Joint Probability distributions

Earlier study of random variables and their probability distributions was restricted to 1-D sample spaces as we recorded outcomes of an experiment as values assumed by a single random variable.

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

For eg: in a study to determine the likelihood of success in college based on high school data, we might use 3-D sample space & record for each individual his/her aptitude test score, high school class rank & grade point average.

$(X, Y) \to$ 2 D random variable

If X and Y are two discrete r.v., the prob. distribution for their simultaneous occurrence can be represented by a function with values $f(x,y)$ for any pair of values (x,y) within the range of the random variables X and Y.

$f(x,y)$ is refered to as Joint prob. distribution of X and Y.

Hence in the discrete case,

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

The values $f(x,y)$ give the probability that outcome x & y occur at the same time.

For eg: If an 18 wheeler is to have its tyres serviced & tyres X represents the no. of miles these tyres have been driven & Y " the no. of tyres that need to be replaced, then $f(30000, 5)$ is the prob that the tyres are used over 30000 mile & the truck needs 5 new tyres.

Def 8 The function $f(x,y)$ is a joint probability distribution or prob. mass function of the discrete r.v. X & Y if

(i) $f(x,y) \geq 0$ for all (x,y)

(ii) $\sum_x \sum_y f(x,y) = 1$

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

For any region in xy plane, $P[(X,Y) \in A] = \sum\sum_A f(x,y)$

Ex 10 Two ball point pens are selected at random from a box that contains 3 blue pens, 2 red pens & 3 green pens. If X is the no. of blue pens selected & Y is the no. of red pens selected, find

a) the joint prob. fn $f(x,y)$

b) $P[(X,Y) \in A]$ where A is the region $\{(x,y) \mid x+y \leq 1\}$

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

a) Now, $f(0,1)$ represents the prob that a red & a green pens are selected.

Total no. of equally likely ways of selecting any 2 pens from 8 is $\binom{8}{2} = 28$.

The no. of ways of selecting 1 red from 2 red pens & 1 green " 3 green pens

is ${}^2C_1 \, {}^3C_1 = 6$

Hence $f(0,1) = \frac{6}{28}$

Similarly others are calculated. The sum of probs equals 1.

Joint prob fn

$$f(x,y) = \frac{{}^3C_x \; {}^2C_y \; {}^3C_{2-x-y}}{{}^8C_2}$$

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

	$f(x,y)$	x: 0	1	2	Row totals
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	$\frac{1}{28}$
	Col. Totals	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

b) $P[(X,Y) \in A] = P[X+Y \le 1] = f(0,0) + f(0,1) + f(1,0)$

$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}$

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

Def 9 The function $f(x,y)$ is a joint density function of the continuous r.v. X & Y if

(i) $f(x,y) \ge 0$ for all (x,y)

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

(iii) $P[(X,Y) \in A] = \iint_A f(x,y)\,dx\,dy$, for any region A in the xy plane.

Ex11 A privately owned business operates both a drive-in facility and walk-in facility. On a randomly selected day, let X & Y, resp. be the proportions of the time that the drive-in & walk-in facilities are in use, & suppose that the joint density fn of these r.v. is

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

i) Verify condition (ii) of Def 9

ii) find $P[(X,Y) \in A]$ where $A = \{(x,y) \mid 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \frac{2}{5}\int_0^1\int_0^1 (2x+3y)\,dx\,dy = \int_0^1 \left(\frac{2}{5} + \frac{6y}{5}\right) dy$$

$$= 1$$

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

$$= \frac{2}{5}\int_{1/4}^{1/2}\int_0^{1/2} (2x+3y)\,dx\,dy = \frac{2}{5}\int_{1/4}^{1/2}\left(\frac{1}{10} + \frac{3y}{5}\right) dy = \frac{13}{160}$$

Given the joint prob. distribution $f(x,y)$ of a discrete r.v. X and Y, the prob. distribution $g(x)$ of X alone is obtained by summing $f(x,y)$ over the values of Y. Similarly, the prob distribution $h(y)$ of Y alone is obtained by summing $f(x,y)$ over the values of X.

