MAT2001-Module-2: Randaom Variables

Dr. Nalliah M

Assistant Professor

Department of Mathematics

School of Advanced Sciences

Vellore Institute of Technology

Vellore,Tamil Nadu,India.

nalliah.moviri@vit.ac.in

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

August 6, 2020

Intoduction

Statistics is concerned with making inferences about populations and population characteristics. Experiments are conducted with results that are subject to chance. The testing of a number of electronic components is an example of a statistical experiment, a term that is used to describe any process by which several chance observations are generated. It is often important to allocate a numerical description to the outcome.

For example, the sample space giving a detailed description of each possible outcome when three electronic components are tested may be written $S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$, where N denotes nondefective and D denotes defective.

One is naturally concerned with the number of defectives that occur. Thus, each point in the sample space will be assigned a numerical value of $0, 1, 2$, or 3. These values are, of course, random quantities determined by the outcome of the experiment. They may be viewed as values assumed by the random variable X, the number of defective items when three electronic components are tested.

A random variable is a variable that associates a real number with each element in the sample space.

Example

Two balls are drawn in succession without replacement from an urn containing 4 red balls(R) and 3 black balls(B). The possible outcomes $S = \{RR, RB, BR, BB\}$. Then we define a function X from S to real number $\mathbb{R}$, where X represent the number of red balls for every sample of S

The variable X, representing the number of red balls in a sample of S.

Define $X : S \longrightarrow \mathbb{R}$ by

$$X(s_1) = 2$$

$$X(s_2) = 1$$

$$X(s_3) = 1$$

$$X(s_4) = 0$$

where $s_i \in S, i = 1, 2, 3, 4$.

Thus the variable X is called a random variable with the values x is $0, 1, 2$.

If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a discrete sample space.

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a continuous sample space.

A random variable is called a discrete random variable if its set of possible outcomes is countable.

When a random variable can take on values on a continuous scale, it is called a continuous random variable.

Discrete Probability Distributions

A discrete random variable assumes each of its values with a certain probability.

In case of tossing a coin two times, the possible outcomes are $S = \{HH, HT, TH, TT\}$. The variable X, representing the number of heads in $s \in S$. Now, we define by $X : S \longrightarrow \mathbb{R}$ as

$$X(s_1) = 2$$

$$X(s_2) = 1$$

$$X(s_3) = 1$$

$$X(s_4) = 0.$$

Thus the random variable X and the values of x with the probability values of x is given below table.

X	0	1	2
P(x)=P(X=x)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

The set of ordered pairs $(x, P(x))$ is called the probability mass function, probability function, or probability distribution of the discrete random variable X.

Probability mass function

The set of ordered pairs $(x, P(x))$ is a probability mass function, probability function, or probability distribution of the discrete random variable X if, for each possible outcome x,

- $P(x) \geq 0$
- $\sum_{x} P(x) = 1.$

Example

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution

Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Now,

- $P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}$
- $P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$
- $P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$

Thus, the probability distribution of X is given below table.

X	0	1	2
P(x)=P(X=x)	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

Continuous Probability Distributions

A continuous random variable has a probability of 0 of assuming exactly any of its values. Consequently, its probability distribution cannot be given in tabular form.

Let us discuss a random variable whose values are the heights of all people over 21 years of age.

Between any two values, say 163.5 and 164.5 centimeters, or even 163.99 and 164.01 centimeters, there are an infinite number of heights, one of which is 164 centimeters.

The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters that you cannot humanly measure the difference is remote, and thus we assign a probability of 0 to the event.
This is not the case, however, if we talk about the probability of selecting a person who is at least 163 centimeters but not more than 165 centimeters tall.
Now we are dealing with an interval rather than a point value of our random variable.

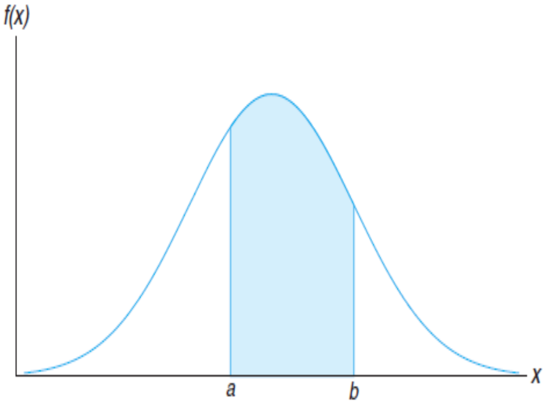
We shall concern ourselves with computing probabilities for various intervals of continuous random variables such as $P(a < X < b), P(W \geq c)$, and so forth. Note that when X is continuous, $P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b)$. That is, it does not matter whether we include an endpoint of the interval or not. This is not true, though, when X is discrete.

A probability density function is constructed so that the area under its curve bounded by the x-axis is equal to 1 when computed over the range of X for which $f(x)$ is defined.

Should this range of X be a finite interval, it is always possible to extend the interval to include the entire set of real numbers by defining $f(x)$ to be zero at all points in the extended portions of the interval.

In the following Figure, the probability that X assumes a value between a and b is equal to the shaded area under the density function between the ordinates at $x = a$ and $x = b$, and from integral calculus is given by

$$P(a < X < b) = \int_a^b f(x)dx.$$


f(x)
a
b
X

Probability density function

The function $f(x)$ is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers, if

- $f(x) \geq 0$, for all $x \in \mathbb{R}$,
- $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1.$

Example

Suppose that the error in the reaction temperature, in C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0. & \textit{elsewhere.} \end{cases}$$

Example Cont...

1. Verify that $f(x)$ is a density function.
2. Find $P(0 < X \leq 1)$.

Solution

For(1),

Obviously, $f(x) \geq 0$. *To verify condition*

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

Example Cont...

Now,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{-1}^{-\infty} f(x)dx + \int_{-1}^{2} f(x)dx + \int_{2}^{\infty} f(x)dx \\
&= 0 + \int_{-1}^{2} \frac{x^2}{3} dx + 0 \\
&= \left[\frac{x^3}{3 \times 3}\right]_{-1}^{2} \\
&= \frac{8}{9} - \frac{(-1)}{9} \\
&= \frac{8}{9} + \frac{(1)}{9} \\
&= 1.
\end{aligned}$$

Example Cont...

