Probability and Statistics

MC-205

Descriptive Statistics

→ **Statistics** → It is the art of learning from data. It is the science of collection, processing and interpretation of data

→ **Descriptive Statistics** - The part of statistics which deals with summarization and description of data.

→ **Variable** - It is a characteristic that varies from individual to individual. Eg. height, weights, body temp. etc.

→ **Population** - The set of mathematics of interest for every experimental unit in the entire collection.

→ **Sample** - A smaller finite subset of the measurements.

→ **Univariant data** - If a single variable is measured on a single experimental unit.

→ **Bivariant data** - When 2 variables are measured on a single experimental unit.

★ Classification of variables

- Qualitative
- Quantitative
  - Discrete
  - Continuous

Page ______
Date / /

Qualitative variable : Measures quality / characteristics of each experimental unit.

Quantitative variable : Measures qty / amount of each experimental unit.

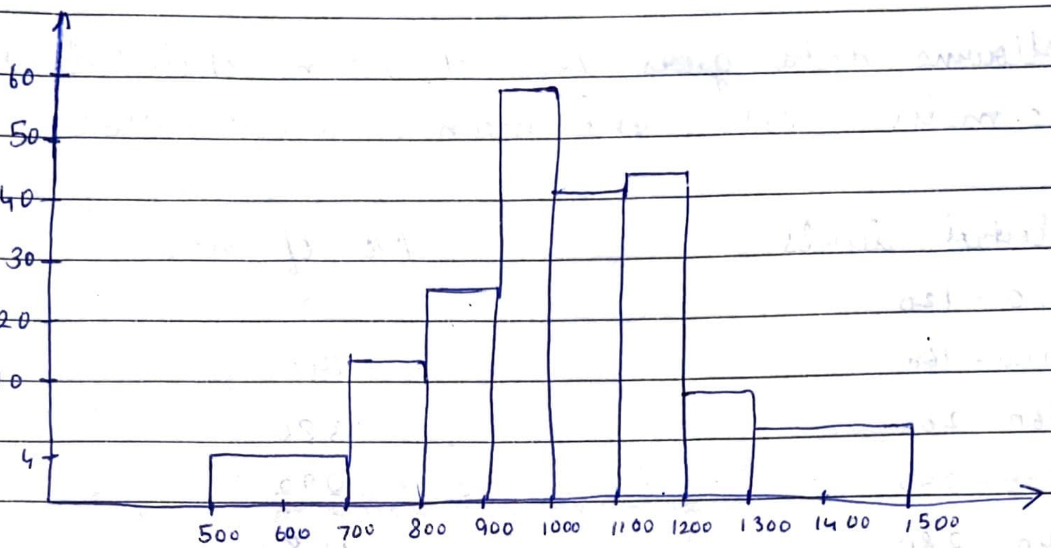
Discrete	Continuous
If variable can assume finite / ~~cont~~ countably ∞ no. of values.	If variable can assume ∞ no. of values b/w 2 specific limits.

- * frequency tables and graphs
- * histogram
- * freq. polygon & curves
- * cumulative frequency curves / graphs

Grouped data

Class interval	No. of lamps
500 - 700	8
700 - 800	11
800 - 900	25
900 - 1000	58
1000 - 1100	41
1100 - 1200	45
1200 - 1300	8
1300 - 1500	6

HISTOGRAM



Mean

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$\bar{x} = a + \left(\frac{\sum f_i u_i}{\sum f_i}\right) h$$

$$u_i = \frac{x_i - a}{h}$$

Median : $l + \left(\frac{n/2 - cf}{f}\right) h$

n : even — median : $\frac{\left(\frac{n}{2}\right)^{th} + \left(\left(\frac{n}{2}\right)+1\right)^{th}}{2}$

n : odd — $\left(\frac{n+1}{2}\right)^{th}$ obs.

