BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction
to Statistical Methods

ISM Team

Session-5
Random Variables
21^{st} & 22^{nd} December 2024

BITS Pilani
Pilani Campus

IMP Note to Self

Important Note for Students

innovate achieve lead

All submissions for graded components must be the result of your original effort.

- It is strictly All submissions for graded components must be the result of your original effort.
- It is strictly prohibited to copy and paste verbatim from any sources, whether online or from your peers.
- The use of unauthorized sources or materials, as well as collusion or unauthorized collaboration to gain an unfair advantage, is also strictly prohibited. Please note that we will not distinguish between the person sharing their resources and the one receiving them for plagiarism, and the consequences will apply to both parties equally.
- In cases where suspicious circumstances arise, such as identical verbatim answers or a significant overlap of unreasonable similarities in a set of submissions, will be investigated, and severe punishments will be imposed on all those found guilty of plagiarism.

Contact Session 5: Module 3: Probability Distributions

Contact Session	List of Topic Title	Reference
CS - 5	Random variables - Discrete & continuous Expectaion of a random variable,mean and variance of a random variable – Sinlge random random variable & Joint distributions	T1 & T2
HW	Problems on random variables	T1 & T2
Lab	Probability Distributions & Sampling	Lab 3

Agenda

- Random variables
- Discrete & continuous random variable,
- Probability Distribution Function
- Mean(Expectation) and variance of a random variable
- Joint distributions

Random Variables

- A **random variable** is a variable that assumes numerical values associated with the random outcome of an experiment, where one (and only one) numerical value is assigned to each sample point.
- In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers.

Random Variables

- A random variable can be classified as being either discrete or continuous depending on the numerical values it assumes:

- A discrete random variable may assume either finite or countably infinite number of values
- A continuous random variable may assume any numerical value in an interval or collection of intervals.
- Continuous random variables are generated in experiments where things are “measured’ as opposed to “counted”.
- Experimental outcomes based on measurement of time, distance, weight, volume etc. generate continuous RV.

Types of Random Variables

- A **discrete random variable** can assume a countable number of values.
  - Number of steps to the top of the Eiffel Tower
  - Number of Heads obtain after tossing two coins simultaneously

- A **continuous random variable** can assume any value along a given interval of a number line.
  - The time a tourist stays at the top once s/he gets there.
  - The height of a person can take any value within a certain range (e.g., 150 cm to 200 cm),

Examples of Random Variables

- **Discrete random variables**
  - Number of sales
  - Number of calls
  - Shares of stock
  - People in line
  - Mistakes per page

- **Continuous random variables**
  - Length
  - Depth
  - Volume
  - Time
  - Weight

Identify which of the following variables are discrete and which are continuous:

- The number of books in a library.
- The weight of a watermelon in kilograms.
- The amount of rainfall in a day (in millimeters).
- The number of passengers on a bus.
- The temperature of a cup of coffee (in °C).
- The number of goals scored in a football match.
- The length of a bridge (in meters).
- The number of steps you take in a day.
- The time it takes to bake a cake (in minutes).

Discrete Probability Distributions

- The probability distribution for a random variable describes how probabilities are distributed over the values of the random variable.
- The probability distribution is defined by a probability function, denoted by $f(x)$ or $p(x)$, which provides the probability for each value of the random variable.

The required conditions for a discrete probability function are:

$$f(x) \geq 0$$

$$\Sigma f(x) = 1$$

- We can describe a discrete probability distribution with a table, graph, or equation.
- Advantage: once the probability distribution is known, it is relatively easy to determine the probability of a variety of events that may be of interest to the decision maker.

Probability Distributions for Discrete Random Variables

Suppose you toss a fair coin 2 times, and you're interested in the number of heads obtained.

Let the random variable X represent the number of heads.

Here, X is a discrete random variable because it can take only integer values (0, 1, 2).

The Probability distribution function for X is:

$$f(0) = P\{TT\} = \frac{1}{4}$$

$$f(1) = P\{HT, TH\} = \frac{2}{4} = \frac{1}{2}$$

$$f(2) = P\{HH\} = \frac{1}{4}.$$

Here we can observe that $f(x) \geq 0 \ \forall x$ and $\sum f(x) = 1$.

Properties of pmf

(i) $p_X(x) = P(X = x) \geq 0$

(ii) $\sum_{x} p_X(x) = 1$

Verify whether $p_X(x)$ is a pmf.

x	0	1	2	3	4
$p(x)$	k	$3k$	$5k$	$7k$	$9k$

(i) $p_X(x) \geq 0$??

(ii) $\sum_{x=0}^{4} p_X(x) = 1$

$= k + 3k + 5k + 7k + 9k = 1$

$\Rightarrow 25k = 1$

$\Rightarrow \therefore k = \frac{1}{25} = 0.04$

x :	0	1	2	3	4
$p(x)$:	0.04	0.12	0.20	0.28	0.36

$$\sum_{x=0}^{4} p_x(x) = 0.04 + 0.12 + 0.20 + 0.28 + 0.36 = 1$$

$$p_X(x) \geq 0$$

Description of data - Mean & SD

Frequency distr

Marks: 75, 81, 64, 80, 70

x	$(x-\bar{x})$	$(x-\bar{x})^2$
75	1	1
81	7	49
64	-10	100
80	6	36
70	-4	16
$\bar{x} = 74$	$\Sigma(x-\bar{x}) = 0$	$\Sigma(x-\bar{x})^2 = 202$

$$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$= \frac{202}{4} = 50.5$$

(74, 50.5)

(74, 7.11)

Expected Value

If X is a discrete random variable with pmf $p_X(x)$ then Expected value of the random variable X is

$$E(X) = \sum_x x\, p_X(x)$$ — 1st order moment (mean

$$E(X^2) = \sum_x x^2 p(x)$$ - 2nd order moments

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

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

$$\bar{X} = 25$$

$n = 10, 10$

$$\sum_{i=1}^{10} \bar{x} = \sum_{i=1} 25 = 25 + 25 + \cdots + 25 = 10 \times 25 = 25\ldots$$

$$(n-1)S^2 = \sum_{1}^{n}(x - \bar{x})^2 = \sum_{1}^{n}\left(x^2 - 2x\bar{x} + \bar{x}^2\right)$$

$$= \sum_{1}^{n} x^2 - 2\bar{x}\sum_{1}^{n} x + \sum_{1}^{n} \bar{x}^2$$

$$\because \bar{x} = \frac{\sum x}{n}$$

$$\therefore \sum x = n\bar{x}$$

$$= \sum_{1}^{n} x^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2$$

$$= \sum_{1}^{n} x^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{1}^{n} x^2 - n\bar{x}^2$$

$$(n-1)s^2 = \sum x^2 - n\bar{x}^2$$

$$= \sum x^2 - \frac{(\sum x)^2}{n}$$

$$= \frac{1}{n-1}\sum x^2 - \frac{1}{n-1}\left(\frac{(\sum x)^2}{n}\right)$$

$$n\bar{x}^2 = n\left(\frac{\sum x}{n}\right)^2 = \frac{(\sum x)^2}{n}$$

$\therefore$ V(x) for prob. distn: $E(x^2) - (E(x))^2$

Properties of E(x) and V(x)

Expected Value

1. If K is a constant and X is a discrete random variable with pmf $P_X(x)$, then

$$E(kX) = kE(X)$$

Proof: $E(kX) = \sum_{x} k\, x\, P_X(x) = k \sum_{x} x\, P_X(x)$

$$= k\,E(X)$$

2. If X is a drv taking the constant value $\underline{k}$ and the pmf of X is $p_X(x)$ then $E(k) = k$, a constant

Proof: $E(X) = \sum x\, p(x)$

$= \sum k\, p(x)$

$= k \left(\sum p(x) \right) = 1$

$= k \cdot 1 = k$

3. If X and Y are 2 drv with pmf $p(x)$

then $E(X+Y) = E(X) + E(Y)$

4. If X and Y are independent drvs

then $E(XY) = E(X)E(Y)$

5. If 'a' and 'b' are constants, then

$E(a+bx) = a + bE(x)$

Properties of Variance:

