[Academic Script]

Probability Distributions

Subject:	Business Economics
Course:	B. A. (Hons.), 1st Semester, Undergraduate
Paper No. & Title:	Paper – 102 Statistics for Business Economics
Unit No. & Title:	Unit – 5 Probability and Distribution
Lecture No. & Title:	Lecture – 2 Probability Distributions

Academic Script

1. Introduction

We may come across to various types of data: discrete or continuous. The random variable corresponding to the characteristic involved in the data may follow any statistical distribution. To study the data and to predict the probability of certain types of events on the basis of the available data it is necessary to know the distribution of the random variable. There are two types of distributions: discrete and continuous. In this talk we will study some discrete distributions.

Probability distribution:

Before defining probability distribution, Let us first discuss the concept of random variable.

2. Random variable:

A variable whose value is determined by the outcomes of a random experiment is called a random variable. Which is a function defined over the sample space of an experiment. It is usually denoted by capital letters X,Y, Z etc. Let U be a sample space associated with a random experiment . A function associating areal number with each outcome of U becomes a random variable , denoted by

$X: U \rightarrow R$.

e.g. A random variable X associated with the sample space in case of families with two children can be defined as follows:

A child in a family is either boy or girl. If we denote a boy by ‘b’ and a girl by ‘g’ then the sample space associated with families having two children is given by

U = {bb, bg, gb, gg}.

For an element u of U let us define

X(u) = number of boys in the element u.

Then X is called random variable denoting the number of boys in a family with two children.

Here X (bb) = 2, X(bg) = 1,

X(gb) = 1 and X(gg) = 0.

Thus the real number X associated with the outcomes of the sample space, which denotes the number of boys in each outcome, assumes the values 0, 1, 2. We shall call X as a random variable.

If a random variable which takes only discrete values is called a discrete random variable and corresponding distribution is called discrete probability distribution.

Suppose X is a random variable and it assumes all values of a finite set $\{ x_1, x_2, ..., x_n\}$ of R.

Suppose that X assumes a value x_i with probability

$P(X = x_i) = p(x_i)$.

If $p(x_i) \geq 0$ for i = 1, 2, ..., n and

$p(x_1) + p(x_2) + ... + p(x_n) =1$

the set of real values $\{ p(x_1), p(x_2), ..., p(x_n)\}$

is called the discrete probability distribution of random variable X.

Example :

In tossing of a die having number 'X' on the sides the number comes on the upper side of the die is a random variable which assumes the values 1,2,3,4,5,6.

Here the die is unbiased i.e. the outcomes are equi-probable; the probability of occurrence of each number becomes 1/6. Thus the probability distribution is given by,

X	1	2	3	4	5	6
P(X)	1/6	1/6	1/6	1/6	1/6	1/6

3. Joint, marginal and conditional distributions:

If (X,Y) takes only a finite or countable number of pairs of values (x,y) in R^2 then (X,Y) is known as discrete two dimensional random variable.

Let $P(X=x_i, Y=y_j) = P(x_i, y_j) = P_{ij}$, then P_{ij} is called joint probability distribution of (X,Y) at the point (x_i,y_i) provided the following conditions are satisfied.

(i) $P_{ij} > 0, \quad \forall i=1,2,\ldots; j = 1,2,\ldots$

(ii) $\sum_i \sum_j p_{ij} = 1$

The marginal probability distribution of a random variable X is given by

$P(X=x_i) = p_{x_i} = \sum_j p(x_i, y_j)$, i=1,2,....

And the marginal probability of a random variable Y is given by

$P(Y=y_i) = p_{y_j} = \sum_i p(x_i, y_j)$, j=1,2,....

The conditional probability distribution of X given Y=y is defined as

$P(X=x \mid Y=y) = \dfrac{p(X=x, Y=y)}{p(Y=y)}$; $P(Y=y) > 0$

or

$P(X=x_i \mid Y=y_i) = \dfrac{p(x_i, y_j)}{p_{y_j}}$; $p_{y_j} > 0$; i=1,2,.... & j=1,2,....

e.g. consider the joint probability distribution of bivariate random variable (X,Y); where X denotes the employment status and Y denotes training status.

Status	Skilled (y_1)	Unskilled ($y_{2\_}$)
Employed (x_1)	0.50	0.30
Unemployed (x_2)	0.15	0.05

Here the marginal probability distribution of a random variable X can be obtained as,

$P(X=x_i)= \sum_{j=1}^{2} p(x_i, y_j)$; i =1, 2

i.e. $P(X=x_1)=P(x_1,y_1) + P(x_1,y_2) = 0.50 + 0.30 = 0.80$

$P(X=x_2)=P(x_2,y_1) + P(x_2,y_2) = 0.15 + 0.05 = 0.20$

Thus the probability distribution of a random variable X can be demonstrated as,

x:	x_1	x_2
P(x):	0.80	0.20

Similarly, the marginal probability distribution of a random variable Y can be demonstrated as,