$g(x)$ & $h(y)$ are called the marginal distributions of X & Y, resp. when X & Y are cts r.v, summations are replaced by integrals.

Def 10 The marginal distributions of X alone & of Y alone are

$$\boxed{g(x) = \sum_y f(x,y), \qquad h(y) = \sum_x f(x,y)}$$

for the discrete case, and

$$\boxed{g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \qquad h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx}$$

for the continuous case.

→ $g(x)$ & $h(y)$ are just the marginal totals of the respective columns & rows when the values of $f(x,y)$ are displayed in the rectangular table.

Ex 12. for Ex 10, find the marginal distributions of X alone & Y alone.

$$g(0) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$$g(1) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28}$$

$$g(2) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = \frac{3}{28}$$

Similarly, $h(y)$ can be obtained by row totals.

x	0	1	2
$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

y	0	1	2
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

Ex 13 find $g(x)$ & $h(y)$ for the joint density function in Ex 11

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{2}{5}\int_0^1 (2x+3y)\,dy = \frac{4xy}{5} + \frac{6y^2}{10}\Big]_0^1 = \frac{4x+3}{5}$$

$$g(x) = \begin{cases} \frac{4x+3}{5} & 0<x<1 \\ 0 & \text{e.w.} \end{cases}$$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \frac{2}{5}\int_0^1 (2x+3y)\,dx = \frac{2(1+3y)}{5}$$

$$h(y) = \begin{cases} \frac{2(1+3y)}{5} & 0<y<1 \\ 0 & \text{e.w.} \end{cases}$$

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

and $P(a<X<b) = P(a<X<b, -\infty<Y<\infty)$

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

* $P(B|A) = \frac{P(A\cap B)}{P(A)}, \quad P(A) \neq 0.$

where A & B are now the events defined by $X=x$ & $Y=y$, resp.

then $P(Y=y|X=x) = \frac{P(X=x, Y=y)}{P(X=x)} = \frac{f(x,y)}{g(x)}, \quad g(x)>0$

here X & Y are discrete r.v.

$\frac{f(x,y)}{g(x)}$ is a fn of y with x fixed, satisfies all the conditions of a prob. distribution.

This is also true when $f(x,y)$ & $g(x)$ are the joint density & marginal dist, resp of a continuous r.v.

Def 11 Let X & Y be 2 r.v., discrete or continuous. The conditional distribution of the r.v. Y given that $X=x$ is

$$\boxed{f(y|x) = \frac{f(x,y)}{g(x)}, \quad g(x)>0}$$

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

$$\boxed{f(x|y) = \frac{f(x,y)}{h(y)}, \quad h(y)>0}$$

* To find

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

when X & Y are continuous,

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

Ex 14 for Ex10, find the conditional dist of X given $Y=1$ & use it to determine $P(X=0 \mid Y=1)$

$f(x|y)$ where $y=1$.

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$$

Now $f(x|1) = \frac{f(x,1)}{h(1)} = \frac{7}{3} f(x,1)$, $x = 0, 1, 2$

$f(0|1) = \frac{7}{3} f(0,1) = \frac{7}{3} \cdot \frac{3}{14} = \frac{1}{2}$

$f(1|1) = \frac{7}{3} f(1,1) = \frac{7}{3} \cdot \frac{3}{14} = \frac{1}{2}$

$f(2|1) = \frac{7}{3} f(2,1) = \frac{7}{3} \times 0 = 0$

and the conditional dist of X given $Y=1$ is

x	0	1	2
$f(x\|1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

$P(X=0 | Y=1) = f(0|1) = \frac{1}{2}$

$\therefore$ If it is known that 1 of the 2 pen refills selected is red, we have prob of $\frac{1}{2}$ that the other refill is not blue.