For(2),

$$\begin{aligned} P(0 < X \leq 1) &= \int_0^1 f(x)dx \\ &= \int_0^1 \frac{x^2}{3}dx \\ &= \left[\frac{x^3}{3 \times 3}\right]_0^1 \\ &= \frac{1}{9} - 0 \\ &= \frac{1}{9}. \end{aligned}$$

Cumulative distribution function

The cumulative distribution function $F(x)$ of a random variable X with probability distribution $P(x)$ and probability desity function $f(x)$ is

$$F(x) = P(X \leq x) = \begin{cases} \sum\limits_{t \leq x} P(X = t), & \text{If } X \text{ is discrete} \\ \int_{-\infty}^{x} f(x)dx, & \text{If } X \text{ is continuous} \end{cases}$$

Let $F(x)$ be a cumulative distribution function of a continuous random variable X. Then $P(a < X < b) = F(b) - F(a)$ and $f(x) = \frac{d(F(x))}{dx}$, if the derivative exists.

Problem

A random variable X has the following probability mass function is given

X	-2	-1	0	1	2	3
P(X)	0.1	k	0.2	$2k$	0.3	k

1. Find k.
2. Evaluate $P(X \leq 2)$ and $P(-1 \leq X \leq 2)$.
3. Find the cumulative distribution function.

Solution

For(1), since $P(x)$ is p.m.f, and $\sum_x P(x) = 1.$, it follows that we get

$$\sum_x P(x) = 1$$

$$0.1 + k + 0.2 + 2k + 0.3 + k = 1$$

$$0.6 + 4k = 1$$

$$4k = 0.4$$

$$k = \frac{0.4}{4}$$

$$k = 0.1$$

Solution Cont...

X	-2	-1	0	1	2	3
P(X)	0.1	0.1	0.2	0.2	0.3	0.1

For(2),

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

$$= 0.1 + 0.1 + 0.2 + 0.2 + 0.3$$

$$= 0.9$$

OR

Solution Cont...

$$\begin{aligned} P(X \leq 2) &= 1 - P(X > 2) \\ &= 1 - P(X = 3) \\ &= 1 - 0.1 \\ &= 0.9 \end{aligned}$$

Now,

$$\begin{aligned} P(-1 \leq X \leq 2) &= P(X = -1) + P(x = 0) + P(x = 1) + P(x = 2) \\ &= 0.1 + 0.2 + 0.2 + 0.3 \\ &= 0.8 \end{aligned}$$

Solution Cont...

$$F(x) = P(X \leq x) = \sum_{t \leq x} P(X = t).$$

$$F(x) = \begin{cases} P(X \leq -2) = 0.1, & \text{If } x = -2 \\ P(X \leq -1) = 0.2, & \text{If } x = -1 \\ P(X \leq 0) = 0.4, & \text{If } x = 0 \\ P(X \leq 1) = 0.6, & \text{If } x = 1 \\ P(X \leq 2) = 0.9, & \text{If } x = 2 \\ P(X \leq 3) = 1, & \text{If } x = 3. \end{cases}$$

Problem

A random variable X has the following probability mass function is given

X	0	1	2	3	4	5	6	7
P(X)	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2$+k

1. Find k.
2. Evaluate $P(X \leq 6)$, $P(X \geq 6)$ and $P((1.5 < X < 4.5)/(X > 2))$
3. Find the cumulative distribution function.

Problem

Let X be a continuous random variable with probability density function (pdf) is

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0. & \text{elsewhere.} \end{cases}$$

1. Find $P(X \leq 0.4)$, $P(X \geq \frac{3}{4})$ and $P(X \geq \frac{1}{2}.)$
2. Evaluate $P(\frac{1}{2} < X \leq \frac{3}{4})$ and $P(X > \frac{3}{4}/X > \frac{1}{2})$.
3. Find the cumulative distribution function.

Solution

$$\text{For}(2), P\left(X > \frac{3}{4}/X > \frac{1}{2}\right) = \frac{P((X > \frac{3}{4}) \cap (X > \frac{1}{2}))}{P(X > \frac{1}{2})}$$

$$= \frac{P(X > \frac{3}{4})}{P(X > \frac{1}{2})}$$

$$= \frac{\displaystyle\int_{\frac{3}{4}}^{1} 2x dx}{\displaystyle\int_{\frac{1}{2}}^{1} 2x dx}$$

$$= \frac{\left[x^2\right]_{\frac{3}{4}}^{1}}{[x^2]_{\frac{1}{2}}^{1}} = \frac{\frac{7}{16}}{\frac{3}{4}} = \frac{7}{12}.$$

Problem

The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b \\ 0. & \text{elsewhere.} \end{cases}$$

Find $F(y)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

Solution

To find $F(y)$,$F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(y)dy$.

For $\frac{2}{5}b \leq y \leq 2b$, then $P(Y \leq y) = \int_{\frac{2}{5}b}^{2b} \frac{5}{8b}dy = \left[\frac{5y}{8b}\right]_{\frac{2}{5}b}^{2b} = \frac{5y}{8b} - \frac{1}{4}$.

Thus,

$$F(y) = \begin{cases} 0, & y \leq \frac{2}{5}b \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5}b \leq y \leq 2b \\ 1. & y \geq 2b. \end{cases}$$

Two dimensional random Variables

Our study of random variables and their probability distributions in the preceding sections was restricted to one-dimensional sample spaces, in that we recorded outcomes of an experiment as values assumed by a single random variable.

There will be situations, however, where we may find it desirable to record the simultaneous outcomes of several random variables.

For example, we might measure the amount of precipitate P and volume V of gas released from a controlled chemical experiment, giving rise to a two-dimensional sample space consisting of the outcomes (p, v), or we might be interested in the hardness H and tensile strength T of cold-drawn copper, resulting in the outcomes (h, t).

Two dimensional random Variables

In a study to determine the likelihood of success in college based on high school data, we might use a threedimensional sample space and record for each individual his or her aptitude test score, high school class rank, and grade-point average at the end of freshman year in college.

For example, if an 18-wheeler is to have its tires serviced and X represents the number of miles these tires have been driven and Y represents the number of tires that need to be replaced, then $p(30000, 5)$ is the probability that the tires are used over 30,000 miles and the truck needs 5 new tires.

Joint probability distribution

Definition

The function $p(x, y)$ *is a joint probability distribution or probability mass function of the discrete random variables* X *and* Y *if*

1. $p(x, y) \geq 0$, *for all* (x, y),
2. $\sum_x \sum_y p(x, y) = 1$,
   where $p(x, y) = P(X = x, Y = y)$.

Problem

Two ballpoint pens are selected at random from a box that contains 3 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

1. the joint probability function $p(x, y)$.
2. $P(X + Y \leq 1)$.

Solution

For(1),

The possible pairs of values (x, y) are $(0, 0), (0, 1), (1, 0), (1, 1), (0, 2)$, and $(2, 0)$.

