Joint probability mass function

If the random variables are discrete, then the **joint probability mass function** consists of probability values $P(X = i , Y = j) = P_{i\,j}$ satisfying

1) $0 \leq P_{i\,j} \leq 1$

2) $\sum\sum P_{i\,j} = 1$

Marginal Probability Distributions: Even though two random variables *X* and *Y* may be jointly distributed, if interest is focused on only one of the random variables, then it is appropriate to consider the probability distribution of that random variable alone. This is known as the **marginal distribution** of the random variable and can be obtained quite simply by summing the joint probability distribution over the values of the other random variable.

For example, for two discrete random variables *X* and *Y*, the probability values of the marginal distribution of *X* are

$P(X = i) = \sum_j Pij$

$i = 1,2,3,4$ and $j = 1,2,3$

$P(X=i) = \sum_{j=1}^{3} P_{ij} = P_{i1} + P_{i2} + P_{i3}$

$P(X=1) = \sum_{j=1}^{3} P_{1j} = P_{11} + P_{12} + P_{13}$

Conditional Probability Distributions : If two random variables *X* and *Y* are jointly distributed, then it is sometimes useful to consider the distribution of one random variable *conditional* on the other random variable having taken a particular value.

The **conditional probability distribution** of random variable *X* conditional on the event *Y* = *j* consists of the probability values

$P(X|Y=j) = \frac{P(X,Y=j)}{P(Y=j)}$

Covariance: Cov(X, Y) = E(XY) − E(X)E(Y)

Correlation: Corr(X,Y) = $\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$

-1 to 1

Question: A company that services air conditioner units in residences. If the random variable X, taking the values 1, 2, 3, and 4, is the service time in hours taken at a particular location, and the random variable Y , taking the values 1, 2, and 3, is the number of air conditioner units at the location, then these two random variables can be thought of as jointly related.

Suppose that their joint probability mass function P_{ij} is given in the following Figure. The figure indicates, for example, that there is a probability of 0.12 that X = 1 and Y = 1, so that there is a probability of 0.12 that a particular location chosen at random has one air conditioner unit that takes a technician one hour to service. Similarly, there is a probability of 0.07 that a location has three air conditioner units that take four hours to service.

		X = service time (hrs)			
		1	2	3	4
Y = number of air conditioner units	1	0.12	0.08	0.07	0.05
	2	0.08	0.15	0.21	0.13
	3	0.01	0.01	0.02	0.07

1) Show that this is a valid joint probability mass function

Solution: $\boldsymbol{\sum_{i=1}^{4}\sum_{j=1}^{3} P_{ij}} = 0.12 + 0.08 + \cdot\ \cdot\ \cdot + 0.07 = 1.00$

2) Calculate the probability that a location has no more than two air conditioner units that take no more than two hours to service.
i = 1,2   j = 1,2

Solution: $P_{11} + P_{12} + P_{21} + P_{22} = 0.12 + 0.08 + 0.08 + 0.15 = 0.43$

3) Find marginal probability mass function.

Solution: P(X = 1) = 0.12 + 0.08 + 0.01 = 0.21
P(X = 2) = 0.08 + 0.15 + 0.01 = 0.24
P(X = 3) = 0.07 + 0.21 + 0.02 = 0.3
P(X= 4) = 0.05 + 0.13+ 0.07 = 0.25

Similarly, you also need to calculate P(Y=1), P(Y=2) and P(Y=3)

X = service time (hrs)

Y = number of air conditioner units	1	2	3	4	Marginal distribution of Y
1	0.12	0.08	0.07	0.05	0.32
2	0.08	0.15	0.21	0.13	0.57
3	0.01	0.01	0.02	0.07	0.11
Marginal distribution of X	0.21	0.24	0.30	0.25	

4) Calculate the expected service time and the variance in the service times

Solution: $E(x) = \sum xp(x)$, $E(x^2) = \sum x^2 p(x)$,

The expected service time is $E(X) = \sum_{i=1}^{4} iP(X = i)$
$=1P(X=1) + 2P(X=2) + 3P(X=3) + 4P(X=4)$
$= (1 \,.\, 0.21) + (2 \,.\, 0.24) + (3 \,.\, 0.30) + (4 \,.\, 0.25) = 2.59$

We know, $V(X) = E(X^2) – (E(X))^2$
Since $E(X^2) = \sum_{i=1}^{4} i^2 P(X = i)$
$=1P(X=1) + 4P(X=2) + 9P(X=3) + 16P(X=4)$
$= (1 \,.\, 0.21) + (4 \,.\, 0.24) + (9 \,.\, 0.30) + (16 \,.\, 0.25) = 7.87$

The variance in the service times is
$Var(X) = E(X^2) - (E(X))^2 = 7.87 – (2.59)^2 = 1.162$

5) Calculate the expected number of units serviced and the variance in the number of units serviced.

Solution: The expected number of units serviced $E(Y) = \sum_{j=1}^{3} jP(Y = j)$
$= 1P(Y=1) + 2P(Y=2) + 3P(Y=3)$
$= 1×0.32 + 2×0.57 + 3×0.11 = 1.79$

We know, $V(Y) = E(Y^2) – (E(Y))^2$
Here, $E(Y^2) = \sum_{j=1}^{3} j^2 P(Y = j)$
$= 1×0.32 + 4×0.57 + 9×0.11 = 3.59$

The variance in the number of units serviced $V(Y) = 3.59 – (1.79)^2$
$= 0.3859$

5) Suppose that a technician is visiting a location that is known to have three air conditioner units, what is the conditional distribution of the service time *X*.

Solution:
We know, $P(X|Y=j) = \frac{P(X,Y=j)}{P(Y=j)}$

$\Rightarrow$P(X|Y=3) = $\frac{P(X,Y=3)}{P(Y=3)}$

Here, P(Y = 3) = 0.01 + 0.01 + 0.02 + 0.07 = 0.11

P(X=1|Y=3) = $\frac{\mathrm{P(X=1,Y=3)}}{P(Y=3)} = \frac{0.01}{0.11}$ = 0.091
P(X=2|Y=3) = $\frac{\mathrm{P(X=2,Y=3)}}{P(Y=3)} = \frac{0.01}{0.11}$ = 0.091
P(X=3|Y=3) = $\frac{\mathrm{P(X=3,Y=3)}}{P(Y=3)} = \frac{0.02}{0.11}$ = 0.182
P(X=4|Y=3) = $\frac{\mathrm{P(X=4,Y=3)}}{P(Y=3)} = \frac{0.07}{0.11}$ = 0.636

6) Calculate the conditional expectation of the service time.

Solution:

E(X|Y = 3) = $\sum_{\mathrm{i=1}}^{4} \mathrm{iP(X|Y = 3)}$
= (1 × 0.091) + (2 × 0.091) + (3 × 0.182) + (4 × 0.636) = 3.36

E(Y|X = 3) = $\sum_{\mathrm{j=1}}^{3} \mathrm{jP(Y|X = 3)}$

7) Calculate the covariance and correlation.

Solution:

Cov(X,Y) = E(X,Y) – E(X)E(Y)

The expected service time is *E(X)* = 2.59 hours, and the expected number of units serviced is *E(Y)* = 1.79.
E(X,Y) = $\sum_{i=1}^{4}\sum_{j=1}^{3} ijPij$ = (1 × 1 × 0.12) + (1 × 2 × 0.08) + ⋯ + (4 × 3 × 0.07) = 4.86

Cov(X,Y) = 4.86 – 2.56 × 1.79 = 0.224

We know,

Corr(X,Y) = $\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{0.224}{\sqrt{1.162 \times 0.384}}$ = 0.34