Mode : $l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) h$

f_1 : freq. of modal elem.
f_0 : " " preceding "
f_2 : " " succeeding "

Mode = 3 Median − 2 Mean

Q. The following data gives freq. of serum cholestrol levels of 1000 males. Calc. the mean, median, mode

Cholestrol levels	No. of males
80 - 120	12
120 - 160	145
160 - 200	380
200 - 240	292
240 - 280	118
280 - 320	35
320 - 360	11
360 - 400	7

C.I.	x_i	f_i	$u_i = \frac{x_i - A}{h}$	$f_i u_i$
80-120	100	12	-2	-24
120-160	140	145	-1	-145
160-200	180	380	0	0
200-240	220	292	1	292
240-280	260	118	2	236
280-320	300	35	3	105
320-360	340	11	4	44
360-400	380	7	5	35
	Σf_i	1000		$\Sigma f_i u_i$: 543

$A = 180$

$h = 40$

$$\text{Mean} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) h$$

$$= 180 + \left(\frac{543}{1000}\right) 40 = 201.72$$

Page ____ Date / /

$$S.D. = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i-\bar{x})^2} \qquad N = \sum_{i=1}^{n} f_i$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i-\bar{x})^2$$

Now, $\frac{1}{N}\sum_{i=1}^{N} f_i(x_i-\bar{x})^2$

$$= \frac{1}{N}\sum_{i=1}^{N} f_i(x_i^2+\bar{x}^2-2x_i\bar{x})$$

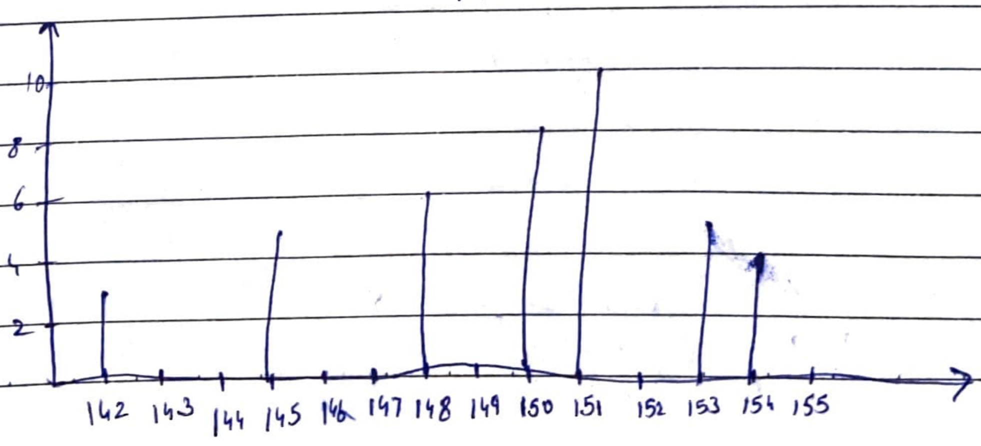
$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 + \frac{1}{N}\sum_{i=1}^{n} f_i\bar{x}^2 - \frac{1}{N}\sum_{i=1}^{n} f_i(2x_i.\bar{x})$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \bar{x}^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} f_i x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{n} f_i x_i\right)^2$$

Q.