Now, $p(0, 1)$, for example, represents the probability that a red and a green pen are selected. The total number of equally likely ways of selecting any 2 pens from the 8 is $\binom{8}{2} = 28$

The number of ways of selecting 1 red from 2 red pens and 1 green from 3 green pens is $\binom{2}{1}\binom{3}{1} = 6$.

Hence, $p(0, 1) = \frac{6}{28} = \frac{3}{14}$.

Similar calculations yield the probabilities for the other cases, which are presented in following table.

Solution Cont...

	$p(x,y)$	x			Row Totals
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

$$\begin{aligned} For(2),\ P(X+Y \le 1) &= p(0,0)+p(0,1)+p(1,0) \\ &= \frac{3}{28}+\frac{3}{14}+\frac{9}{28} \\ &= \frac{9}{14}. \end{aligned}$$

Joint probability density function

Definition

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

1. $f(x, y) \geq 0$, *for all* (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1.$

Problem

A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

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

1. Verify $f(x,y)$ is a probability density function
2. Find $P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$

Solution

For(1), $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= \int_{0}^{1}\int_{0}^{1} \frac{2}{5}(2x+3y)dxdy \\
&= \int_{0}^{1}\left[\frac{2x^2}{5}+\frac{6xy}{5}\right]_{x=0}^{x=1} dy \\
&= \int_{0}^{1}\left(\frac{2}{5}+\frac{6y}{5}\right) dy \\
&= \left[\frac{2y}{5}+\frac{3y^2}{5}\right]_{y=0}^{y=1} \\
&= \frac{2}{5}+\frac{3}{5} \\
&= 1.
\end{aligned}$$

Solution Cont...

For(2), $P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$

$$
\begin{aligned}
P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right) &= \int_{\frac{1}{4}}^{\frac{1}{2}} \int_{0}^{\frac{1}{2}} \frac{2}{5}(2x+3y)\,dx\,dy \\
&= \int_{\frac{1}{4}}^{\frac{1}{2}} \left[\frac{2x^2}{5} + \frac{6xy}{5}\right]_{x=0}^{x=\frac{1}{2}} dy \\
&= \int_{\frac{1}{4}}^{\frac{1}{2}} \left(\frac{1}{10} + \frac{3y}{5}\right) dy \\
&= \left[\frac{y}{10} + \frac{3y^2}{10}\right]_{y=\frac{1}{4}}^{y=\frac{1}{2}} \\
&= \frac{1}{10}\left[\left(\frac{1}{2} + \frac{3}{4}\right) - \left(\frac{1}{4} + \frac{3}{16}\right)\right] = \frac{13}{160}.
\end{aligned}
$$

The marginal distributions

The marginal distributions of X alone and of Y alone are

$P_X(x) = \sum_y p(x, y)$ and $P_Y(y) = \sum_x p(x, y)$, if (X, Y) discrete case, and

$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$,if (X, Y) continuous case.

The conditional distributions for discrete case

Let X and Y be two discrete random variables. The conditional distribution of the random variable Y given that $X = x$ is

$$P(Y/X) = \frac{p(x,y)}{P_X(x)}, \text{ provided } P_X(x) > 0$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$P(X/Y) = \frac{p(x,y)}{P_Y(y)}, \text{ provided } P_Y(y) > 0.$$

The conditional distributions for continuous case

Let X and Y be two continuous random variables. The conditional distribution of the random variable Y given that $X = x$ is
$f(y/x) = \dfrac{f(x,y)}{f_X(x)}$, provided $f_X(x) > 0$

Similarly, the conditional distribution of X given that $Y = y$ is
$f(x/y) = \dfrac{f(x,y)}{f_Y(y)}$, provided $f_Y(y) > 0$.

Remark

If we wish to find the probability that the discrete random variable X falls between a and b when it is known that the discrete variable $Y = y$, we evaluate

$$P(a < X < b|Y = y) = \sum_{a<x<b} P(X|Y),$$

where the summation extends over all values of X between a and b.

When X and Y are continuous, we evaluate

$$P(a < X < b|Y = y) = \int_a^b f(x|y)dx. \text{ Similarlly,}$$

$$P(a < Y < b|X = x) = \sum_{a<y<b} P(Y|X), \text{ for } (X, Y) \text{ is discrete case}$$

$$P(a < Y < b|X = x) = \int_a^b f(y|x)dy, \text{ for } (X, Y) \text{ is continuous case}$$

Problem

The joint probability mass function of two dimensional random variables (X, Y) is

	$p(x, y)$	x		
		0	1	2
	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
y	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

1. Find the marginal distributions functions of X and Y
2. Find the conditional distribution function of X given $Y = 1$.

Solution

	$p(x,y)$	x			$P_Y(y)$
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
$P_X(x)$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

The marginal distributions functions of X is $P_X(x) = \begin{cases} \frac{5}{14}, & x = 0 \\ \frac{15}{28}, & x = 1 \\ \frac{3}{28}, & x = 2. \end{cases}$

The marginal distributions functions of Y is $P_Y(y) = \begin{cases} \frac{15}{28}, & y = 0 \\ \frac{3}{7}, & y = 1 \\ \frac{1}{28}, & y = 2. \end{cases}$

For(2),To find the conditional distribution of X, given that $Y = 1$.

We need to find $P(X/Y) = \dfrac{p(x,y)}{P_Y(y)}$, provided $P_Y(y) > 0.$, where $y = 1$.

That is, to find $P(X/1) = \dfrac{p(x,1)}{P_Y(1)}$

First we have find $P_Y(1) = \sum_{x=0}^{x=2} p(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$.

Solution Cont...

Now, $P(X/1) = \dfrac{p(x,1)}{P_Y(1)} = \frac{3}{7}p(x,1), x = 0,1,2$

Therefore,

$$P(0/1) = \frac{p(0,1)}{P_Y(1)} = \frac{7}{3}p(0,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}$$

$$P(1/1) = \frac{p(1,1)}{P_Y(1)} = \frac{7}{3}p(1,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}$$

$$P(2/1) = \frac{p(2,1)}{P_Y(1)} = \frac{7}{3}p(2,1) = \left(\frac{7}{3}\right)(0) = 0.$$

Solution Cont...

The conditional distribution of X, given that $Y = 1$, is

$$P(X/1) = \begin{cases} \frac{1}{2}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ 0, & x = 2. \end{cases}$$

Problem

The joint probability mass function of two dimensional random variables (X, Y) is

	$p(x,y)$	y			
		1	2	3	4
x	1	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{5}{36}$	$\frac{1}{36}$
	2	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{2}{36}$
	3	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
	4	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{5}{36}$

1. Find the marginal distributions functions of X and Y
2. Find the conditional distribution function of X given $Y = y$
3. Find the conditional distribution function of Y given $X = x$.