Ex 15 Given the joint density function

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

find $g(x)$, $h(y)$, $f(x|y)$ & evaluate $P(\frac{1}{4} < X < \frac{1}{2} | Y = \frac{1}{3})$

Soln for $0<x<2$

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{x(1+3y^2)}{4}\,dy = \frac{x}{2}$$

for $0<y<1$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^2 \frac{x(1+3y^2)}{4} = \frac{1+3y^2}{2}$$

Ex The joint density function of X & Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & 0<x<\infty, \quad 0<y<\infty \\ 0 & \text{o.w.} \end{cases}$$

Compute a) $P[X>1, Y<1]$

$P[(X,Y)\in A] = \iint_A f(x,y)\,dx\,dy$ b) $P[X<Y]$

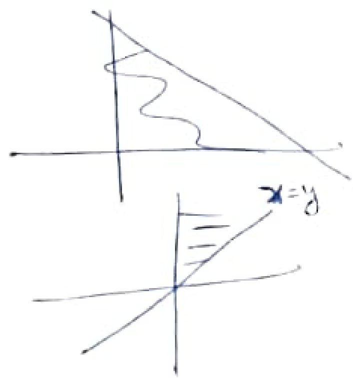
c) $P[X<a]$

a) $P[X>1, Y<1] = 2\int_{y=0}^{1}\int_{x=1}^{\infty} e^{-x}e^{-2y}\,dx\,dy$

$= -2\int_{y=0}^{1} e^{-x}\Big|_1^{\infty} e^{-2y}\,dy = -2\int_{y=0}^{1} -e^{-1}e^{-2y}\,dy$

$= -\frac{2}{2}e^{-1}e^{-2y}\Big|_0^1 = -e^{-1}(e^{-2}-1) = e^{-1}(1-e^{-2})$

b) $P(X<Y) = 2\iint_{(x,y):\,(x<y)} e^{-x}e^{-2y}\,dx\,dy$



$= 2\int_{y=0}^{\infty}\int_{x=0}^{y} e^{-x}e^{-2y}\,dx\,dy$

$= -2\int_{y=0}^{\infty} e^{-x}\Big)_0^y e^{-2y}\,dy = -2\int_{y=0}^{\infty}(e^{-y}-1)e^{-2y}\,dy$

$= -2\int(e^{-3y}-e^{-2y})dy = -2\left[\frac{e^{-3y}}{-3}+\frac{e^{-2y}}{2}\right]_0^{\infty}$

$= \frac{2}{3}(0-1)-1(0-1) = -\frac{2}{3}+1 = \frac{1}{3}$

c) $P[X<a] = 2\int_{x=0}^{a}\int_{y=0}^{\infty} e^{-x}e^{-2y}\,dy\,dy = -\frac{2}{2}\int_{x=0}^{a} e^{-x}e^{-2y}\Big]_0^{\infty}dx$

$= \int_0^a e^{-x}dx = -e^{-x}\Big|_0^a = -(e^{-a}-1) = 1-e^{-a}$

Ex Suppose that $f(x,y)$, the joint pmf of X & Y is given by

$f(0,0) = 0.4$, $f(0,1) = 0.2$, $f(1,0) = 0.1$ $f(1,1) = 0.3$

Calculate the conditional pmf of X given that $Y=1$.

$$P[Y=1] = \sum_{x} f(x,1) = f(0,1) + f(1,1) = 0.2+0.3 = 0.5$$

Hence $$P[X=0 \mid Y=1] = \frac{f(0,1)}{P[Y=1]} = \frac{0.2}{0.5} = \frac{2}{5}$$

& $$P[X=1 \mid Y=1] = \frac{f(1,1)}{P[Y=1]} = \frac{0.3}{0.5} = \frac{3}{5}$$

Ex The joint density of X & Y is given by

$$f(x,y) = \begin{cases} \frac{12}{5} x(2-x-y), & 0<x<1,\ 0<y<1 \\ 0 & \text{o.w.} \end{cases}$$

Compute the conditional density of X given that $Y=y$, where $0<y<1$.

Soln For $0<x<1$, $0<y<1$,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dx}$$

$$= \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx} = \frac{x(2-x-y)}{\frac{2}{3} - y/2} = \frac{6x(2-x-y)}{4-3y}$$