Height (in cm)	frequency
142	3
145	5
148	6
150	8
151	10
153	5
154	4
	41

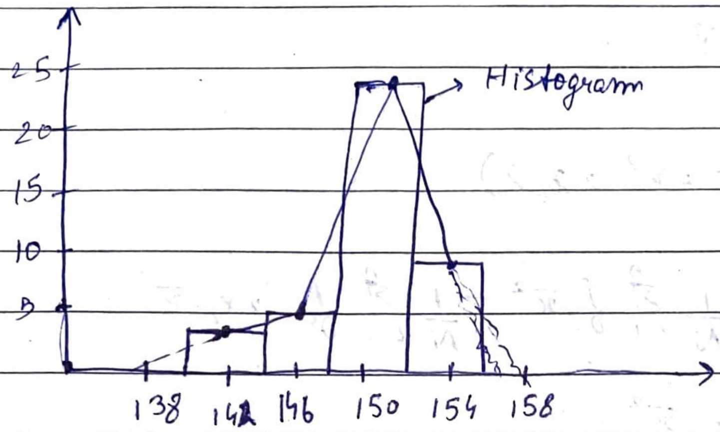


140 - 144 - 3

144 - 148 - 5

148 - 152 - 24

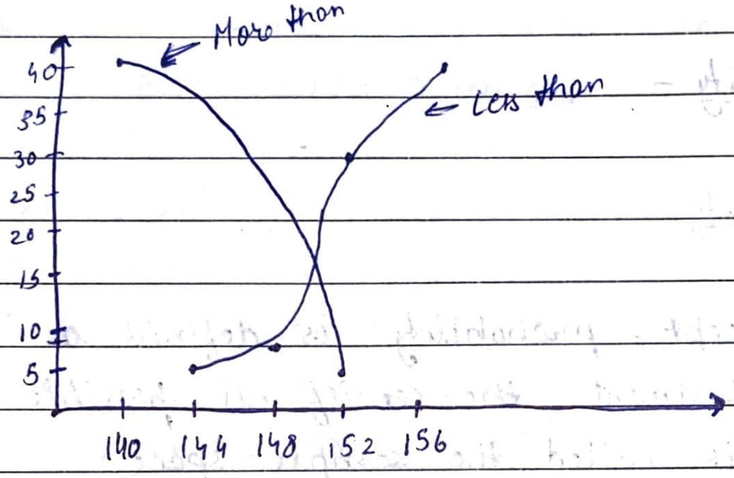
152 - 156 - 9



freq. polygon

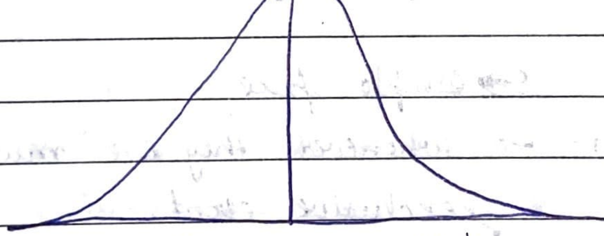
Less than	f
144	3
148	8
152	32
156	41

More than	F
140	41
144	38
148	33
152	9

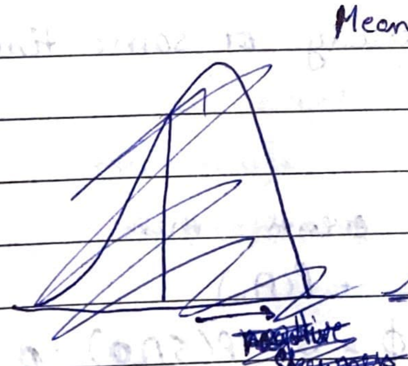


Skewness & Kurtosis –

Skewness – Lack of symmetry



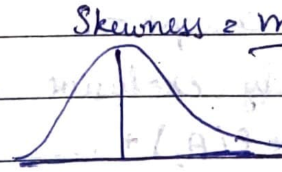
Mean = Median = Mode

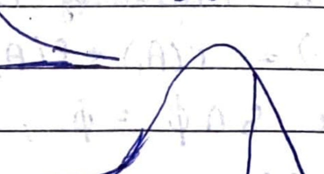


This Skewness is calculated using

$$\text{Skewness} = \frac{\text{mean} - \text{mode}}{\text{S.D.}}$$

Kart pearson's co-eff. of skewness





~~negative skewness~~

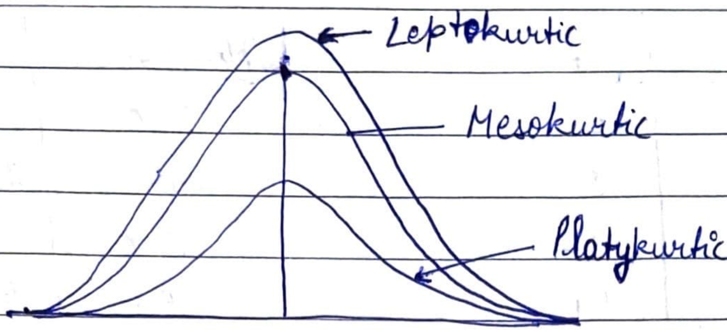
Mode < Median < Mean

positively skewed

~~positive~~ negatively

Mean < Median < Mode

Kurtosis



Kurtosis represent flatness or peakness of curve

Axiomatic Probability -

Axioms of Probability

* In axiomatic concept, probability is defined as set fn
* In a random experiment, the set of all possible outcomes of the experiment is called the sample space.
* Any subset A of the S is known as an event. An event is a set consisting of some or all possible outcomes of the exp.