1. If $X = k$, a constant, then

$$V(k) = 0$$

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

$$= \sum x^2 p(x) - \left(\sum k\, p(x)\right)^2$$

$$= \sum k^2 p(x) - k^2 \left(\sum p(x)\right)^2$$

$$= k^2 \sum p(x) - k^2 \left(\sum p(x)\right)^2$$

$$= k^2 \cdot 1 - k^2 \cdot 1^2 = k^2 - k^2 = 0$$

X:	0	1	2	3
$p(x)$:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

If $Y = a + bX$

Find $E(a + bX)$

where $a = 3$ & $b = 5$

$E(Y) = ?$

2. If X is a discrete random variable with pmf $p_X(x)$ and $Y = bX$, then $V(Y) = V(bX)$
$= b^2 V(X)$

Proof: $V(Y) = E(Y^2) - (E(Y))^2$

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

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

$= a^2 E(X^2) - a^2 (E(X))^2 = a^2 [E(X^2) - (E(X))^2]$

$= a^2 V(X)$

3. If a and b are constants, X is a d.r.v with pmf $p_X(x)$, then $V(a+bX) = b^2 V(X)$

4. If X and Y are two d.r.vs, with their respective pmfs, then $V(X \pm Y) = V(X) \pm V(Y)$

5. If X and Y are two d.r.vs, with their respective pmfs, $Cov(X,Y) = E(XY) - E(X)E(Y)$

However, if x and y are independent, then

$\operatorname{cov}(x,y) = 0$

because $E(xy) = E(x)E(y)$

$\therefore \operatorname{cov}(x,y) = E(x)E(y) - E(x)E(y) = 0$

Examples:

1. X : 0 1 2 3

$p(x)$: 0.125 0.375 0.375 0.125

Find $E(X)$ and $V(X)$

Solution:

$$E(X) = \sum_{x=0}^{3} x\,p(x) = 0\times 0.125 + 1\times 0.375 + 2\times 0.375 + 3\times 0.125$$

$$= 0 + 0.375 + 0.750$$

$$= \underline{\underline{1.5}}$$

$$E(x^2) = \sum_{x=0}^{3} x^2 p(x)$$

$$= 0^2 \times 0.125 + 1^2 \times 0.375 + 2^2 \times 0.375 + 3^2 \times 0.125$$

$$= 0 + 0.375 + 1.5 + 1.125$$

$$= 3$$

$$\therefore V(x) = E(x^2) - (E(x))^2 = 3 - (1.5)^2 = 3 - 2.25$$

$$= \underline{\underline{0.75}}$$

Verify whether $p_X(x)$ is a pmf.

x	0	1	2	3	4
$p(x)$	k	$3k$	$5k$	$7k$	$9k$

$p(x)$: 0.04 0.12 0.20 0.28 0.36

$$E(X) = \sum_{x=0}^{4} x\, p(x)$$

$$E(X^2) = \sum_{x=0}^{4} x^2 p(x)$$

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

Age(yrs)-X	freq(f)	Cumulative freq.
0 - 1	10	10
2 - 5	13	23
6 - 12	26	49
13 - 19	31	80
20 - 29	38	118
30 - 39	27	145
40 - 49	16	161
50 - 59	19	180
60 and above	10	190

Probability distribution

→ Cumulative Probability distribution Functi

(cdf): $F_X(x) = P(X \leq x), \forall x$

X	$p(x)$	$F_X(x) = P(X \leq x)$
0	0.125	0.125 $(P(X=0))$
1	0.375	0.500 $(P(X \leq 1) = P(X=0) + P(X=1))$
2	0.375	0.875 $(P(X \leq 2) = P(X=0) + P(X=1) + P(X=$
3	0.125	1.000 $(P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) +$ $P(X=3))$

Verify whether $p_X(x)$ is a pmf.

x	0	1	2	3	4
$p(x)$	k	$3k$	$5k$	$7k$	$9k$
$p(x)$:	0.04	0.12	0.20	0.28	0.36

x	$p(x)$	$F(x) = P(X \leq x)$
0	0.04	0.04 $(F(0) = P(X \leq 0))$
1	0.12	0.16 $(F(1) = P(X \leq 1))$
2	0.20	0.36 $(F(2) = P(X \leq 2))$
3	0.28	0.64 $(F(3) = P(X \leq 3))$
4	0.36	1.00 $(F(4) = P(X \leq 4))$

Relationship between pmf and cdf

→ Given $p(x)$, then $F_X(x)$ can be computed by $F_X(x) = P(X \leq x)$, $\forall x$

→ Given $F_X(x)$, then $p(x)$ can be computed by $p_X(x) = F_X(x) - F_X(x-1)$

Example:

Note: $F(0) = P(0) - P(0-1) = 0$

X	$F_X(x)$	$p_X(x) = F_X(x) - F_X(x-1)$
0	0.04	$p(0) = F(0) = 0.04$
1	0.16	$p(1) = F(1) - F(0) = 0.16 - 0.04 = 0.12$
2	0.36	$p(2) = F(2) - F(1) = 0.36 - 0.16 = 0.20$
3	0.64	$p(3) = F(3) - F(2) = 0.64 - 0.36 = 0.28$
4	1.00	$p(4) = F(4) - F(3) = 1 - 0.64 = 0.36$

x	$p(x)$
0	0.125
1	0.375
2	0.375
3	0.125

Other than 0, 1, 2, & 3 prob. are zero

Example

At a shooting range, a shooter is able to hit a target in either 1, 2 or 3 shots. Let x be a random variable indicating the number of shots fired to hit the target. The following probability function was proposed:

$$f(x) = \frac{x}{6}, \; x = 1,2,3.$$

Is this probability function valid?

Expected Value

- Mean is a measure of location or central tendency in the sense that it roughly locates a middle or average value of the random variable

- The mean or Expected value of a discrete random variable:

$$E(X) = \mu = \sum x\, f(x)$$

Rules of Expected Value

- Multiplying RV by a constant a, $E(aX) = aE(X)$
- Adding a constant b, $E(X + b) = E(X) + b$
- Therefore, E(aX + b) = ?
- E(X/Y)$\neq \frac{E(X)}{E(Y)}$
- E(X.Y)$\neq E(X).E(Y)$ unless they are independent

Variability of Discrete Random Variables

- The **variance** of a **discrete random variable** *x* is

$$var(X) = \sigma^2 = \sum E((X-\mu)^2)$$

- Or $$var(X) = \sigma^2 = E(X^2) - \big(E(X)\big)^2$$

- **The <u>standard deviation</u>(** ***σ*)**, is defined as the positive square root of the variance.

Rules of variability

- $V(b) = 0$, b is the constant
- Multiplying RV by a constant $a, V(aX) = a^2V(X)$
- Therefore, V(aX+b) = ?
- $V(aX + bY) = a^2V(X) + b^2V(Y)$, whenever X and Y are independent
- $\sigma_{aX} = |a|\sigma_x$

Example:

- Discrete random variable with a finite number of values
  - Let *x* = number of TV sets sold at the store in one day

    where *x* can take on 5 values (0, 1, 2, 3, 4)

- Discrete random variable with an infinite sequence of values
  - Let *x* = number of customers arriving in one day

    where *x* can take on the values 0, 1, 2, . . .
  - We can count the customers arriving, but there is no finite upper limit on the number that might arrive.