Solution

For(1),

	$p(x,y)$	y = 1	2	3	4	$P_X(x)$
x	1	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{12}{36}$
	2	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{7}{36}$
	3	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{8}{36}$
	4	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{5}{36}$	$\frac{9}{36}$
$P_Y(y)$		$\frac{10}{36}$	$\frac{9}{36}$	$\frac{9}{36}$	$\frac{8}{36}$	1

Solution Cont...

The marginal distribution function of X is

$$P_X(x) = \begin{cases} \frac{12}{36}, & x = 1 \\ \frac{7}{36}, & x = 2 \\ \frac{8}{36}, & x = 3 \\ \frac{9}{36}, & x = 4. \end{cases}$$

The marginal distribution function of Y is

$$P_Y(y) = \begin{cases} \frac{10}{36}, & y = 1 \\ \frac{9}{36}, & y = 2 \\ \frac{9}{36}, & y = 3 \\ \frac{8}{36}, & y = 4. \end{cases}$$

Problem

The joint probability mass function of two dimensional random variables (X, Y) is

	$p(x, y)$	x		
		-1	0	1
	0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
y	1	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
	2	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{2}{15}$

1. Find the marginal distributions functions of X and Y
2. Find the conditional distribution function of X given $Y = y$
3. Find the conditional distribution function of Y given $X = x$.

Problem

The joint probability mass function of two dimensional random variables (X, Y) is $p(x, y) = k(2x + 3y), x = 0, 1, 2 \quad \text{and} \quad y = 1, 2, 3.$

1. Find k.
2. Find the marginal distributions functions of X and Y
3. Find the conditional distribution function of X given $Y = y$
4. Find the conditional distribution function of Y given $X = x$.
5. Find $P(X + Y > 3)$.

Problem

The joint probability mass function of two dimensional random variables (X, Y) is $p(x, y) = \frac{x+y}{21}, x = 1, 2, 3 \quad \text{and} \quad y = 1, 2$

1. Find the marginal distributions functions of X and Y
2. Find the conditional distribution function of X given $Y = y$
3. Find the conditional distribution function of Y given $X = x$.
4. Find $P(X + Y > 3)$. and $P(X < 3)$.

Problem

The joint density for the random variables (X, Y), where X is the temperature change and Y is the proportion of the spectrum that shifts for a certain atomic particle, is

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0, & \text{elesewhere.} \end{cases}$$

1. Find $f_x(x), f_Y(y)$ and $f(y|x)$.
2. Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.

Solution

For(1),The marginal desity function of X is

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \int_x^1 10xy^2 dy \\ &= \left[\frac{10xy^3}{3}\right]_{y=x}^{y=1} \\ &= \frac{10x}{3} - \frac{10xx^3}{3} \\ &= \frac{10x}{3} - \frac{10x^4}{3} \\ &= \frac{10x(1-x^3)}{3}, 0 < x < 1. \end{aligned}$$

Solution Cont...

The marginal desity function of Y is

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx \\
&= \int_0^y 10xy^2 dx \\
&= \left[\frac{10x^2y^2}{2}\right]_{x=0}^{x=y} \\
&= \frac{10y^4}{2} - 0 \\
&= 5y^4, 0 < y < 1.
\end{aligned}
$$

Solution Cont...

The conditional distribution functions of Y and X is

$$\begin{aligned}
f(y|x) &= \frac{f(x,y)}{f_X(x)} \\
&= \frac{10xy^2}{\frac{10x(1-x^3)}{3}} \\
&= \frac{3y^2}{1-x^3}, 0 < x < y < 1. \\
f(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{10xy^2}{5y^4} = \frac{2x}{y^2}, 0 < x < y < 1.
\end{aligned}$$

Solution Cont...

For(2),

$$
\begin{aligned}
P\left(Y > \frac{1}{2}|X = 0.25\right) &= \int_{\frac{1}{2}}^{1} f(y|x = 0.25)dy \\
&= \int_{\frac{1}{2}}^{1} \frac{3y^2}{1-(0.25)^3}dy \\
&= \left[\frac{3y^3}{3(1-0.016)}\right]_{\frac{1}{2}}^{1} \\
&= \left[\frac{y^3}{0.98}\right]_{\frac{1}{2}}^{1} \\
&= \frac{1}{0.98}\left[1-\frac{1}{8}\right] = \frac{7}{7.84} = 0.89
\end{aligned}
$$

Problem

Given the joint density function

$$f(x,y)=\begin{cases}\frac{x(1+3y^2)}{4}, & 0\le x\le 2, 0\le y\le 1\\ 0, & \text{elesewhere.}\end{cases}$$

1. Find $f_x(x), f_Y(y)$, $f(x|y)$ and $f(y|x)$.
2. Evaluate $P\left(\frac{1}{4}<X<\frac{1}{2}|Y=\frac{1}{3}\right)$.

Problem

The joint probability density function of two dimensional continuous random variables (X, Y) is

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

1. Find $f_x(x)$ and $f_Y(y)$.
2. Find $f(x|y)$ and $f(y|x)$.

Problem

Given the joint density function

$$f(x,y) = \begin{cases} xy^2 + \frac{x^2}{8}, & 0 \leq x \leq 2, 0 \leq y \leq 1 \\ 0, & \text{elesewhere.} \end{cases}$$

1. Compute $P(X > 1), P(Y < \frac{1}{2})$, $P(X > 1 | Y < \frac{1}{2})$ and $P(Y < \frac{1}{2} | X > 1)$
2. Evaluate $P(X < Y)$ and $P(X + Y \leq 1)$.

Solution

$$
\begin{aligned}
P(X > 1) &= \int_0^1 \int_1^2 f(x, y) dxdy \\
&= \int_0^1 \int_1^2 \left(xy^2 + \frac{x^2}{8} \right) dxdy \\
&= \int_0^1 \left[\frac{x^2 y^2}{2} + \frac{x^3}{24} \right]_1^2 dy \\
&= \int_0^1 \left[\left(2y^2 + \frac{1}{3} \right) - \left(\frac{y^2}{2} + \frac{1}{24} \right) \right] dy \\
&= \frac{19}{24}
\end{aligned}
$$

Solution Cont...

$$P(Y < \frac{1}{2}) = \int_0^{\frac{1}{2}} \int_0^2 f(x,y)dxdy$$

$$= \int_0^{\frac{1}{2}} \int_0^2 \left(xy^2 + \frac{x^2}{8} \right) dxdy$$

$$= \int_0^{\frac{1}{2}} \left[\frac{x^2y^2}{2} + \frac{x^3}{24} \right]_0^2 dy$$

$$= \int_0^{\frac{1}{2}} \left[2y^2 + \frac{1}{3} \right] dy$$

$$= \left[\frac{2y^3}{3} + \frac{y}{3} \right]_0^{\frac{1}{2}} = \frac{1}{4}$$

Solution Cont...

