9 Multiple RVs

Wednesday, February 29, 2012
1:25 PM

Previously, we talked about mean (average)
variance
standard deviation

of one set of data.

Now, we want to extend the calculation to > 1 sets of data.

Correlation
Covariance
correlation coefficient

One big example...

Suppose we have 10 students in a class...

student id	Midterm score (x)	Final score (y)
1	10	40
2	10	40
3	10	50
4	20	40
5	20	40
6	20	50
7	20	50
8	20	50
9	20	50
10	20	50

To summarize these data, use a table

x \ y	40	50
10	2	1
20	2	5

Randomly select one of the students

X = his/her midterm score

Y = his/her final score

What is the probability that $X = 10$ and $Y = 40$?

$$P[X = 10 \text{ and } Y = 40] = \frac{2}{10}$$

$$\parallel\!\!|$$

$$p_{X,Y}(10, 40)$$

Defn $p_{X,Y}(x,y) = P[X = x \text{ and } Y = y]$

joint pmf

$p_{X,Y}(10, 20) = 0$

$p_{X,Y}(x,y)$

$x \backslash y$	40	50
10	1/5	1/10
20	1/5	1/2

What if my interest is only on the midterm score?

For example, randomly select a student
what is the probability that
his/her midterm score = 10?

$p_X(10) = P[X = 10] = \frac{2+1}{10} = \frac{3}{10}.$

$= \frac{2}{10} + \frac{1}{10} = \frac{3}{10}$ same

$= p_{X,Y}(10,40) + p_{X,Y}(10,50)$

$= \sum_y p_{X,Y}(10,y)$

$p_X(20) = P[X = 20] = \sum_y p_{X,Y}(20,y) = \frac{2}{10} + \frac{5}{10} = \frac{7}{10}$

Formula :

$$\left.\begin{aligned} p_X(x) &= \sum_y p_{X,Y}(x,y) \\ p_Y(y) &= \sum_x p_{X,Y}(x,y) \end{aligned}\right\} \text{marginal pmf's.}$$

$$p_X(x) = \begin{cases} 3/10, & x = 10 \\ 7/10, & x = 20 \\ 0, & \text{otherwise} \end{cases} \qquad p_Y(y) = \begin{cases} 2/5, & y = 40 \\ 3/5, & y = 50 \\ 0, & \text{otherwise.} \end{cases}$$

$\mathbb{E}X = \sum_x x\, p_X(x) = \frac{3}{1\cancel{0}} \times 1\cancel{0} + \frac{7}{1\cancel{0}} \times 2\cancel{0} = 17 \qquad \mathbb{E}Y = \sum_y y\, p_Y(y) = \frac{2}{5} \times 40 + \frac{3}{5} \times 50 = 46$

Given that $X = 10$, what is the probability that $Y = 50$?

$= \frac{1}{3}$

$$P[\overbrace{Y = 50}^{A} \mid \overbrace{X = 10}^{B}] = \frac{P(A \cap B)}{P(B)} = \frac{P[X = 10 \text{ and } Y = 50]}{P[X = 10]}$$

$$P[\underbrace{Y = 50}_{A} \mid \underbrace{X=10}_{B}] = \frac{P(A\cap B)}{P(B)} = \frac{P[X=10 \text{ and } Y=50]}{P[X=10]}$$

$$p_{Y|X}(50|10) = \frac{p_{X,Y}(10,50)}{p_X(10)} = \frac{1/10}{3/10} = \frac{1}{3}$$

Formula: $p_{Y|X}(y|x) \equiv P[Y=y \mid X=x] = \dfrac{p_{X,Y}(x,y)}{p_X(x)}$

conditional pmf

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

To calculate the total score, we use

$$t = \frac{x}{20}\times 40 + \frac{y}{50}\times 40 + 20$$

midterm = 40% of total score

final = 40% of total score

HW/quiz/participation

max final score

max midterm score

t		
$x \backslash y$	40	50
10	72	80
20	92	100

Let T be the total score of a randomly selected student.

$$p_T(t) = \begin{cases} 1/5 & t = 72, \\ 1/10 & t = 80, \\ 1/5 & t = 92, \\ 1/2 & t = 100, \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbb{E}T = \sum_t t\, p_T(t) = 90.8.$$

Alternatively,

$$\mathbb{E}T = \mathbb{E}\left[\frac{X}{20}40 + \frac{Y}{50}40 + 20\right] = \mathbb{E}\left[2X + \frac{4}{5}Y + 20\right]$$

$$= 2\,\mathbb{E}X + \frac{4}{5}\mathbb{E}Y + 20$$

$$= 2(17) + \frac{4}{5}(46) + 20 = 90.8.$$

In general, if $T = g(X,Y)$

$$\mathbb{E}T = \mathbb{E}\left[g(X,Y)\right] = \sum_{x}\sum_{y} g(x,y)\, p_{X,Y}(x,y)$$