Example : KPN Appliances

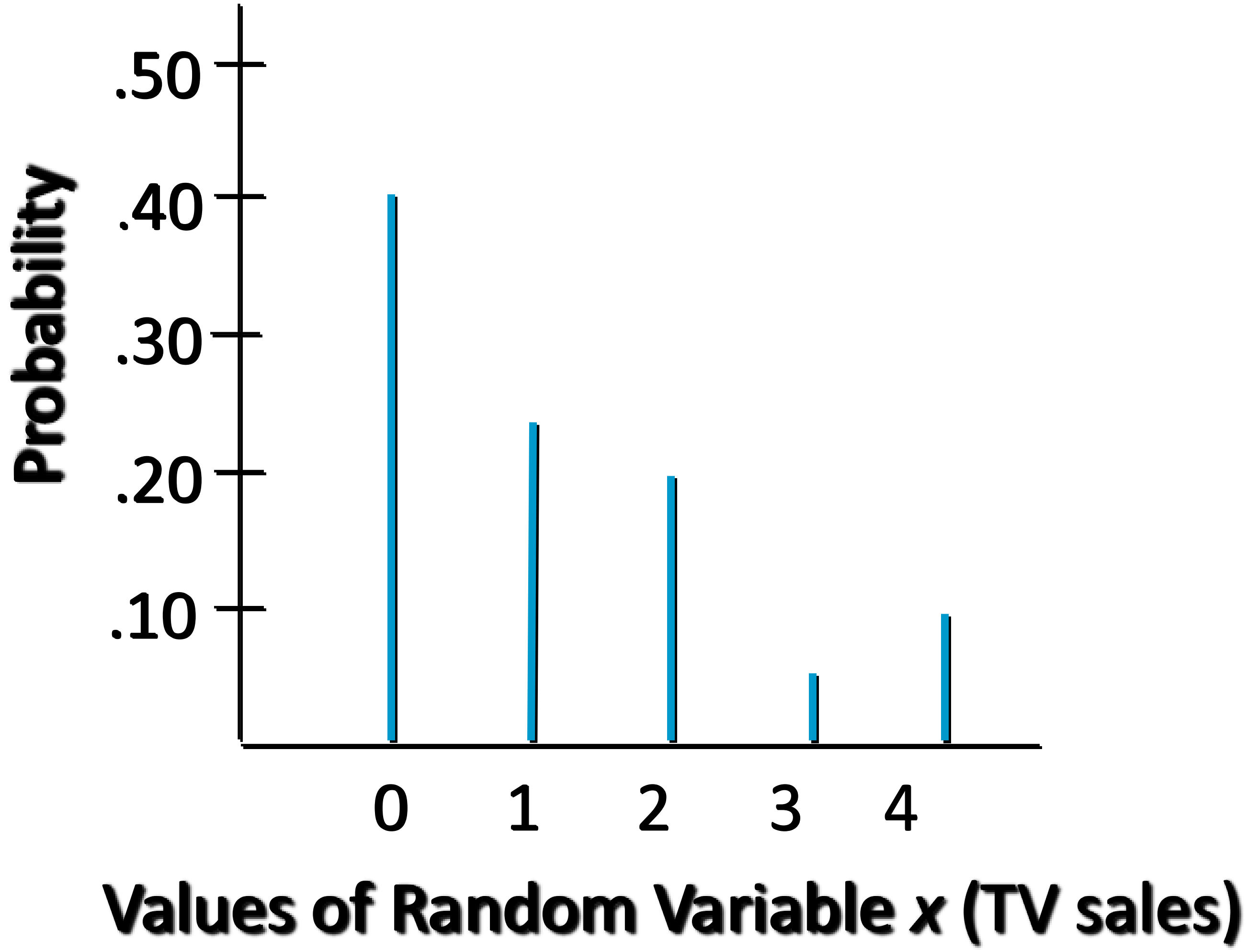
Using past data on TV sales (below left), a tabular representation of the probability distribution for TV sales (below right) was developed.

Units Sold	No of days	x	f(x)
0	80	0	0.4
1	50	1	0.25
2	40	2	0.2
3	10	3	0.05
4	20	4	0.1
Total	200	1	

— pmf → $p_X(x)$

Example:

- Graphical Representation of the Probability Distribution



Example:

Expected Value of a Discrete Random Variable

x	f(x)	xf(x)
0	0.4	0.00
1	0.25	0.25
2	0.2	0.40
3	0.05	0.15
4	0.1	0.40

E(x) = 1.20

The expected number of TV sets sold in a day is 1.2

Example: KPN Appliances

- Variance and Standard Deviation of a Discrete Random Variable

x	$x - \mu$	$(x - \mu)^2$	$f(x)$	$(x - \mu)^2 f(x)$
0	-1.2	1.44	.40	.576
1	-0.2	0.04	.25	.010
2	0.8	0.64	.20	.128
3	1.8	3.24	.05	.162
4	2.8	7.84	.10	.784
				$1.660 = \sigma^2$

- The variance of daily sales is 1.66 TV sets squared.
- The standard deviation of sales is 1.2884 TV sets.

Cumulative Distribution for Discrete Data

Cumulative probability Distribution:

It is given by $F(x) = P(X \leq x)$.

Example: If two dices are rolled. Find the probability distribution and cumulative probability distribution. Also represent in graph.

Solution: X- random variable is sum of the two numbers

X	1	2	3	4	5	6	7	8	9	10	11	12
P(x)	-	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
F(x)	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

Example 1

Two cards are drawn simultaneously (or successively without replacement) from a well shuffled pack of 52 cards. Find the mean, variance and standard deviation of the number of kings.

Solution: Let X denote the number of kings in a draw of two cards. X is a random variable which can assume the values 0, 1 or 2.

$$f(0) = P(no\ king) = 48_{C_2}/52_{C_2} = \frac{188}{221}.$$

$$f(1) = P(1\ king) = \frac{4_{C_1}.48_{C_1}}{52_{C_2}} = \frac{32}{221}.$$

$$f(2) = P(2\ kings) = \frac{4_{C_2}}{52_{C_2}} = \frac{1}{221}$$

Solution:

Mean of X:

$$E(X) = \sum x\, f(x) = 0 \times \frac{188}{221} + 1 \times \frac{32}{221} + 2 \times \frac{1}{221} = \frac{34}{221}$$

Var of X:

$$E(X^2) = \sum x^2 f(x) = 0^2 \times \frac{188}{221} + 1^2 \times \frac{32}{221} + 2^2 \times \frac{1}{221} = \frac{36}{221}.$$

$$var(X) = E(X^2) + \left(E(X)\right)^2 = \frac{36}{221} - \left(\frac{34}{221}\right)^2 = \frac{6800}{(221)^2}$$

$$\sigma_x = \sqrt{\frac{6800}{(221)^2}} = 0.37.$$

Example 2

A random variable X has the following probability function :

x	0	1	2	3	4
$p(x)$	k	$3k$	$5k$	$7k$	$9k$

Find $i)$ $the\ value\ of$ k, $ii)\ P(X < 3),\ iii)\ P(0 < X < 4)$, (iv) the distribution function of X, $v)\ E(X),\ vi)\ Var(X)$

Solution: $i)$ We Know that $\sum p(x) = 1$, therefore, $k + 3k + 5k + 7k + 9k = 1$. This gives $25k = 1\ \rightarrow k = 1/25$.

$$ii)\ P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 9k = \frac{9}{25}.$$

Solution

$iii)$ The distribution function of X is

$$\begin{array}{ll} For\ x < 0, & F(x) = 0\ , \\ For\ 0 \le x < 1, & F(x) = 1/25 \\ For\ 1 \le\ x < 2, & F(x) = 4/25 \\ For\ 2 \le x < 3\ , & F(x) = 9/25 \\ For\ x \ge 1\ , & F(x) = 1. \end{array}$$

$iv)\ E(x) = \sum x.f(x) = 0.\frac{1}{25} + 1.\frac{3}{25} + 2.\frac{5}{25} + 3.\frac{7}{25} + 4.\frac{9}{25} = \frac{70}{25}$

$v)\ \ var(x) = E(x^2) - \left(E(x)\right)^2.$

$$E(x^2) = \sum x^2 f(x) = 0.\frac{1}{25} + 1.\frac{3}{25} + 4.\frac{5}{25} + 9.\frac{7}{25} + 16.\frac{9}{25} = \frac{230}{25}$$

$$\therefore var(x) = \frac{230}{25} - \left(\frac{70}{25}\right)^2$$

Discrete Uniform Probability Distribution

- The discrete uniform **probability** distribution is the simplest example of a discrete probability distribution given by a formula.
- The discrete uniform probability function is