Given a finite sample space S and an event A belongs to S, then probability of A satisfy following three axioms.

1. $0 \le P(A) \le 1$
2. $P(S) = 1$ $\quad S \rightarrow$ sample space
3. $P(A \cup B) = P(A) + P(B)$ $\rightarrow$ whenever they are mutually exclusive event.

No two events can happen simultaneously at same time

Results based on axiomatic concept -

1. If there are N events mutually exclusive events, then
$P(A_1 \cup A_2 \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$
2. ~~If~~ the sample space $S \cap \phi = \phi$, $S \cup \phi = S$; $P(S \cap \phi) = 0$
$P(S \cup \phi) = P(S) = 1$

③ If A is an event, then $A^c, \bar{A}$

$$P(A) = 1 - P(A^c)$$

$$P(A) + P(A^c) = 1$$

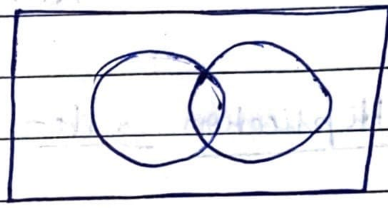
④ If $A \subset B$, then $P(A) \leq P(B)$

⇒ Additional rule of probability

If A & B are two events in a sample space S, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof-



$\cancel{A \cup B = A \cup (A^c \cap B)}$

$$P(A \cup B) = P(A) + P(A^c \cap B)$$

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

$$P(A^c \cap B) = P(B) - P(A \cap B) \quad -②$$

Substitute ② in ①

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

② $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$

③ $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$, Boole's inequality

Compount event.

When 2 or more events occur in collection with each other, their simultaneous occurance is known as compound event.

Independent event -

If the happening or non-happening of an event does not affect the happening or non-happening of another event.

11 12 13 14 15 16
21

1, 2 → 1, 2
1 2

Page
Date / /

Conditional Probability -

The probability for the event A to hold when it is known that event B has already occured is called the conditional probability.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Multiplication rule-

For two events A & B in S then

$$P(A \cap B) = \begin{cases} P(A|B)\, P(B) & , P(B) > 0 \\ P(B|A) \cdot P(A) & , P(A) > 0 \end{cases}$$

Q1.) Two dice ~~one~~ 1 green, 1 red are thrown. Let A be the event that the sum of the points on the faces are odd & B be the event that atleast one die show 1.

(i) Describe the complete sample space

(ii) Event $A, B, \bar{B}, A \cap B, A \cup B, A \cap \bar{B}$ and find their probabilities assuming that all the 36 events have equal probabilities.

(iii) Find the probabilities for $\bar{A} \cup \bar{B}$, $\bar{A} \cap \bar{B}$, $A \cap \bar{B}$, $\bar{A} \cap B$, $A \cup (\bar{A} \cap B)$, $\overline{A \cup B}$, $(A|B)$, $(B|A)$

Page ___ Date / /

Soln.)

$$S = \begin{cases} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6) \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{cases}$$

$\cancel{P(A) = \frac{1}{2}}$ $A = \{ (1,2), (2,1), (1,4) \ldots\ldots \}$

$P(A) = \frac{1}{2}$, $P(B) = \frac{11}{36}$, $P(\overline{B}) = \frac{25}{36}$

$P(A \cap B) = \frac{6}{36}$ $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{23}{36}$

$P(A \cap \overline{B}) = \frac{12}{36}$ $\quad P(\overline{A} \cup \overline{B}) = \frac{30}{36}$

$P(\overline{A} \cap \overline{B}) = \frac{13}{36}$

$P(A \cap \overline{B}) = \frac{1}{3}$

$P(\overline{A} \cap B) = \frac{5}{36}$

$P(A \cup (\overline{A} \cap B)) = \frac{23}{36}$

$P(\overline{A \cup B}) = \frac{13}{36}$

$P(A | B) = \frac{18}{33}$

$P(B | A) =$

$P(\overline{A} | \overline{B}) = \frac{13}{25}$

$P(\overline{B} | \overline{A}) = \frac{13}{18}$

Page ______
Date / /

Q. A card is drawn from a well shuffled pack of playing cards. What is the prob. that the card is a spade or an ace?

$$P(A) + P(S) - P(A \cap S)$$

Bayes' Theorem

If $B_1, B_2, B_3 \dots B_k$ are the mutually exclusive events in the sample space S of which one must occur and $P(B_i) \neq 0$ for $i = 1, 2, \dots k$, And A is subset of $\bigcup_{i=1}^{k} B_i$ then for any event B_i in S such that $P(A) \neq 0$

$$P(B_i | A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{i=1}^{k} P(B_i) \cdot P(A|B_i)}, \quad \text{for } i = 1, 2, 3 \dots k$$

