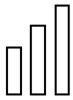
5. Joint probability Distribution:

Defines the simultaneous behavior of more than one random variable:

$$P(x_1 \le X \le x_2 \text{ AND } y_1 \le Y \le y_2)$$

Ex: length and diameter of a piece of plastic material from injection molding machine

Ex: X: no. of bars on cell phone 1, 2, 3
Y: no. of times city name is stated 1,2,3,4



y \ x	1	2	3	Sum
4	.15	.1	.05	.30
3	.02	.1	.05	.17
2	.02	.03	.2	.25
1	.01	.02	.25	.28
Sum	.2	.25	.55	1.0

- Each entry in the table specifies the joint probability distribution of X and Y
- Range of (X,Y): set of points (x,y) in 2-Dimeusional space for which P(X=x) and P(Y=y) is positive
- The joint probability of two variables is called bivariate probability distribution OR bivariate distribution.

Discrete Variates X and Y :

Joint probability distribution of X and Y is a distribution of the points (x,y) in the range of (X,Y) along with the probability of each point.

Joint probability Mass function of (X,Y) is $f_{XY}(x,y)$ and must satisfy

1) $f_{XY}(x,y) \ge 0$

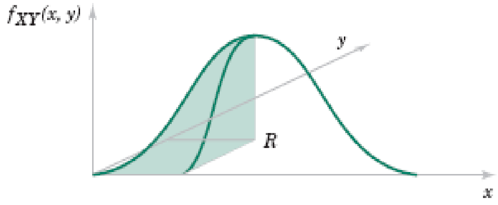
2) $\sum_x \sum_y f_{XY}(x,y) = 1.0$

3) $f_{XY}(x,y) = P(X = x, Y = y)$

$f_{XY}(x,y) = 0.0$ for values of (x,y) outside the range of X and Y

Continuous Variates X and Y

Probability distribution of two continuous random variables X,Y can be specified by providing a method for calculating the probability that X and Y assume a value in the region R of two dimensional space.



Probability that (X, Y) is in the region R is determined by the volume of $f_{XY}(x, y)$ over the region R.

Joint Probability density function f_{XY}(x,y) must satisfy

1) $f_{XY}(x, y) \geq 0$ for all x and y

2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x, y)dxdy = 1.0$

3) For any region R of two dimensional space

$$P\big((X, Y) \in R\big) = \iint_R f_{XY}(x, y)dxdy$$

= Probability X,Y assume values in R

=Vol under surface $f_{XY}(x, y)$ over R

Marginal Probability Distributions:

It is the individual Probability distribution of a random variable in cases in which more than one variable is defined in a random experiment.

Discrete Random Variable

Mass function for X and Y is f_{XY}(x,y)

$f_{XY}(x_i, y_j)$ →

$f_X(x_i) = \sum_j f_{xy}(x_i, y_j)$

$f_Y(y_j) = \sum_i f_{xy}(x_i, y_j)$

$f_X(x_i)$ = P(X=x_i) = Sum of P (X=x_i ;Y=y_j) for all points in the range of X,Y for which X= x_i

y \ x	1	2	3	$f_Y(y)$
4	.15	.1	.05	0.3
3	.02	.1	.05	0.17
2	.02	.03	.2	0.25
1	.01	.02	.25	0.28
$f_X(x)$	0.2	0.25	0.55	1.0 \ 1.0

Continuous Random Variable

Marginal Probability density function

$f_X(x) = \int_y f_{xy}(x, y)\, dy$ ↑ x is constant ; $f_Y(y) = \int_x f_{xy}(x, y)\, dx$ ↑ y is constant

$$P(a < X < b) = \int_a^b f_X(x)\, dx = \int_a^b \left[\int_{all\, y} f_{XY}(x, y)\, dy\right] dx$$

Example:

Let X denote the temperature (°C) and let Y denote the time in minutes that it takes for a diesel engine on an automobile to get ready to start. Assume that the joint density for (X,Y) is given by:

$f_{xy}(x,y) = c(4x + 2y + 1)$ $0 \leq x \leq 40$

$0 \leq y \leq 2$

a. Find the value of c that makes this a density.

b. Find the probability that on a randomly selected day, the air temperature will exceed 20°C and it will take at least one minute for the car to be ready to start.

c. Find the marginal densities for X and Y.

d. Find the probability that on a randomly selected day it will take at least one minute for the car to be ready to start.

e. Are X and Y independent? explain on a mathematical basis.