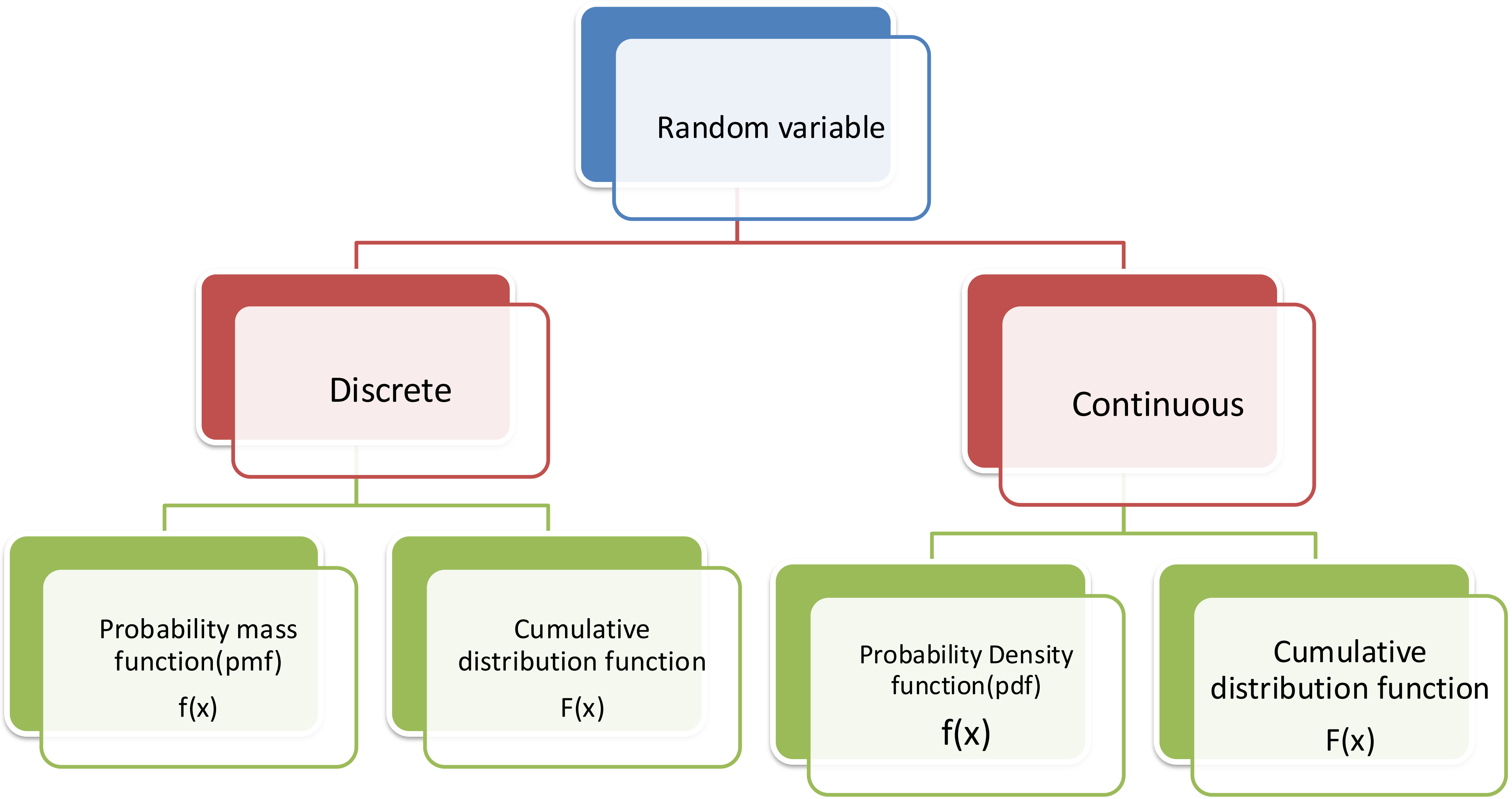
$$f(x) = 1/n$$

where:

n = the number of values the random variable may assume

Note that the values of the random variable are equally likely.

Classification of Random Variables



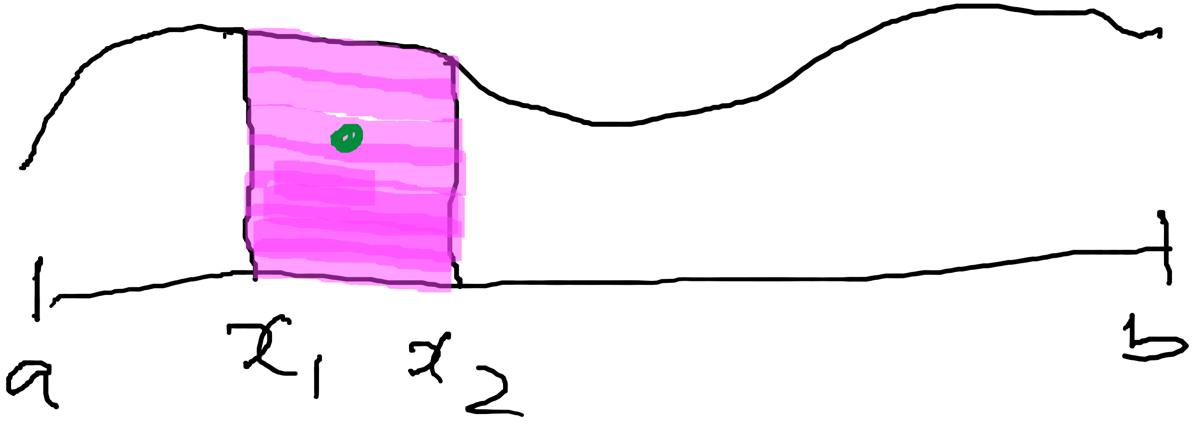
Continuous Probability Distributions

- A continuous random variable can assume any value in an interval on the real line or in a collection of intervals.
- It is not possible to talk about the probability of the random variable assuming a particular value.
- Instead, we talk about the probability of the random variable assuming a value within a given interval.
- The probability of the random variable assuming a value within some given interval from x_1 to x_2 is defined to be the area under the graph of the probability density function between x_1 and x_2.

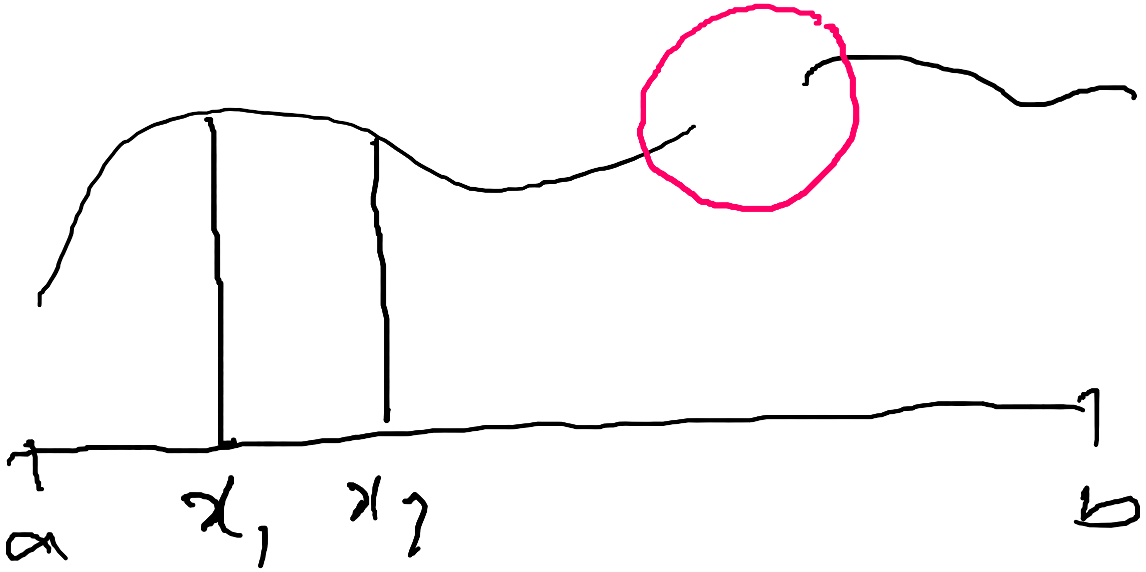
infinitesimal interval

continuous

(x_1, x_2)



Discontinuous



$$\int_{x_1}^{x_2} \boxed{f(x)}\, dx \longrightarrow \text{Probability density function}$$

Continuous Random Variables

A continuous random variable can assume any value in an interval on the real line or in a collection of intervals.

$$\int_{x_1}^{x_2} f(x)\,dx = 0$$

$$f(x_2) - f(x_1) = 0$$

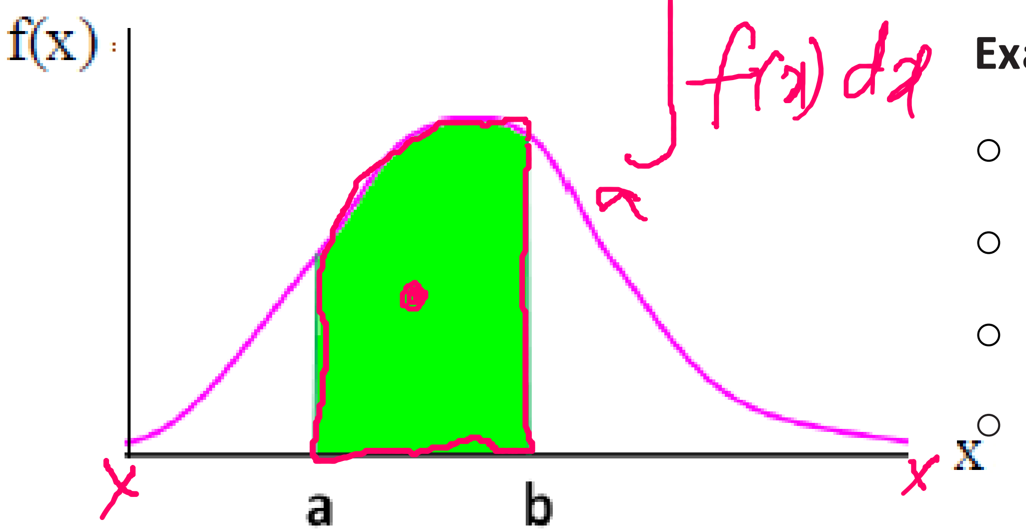
The probability of the random variable assuming a value within some given interval from x1 to x2 is defined to be the area under the graph of the **probability density function** between x1 and x2

