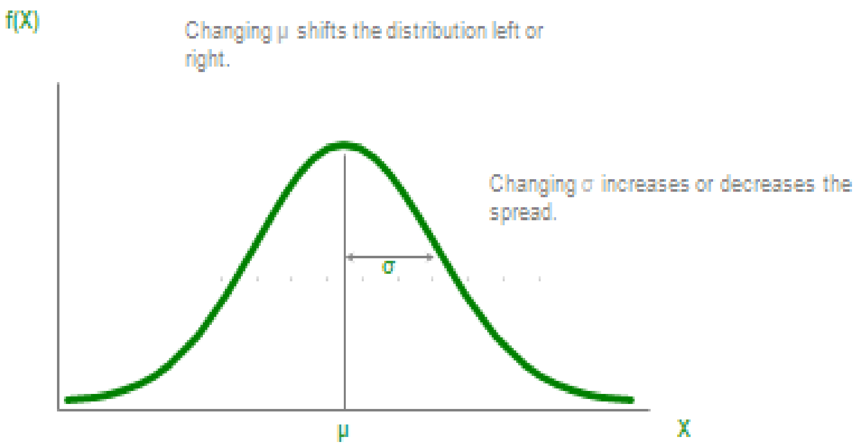
Normal Distribution

- Definition : A continuous random variable X is said to be normally distributed if it has the probability density function represented by the equation:

$$p(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(X - \mu)/\sigma]^2}$$

  - $-\infty \leq x \leq \infty$
  - Where μ and σ, the mean and the standard deviation are known as two parameters and Π = 3.14159 e = 2.7183 are two constant.

Normal Distribution



Minal Shah

23

Normal Distribution

- Now, standard normal distribution or Z – distribution is

$$p(Z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)Z^2}$$

$$Z = \frac{X - \mu}{\sigma}$$

  – Where
  – e = the mathematical constant approximated by 2.71828
  – π = the mathematical constant approximated by 3.14159
  – Z = any value of the standardized normal distribution
  – The Z distribution always has mean = 0 and standard deviation = 1

Minal Shah

24

Fitting A Normal Curve

- There are two main objects of fitting a normal curve to sample data.
  - To provide a visual device for judging whether or not the normal curve is a good fit to the sample data, and
  - To use the smoothed normal curve, instead of the irregular curve representing the sample data, to estimate the characteristics of the population.

Methods Of Fitting

- There are two methods of fitting a normal curve.
  - Method of ordinates,
  - Method of area.

Minal Shah