We know that $P(X > 1 | Y < \frac{1}{2}) = \frac{P(X>1, Y<\frac{1}{2})}{P(Y<\frac{1}{2})}$.

$$
\begin{aligned}
P(X > 1, Y < \frac{1}{2}) &= \int_0^{\frac{1}{2}} \int_1^2 f(x,y)\,dxdy \\
&= \int_0^{\frac{1}{2}} \int_1^2 \left(xy^2 + \frac{x^2}{8} \right) dxdy \\
&= \int_0^{\frac{1}{2}} \left[\frac{x^2y^2}{2} + \frac{x^3}{24} \right]_1^2 dy \\
&= \int_0^{\frac{1}{2}} \left[\left(2y^2 + \frac{1}{3} \right) - \left(\frac{y^2}{2} + \frac{1}{24} \right) \right] dy \\
&= \frac{5}{24}
\end{aligned}
$$

Solution Cont...

$$\text{Therefore,} \quad P(X > 1 | Y < \frac{1}{2}) = \frac{P(X > 1, Y < \frac{1}{2})}{P(Y < \frac{1}{2})} = \frac{\frac{5}{24}}{\frac{1}{4}} = \frac{5}{6},$$

$$\text{and} \quad P(Y < \frac{1}{2} | X > 1) = \frac{P(X > 1, Y < \frac{1}{2})}{P(X > 1)} = \frac{\frac{5}{24}}{\frac{19}{24}} = \frac{5}{19}.$$

Solution Cont...

For(2),

$$
\begin{aligned}
P(X < Y) &= \int_0^1 \int_0^y f(x,y)dxdy \\
&= \int_0^1 \int_0^y \left(xy^2 + \frac{x^2}{8}\right) dxdy \\
&= \int_0^1 \left[\frac{x^2y^2}{2} + \frac{x^3}{24}\right]_0^y dy \\
&= \int_0^1 \left[\frac{y^4}{2} + \frac{y^3}{24}\right] dy \\
&= \left[\frac{y^5}{10} + \frac{y^4}{96}\right]_0^1 = \frac{53}{480}
\end{aligned}
$$

Solution Cont...

$$
\begin{aligned}
P(X+Y\le 1) &= \int_0^1\int_0^{1-y} f(x,y)\,dxdy \\
&= \int_0^1\int_0^{1-y}\left(xy^2+\frac{x^2}{8}\right)dxdy \\
&= \int_0^1\left[\frac{x^2y^2}{2}+\frac{x^3}{24}\right]_0^{1-y} dy \\
&= \int_0^1\left[\frac{(1-y)^2y^2}{2}+\frac{(1-y)^3}{24}\right] dy \\
&= \frac{13}{480}
\end{aligned}
$$

Problem

Given the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 < x < 1, 0 < y < x \\ 0, & \text{elesewhere.} \end{cases}$$

1. Find $f_x(x), f_Y(y)$, $f(x|y)$ and $f(y|x)$.
2. Evaluate $P\left(Y < \frac{1}{8} | X < \frac{1}{2}\right)$.

Statistical Independence

Let X and Y be two random variables, discrete (continuous), with joint probability mass (density) function) $p(x, y)$ $(f(x, y)$ and marginal distributions $P_X(x)$ and $P_Y(y)$, $(f_X(x)$ and $f_Y(y))$ respectively. The random variables X and Y are said to be statistically independent if and only if $p(x, y) = P_X(x)P_Y(y)$,for all (x, y) within their range.
$(f(x, y) = f_X(x)f_Y(y)$,for all (x, y) within their range.)

Problem

Given the joint density function

$$f(x, y) = \begin{cases} \frac{x^3 y^3}{16}, & 0 \leq x \leq 2, 0 \leq y \leq 2. \\ 0, & \text{elesewhere.} \end{cases}$$

Show that the random variables X and Y are independent.

Problem

The joint probability mass function of two dimensional random variables (X, Y) is

	$p(x,y)$	x		
		0	1	2
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

Show that the random variables X and Y are not independent.

Mean of a Random Variable (Mathematical Expectation)

Assuming that one fair coin was tossed twice, we find that the sample space for our experiment is $S = \{HH, HT, TH, TT\}$.

Denote by X the number of heads. Since the 4 sample points are all equally likely, it follows that

$P(X = 0) = P(TT) = \frac{1}{4}, P(X = 1) = P(TH) + P(HT) = \frac{1}{2}$, and $P(X = 2) = P(HH) = \frac{1}{4}$, where a typical element, say TH, indicates that the first toss resulted in a tail followed by a head on the second toss. Now, these probabilities are just the relative frequencies for the given events in the long run. Therefore, Mean $= (0)\left(\frac{1}{4}\right) + (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) = 1$.

This result means that a person who tosses 2 coins over and over again will, on the average, get 1 head per toss.

Definition:

Let X be a random variable with probability distribution. The mean, or expected value, of X is $E(X) = \sum xp(x)$ if X is discrete, and $E(X) = \int_{-\infty}^{\infty} xf(x)dx$ if X is continuous.

Problem

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Solution : Let X represent the number of good components in the sample. The probability distribution of X is

$p(0) = 1/35, p(1) = 12/35, p(2) = 18/35$, and $p(3) = 4/35$.

$$E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right)$$
$$= \frac{1}{7}.$$

Problem

Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \frac{20{,}000}{x^3}, & x > 100 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected life of this type of device.

Solution : $E(X) = \int_{100}^{\infty} x f(x) dx.$

$E(X) = \int_{100}^{\infty} (x) \left(\frac{20,000}{x^3} \right) dx = 200.$ Therefore, we can expect this type of device to last, on average, 200 hours.

Let X be a random variable with probability distribution.

Then $E(X^2)$ is

$$E(X^2) = \sum x^2 p(x), \qquad \text{if } X \text{ is discrete, and}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx, \quad \text{if } X \text{ is continuous.}$$

Similarly, we can define $E(Y^2)$. In generally, we can define $E(X^r), r \geq 1$ and $E(Y^r), r \geq 1$. for the discrete and continuous random variables X and Y.