$= p(3)$

$2) = 2/8$

x	$p(x)$
0	$\frac{1}{8}$
1	$\frac{2}{8}$
2	$2/8$
3	$1/8$

$$\int_a^b f(x)\,dx$$



Example:

- Height of students in a class
- Amount of ice tea in a glass
- Change in temperature throughout a day
- Price of a car in next year

Continuous Random Variables

Probability Density Function

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx$ = area under $f(x)$ from a to b for any a and b (4.1)

Pmf

(1) $p_X(x) \geq 0$

(2) $\sum_x p_X(x) = 1$

Continuous Random Variables

innovate achieve lead

Cumulative Distribution Function

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

for $-\infty < x < \infty$.

Probability Density Function from the Cumulative Distribution Function

Given $F(x)$,

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

cdf for pmf

$F(x) = P(X \leq x),\ \forall x$

$= \sum_{-\infty}^{x} p_X(x)$

$p_X(x) = F(x) - F(x^-)$

Entire range of $(-\infty, \infty)$

If you are finding area between (a, b), then $\int_a^b f(x)dx \neq 1$

Continuous Random Variables

Mean and Variance

Suppose that X is a continuous random variable with probability density function $f(x)$. The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

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

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

pmf

$$\mu = E(X) = \sum_x x\, p(x)$$

$$E(X^2) = \sum_x x^2 p(x)$$

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

$$\sigma = \sqrt{V(X)}$$

Recall: Integration Formulas

$$\int kf(u)du = k\int f(u)du \quad = K$$

$$\int u^n du = \frac{u^{n+1}}{n+1}$$

$$\int e^u du = e^u$$

$$\int \sin u \, du = -\cos u$$

$$\int \cos u \, du = \sin u$$

$$\int [f(u) \pm g(u)]du = \int f(u)du \pm \int g(u)du$$

EXAMPLE 1 Calculating probabilities from the probability density function

If a random variable has the probability density

$$f(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$ (Elsewhere/otherwise)

find the probabilities that it will take on a value

(a) between 1 and 3;

(b) greater than 0.5.

$2\int_0^\infty e^{-2x}\,dx = \cancel{2}\left[\frac{e^{-2x}}{-\cancel{2}}\right]_0^\infty$

$= \left(-e^{-2x}\right)_0^\infty$

$= -e^{-2(\infty)} - \left(-e^{-2(0)}\right) = 0 + 1 \Rightarrow \int_0^\infty 2e^{-2x}\,dx = 1$

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{Elsewhere (otherwise)} \end{cases}$$

$$p(x) = \begin{cases} \frac{x}{6}, & x = 1, 2, 3 \\ 0, & \text{Elsewhere} \end{cases}$$

Solution Evaluating the necessary integrals, we get

(a) $$\int_1^3 2e^{-2x}\,dx = e^{-2} - e^{-6} = 0.133$$

(b) $$\int_{0.5}^\infty 2e^{-2x}\,dx = e^{-1} = 0.368$$

$$\left.\begin{aligned} p(1) &= \frac{1}{6} \\ p(2) &= \frac{2}{6} \\ p(3) &= \frac{3}{6} \end{aligned}\right\} \sum_{x=1}^{3} p(x) = 1$$

Note: $\int_0^\infty e^{-t}\,dt = \left(\frac{e^{-t}}{-1}\right)_0^\infty = \left(-e^{-x}\right)_0^\infty = -e^{-\infty} - \left(-e^{-0}\right)$

$= 0 + 1 = 1$

(1) $\int_1^3 f(x)\,dx = 2\int_1^3 e^{-2x}\,dx$

$= \cancel{2}\left(\frac{e^{-2x}}{-\cancel{2}}\right)_1^3$

$= -\left(e^{-2x}\right)_1^3$

$= -e^{-2(3)} - \left(-e^{-2(1)}\right)$

$= e^{-2} - e^{-6}$

(2) $P(X \geq 0.5) = \int_{0.5}^{\infty} f(x)\,dx$

$= 2\int_{0.5}^{\infty} e^{-2x}\,dx$

$= \cancel{2}\left(\frac{-e^{-2x}}{-\cancel{2}}\right)_{0.5}^{\infty}$

$= -e^{-2(\infty)} + e^{-2(0.5)}$

$= 0 + e^{-1} =$

With reference to the preceding example, find the distribution function and use it to determine the probability that the random variable will take on a value less than or equal to 1.

Performing the necessary integrations, we get

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \int_0^x 2e^{-2t}\,dt = 1 - e^{-2x} & \text{for } x > 0 \end{cases}$$

and substitution of $x = 1$ yields

$$F(1) = 1 - e^{-2} = 0.865$$

$$F(1) = 1 - e^{-2(1)} =$$

$$F(x) = P(X \le x)$$

$$F(x) = \int_0^x 2e^{-2t}\,dt, \; x > 0$$

$$= 2\int_0^x e^{-2t}\,dt$$

$$= \cancel{2}\left(\frac{e^{-2t}}{-\cancel{2}}\right)_0^x$$

$$= \left(-e^{-2t}\right)_0^x$$

$$= -e^{-2x} - \left(-e^{-2(0)}\right)$$

$$= 1 - e^{-2x}$$

Determining the mean and variance using the probability density function

With reference to Example 1, find the mean and the variance of the given probability density.

Performing the necessary integrations, using integrations by parts, we get

$$\mu = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^{\infty} x \cdot 2e^{-2x}\, dx = \frac{1}{2}$$

Alternatively, the expectation of x is $E(X) = 0.5$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_0^{\infty} \left(x - \frac{1}{2}\right)^2 \cdot 2e^{-2x}\, dx = \frac{1}{4}$$

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

$$\mu = E(x) = \int_0^{\infty} x \left(2e^{-2x}\right) dx$$

$$= 2\int_0^{\infty} \underline{x e^{-2x}}\, dx$$

$$= ?$$

A probability density function assigns probability one to $(-\infty, \infty)$

Find k so that the following can serve as the probability density of a random variable:

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ kxe^{-4x^2} & \text{for } x > 0 \end{cases}$$

Solution

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} kxe^{-4x^2}\,dx = \int_{0}^{\infty} \frac{k}{8} \cdot e^{-u}\,du = \frac{k}{8} = 1$$

so that $k = 8$.