Proof - Since $A = \bigcup_{i=1}^{k} (A \cap B_i)$, $A \cap B_i$ are M.E.

$$A \subset \bigcup_{i=1}^{k} B_i$$

$$A = A \cap \left(\bigcup_{i=1}^{k} B_i\right)$$

$$A = \bigcup_{i=1}^{k} (A \cap B_i)$$

then

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i)$$

$$= \sum_{i=1}^{k} P(B_i) P(A|B_i)$$

and $P(A \cap B_i) = P(A) \cdot P(B_i | A)$

$\cancel{(A \cap B_1) \cup (A \cap B_2)}$

$[\cancel{(A \cap B_1)} \cup A] \cap [(A \cap B_1) \cup \cancel{B_2}]$

$\cancel{(A \cup A) \cup (B_1 \cap A)}$

$\cancel{(A \cup A) \cap (B_1 \cup A)}$

$\cancel{(A \cap A)}$

Now

$$P(B_i | A) = \frac{P(A \cap B_i)}{P(A)}$$

$$= \frac{P(A|B_i) P(B_i)}{P(A)}$$

Page
Date / /

Q.) In 1989, there were 3 candidates for the position of principal Mr. A, Mr. B, Mrs. C whose chances of getting the appointment are in the proportion 4 : 2 : 3. The prob. that Mr. A if selected will introduce co-education is 0.3. The prob. of Mr. B & Mrs. C doing the same work respectively 0.5 and 0.8. What is the probability that there was co-education in the college in 1989.

$P(A) = \frac{4}{9}$, $P(B) = \frac{2}{9}$, $P(C) = \frac{3}{9}$

$P(D) = P[(D \cap A) \cup (D \cap B) \cup (D \cap C)]$

$= P(D \cap A) + P(D \cap B) + P(D \cap C)$

$P(A/D) = 0.3$

$P(B/D) = 0.5$

$P(C/D) = 0.8$

$P(D) = P(A/D) \times P(A) + P(B/D) \times P(B) + P(C/D) \times P(C)$

Q.) A bag contains 4 tickets numbered 1, 2, 3, 4 & another contains 6 tickets numbered 2, 4, 6, 7, 8, 9. If one of the two bags is chosen at random & a ticket is drawn at random from the chosen bag. Find the probs. that the ticket drawn have the number (i) 2 or 4

(ii) 3

(iii) 1 or 9

Page ____
Date / /

Random Variables

A random variable is a real number X connected with the outcome of a random experiment. For eg. If E consist of 2 faces of a coin, then consider the random variable which is the no. of heads

outcome - (a)	HH	HT	TH	TT
Value of X - $X(a)$	2	1	1	0

Thus, to each outcome a, there corresponds a number $X(a)$, since ~~these to~~ points of the sample space S corresponds to outcomes. This means that a real number which is denoted by $X(a)$ is defined for each a belongs to S.

Random Variable

A random variable is a real fn. on S and it can be defined as

1. Let S be the sample space associated with a given random experiment. A real valued fn defined on S and taking values in R is called one-dimensional random variable.
2. If the functions are ordered pairs of real nos. the fn. is said to be 2-dimensional random variable.
3. Prob. on a random variable X is
$$P(X \le b) = P\{a, X(a) \le b\}$$

outcome - (a)	HH	TH	HT	TT
Value of X (X(a)) -	2	1	1	0
Probability	2/4	1/4	1/4	0/4

$\underbrace{\tfrac{1}{4} \quad \tfrac{1}{4}}_{1/2}$

Discrete random variables -

Let X be a random variable. If the no. of possible values of X is finite or countably infinite, then it is called as discrete random variable.

The possible values of X may be listed as x_1, x_2 ---- in the finite, the list terminates & in infinite list continues indefinitely.