Expected Value of a new random variable $Y = g(X)$

Theorem:Let X be a random variable with probability distribution. The expected value of the random variable $Y = g(X)$ is

- $E(Y = g(X)) = \sum g(X)P(X = x)$, if X is discrete, and
- $E(Y = g(X)) = \int_{-\infty}^{\infty} g(X)f(x)dx$, if X is continuous.

Problem

Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

X	4	5	6	7	8	9
P(X=x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $Y = 2X - 1$ represent the amount of money, in INR (Rs), paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution

By the above Theorem, the attendant can expect to receive

$$
\begin{aligned}
E(Y = 2X - 1) &= \sum_{x=4}^{x=9} (2x - 1) P(X = x) \\
&= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) \\
&+ (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) \\
&= \text{Rs.}12.67
\end{aligned}
$$

Problem

Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of $Y = 4X + 3$.

Solution:

By the above Theorem,

$$E(Y = 4X + 3) = \int_{-1}^{2} \frac{(4x+3)(x^2)}{3} dx = \int_{-1}^{2} \frac{1}{3}(4x^3 + 3x^2) dx = 8.$$

Mathematical Expectation Properties

- $E[(aX + b)] = aE(x) + b$
- $Var[(aX + b)] = a^2 Var(X)$
- The variables X and Y are independent if $E(XY) = E(X)E(Y)$
- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- Correlation between X and Y is $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$, where σ_x is a standard deviation of X and σ_y is a standard deviation of Y.
- If the variables X and Y are independent, then $Cov(X, Y) = E(XY) - E(X)E(Y) = 0$ and hence $\rho(X, Y) = 0$.
- $-1 \leq \rho(X, Y) \leq 1$.

Definition

Let X and Y be random variables with joint probability distribution. The mean, or expected value, of the random variable $Z = g(X, Y)$ is

- $E(Z) = \sum_{X=x} \sum_{Y=y} g(X, Y)p(x, y), \qquad$ if (X, Y) is discrete, and
- $E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(X, Y)f(x, y)dxdy,$ if (X, Y) is continuous.

Note-1

If $g(X, Y) = X$ in the above Definition, we have

$$E(X) = \begin{cases} \sum\limits_{X=x} \sum\limits_{Y=y} xp(x, y) & = \sum\limits_{X=x} xP_x(x), \quad \text{discrete case,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)dydx & = \int_{-\infty}^{\infty} xf_X(x)dx, \quad \text{continuous case.} \end{cases}$$

where $P_X(x)(f_X(x))$ is the marginal distribution of X(marginal density function of the random variable X) Therefore, in calculating $E(X)$ over a two-dimensional space, one may use either the joint probability distribution of X and Y or the marginal distribution of X. (marginal density function of the random variable x).

Note-2

Similarly, we define If $g(X, Y) = Y$ in the above Definition, we have

$$E(Y) = \begin{cases} \sum_{X=x}\sum_{Y=y} yp(x,y) = \sum_{Y=y} yP_Y(y), & \text{discrete case,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy = \int_{-\infty}^{\infty} yf_Y(y)dy, & \text{continuous case.} \end{cases}$$

where $P_Y(y)(f_Y(y))$ is the marginal distribution of the random variable Y (marginal density function of the random variable Y).

The expectation of X^2 is

$$E(X^2) = \begin{cases} \sum\limits_{X=x}\sum\limits_{Y=y} x^2 p(x,y) & = \sum\limits_{X=x} x^2 P_x(x), \quad \text{discrete case,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)dydx & = \int_{-\infty}^{\infty} x^2 f_X(x)dx, \quad \text{continuous case.} \end{cases}$$

Similarly, we can define $E(Y^2)$. In generally, we can define $E(X^r), r \geq 1$

and $E(Y^r), r \geq 1$. for the discrete and continuous random variables X and Y.

Problem

Let X and Y be the random variables with joint probability distribution indicated in the following Table.

	$p(x,y)$	x		
		0	1	2
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

Find $E(X), E(Y)$ and $E(XY)$.

Solution

The marginal distribution function of X and Y is

	$p(x,y)$	x			$P_Y(y)$
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
$P_X(x)$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

To find $E(X)$ and $E(Y)$ using joint probality mass function and marginal distribution of X and Y (Exercise.)

Solution Cont...

$$
\begin{aligned}
E(XY) &= \sum_{x=0}^{x=2}\sum_{y=0}^{y=2} xyp(x,y) \\
&= (0)(0)p(0,0) + (0)(1)p(0,1) + (0)(2)p(0,2) \\
&+ (1)(0)p(1,0) + (1)(1)p(1,1) + (1)(2)p(1,2) \\
&+ (2)(0)p(2,0) + (2)(1)p(2,1) + (2)(2)p(2,2) \\
&= 0 + 0 + 0 + 0 + p(1,1) + 2p(1,2) + 0 + 2p(2,1) + 4p(2,2) \\
&= 0 + \frac{3}{14} + (2)(0) + 0 + (2)(0) + 4(0) \\
&= \frac{3}{14}.
\end{aligned}
$$

Problem

Find $E\left(\frac{Y}{X}\right)$ for the density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, 0 < y < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Solution: We have

$$\begin{aligned} E\left(\frac{Y}{X}\right) &= \int_0^2 \int_0^1 \left(\frac{y}{x}\right) f(x,y) dxdy \\ &= \int_0^1 \int_0^2 \left(\frac{y(1+3y^2)}{4}\right) dxdy \\ &= \int_0^1 \left(\frac{y+3y^3}{2}\right) dy = \frac{5}{8}. \end{aligned}$$

Problem

The joint probability mass function of two dimensional random variables (X, Y) is $p(x, y) = \frac{x+y}{21}, x = 1, 2, 3$ and $y = 1, 2.$

1. Find the mean of X and Y
2. Find the variance of X and Y
3. Find the standard deviation of X and Y.
4. Find the covariance of X and Y.
5. Find the correlation bewteen X and Y.