Cumulative Distribution for continuous Data

The cumulative distribution function of the random variable Y is

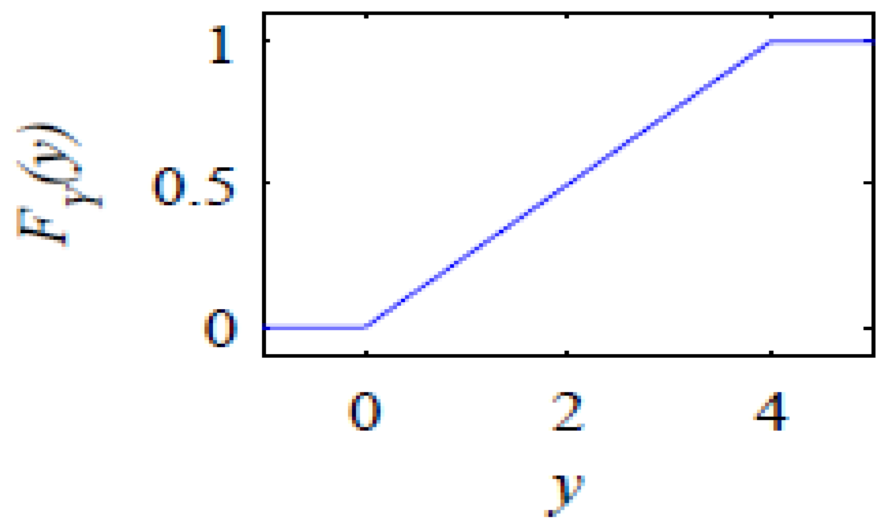
$$F_Y(y) = \begin{cases} 0 & y < 0, \\ y/4 & 0 \le y \le 4, \\ 1 & y > 4. \end{cases}$$

Sketch the CDF of Y and calculate the following probabilities:

(1) $P[Y \le -1]$ (2) $P[Y \le 1]$

(3) $P[2 < Y \le 3]$ (4) $P[Y > 1.5]$

The CDF of Y is



From the CDF $F_Y(y)$, we can calculate the probabilities:

(1) $P[Y \le -1] = F_Y(-1) = 0$

(2) $P[Y \le 1] = F_Y(1) = 1/4$

(3) $P[2 < Y \le 3] = F_Y(3) - F_Y(2) = 3/4 - 2/4 = 1/4$

(4) $P[Y > 1.5] = 1 - P[Y \le 1.5] = 1 - F_Y(1.5) = 1 - (1.5)/4 = 5/8$

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

innovate achieve lead

Problem:1

If the probability density of a random variable X is given by

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & otherwise \end{cases}$$

$\rightarrow \int_0^1 x\,dx + \int_1^2 2-x\,dx$

Find $i)\ P(0.2 < X < 0.8)$ $\qquad ii)\ P(0.6 < X < 1.2)$

Ans: $i)$ 0.3 $ii)$ 0.5

$\int_{0.2}^{0.8} x\,dx$

$\int_0^{0.6} x\,dx + \int_{0.6}^{1.2} (2-x)\,dx$

Problem:2

Let $f(x) = \frac{k}{1+x^2}$, $-\infty < x < \infty$ be a probability density function then find the value of k.

Ans: $k = \frac{1}{\pi}$.

$\int_{-\infty}^{\infty} f(x)\,dx = k\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = 1$

Problem:3

The length of satisfactory service (years) provided by a certain model of laptop computer is random variable having the probability density function :

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{cases} \frac{1}{4.5} e^{-\frac{x}{4.5}} & x > 0 \\ 0 & otherwise \end{cases}$$

Find the probabilities that one of these laptops will provide satisfactory service for

a) At most 2.5 years (Ans: 0.4262)

b) Anywhere from 4 to 6 years (Ans: 0.1475)

c) At least 6.75 years (Ans:0.2231)

Problem:4

For following probability density function :

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & otherwise \end{cases}$$. Find the mean and variance.

Ans: Mean =1, Variance =$\frac{1}{6}$

Problem:5

If $P(X = x) = \begin{cases} kx & x = 1,2,3,4,5 \\ 0 & otherwise \end{cases}$ represents a probability function , find $i)\ k,\quad ii)\ P(X\ \ being\ a\ prime\ number),\quad iii)$ $P\left(\frac{1}{2} < X < \frac{5}{2}\right)$.

(Answer: $i)\ k = \frac{1}{15},\quad ii)\frac{11}{15}\quad iii)\frac{1}{5}$)

Problem:6

From a lot of 10 items containing 3 defective items, a sample of 4 items is drawn at random. Let the random variable X denote the number of defective items in the sample. Answer the following when the sample is drawn without replacement:

a) Find the probability distribution of X.

b) Find $P(X \leq 1)$

c) Find $P(0 < X < 2)$

Ans: a)

x	0	1	2	3
$P(X)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

b) $\frac{2}{3}$ c) $\frac{1}{2}$

Problem:7

The probability mass function of a random variable is zero except at the points $x = 0,1,2$. At these points it has the values $P(0) = 3a^3,\ P(1) = 4a - 10a^2,\ \ P(2) = 5a - 1$ for some $a > 0.$

a) Determine the value of a

b) Compute the probabilities $P(X < 2)$

c) Find the largest x such that $F(x) < \frac{1}{2}$, where $F(x)$ is cumulative distribution function.

Ans: a) $a = \frac{1}{3}$, b) $\frac{1}{3}$ c) $x = 1.$

Joint Probability Distribution

Discrete Joint Probability Mass Function

A bag contains 3 Red balls, 4 Green balls and 2 Blue balls. 2 balls are selected at random from the bag after shuffling well.
Let X denote the number red balls and Y denote the number Green balls. construct the probability distribution X and Y.

innovate achieve lead

Range of both X and Y is 0, 1, 2

X \ Y	0	1	2
0	$\frac{1}{36}$	$\frac{8}{36}$	$\frac{6}{36}$
1	$\frac{6}{36}$	$\frac{12}{36}$	0
2	$\frac{3}{36}$	0	0

(X, Y)
(0, 0)
(0, 1)
(0, 2)
(1, 0)
(1, 1)
(1, 2) ✗

(2, 0)
(2, 1) ✗
(2, 2) ✗

R G B

$$\overset{x\ \ y}{(0,0)} = \frac{{}^3C_0 \times {}^4C_0 \times {}^2C_2}{{}^9C_2} = \frac{1 \times 1 \times 1}{36} = \frac{1}{36}$$

$$(0,1) = \frac{{}^3C_0 \times {}^4C_1 \times {}^2C_1}{{}^9C_2} = \frac{1 \times 4 \times 2}{36} = \frac{8}{36}$$

$$(0,2) = \frac{{}^3C_0 \times {}^4C_2 \times {}^2C_0}{{}^9C_2} = \frac{1 \times 2 \times 3}{36} = \frac{6}{36}$$

$$(1,0) = \frac{{}^3C_1 \times {}^4C_0 \times {}^2C_1}{{}^9C_2} = \frac{3 \times 1 \times 2}{36} = \frac{6}{36}$$

$$(1,1) = \frac{{}^3C_1 \times {}^4C_1 \times {}^2C_0}{{}^9C_2} = \frac{3 \times 4 \times 1}{36} = \frac{12}{36}$$

$$(2,0) = \frac{{}^3C_2 \times {}^4C_0 \times {}^2C_0}{{}^9C_2} = \frac{3}{36}$$

$${}^9C_2 = \frac{9!}{2!\,(9-2)!} = \frac{9 \times \overset{4}{\cancel{8}} \times \cancel{7!}}{\cancel{2} \times 1 \times \cancel{7!}} = 36$$

$${}^4C_2 = \frac{4!}{2!\,2!} = \frac{\overset{2}{\cancel{4}} \times 3 \times \cancel{2!}}{\cancel{2} \times \cancel{2!}} = 6$$

$${}^3C_2 = \frac{3!}{2! \times 1!} = \frac{3 \times \cancel{2!}}{\cancel{2!} \times 1} = 3$$

Range of both X and Y is 0, 1, 2

X	Y = 0	Y = 1	Y = 2	$g_X(x)$
0	$\frac{1}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{15}{36}$
1	$\frac{6}{36}$	$\frac{12}{36}$	0	$\frac{18}{36}$
2	$\frac{3}{36}$	0	0	$\frac{3}{36}$
$h_Y(y)$	$\frac{10}{36}$	$\frac{20}{36}$	$\frac{6}{36}$	1

$$g(0) = \sum_{y=0}^{2} p(x, y)$$

Marginal prob. mass function of X denoted by $g_X(x)$

Marginal prob. mass function of Y denoted by $h_Y(y)$

Mathematically,

If X and Y are 2 discrete random variables, their joint probability mass function is given by

$$\boxed{P(X=x, Y=y) = p(x,y)}$$

Marginal probabilities

1. The marginal probability mass function X is

$$g_X(x) = \sum_y p(x,y)$$

2. The marginal probability mass function of Y is

$$h(y) = \sum_x p(x,y)$$

Conditional Probability distribution

If X and Y have the joint prob. mass function $p(x,y) = P(X=x, Y=y)$; their respective mpmf are $g_X(x) = \sum_y p(x,y)$ and $h_Y(y) = \sum_x p(x,y)$, then the conditional probability function of $X=x$ given $Y=y$ is

$$P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{h_Y(y)}, \quad h_Y(y) > 0$$

