Lec 21

5 Joint Probability Distributions and Random Samples

5.1 Jointly Distributed Random Variables

There are many experimental situations in which more than one random variables will be of interest to an investigator. For example, X and Y might be the height and weight, respectively, of a randomly selected individual.

5.1.1 Two Discrete Random Variables

Recall: The probability mass function (pmf) of a single discrete random variable X specifies how much probability mass is placed on each possible X value.

The joint pmf of two discrete random variables X and Y describes how much probability

mass is placed ____.

Definition 24. Let X and Y be two discrete random variables defined on the sample space S. The joint probability mass function $p(x, y)$ is defined for each pair of numbers (x, y) by

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

To make it a valid pmf:

1. $0 \le p(x,y) \le 1$

2. $\sum_{x,y} p(x,y) = 1$

Let A be any particular set consisting of pairs of (x, y) values. Then the probability $P[(X, Y) \in A]$ is

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

Example 69. The joint pmf is given as

$X \in \{0,1,2\}$

$Y \in \{0,1,2\}$

$p(x, y)$	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0.1	0.04	0.02
$x = 1$	0.08	0.2	0.06
$x = 2$	0.06	0.14	0.3

$x \backslash y$	0	1	2
0			
1			
2			

$P(1,0) = 0.08$

$P(X=1, Y=0) = 0.8$

Question:

a. Is it a valid pmf?

$0.1 + 0.04 + 0.02 + 0.08 + \cdots + 0.3 = 1$

so yes

b. Find $P(X = 1 \text{ and } Y = 1) = p(1,1) = 0.2$

c. Find $P(X \leq 1 \text{ and } Y \leq 1)$.

$= 0.1 + 0.04 + 0.08 + 0.2 = 0.42$

d. Find $P(X \neq 0 \text{ and } Y \neq 0)$.

$= 0.2 + 0.06 + 0.14 + 0.3$

$P_X(x) = P(X = x) = \sum_y P(X = x, Y = y)$

$\{X = x\} = \bigcup_y \{X = x, Y = y\}$

> **Definition 25.** The **marginal probability mass function** of X, denoted by $p_X(x)$, is given by
>
> $$P_X(x) = \sum_y p(x, y)$$
>
> for each possible value x. Similarly, the marginal probability mass function of Y is
>
> $$P_Y(y) = \sum_x p(x, y)$$
>
> for each possible value y.

The marginal pmf of X gives the distribution of X if it is observed without Y, and the marginal pmf of Y gives the distribution of Y if it is observed without X,

Example 70. (Example 69 continued)

e. Find the marginal pmf of X.

x	0	1	2
p_X	0.16	0.34	0.5

← sum to 1

$0.1 + 0.04 + 0.02$

f. Find $P(X \leq 1)$. $= p_X(0) + p_X(1) = 0.16 + 0.34 = 0.5$

or $p(0,0) + p(0,1) + p(0,2) + p(1,0) + p(1,1) + p(1,2)$

g. Find the marginal pmf of Y.

y	0	1	2
p_Y	0.24	0.38	0.38

if X & Y are indep.

$P(X=x, Y=y) \stackrel{\text{if}}{=} P(X=x)\, P(Y=y)$

Definition 26. Two discrete random variables are independent if **for every pair** of (x, y), we have

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

If $p(x, y) \neq p_X(x) \cdot p_Y(y)$ for **at least one pair** of (x, y), then X and Y are dependent.

Example 71. (Example 69 continued)

h. Are X and Y independent?

$P(0,0)=0.1$

$P_X(0)\,P_Y(0)$
$\downarrow \quad \downarrow$
$0.16 \quad 0.24$

$0.1 \neq 0.16 \times 0.24$

$\Rightarrow$ not indep.

if = move to next pair of val.

i. Find $P(X < Y)$.

$(0,1),(0,2),(1,2)$: $0.04+0.02+0.06=0.12$

j. Find $P(5X + Y \leq 7)$.

$0,0 \rightarrow 5\times0+0=0\leq7$ ✓

$0,1 \rightarrow 5\times0+1=1\leq7$ ✓

$\vdots$

$2,0 \rightarrow$ — — —

$2,1 \rightarrow$

$2,2 \rightarrow 5\times2+2=12>7$ ✗

$+$

prob. $=0.5$

k. Find $E(X)$. $=\sum x P_X(x) = 0\times0.16+1\times0.34+2\times0.5$
$=1.34$

l. Find $E(Y)$. $=\sum y P_Y(y) = 0\times0.24+1\times0.38+2\times0.38$

$=1.14$

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

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

$$= \sum_{x} x \sum_{y} p(x,y)$$

$$= \sum_{x} x\, p_X(x)$$

5.2 Expected Values, Covariance, and Correlation

Let X and Y be jointly distributed random variables with pmf $p(x,y)$. Let $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be any function. Then the expected value of $h(X,Y)$, denoted by $E[h(X,Y)]$ is given by

$$E[h(x,y)] = \sum_{x,y} h(x,y)\, P(x,y)$$

For example, $E(XY) =$ $\sum_{x,y} xy\, p(x,y)$

5.2.1 Covariance

When two random variables X and Y are not independent, it is frequently of interest to assess if the two random variables are linearly dependent and if so, how strongly they are related to one another.

Definition 27. The **covariance** is a measure of the strength between two random variables X and Y of linear association, which is defined as

$$E[XY] - E[X]E[Y]$$

Example 72. The joint and marginal pmf's for X = automobile policy deductible amount and Y = homeowner policy deductible amount are

	$p(x,y)$	y = 500	1000	5000
	100	0.3	0.05	0
x	500	0.15	0.20	0.05
	1000	0.10	0.10	0.05

x	100	500	1000
$p_X(x)$	0.35	0.4	0.25

y	500	1000	5000
$p_Y(y)$	0.55	0.35	0.10

Find the $\text{Cov}(X,Y)$.

Solution.