Solution

The joint probability mass function of two dimensional random variables (X, Y) is

	$p(x, y)$	y	
		1	2
	1	$\frac{2}{21}$	$\frac{3}{21}$
x	2	$\frac{3}{21}$	$\frac{4}{21}$
	3	$\frac{4}{21}$	$\frac{5}{21}$

Solution Cont...

$$For(1), E(X) = \sum_{x=1}^{3}\sum_{y=1}^{2} xp(x,y)$$

$$= (1)p(1,1) + (1)p(1,2) + (2)p(2,1) + (2)p(2,2)$$

$$+ (3)p(3,1) + (3)p(3,2)$$

$$= (1)\left(\frac{2}{21}\right) + (1)\left(\frac{3}{21}\right) + (2)\left(\frac{3}{21}\right) + (2)\left(\frac{4}{21}\right)$$

$$+ (3)\left(\frac{4}{21}\right) + (3)\left(\frac{5}{21}\right)$$

$$= \frac{46}{21}.$$

Solution Cont...

$$
\begin{aligned}
E(Y) &= \sum_{x=1}^{3} \sum_{y=1}^{2} y p(x, y) \\
&= (1)p(1,1) + (1)p(2,1) + (1)p(3,1) \\
&+ (2)p(1,2) + (2)p(2,2) + (2)p(3,2) \\
&= (1)\left(\frac{2}{21}\right) + (1)\left(\frac{3}{21}\right) + (1)\left(\frac{4}{21}\right) \\
&+ (2)\left(\frac{3}{21}\right) + (2)\left(\frac{4}{21}\right) + (2)\left(\frac{5}{21}\right) \\
&= \frac{33}{21}.
\end{aligned}
$$

Solution Cont...

$$\textcolor{red}{For(2)}, E(X^2) = \sum_{x=1}^{3}\sum_{y=1}^{2} x^2 p(x,y)$$

$$= (1^2)p(1,1) + (1^2)p(1,2) + (2^2)p(2,1) + (2^2)p(2,2)$$

$$+ (3^3)p(3,1) + (3^2)p(3,2)$$

$$= (1)\left(\frac{2}{21}\right) + (1)\left(\frac{3}{21}\right) + (4)\left(\frac{3}{21}\right) + (4)\left(\frac{4}{21}\right)$$

$$+ (9)\left(\frac{4}{21}\right) + (9)\left(\frac{5}{21}\right)$$

$$= \frac{114}{21}.$$

Solution Cont...

$$
\begin{aligned}
E(Y^2) &= \sum_{x=1}^{3}\sum_{y=1}^{2} y^2 p(x,y) \\
&= (1^2)p(1,1) + (1^2)p(2,1) + (1^2)p(3,1) \\
&+ (2^2)p(1,2) + (2^2)p(2,2) + (2^2)p(3,2) \\
&= (1)\left(\frac{2}{21}\right) + (1)\left(\frac{3}{21}\right) + (1)\left(\frac{4}{21}\right) \\
&+ (4)\left(\frac{3}{21}\right) + (4)\left(\frac{4}{21}\right) + (4)\left(\frac{5}{21}\right) \\
&= \frac{57}{21}.
\end{aligned}
$$

Solution Cont...

$$
\begin{aligned}
\sigma_x^2 = \mathrm{Var(X)} &= E(X^2) - [E(X)]^2 \\
&= \frac{114}{21} - \left[\frac{46}{21}\right]^2 = \frac{2394}{441} - \frac{2116}{441} \\
&= \frac{278}{441}.
\end{aligned}
$$

$$
\begin{aligned}
\sigma_y^2 = \mathrm{Var(Y)} &= E(Y^2) - [E(Y)]^2 \\
&= \frac{57}{21} - \left[\frac{33}{21}\right]^2 = \frac{1197}{441} - \frac{1089}{441} \\
&= \frac{108}{441}.
\end{aligned}
$$

Solution Cont...

$$\text{For}(3), \sigma_x = \sqrt{Var(X)} = \sqrt{\frac{278}{441}} = 0.794$$
$$\sigma_y = \sqrt{Var(Y)} = \sqrt{\frac{108}{441}} = 0.495$$

For(4), Cov(X,Y) $= E(XY) - E(X)E(Y)$.

First we have to find $E(XY)$.

Solution Cont...

$$E(XY) = \sum_{x=1}^{3}\sum_{y=1}^{2} xyp(x,y)$$

$$= (1)(1)p(1,1) + (1)(2)p(1,2) + (2)(1)p(2,1) + (2)(2)p(2,2)$$

$$+ (3)(1)p(3,1) + (3)(2)p(3,2)$$

$$= (1)p(1,1) + (2)p(1,2) + (2)p(2,1) + (4)p(2,2)$$

$$+ (3)p(3,1) + (6)p(3,2)$$

$$= (1)\left(\frac{2}{21}\right) + (2)\left(\frac{3}{21}\right) + (2)\left(\frac{3}{21}\right) + (4)\left(\frac{4}{21}\right)$$

$$+ (3)\left(\frac{4}{21}\right) + (6)\left(\frac{5}{21}\right) = \frac{72}{21}.$$

Solution Cont...

$$
\begin{aligned}
\text{Therefore,} \quad \text{Cov(X,Y)} &= E(XY) - E(X)E(Y) \\
&= \frac{72}{21} - \left(\frac{46}{21}\right)\left(\frac{33}{21}\right) \\
&= \frac{1512}{441} - \frac{1518}{441} \\
&= \frac{-6}{441} \\
&= -0.0136
\end{aligned}
$$

Solution Cont...

$$\textit{For}(5), \quad \rho(X,Y) = \frac{\text{Cov(X,Y)}}{(\sigma_x)(\sigma_y)} = \frac{-0.0136}{(0.794)(0.495)} = \frac{-0.0136}{0.393} = -0.035$$

Problem

Let X and Y be the random variables with joint probability distribution indicated in the following Table.

	$p(x,y)$	x		
		0	1	2
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

- Find the covariance of X and Y.
- Find the correlation between X and Y.

Problem

The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \leq y \leq x \leq 1 \\ 0, & \text{elesewhere.} \end{cases}$$

1. Find the covariance of X and Y.
2. Find the correlation between X and Y.

Solution

We first compute the marginal density functions of X and Y(Exercise).

They are

$$f_X(x) = \begin{cases} 4x^3, & 0 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_Y(y) = \begin{cases} 4y(1-y^2), & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Solution Cont...

From these marginal density functions, we compute

$E(X) = \int_0^1 x4x^3 dx = \frac{4}{5}$ and

$E(Y) = \int_0^1 y4y(1-y^2) dy = \frac{8}{15}$.

From the joint density function given above, we have

$E(XY) = \int_0^1 \int_y^1 xy8xydxdy = \frac{4}{9}$.

$$
\begin{aligned}
For(1) \quad \text{Cov(X,Y)} &= E(XY) - E(X)E(Y) \\
&= \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) \\
&= \frac{4}{225}
\end{aligned}
$$

Solution Cont...