$P(Y=y_j) = \sum_{i=1}^{2} p(x_i, y_j)$; j=1, 2

i.e. $P(Y=y_1)=P(x_1,y_1) + P(x_1,y_2) = 0.50 + 0.15 = 0.65$

$P(Y=y_2)=P(x_1,y_2) + P(x_2,y_2) = 0.30 + 0.05 = 0.35$

Thus we can state the probability distribution of a random variable Y as,

y	y_1	y_2
P(y)	0.65	0.35

The conditional probability distribution of Y given $X=x_i$ can be obtained as,

$$P(Y=y_j \mid X=x_i)=\frac{p(Y=y_j, X=x_i)}{p(X=x_i)} \quad ; \quad P(X=x_i) > 0$$

The particular conditional probabilities can be derived as follows,

$$P(Y=y_1 \mid X=x_1)=\frac{p(Y=y_1, X=x_1)}{p(X=x_1)} = \frac{0.50}{0.80} = \frac{5}{8}$$

$$P(Y=y_2 \mid X=x_1)=\frac{p(Y=y_2, X=x_1)}{p(X=x_1)} = \frac{0.30}{0.80} = \frac{3}{8}$$

Similarly

$$P(Y=y_1 \mid X=x_2)=\frac{p(Y=y_1, X=x_2)}{p(X=x_2)} = \frac{0.15}{0.20} = \frac{3}{4}$$

$$P(Y=y_2 \mid X=x_2)=\frac{p(Y=y_2, X=x_2)}{p(X=x_2)} = \frac{0.05}{0.20} = \frac{1}{4}$$

Similar way one can obtain the conditional probability distribution For $X=x_i$ given $Y=y_j$ using the formula,

$$P(X=x_i \mid Y=y_i)=\frac{p(x_i, y_j)}{p_{y_j}} \text{ ; } i=1,2 \ \& \ j=1, 2.$$

4. Mathematical expectation:

i) Expectation of random variable X:

Suppose X is a discrete random variable which assumes the possible values $X_1, X_2, \ldots\ldots\ldots, X_k$ with corresponding probabilities $P_1, P_2, \ldots\ldots\ldots, P_k$, where $P_i = P(X=X_i)$, then mathematical expectation as a random variable X is defined as

$$E(X)=\sum_{i=1}^{K} X_i P_i = X_1 P_1 + X_2P_2 + \ldots\ldots + X_kP_k$$

If $\phi(X)$ is any function of a random variable X then expectation of $\phi(X)$ is defined as,

$$E[\phi(X)] = \sum_{i=1}^{K} f(X) P_i$$

e.g. If $\phi(X)$ =X^2 then E(X^2)= $\sum_{i=1}^{K} X_i^2 P_i$

ii) Variance of random variable X:

The variance of a random variable X is defined as

$$V(X)=E(X-E(X))^2$$

$$=E(X^2)-(E(X))^2$$

$$= \sum_{i=1}^{K} X_i^2 P_i - \left(\sum_{i=1}^{K} X_i P_i\right)^2$$

Which is denoted by σ_X^2 also.

Example:

A bakery has the following schedule of daily demand for cakes.

Number of cakes demanded(X)	0	1	2	3	4	5	6	7
Probability P(X)	0.05	0.07	0.10	0.30	0.25	0.10	0.08	0.05

(i) The expected number of cakes demand per day,

= E(X)

= $\sum XP(X)$

= 0(0.05) +1(0.07)+ 2(0.10)+ 3(0.30)+ 4(0.25)
+ 5(0.10)+ 6(0.08)+ 7(0.05)

= 3.50

(ii) The variance of the daily demand for cakes

Here E(X^2)= $\sum X^2 P(X)$

= $0^2(0.05) + 1^2(0.07) + 2^2(0.10) + 3^2(0.30)$
$+ 4^2(0.25) + 5^2(0.10) + 6^2(0.08) + 7^2(0.05)$

$= 15$

Thus, $V(X) = E(X^2) - (E(X))^2$

$= 15 - (3.5)^2 = 2.75$

5. Discrete distributions:

In this section we will discuss discrete distributions like: uniform distribution, Bernoulli distribution, binomial distribution and Poisson distribution

(1)Uniform distribution

Let a random variable X assumes the values 1,2,………N, then the discrete uniform distribution is given by,

$P(X)=\frac{1}{N}$; X=1,2,….,N

Here probability remains equal for every value assumed by the random variable X.

Its mean and variance are given by,

$E(X)=\frac{N+1}{2}$ and $V(X)= \frac{N^2-1}{12}$

Example: If a random variable X follows discrete uniform distribution with parameter 5 i.e. If P(X)=1/5 ; X=1,2,3,4,5; find (i) mean and variance of the distribution (ii) P(X < 3) (iii) P (X > 2)

(i) Then mean and variance of the distribution are

Mean = $E(X)= \frac{N+1}{2}$

$= (5 + 1) /2 = 3$

and variance = $\frac{N^2-1}{12}$

$= (5^2 -1)/12 = 2$

(ii) P(X < 3) = P(1) + P(2) = 1/5 + 1/5 = 2/5

(iii) P(X > 2) = P(3) + P(4) + P(5) = 1/5 + 1/5 + 1/5

= 3/5

(2) Bernoulli distribution

Suppose an experiment exhibition only two mutually exclusive and exhaustive outcomes.

