Stat 230 Introductory Statistics
Joint Distribution

University of British Columbia Okanagan

Joint Distributions

- So far, we have mostly looked at univariate probability.
- That is, distributions of a single random variable.
- Now we jump into multiple variables (two to start) .

Discrete: Joint

Definition 5.1 (Joint Probability Mass Function)

Let X and Y be discrete random variables. The joint PMF of X and Y describes how much mass is placed on each possible pair of values (x, y).

$$p(x, y) = P([X = x] \cap [Y = y])$$

Definition 5.2 (Joint Probability Function)

Let A be a set consisting of (x, y) pairs. Then the probability that a random pair (X, Y) lies in A is obtained by

$$P[(X, Y) \in A] = \sum \sum_{(x,y) \in A} p(x, y)$$

Discrete: Joint

Definition 5.3 (Marginal Pability Mass Function)

The marginal PMFs of X and Y, called $p_X(x)$ and $p_Y(y)$, can be found by

$$p_X(x) = \sum_y p(x, y) \qquad p_Y(y) = \sum_x p(x, y)$$

Theorem 5.4 (Joint Probability Mass Function)

For $p(x, y)$ to be a joint PMF, then $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$ must hold.

Discrete: Joint PMF

Example 5.1

Suppose that the random variable X can take only the values1,2 and 3; that the random variable Y can take only the value 1,2,3 and 4; that the joint *pmf* of X and Y is as specified in the follow table:

		Y			
		1	2	3	4
	1	0.1	0	0.1	0
X	2	0.3	0	0.1	0.2
	3	0	0.2	0	0

Find the value below:

1. $p(X = 1, Y = 3)$
2. $P((X, Y) \in \{(X, Y) | X \leq 2, Y \geq 3\})$
3. $P(X = 1)$
4. List probability distribution of X
5. $P((X, Y) \in \{(X, Y) | X \leq 2, Y \geq 3 \text{ and } X + Y = 5\})$

Discrete: Joint PMF

Continuous

Definition 5.5

Let X and Y be continuous random variables. The joint PDF for X and Y for any two-dimensional set A is $f(x, y)$ if for any two-dimensional set A

$$P[(X, Y) \in A] = \int_A \int f(x, y) dxdy$$

- If A is a rectangle, $\{(x, y) \mid a \leq x \leq b, c \leq y \leq d\}$

$$P[(X, Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dydx$$

Continuous

- From textbook...

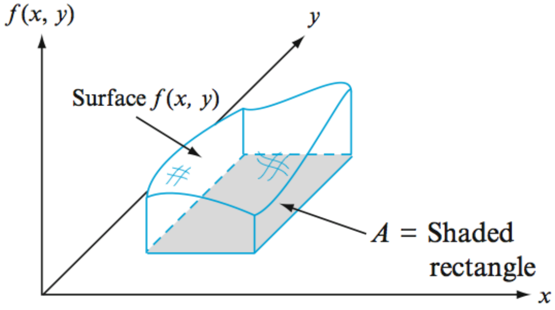


Figure 5.1 $P[(X, Y) \in A]$ = volume under density surface above A

Continuous

Definition 5.6

Marginal PDFs of X and Y, denoted $f_X(x)$ and $f_Y(y)$, can be found by

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

Continuous

Example 5.2

Let X and Y be two continuous random variables defined over the unit square and with joint PDF $f(x, y) = c(x^2 + y^2)$.

1. Find c.
2. Find the marginal PDF of X.

Independence

- We have discussed independence of events.
- We learned, for instance, that if A and B are independent events, then $P(A \cap B) = P(A)P(B)$.
- Let's now extend this idea to independence among probability distributions.

Independence

- The random variables X and Y are said to be independent if for any two sets of real numbers A and B,

$$P(X \in A, Y \in B) = P\{X \in A\}P\{Y \in B\}$$

- In other words, X and Y are independent if , for all A and B, the events $E_A = \{X \in A\}$ and $F_B = \{Y \in B\}$ are independent.