First we have to find $E(X^2), E(Y^2), \sigma_x^2$ and σ_Y^2. $E(X^2) = \int_0^1 4x^5 dx = \frac{2}{3}$ and $E(Y^2) = \int_0^1 4y^3(1-y^2)dy = \frac{1}{3}$.

$$\begin{aligned}
\sigma_x^2 = \text{Var(X)} &= E(X^2) - [E(X)]^2 \\
&= \frac{2}{3} - \left[\frac{4}{5}\right]^2 = \frac{2}{3} - \frac{16}{25} = \frac{2}{75}. \\
\sigma_y^2 = \text{Var(Y)} &= E(Y^2) - [E(Y)]^2 \\
&= \frac{1}{3} - \left[\frac{8}{15}\right]^2 = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}.
\end{aligned}$$

Solution Cont...

$$\text{Therefore,} \quad \rho(X, Y) = \frac{\text{Cov(X,Y)}}{\sqrt{(\sigma_x^2)(\sigma_y^2)}}$$

$$= \frac{\frac{4}{225}}{\sqrt{\left(\frac{2}{75}\right)\left(\frac{11}{225}\right)}}$$

$$= \frac{4}{\sqrt{66}}.$$

Moments

Let X be a discrete random variable and X takes the values $x_1, x_2, x_3, \ldots, x_n$ with probabilities $p_1, p_2, p_3, \ldots, p_n$ Then r^{th} moment is

- $\mu_r^{'}$(about origin) $= \sum_{i=1}^{n} x_i^r p_r$
- $\mu_r^{'}$(about any point x=A) $= \sum_{i=1}^{n} (x_i - A)^r p_i$
- $\mu_r^{'}$(about mean $\bar{x}$) $= \sum_{i=1}^{n} (x_i - \bar{x})^r p_i$

Similarly, for continuous case on the interval (a, b), the r^{th} moment is

- $\mu_r^{'}$(about origin) $= \int_a^b x^r f(x) dx$
- $\mu_r^{'}$(about any point x=A) $= \int_a^b (x - A)^r f(x) dx$
- $\mu_r^{'}$(about mean $\bar{x}$) $= \int_a^b (x - \bar{x})^r f(x) dx$

Note

mean $= E(X) = \mu_1^{'}$ and

Variance $= \sigma^2 = E(X^2) - (E(X))^2 = {\mu_2}' - (\mu_1^{'})^2$.

Characteristic Function

The expected values $E(X), E(X^2), E(X^3), \ldots,$ and $E(X^r)$ are called moments. As you have already experienced in some cases, the mean $\mu = E(X)$ and the variance $\sigma^2 =$ Var(X)$=E(X^2)\mu_2$, which are functions of moments, are sometimes difficult to find. Special functions, called moment-generating functions can sometimes make finding the mean and variance of a random variable simpler.

Although higher order moments of a random variable X may be obtained directly by using the definition of $E(X^n)$, it will be easier in many problems to compute them through the characteristic function or equivalently through the moment generating function of the random variable X.

While the characteristic function analysis exists, the moment generating function need not . Moment Generating Function (MGF) of a random variable(discrete or continuous) X is defined as $E(e^{tx})$, where t is real variable, and denoted as $M_X(t)$.

If X is discrete, then $M_X(t) = E(e^{tx}) = \sum_{i=1}^{n} e^{tx_i} p(x_i)$,
where X takes the values $x_1, x_2, x_3, \ldots, x_n$ with probabilities
$p_(x_1), p_(x_2), p_(x_3), \ldots p(x_n)$

If X is continuous, then $M_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$.

Properties of MGF

- The r^{th} moment $\mu_r^{'}$ is the coefficient of $\frac{t^r}{r!}$ in the expansion of $M_X(t)$ in series of powers of t.

$$M_X(t) = E(e^{tx}) = E\left[1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \cdots + \frac{(tx)^r}{r!} + \dots\right]$$
$$= 1 + tE(x) + \frac{t^2}{2!}E(x^2) + \cdots + \frac{t^r}{r!}E(x^r) + \dots$$
$$= 1 + t\mu_1^{'} + \frac{t^2}{2!}\mu_2^{'} + \cdots + \frac{t^r}{r!}\mu_r^{'} + \dots$$

Therefore, $\mu_r^{'}$ is the coefficient of $\frac{t^r}{r!}$ in the expansion of $M_X(t)$ in series of powers of t.

Properties of MGF

- $$\mu_r^{'} = E(X^r) = \left[\frac{d^r}{dt^r}(M_X(t))\right]_{t=0}.$$

  In particular, if $r = 1$ then $\mu_1^{'} = \left[\frac{d(M_X(t))}{dt}\right]_{t=0}$ is the mean of X.
- $M_{cX}(t) = M_X(ct)$
- $M_{aX+b}(t) = e^{bt} M_X(at)$
- $M_{X+Y}(t) = M_X(t) M_Y(t)$

Characteristic function

Let X be a discrete random variable and X takes the values $x_1, x_2, x_3, \ldots$ with probabilities $p_1, p_2, p_3, \ldots$ Then the characteristic function is defined as $\phi_X(t) = E(e^{itx}) = \sum_r e^{itx_r} p(x_r)$.

Similarly, for continuous case $\phi_X(t) = E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$.

Properties of Characteristic function

- $\mu_r^{'} = E(X^r)$=the coefficient of $\frac{i^r t^r}{r!}$ in the expansion of $\phi_X(t)$ in series of ascending powers of it.
-
$$\mu_r^{'} = E(X^r) = \frac{1}{i^r}\left[\frac{d^r}{dt^r}(\phi_X(t))\right]_{t=0}.$$

  In particular, if $r = 1$ then $\mu_1^{'} = \frac{1}{i}\left[\frac{d(\phi_X(t))}{dt}\right]_{t=0}$ is the mean of X.
- $\phi_{aX+b}(t) = e^{ibt}\phi(at)$
- $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$

Problem

Find the moment generating function of the random variable X whose probability mass function $P(X = x) = \frac{1}{2^x}, x = 1, 2, 3, \ldots,$ Deduce the mean and variance from moment generating function.

Answer: $M_X(t) = \frac{e^t}{2-e^t}$ and mean is 2.

Problem

The density function of a continous random variable X is given

$f(x) = kx(2-x), 0 \leq x \leq 2$. *Find* k, *mean,variance and* r^{th} *moment.*

Problem

Find the characteristic function of the random variable X whose probability mass function $P(X = r) = q^r p, r = 0, 1, 2, 3, \ldots$ and $p + q = 1$. Deduce the mean and variance from characteristic function.

Answer: $\phi_X(t) = p(1 - qe^{it})^{-1}$, mean is $\frac{q}{p}$ and variance is $\frac{q}{p^2}$.

Thank you