Probability Mass Function -

Let X is a discrete random variable with each possible outcome x_i in R. A number $P(x_i) = P(x-x_i)$ gives the probability that the random variable = the value of x_i.

The no. $P(x_i)$ satisfies - ① $P(x_i) \geq 0$ for all i

② $\Sigma P(x_i) = 1$

The collection pair $(x_i, P(x_i))$ is called the probability distribution of x & $P(x_i)$ is prob. mass fn. of x.

Continuous random Variable -

If the range space of random variable x is a interval then x is called as continuous random variable.

For a continuous random variable x, the prob. that x lies in the interval I is given by $I(a,b)$, then

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

This fn. $f(x)$ is prob. density function of random variable x and it satisfies ① $f(x) \geq 0$ $\forall$ x in R_x

② $\int_{R_x} f(x)\,dx = 1$

③ $\int f(x)\,dx = 0$, if $x \notin R_x$

(4) $P(x=x_0) = \int_{x_0}^{x_0} f(x)\,dx = 0$

Q.) The life of a device used to inspect cracks in aircraft ~~wheels~~ is given by a continuous random variable. Assuming all values in the range $x \geq 0$, the pdf of the life time in years is as follows

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & x \geq 0 \\ 0, & \text{else} \end{cases}$$

Calculate (1) $P(2 \leq x \leq 3)$

Joint Probability Law -

2 random variable x & y are said to be jointly distributed if they are defined on the same prob. space, the sample points consist of 2 doubles if the joint prob. ~~form sum~~ fn. is denoted by $P_{xy}(X,Y)$, then Prob. of a certain event is given by $P_{xy}(X,Y) = P[(X,Y) \in E]$

Joint Prob. Mass Fn. -

Let x & y be the random variable on the sample space S with respective image set

$$X(S) = \{x_1, x_2 \cdots x_n\}$$
$$Y(S) = \{y_1, y_2 - - y_n\}$$

then the product

$X(S) \times Y(S)$ ~~S~~ into a prob. space by defining the prob. of the ordered pair (x_i, y_j) is $P(x_i, y_j) = P_{ij}$

The function P on $X(S) \times Y(S)$ is defined by

$P_{ij} = P(X = x_i \cap Y = y_j)$ is called joint prob. fn. of X & Y.

X/Y	y_1, y_2 ...		y_n	Total
x_1	P_{11}	P_{12} ---	P_{1n}	P_{1j}
x_2	P_{21}	P_{22} ---	P_{2n}	P_{2j}
⋮	⋮	⋮	⋮	⋮
x_n	P_{n1}	P_{n2} ---	P_{nn}	P_{nj}
Total	P_{i1}	P_{i2} ---	P_{in}	1

Joint Marginal Prob. Fn.

Suppose the joint distribution of 2 random variables x & y is given, then prob. $P_x(x_i) = P(X=x_i) = P[X=x_i \cap Y=y_1]$
$+ P[X=x_i \cap Y=y_2] \ldots\ldots P[X=x_i \cap Y=y_n]$

$= P_{i1} + P_{i2} + \ldots + P_{in}$

$= \sum_{j=1}^{n} P_{ij}$

and this is known as Marginal Prob. Fn. of x.

Joint Conditional Prob. Fn.

The fn. $P[x=x_i \mid Y=y_j] = \dfrac{P[x_i \cap y_j]}{P(y_j)}$ is known as conditional prob. fn. of x when random variable y is given.

~~Similarly,~~

Q. For the joint P.D.F. of two variables x & y

$P(X=0, Y=1) = 1/3$

$P(X=1, Y=-1) = 1/3$

$P(X=1, Y=1) = 1/3$

① Calculate marginal dis. of x & y.

② " conditional dis. of x given by when $y=1$

①

X/Y	-1	1	X
0	0	1/3	1/3
1	1/3	1/3	2/3
Y	1/3	2/3	1

(ii) $P(X=0, Y=1) = \frac{P(X \cap Y)}{P(Y)} = \frac{1/3}{2/3} = \frac{1}{2}$

$P(X=1, Y=1) = \frac{P(X \cap Y)}{P(Y)} = \frac{1/3}{2/3} = \frac{1}{2}$