Solution:

a) For $f_{XY}(x,y)$ to be a density, the condition $\int_0^2 \int_0^{40} c(4x + 2y + 1)dxdy = 1$ must be satisfied. Therefore, c can be found upon integration and solving for it.
Upon integration we obtain:

$$(2x^2y + xy^2 + xy)\Big|_{x=0}^{x=40}\Big|_{y=0}^{y=2} = \frac{1}{c}$$

$$(2x40^2x2 + 40x2^2 + 40x2) = \frac{1}{c}$$

$$(6400 + 160 + 80) = \frac{1}{c} \quad ; \quad c = \frac{1}{6640} \quad ; \quad = 1.506x10^{-4}$$

b) P(X>20 and Y≥1) =?

$$P(\text{X} > 20 \text{ and Y} \geq 1) = \int_1^2 \int_{20}^{40} \frac{1}{6640} (4x + 2y + 1)dxdy$$

$$= \frac{2480}{6640} = 0.3735$$

c) $f_X(x) = \int_0^2 f_{XY}(x,y)dy$

$$f_X(x) = \int_0^2 c(4x + 2y + 1)dy$$

$$= c\,[4xy + y^2 + 2]_0^2$$

$$f_X(x) = c(8x + 6)$$

$$f_Y(y) = \int_0^{40} f_{XY}(x,y)dx$$

$$f_Y(y) = \int_0^{40} c(4x + 2y + 1)dx$$

$$= c\,[2x^2 + 2xy + x]_0^{40}$$

$$f_Y(y) = c(3240 + 80y)$$

d) $P(Y \geq 1) = \int_1^2 f_Y(y)dy$

$$= \int_1^2 c(3240 + 80y)dy$$

$$= c[3240y + 40y^2]_1^2$$

$$= \frac{3360}{6640}$$

e) $check\ if\ f_{XY}(x,y) = f_X(x)f_Y(y)$

$$f_X(x)f_Y(y) = c(8x + 6)\,c(3240 + 80y)$$

$$= c^2\,(25920\,x + 640xy + 480y + 19440) \neq f_{XY}(x,y)$$

Therefore, X and Y are NOT independent

Conditional Probability Distribution:

For two events A and B

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

This is the conditional probability of B given A provided A and B are dependent

Discrete RV:

A: event defined to be X=x
B: event defined to be Y=y

P(Y=y | X=x) = $\frac{P(X=x,Y=y)}{P(x)} = \frac{Joint}{Marginal}$

This is a new Probability Distribution of Y defined for each X.

y \ x	1	2	3	$f_Y(y)$
4	.15	.1	.05	0.3
3	.02	.1	.05	0.17
2	.02	.03	.2	0.25
1	.01	.02	.25	0.28
$f_X(x)$	0.2	0.25	0.55	1.0 \ 1.0

Conditional

$f_{Y|x}(y) = \frac{Joint}{Marginal}$

y \ x	1	2	3
4	.75	.4	.091
3	.1	.4	.091
2	.1	.12	.364
1	.05	.08	.454
$f_{Y\|x}(y)$	1.0	1.0	1.0

Each column represents a conditional Probability Mass Function

Ex: Signal bars

P(Y=1 | X=3) = $\frac{P(X=3,Y=1)}{P(X=3)} = \frac{f_{XY}(3,1)}{f_X(3)} = \frac{0.25}{0.55}$ = 0.454

The conditional PMF must satisfy the condition:

$$\sum_j P(Y = y_j | X = x_i = 1.0$$

Continuous RV

$f_{Y|x}(y) = \dfrac{f_{xy}(x,y)}{f_X(x)}$ for $f_X(x) > 0$

It is a probability density function for all Y in Rx

1) $f_{Y|x}(y) \geq 0$
2) $\int f_{Y|x}(y)\, dy = 1$
3) $P(Y \in B|_{X=x}) = \int_B (f_{Y|x}(y)\, dy$ For any set B in the range of Y.

Similar results are for X can be written
Using the results in the example of the diesel engine:

$$f_{Y|x}(y) = \frac{f_{xy}(x,y)}{f_X(x)}$$

$$= \frac{c(4x + 2y + 1)}{c(8x + 6)}$$

$$= \frac{(4x + 2y + 1)}{(8x + 6)}$$

Find probability Y>1 given X=20

$$P(Y > 1, X = 20) = \int_1^2 \frac{(4x + 2y + 1)}{(8x + 6)}\, dy \quad ; \quad for\ x = 20$$

$$= \frac{1}{166} \int_1^2 (4x + 2y + 1)\, dy$$

$$= \frac{1}{166} (84) \quad = \quad 0.51$$

Conditional Mean and Variance:

Mean:

Discrete: $E(Y|_x) = \mu_{Y|x} = \sum_y f_{y|x}(y).y$

Continuous: $E(Y|x) = \mu_{Y|x} = \int_y f_{Y|x}(y).y\, dy$

Variance:

Discrete: $V(Y|_x) = \sigma^2_{Y|_x} = \sum_y f_{Y|_x}(y)\,(y - \mu_{Y|_x})^2$

Continuous: $V(Y|_x) = \sigma^2_{Y|_x} = \int_y f_{Y|x}(y)\,.\,(y - \mu_{Y|_x})^2 dy$

Ex:

$E(Y|_1) = \mu_{Y|1} = 1\ x\ 0.05 + 2x\ 0.1 + 3\ x\ 0.1 + 4\ x\ 0.75 = 3.55$

$V(Y|_1) = \sigma^2_{Y|_1} = (1 - 3.55)^2\ x\ 0.05 + (2 - 3.55)^2 x\ 0.1 + (3 - 3.55)^2 x\ 0.1 + (4 - 3.55)^2\ x\ 0.75$