In relation of the experiment, the probability of occurrence of these two outcomes remains unchanged and the trials are independent of one onther then such trials of the experiment are called Bernoulli trials and corresponding distribution for a single trail is known as Bernoulli distribution.

Let X be a variable satisfied the above condition and p be the probability of happening of the particular event(say 'success') and q=1-p be the probability of not happening of the above event (say 'failure') then the distribution of X is defined as,

P(X)= p, if X=1(i.e. Success)

1-p, if X=0 (i.e. Failure)

That is $P(X= x) = p^x(1-p)^{1-x}$, x = 0, 1; 0 < p < 1.

This distribution is known as Bernoulli distribution

(3) Binomial distribution:

If the Bernoulli trials are repeated n times and P be the probability of success then the probability of getting x success in n such trials is given by,

$P(X)= \binom{n}{x} p^x q^{n-x}$; X=0,1,2……..,n; 0 < P < 1; q=1-p.

This distribution is called binomial distribution.

Its mean=E(X)= np and variance =V(X)= npq.

Here mean > variance, n and p are known as the parameter of the binomial distribution.

Example 1.

If 2% of the electric bulbs produced by a certain company are defective, then the probability that in a random sample of 10 bulbs, 7 bulbs getting non defective can be obtained using the binomial distribution.

Here n=10; p= probability of defective bulbs=0.02;

x= number of defective bulbs in a sample (of size 10) = 3

Hence P(X=3)= probability of getting 3 defective bulbs in a sample of size 10.

$$= \binom{10}{3}(0.02)^3(0.98)^{10-3}$$

$$= 0.0008334$$

Here the mean number of defective bulbs

= np =10(0.02) = 0.2

& variance of the number of defective bulbs

= npq = 10(0.02)(0.98) = 0.196

Example 2.

Let mean and variance of a binomial distribution are 3 and 3/2, find the probability of getting 2 success.

Here $\frac{variance}{mean}$ = q = $\frac{3/2}{3}$ = $\frac{1}{2}$. $\quad \therefore p = 1 - q = 1/2$

Now mean = np = 3. $\quad \Rightarrow n\left(1/2\right) = 3 \Rightarrow n = 6$.

Hence the probability of getting 2 success = P(X=2)

$= \binom{6}{2}\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{6-2}$ = 15/64

(4) Poisson distribution:

In binomial distribution number of trials or a sample size is precisely known and finite. But, there are certain situations where this may not be possible. Since some times the event is rare and casual like accidents on a road, goals scored in a football match etc.

In such events, we know the number of times an event occurs but do not how many times it does not occur. Obviously the total number of trials in regard to a given experiment is not precisely known.

In such situation Poisson distribution is used in place of binomial distribution assuming n is very large and p is very small.

The Poisson distribution of happening of the event x times is given by

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2, \dots, \lambda > 0.$$

Where λ = mean number of times the event happened

= np, which is finite.

For Poisson distribution mean = variance = λ.

Example 1. Let the probability that a price of a certain share falls down 30% within the 3 hours of a day is 0.001, determine the probability that out of 1000 days in (i) exactly three (ii) more than two days the price of the share falls down 30% within the 3 hours of a day.

Here n = 1000, p = 0.001, λ = np = 1

(i) $P(X = 3) = \frac{e^{-1}1^3}{3!} = 0.061313$

(ii) $P(X>2) = 1-P(X\leq 2) = 1-[P(0) + P(1)+ P(2)]$

$$= 1 - e^{-1}[\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!}]$$

$$= 1 - e^{-1}(2.5)$$

$$= 0.080301$$

Example 2. A hospital telephone receiving counter receives on average 2 emergency calls per minute. What is the probability of receiving (i) no calls in a one minute interval (ii) at the most 3 calls in a 2 minute interval?

Here we use Poisson distribution for the number of emergency calls in a given time period.

Let X = number of emergency calls received at the counter per minute. We have given that the average number of emergency calls per minute i.e. $\lambda = 2$.

Then,

(i) P(no calls in a one minute)

= P(X = 0)

$= \frac{e^{-2}2^0}{0!} = e^{-2} = 0.135335$

(ii) P(at the most three calls in a 2-minute interval)

= P(X≤3)

Here we have to find probability of calls per 2 minute interval, hence average number of calls for 2 minute interval becomes $2\lambda = 2(2) = 4$. Therefore the required probability will be given by

$$P(X\leq 3) = e^{-4}\left(\frac{4^0}{0!} + \frac{4^1}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!}\right) = 0.43347$$

6. Summary

A random variable always follows some distribution. The distribution may be univariate or bivariate. If the distribution is bivariate we can determine the marginal and condition distribution of the given two random variables. Also the distribution of a random variable may be discrete or continuous. The discrete distributions we have discussed discrete uniform, Bernoulli, binomial and Poisson distributions.