Let X and Y be jointly distributed rv's with pmf $p(x,y)$ or pdf $f(x,y)$ according to whether the variables are discrete or continious. Then the expected value of a function $h(X,Y)$, denoted by $E[h(X,Y)]$, is given by

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y) \cdot p(x,y) & \text{if X and Y are discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) \cdot f(x,y)\, dx\, dy & \text{if X and Y are continious.} \end{cases}$$

The covariance between two rv's X and Y is

$$\text{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))]$$

$$= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)\, p(x,y) & X, Y = \text{discrete.} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f(x,y)\, dx\, dy & X, Y = \text{continious.} \end{cases}$$

(*) $\text{Cov}(X,Y) = E(XY) - \mu_X \cdot \mu_Y$

Correlation

The correlation coefficient of X and Y, denoted by $\text{Corr}(X,Y)$, $\rho_{X,Y}$ or just ρ, is defined by

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Proposition

1. If 'a' and 'c' are either both positive or both negative,

$$Corr(aX+b, cX+d) = Corr(X,Y).$$

2. For any two rv's X and Y,

$$-1 \leq Corr(X,Y) \leq 1.$$

3. If X and Y are independent, then $\rho_{X,Y} = 0$ but converse not true.

4. $\rho_{X,Y} = 1$ or -1 if and only if $Y = aX + b$ for some nos. a & b with $a \neq 0$.

E.g. An instructor has given a short quiz consisting of two parts. For a randomly selected student, let X = the number of points earned on the 1st part and Y = the number of points earned on the 2nd part. Suppose that the joint pmf of X and Y is given in the table

$p(x,y)$		y = 0	5	10	15
x	0	0.02	0.06	0.02	0.1
	5	0.04	0.15	0.2	0.1
	10	0.01	0.15	0.14	0.01

a) If the score recorded in the grade book is the total no. of points earned on the two parts, what is the expected recorded score $E(X+Y)$?

let $h(x,y) = X+Y$

$$E[h(x,y)] = \sum_x \sum_y (x+y) \cdot p(x,y)$$

$$= (0+0) \cdot 0.02 + (0+5)\, 0.06 + (0+10) \cdot 0.02 + (0+15) \cdot 0.1$$
$$+ \cdots + (10+10) \cdot 0.14 + (10+15) \cdot 0.01$$

$= 14.1$

(b) If the max^m of the two scores is recorded, what is the expected recorded score ?

$$E[\underbrace{\max(X,Y)}_{h(x,y)}] = \sum_x \sum_y h(x,y)\; p(x,y)$$

$$= \max(0,0) \cdot 0.02 + \max(0,5) \cdot 0.06 + \cdots + \max(10,10) \cdot 0.14 + \max(10,15) \cdot 0.06$$

$= 9.6$ //

(c) Covariance for X and Y

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$
$$= \sum xy\, p(x,y) - \left[\left(\sum x P_X(x)\right)\left(\sum y P_Y(y)\right)\right]$$

$E(x) = \sum x P_X(x) = 0(0.02+0.06+0.02+0.1) + 5(0.04+0.15+0.2+0.1) + 10(0.01+0.15+0.01+0.14).$

$= 5.55.$

$E(Y) = \sum y P_Y(y) = 8.55.$

$E(XY) = \sum xy\, p(x,y) = 0.0 \cdot (0.02) + 0.5 \cdot (0.06) + \cdots + 10 \cdot 10 \cdot 0.14 + 10 \cdot 15 \cdot 0.01$

$= 44.25.$

$Cov(X,Y) = 44.25 - 5.55\,(8.55) = -3.2025.$ //

(d) Correlation for X and Y

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$\therefore \sigma_X^2 = \sum (x-\mu)^2 P_X(x) \qquad \mu = E(X) = 5.55$$

$$= (0-5.55)^2 (0.02+0.06+0.02+0.1) +$$

$$(5-5.55)^2 (0.04+0.15+0.2+0.1) + (10-5.55)^2 (0.01+0.15+0.14+0.01)$$

$$= 12.4475$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{12.4475} = \cdot$$

$$\therefore \sigma_Y^2 = \sum (y-\mu)^2 P_Y(y) = (0-8.55)^2 (0.02+0.04+0.01) + \cdots$$

$$= 19.1475.$$

$$\sigma_Y = \sqrt{\sigma_Y^2} = 19.1475.$$

$$\rho = \frac{-3.2025}{\sqrt{12.4475}\sqrt{19.1475}} \approx -0.2074$$

35. (a) Show that $Cov(ax+b, cY+d) = ac\, Cov(X,Y)$.

$$Cov(ax+b, cY+d) = E([ax+b]\cdot[cY+d]) - E(ax+b)\cdot E(cY+d)$$

$$= E(acXY + adX + bcY + bd) - (aE(X)+b)(cE(Y)+d)$$

$$= ac\,E(XY) + ad\,\cancel{E(X)} + bc\,\cancel{E(Y)} + \cancel{bd}$$

$$- (ac\,E(X)E(Y) + \cancel{ad}\,E(X) + \cancel{bc}\,E(Y) + \cancel{bd})$$

$$= ac(E(XY) - E(X)E(Y))$$

$$= ac\, Cov(X,Y).$$

(b) Show that $Corr(ax+b, cY+d) = Corr(X,Y)$.

$$Corr(ax+b, cY+d) = \frac{Cov(ax+b, cY+d)}{\sigma_{ax+b} \cdot \sigma_{cY+d}} = \frac{ac\, Cov(X,Y)}{|a|\sigma_X\, |c|\,\sigma_Y}$$

$$= \frac{ac}{|ac|} \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = Corr(X,Y)$$

(e) If a and c has opposite sign, then

$$Corr(aX+b, cY+d) = -Corr(X,Y).$$

36 Show that if $Y = aX+b$ $(a \neq 0)$, then $Corr(X,Y) = +1$ or -1.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{Cov(X, aX+b)}{\sigma_X \sigma_{aX+b}}$$

$$Cov(X,Y) = E[XY] - E(X) \cdot E(Y)$$

$$= E[aX^2 + bX] - E(X) \cdot E[aX+b]$$

$$= aE(X^2) + b\cancel{E(X)} - a(E(X))^2 - bE\cancel{(X)}$$

$$= a\left[E(X^2) - (E(X))^2\right]$$

$$Cov(X, aX+b) = a\,\sigma_X^2$$

$$\therefore Corr(X,Y) = \frac{a\,\sigma_X^2}{\sigma_X \cdot |a|\,\sigma_X} = \frac{a\,\sigma_X^2}{|a|\,\sigma_X^2} = \pm 1.$$