Similarly

$$P(Y=y \mid X=x) = \frac{P(X=x, Y=y)}{g_X(x)}, \quad g_X(x) > 0$$

Range of both X and Y is 0, 1, 2

X \ Y	0	1	2	$g_X(x)$
0	$\frac{1}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{15}{36}$
1	$\frac{6}{36}$	$\frac{12}{36}$	0	$\frac{18}{36}$
2	$\frac{3}{36}$	0	0	$\frac{3}{36}$
$h_Y(y)$	$\frac{10}{36}$	$\frac{20}{36}$	$\frac{6}{36}$	1

$$P(X=1 \mid Y=1) = \frac{P(X=1, Y=1)}{P(Y=1)} = \frac{p(1,1)}{h(1)} = \frac{\frac{12}{36}}{\frac{20}{36}} = \frac{12}{20}$$

$$P(Y=0 \mid X=2) = \frac{p(0,2)}{g(2)} = \frac{\frac{3}{36}}{3/36} = 1$$

Joint Probability Distribution Function

❖ Let $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ be two discrete random variables. Then $P(x, y) = J_{ij}$ is called joint probability function of X and Y if it satisfies the conditions:

$$(i)\ J_{ij} \geq 0 \quad (ii) \sum_{i=1}^{m} \sum_{j=1}^{n} J_{ij} = 1$$

❖ Set of values of this joint probability function J_{ij} is called joint probability distribution of X and Y.

	y_1	y_2	$\dots$	y_n	Sum
x_1	J_{11}	J_{12}	$\dots$	J_{1n}	$f(x_1)$
x_2	J_{21}	J_{22}	$\dots$	J_{2n}	$f(x_2)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_m	J_{m1}	J_{m2}	$\dots$	J_{mn}	$f(x_m)$
Sum	$g(y_1)$	$g(y_2)$	$\dots$	$g(y_n)$	$Total = 1$

If X and Y are discrete random variables, the joint probability distribution of X and Y is a description of the set of points (x,y) in the range of (X,Y) along with the probability of each point.

The joint probability distribution of two discrete random variables is sometimes referred to as the **bivariate probability distribution** or **bivariate distribution.**

Thus we can describe the joint probability distribution of two discrete random variables is through a **joint probability mass function**

$$f(x,y)=P(X=x, Y=y)$$

➢We often want to determine the joint probability of two variables, such as X and Y. Suppose we are able to determine the following information for education (X) and age (Y) for all Indian citizens based on the census.

$X = 30, 45, 70$

$Y = 0, 1, 2, 3$

$(X=30, Y=3) = 0.07$

Age (Y): Education (X)	Age : 25-35 30	Age: 35-55 45	Age: 55-85 70
None 0	.01	.02	.05
Primary 1	.03	.06	.10
Secondary 2	.18	.21	.15
College 3	.07	.08	.04

class interval

→ Mid-point

$$\text{Mid-point}(x) = \frac{LL + UL}{2}$$

$$P(X=45 \mid Y=2) = \frac{P(45,2)}{h(2)} = \frac{0.21}{0.18+0.21+0.15} =$$

Each cell is the relative frequency (f/N).

We can define the joint probability distribution as:
$p(x, y) = Pr(X = x \text{ and } Y = y)$

Example: what is the probability of getting a 30 year old college graduate?
$p(x,y) = Pr(X=3 \text{ and } Y=30) = .07$

We can see that: $p(x) = \Sigma_y\, p(x,y)$
$p(x=1) = .03 + .06 + .10 = .19$

Education (X)	Age (Y):	Age : 25-35 30	Age: 35-55 45	Age: 55-85 70
None	0	.01	.02	.05
Primary	1	.03	.06	.10
Secondary	2	.18	.21	.15
College	3	.07	.08	.04

Marginal Probability

- We call this the **marginal probability** because it is calculated by summing across rows or columns and is thus reported in the margins of the table.

We can calculate this for our entire table.

Age (Y): / Education (X)	30	45	70	p(x)
None: 0	.01	.02	.05	**.08**
Primary: 1	.03	.06	.10	**.19**
Secondary: 2	.18	.21	.15	**.54**
College: 3	.07	.08	.04	**.19**
p(y)	**.29**	**.37**	**.34**	**1**

Joint Density Function

When X and Y are continuous random variables, the **joint density function** $f(x, y)$ is a surface lying above the xy plane, and $P[(X, Y) \in A]$, where A is any region in the xy plane, is equal to the volume of the right cylinder bounded by the base A and the surface.

The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\ dx\ dy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y)\ dx\ dy$, for any region A in the xy plane.

Marginal Distributions

The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_{y} f(x, y) \quad \text{and} \quad h(y) = \sum_{x} f(x, y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

for the continuous case.

Independent Random Variables

$f(x,y) = g(x)h(y)$

Let X and Y are two random variables with joint probability function $f(x, y)$ are said to be independent if following condition satisfied:

$$f(x,y) = g(x).h(y) \textit{ for all } x \textit{ and } y,$$

where $g(x)$ is marginal probability function of X and $h(y)$ is marginal probability function of Y.

Example: Suppose $f(x,y) = e^{-(x+y)}, x \geq 0, y \geq 0$. For this probability function, marginal probability function for X is $g(x) = e^{-x}, x \geq 0$ and

marginal probability function for Y is $h(y) = e^{-y}, y \geq 0$.

Clearly $f(X,Y) = g(X).h(Y)$. So X and Y are independent variables.

Consider the joint distribution of X and Y.
Compute the following probabilities:
(i) $P(X=1, Y=2)$ (ii) $P(X \geq 1, Y \geq 2)$
(iii) $P(X \leq 1, Y \leq 2)$ (iv) $P(X+Y \geq 2)$ (v) $P(X \geq 1, Y \leq 2)$.

X \ Y	0	1	2	3
0	0	1/8	1/4	1/8
1	1/8	1/4	1/8	0

Solution:

(i) $X = \{0,1\}, Y = \{0,1,2,3,4\}$

$$P(X=1, Y=2) = P(1,2) = \frac{1}{8}$$

(ii) If $X \geq 1, X = \{1\}$. If $Y \geq 2, Y = \{2,3\}$

$$P(X \geq 1, Y \geq 2) = P(1,2) + P(1,3) = \frac{1}{8} + 0 = \frac{1}{8}$$

(iii) If $X \leq 1, X = \{0,1\}$. If $Y \leq 2, Y = \{0,1,2\}$

$$P(X \leq 1, Y \leq 2) = P(0,0) + P(0,1) + P(0,2) + P(1,0) + P(1,1) + P(1,2)$$

$$= 0 + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$$

Cont.

(iv) If $X + Y \geq 2$ then

$$X + Y = 0 + 2 \; or \; 0 + 3 \; or \; 1 or \; 1 + 2 \; or \; 1 + 3$$

$$P(X + Y \geq 2) = P(0,2) + P(0,3) + P(1,1) + P(1,2) + P(1,3)$$

$$= \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + 0 = \frac{3}{4}$$

(v) If $X \geq 1, X = \{1\}$. If $Y \leq 2, Y = \{0, 1, 2\}$

$$P(X \geq 1, Y \leq 2) = P(1,0) + P(1,1) + P(1,2)$$

$$= \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = \frac{1}{2}$$

Example:

Two ballpoint pens are selected at random form a box that contains blue pens, 2 red pens and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find the joint probability function $f(x, y)$.

- **Solution:**

The possible pairs of values $(X, Y)$are (0,0),(0,1),(1,0),(1,1),(0,2),(2,0)
The joint probability distribution can be represented by the formula

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}},$$

for $x = 0, 1, 2$; $y = 0, 1, 2$; and $0 \le x + y \le 2$.