Independence

Theorem 5.7

When X and Y are two discrete random variables, the condition of independence is equivalent to

$$p(x, y) = p_X(x)p_Y(y)$$

- In the jointly continuous case the condition of independent is equibalent to
$$f(x, y) = f_X(x)f_Y(y)$$
for all x, y
- If these properties do not hold, then X and Y are dependent.

Examples

Example 5.3

Let X and Y be two continuous random variables defined over the unit square and with joint PDF $f(x, y) = \frac{3}{2}(x^2 + y^2)$. Are X and Y independent?

Examples

Example 5.4

Let X and Y be two discrete random variables with the following joint probability distribution:

$$p(x,y) = \begin{cases} \frac{3^x 2^y e^{-5}}{x!y!} & x = 0,1,\ldots \quad y = 0,1,\ldots \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent?

Expected Values for Joint variables

Definition 5.8 (Joint Distribution)

When h is a function of two variables

- If X and Y are discrete random variables with joint probability mass function$p(x, y)$, then

$$\mathbb{E}[h(X, Y)] = \sum_{x} \sum_{y} h(x, y) p(x, y)$$

- If X and Y are continuous random variables with joint probability density function $f(x, y)$, then

$$\mathbb{E}[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dxdy$$

Expected Values for Joint variables

Example 5.5

Consider X and Y are discrete random variables with joint distribution $p(x, y)$. If X and Y are independent, find $\mathbb{E}[XY]$.

Covariance

Definition 5.9

The covariance between random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mu_x)(Y - \mu_y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- It is a measure of how linearly associated two variables are.
  Negative covariance means that an increase in one variable results in a decrease of the other.
  Positive covariance means that the variables increase/decrease together.

- If X and Y are independent, this implies $\mathrm{Cov}(X, Y) = 0$.
  THE REVERSE IS NOT NECESSARILY TRUE!

Variance

Theorem 5.10

Let a, b and c be constants, then:

$$\mathbb{V}ar[X+Y] = \mathbb{V}ar[X] + \mathbb{V}ar[Y] + 2Cov(X,Y)$$
$$\mathbb{V}ar[X-Y] = \mathbb{V}ar[X] + \mathbb{V}ar[Y] - 2Cov(X,Y)$$

Theorem 5.11

If X and Y are independent then $Cov(X,Y) = 0$ and so

$$\mathbb{V}ar[X \pm Y] = \mathbb{V}ar[X] + \mathbb{V}ar[Y].$$

Covariance

- **Theorem 5.12**

  for a and b constants:

  $$Cov(aX + bY, Z) = aCov(X, Z) + bCov(Y, Z)$$

- Note: covariance needs context. If you measure a covariance of $-150,348$, is this a strong negative relationship?

- That will depend on the scales of X and Y...

- Hence, more often we would like correlation, which is a standardized measure.

Correlation

- The correlation between two random variables X and Y is

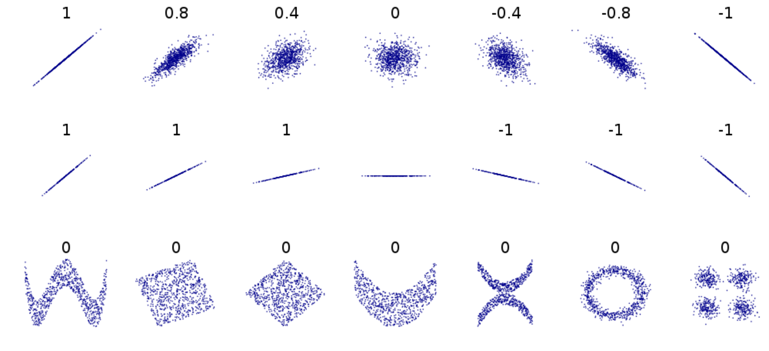
$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y}$$

- Correlation is bounded:

$$-1 \leq \rho \leq 1$$

- If X and Y independent, then $\rho = 0$.
  THE REVERSE IS NOT NECESSARILY TRUE!

Correlation



upload.wikimedia.org/wikipedia/commons/d/d4/Correlation_examples2.svg