Joint distribution

$f(x,y)$		X			Rows Total
		0	**1**	**2**	
y	**0**	3/28	9/28	3/28	15/28
	1	3/14	3/14	0	3/7
	2	1/28	0	0	1/28
Columns Total		5/14	15/28	3/28	1

Example

- Find the joint distribution of X and Y which are the independent random variables with the following respective distributions.

x_i	1	2
$f(x_i)$	0.7	0.3

y_j	-2	5	8
$g(y_j)$	0.3	0.5	0.2

Solution:

Since X and Y are independent random variables,

$$J_{ij} = f(x_i)g(y_j)$$

$x\backslash y$	-2	5	8	$f(x)$
1	0.21	0.35	0.14	0.7
2	0.09	0.15	0.06	0.3
$g(y)$	0.3	0.5	0.2	$Total = 1$

Example:

A candy company distributed boxes of chocolates with a mixture of creams, toffees, and nuts coated in both light and dark chocolate. For a randomly selected box, let X and Y, respectively, be the proportions of the light and dark chocolates that are creams and suppose that the joint density function is

$$f(x,y)=\begin{cases}\frac{2}{5}(2x+3y), & 0\le x\le 1, 0\le y\le 1\\ 0, & \text{elsewhere}\end{cases}$$

a) Verify whether

b) Find P[(X,Y) $\in$ A], where A is the region {(x,y) | 0<x<½,¼<y<½}

c) Find g(x) and h(y) for the joint density function.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

Solution

a) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{0}^{1}\int_{0}^{1}\frac{2}{5}(2x+3y)dxdy$$

$$= \int_{0}^{1} \frac{2x^2}{5} + \frac{6xy}{5}\Bigg|_{x=0}^{x=1} dy$$

$$= \int_{0}^{1}\left(\frac{2}{5}+\frac{6y}{5}\right)dy = \frac{2y}{5}+\frac{3y^2}{5}\Bigg|_{0}^{1}$$

$$= \frac{2}{5}+\frac{3}{5} = 1$$

Solution..

b)

$$P[(X,Y)\in A] = P(0 < X < \tfrac{1}{2}, \tfrac{1}{4} < Y < \tfrac{1}{2})$$

$$= \int_{\frac{1}{4}}^{\frac{1}{2}}\int_{0}^{\frac{1}{2}} \frac{2}{5}(2x+3y)dxdy$$

$$= \int_{\frac{1}{4}}^{\frac{1}{2}} \left.\frac{2x^2}{5}+\frac{6xy}{5}\right|_{x=0}^{x=\frac{1}{2}} dy$$

$$= \int_{\frac{1}{4}}^{\frac{1}{2}} \left(\frac{1}{10}+\frac{3y}{5}\right)dy = \left.\frac{y}{10}+\frac{3y^2}{10}\right|_{\frac{1}{4}}^{\frac{1}{2}}$$

$$= \frac{1}{10}\left[\left(\frac{1}{2}+\frac{3}{4}\right)-\left(\frac{1}{4}+\frac{3}{16}\right)\right] = \frac{13}{160}$$

Solution

By definition,

$$g(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_{0}^{1}\frac{2}{5}(2x+3y)dy=\frac{4xy}{5}+\frac{6y^2}{10}\bigg|_{y=0}^{y=1}=\frac{4x+3}{5}$$

For 0≤x ≤1, and g(x)=0 elsewhere.

Similarly, $$h(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_{0}^{1}\frac{2}{5}(2x+3y)dx=\frac{4(1+3y)}{5}$$

For 0 ≤y ≤1, and h(y)=0 elsewhere.

Example

The joint probability distribution of X and Y is given by $f(x,y) = c(x^2 + y^2)$ for $x = -1, 0, 1, 3.\ and\ y = -1, 2, 3.$ (i) Find the value of c. (ii) $P(x = 0, y \leq 2)$ (iii) $P(x \leq 1, y > 2)$ (iv) $P(x \geq 2 - y)$

Solution:

By data, $X = \{-1, 0, 1, 3\}$ and $Y = \{-1, 2, 3\}$

$$f(x,y) = c(x^2 + y^2)$$

The joint probability distribution of X and Y:

X \ Y	-1	2	3	$f(X)$
-1	2c	$5c$	$10c$	$17c$
0	c	$4c$	$9c$	$14c$
1	$2c$	$5c$	$10c$	$17c$
3	$10c$	$13c$	$18c$	$41c$
g(Y)	$15c$	$27c$	$47c$	89c

(i) **Find c:** $1 = \Sigma f(x,y) = 89c$

$$c = \frac{1}{89}$$

(ii) $x = 0, y = \{-1, 2\}$

$$P(x = 0, y \leq 2)$$
$$= P(0,-1) + P(0,2)$$
$$= c + 4c = 5c$$
$$= 5/89$$

(iii) $x = \{-1, 0, 1\},\ y = \{3\}$

$$P(x \leq 1, y > 2)$$
$$= P(-1,3) + P(0,3) + P(1,3)$$
$$= 10c + 9c + 10c$$
$$= 29c = 29/89$$

Cont.

By data, $X = \{-1, 0, 1, 3\}$ and $Y = \{-1, 2, 3\}$

(iv) $P(x \geq 2 - y) = P(x + y \geq 2)$

$$= P(-1,3) + P(0,2) + P(0,3) + P(1,2) +$$

$$P(1,3) + P(3,-1) + P(3,2) + P(3,3)$$

$$= 10c + 4c + 9c + 5c + 10c + 10c + 13c + 18c$$

$$= 79c = 79/89$$

Glossary

1. **Random Variable**: A variable that takes on numerical values based on the outcomes of a random experiment.
   - **Discrete Random Variable**: A random variable that can take on a countable number of values.
   - **Continuous Random Variable**: A random variable that can take on an uncountable range of values, often over an interval.
2. **Probability Distribution**: A function that describes the likelihood of different outcomes for a random variable.
   - **Probability Mass Function (PMF)**: A function that gives the probability of each value of a discrete random variable.
   - **Probability Density Function (PDF)**: A function that represents the likelihood of different values of a continuous random variable.
3. **Expected Value (Mean)**: The average value a random variable is expected to take, calculated as a weighted sum or integral of its possible values.
4. **Variance**: A measure of how spread out the values of a random variable are around the mean.

Glossary

5. **Joint Probability Distribution**:A probability distribution that represents the likelihood of two or more random variables occurring simultaneously.

- **Joint Probability Mass Function (PMF)**: For discrete random variables, a function P(X=x,Y=y) that gives the probability of X and Y taking on specific values simultaneously.
- **Joint Probability Density Function (PDF)**: For continuous random variables, a function $f(X,Y)$that represents the likelihood of X and Y taking values in an infinitesimally small region around (x,y).

6. **Marginal Distribution**: The probability distribution of a subset of variables within the joint distribution, found by summing or integrating over the other variable(s).

- Discrete: $f(x) = \sum_y P(x,y)$, where P(x,y) joint probability mass function.
- Continuous: $f(x) = \int P(x,y)dy$ where P(x,y) joint probability density function.

Problem:1

The joint probability mass function of $(X, Y)$is given by $P(x, y) = K(2x + 3y)$, $x = 0,1,2, y = 1,2,3$. Find i) the value of K, ii) The marginal probability function of X and Y, iii) $P(X = 2, Y \leq 2)$, iv) $P(X = 2)$.

Ans : i) $\frac{1}{72}$ ii) $P(X = 0) = \frac{18}{72}$, $P(X = 1) = \frac{24}{72}$, $P(X = 2) = \frac{30}{72}$, $P(Y = 1) = \frac{15}{72}$, $P(Y = 2) = \frac{24}{72}$, $P(Y = 3) = \frac{33}{72}$ ii) $\frac{17}{72}$ iv) $\frac{30}{72}$.

Problem: 2

Let X and Y have the following joint probability distribution

X / Y	2	4
1	0.10	0.15
3	0.20	0.30
5	0.10	0.15

Show that X and Y are independent variables.

Problem: 3

For the bivariate probability distribution of (X, Y) given below.

Find i) $P(X \leq 1)$, ii) $P(Y \leq 3)$, iii) $P(X \leq 1, Y \leq 3)$

Y / X	1	2	3	4	5	6
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$

Ans : i) $\frac{7}{8}$, ii) $\frac{23}{64}$ iii) $\frac{9}{32}$

Problem: 4 If $f(x, y) = e^{-(x+y)}, x \geq 0 , y \geq 0$ is the joint probability density function of X and Y , find $P(X < 1)$. (Ans: 0.6321)

Problem: 5

The joint probability density function of X and Y given by

$$f(x,y) = \begin{cases} Cxy & 0 < x < 2, 0 < y < 2 \\ 0 & otherwise \end{cases}.$$

$i)$ Find constant C $ii)$ $P(\frac{1}{2} < X < \frac{3}{2}, 1 < Y < 2)$.

$Ans : i) \frac{1}{4}$ ii) $\frac{3}{8}$

Problem: 6

The joint probability density function of X and Y given by

$$f(x,y) = \begin{cases} x + y & 0 < x < 1, 0 < y < 1 \\ 0 & otherwise \end{cases}.$$

Find marginal probability function of X and Y. Also, check whether X and Y are independent or not.

Ans: Marginal prob function for X: $x + \frac{1}{2}$,

Marginal probability function of Y : $\frac{1}{2} + Y$; Not independent .

IMP Note to Self

Thank You !