3.3 Multiple Random Variables

3.3.1 Joint and conditional probability distributions

X and Y are **<u>discrete</u>**:

Joint Probability Mass Function (joint PMF)

$$p_{X,Y}(x,y) = P(X = x, Y = y) = P[(X = x) \cap (Y = y)] \qquad (3.57)$$

Joint (Cumulative) Distribution Function (joint CDF)

$$F_{X,Y}(x,y) = P(X \le x, Y \le y) = P[(X \le x) \cap (Y \le y)] \qquad (3.56)$$

$$= \sum_{\{x_i \le x,\, y_j \le y\}} p_{X,Y}(x_i, y_j) \qquad (3.58)$$

(a) $F_{X,Y}(-\infty,-\infty) = 0; \quad F_{X,Y}(\infty,\infty) = 1$

(b) $F_{X,Y}(-\infty, y) = 0; \quad F_{X,Y}(\infty, y) = F_y(y)$

$F_{X,Y}(x,-\infty) = 0; \quad F_{X,Y}(x,\infty) = F_x(x)$

(c) $F_{X,Y}(x,y)$ is nonnegative and a non-decreasing function

Conditional Probability Mass Function

$$p_{X|Y}(x \mid y) \equiv P(X = x \mid Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \tag{3.61}$$

$$p_{Y|X}(y \mid x) \equiv P(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)} \tag{3.61a}$$

<u>If X and Y are statistically independent</u>

$$p_{X|Y}(x \mid y) = p_X(x) \qquad \text{and} \qquad p_{Y|X}(y \mid x) = p_Y(y)$$

$$p_{X,Y}(x, y) = p_X(x) p_Y(y) \tag{3.63}$$

Marginal Probability Mass Function

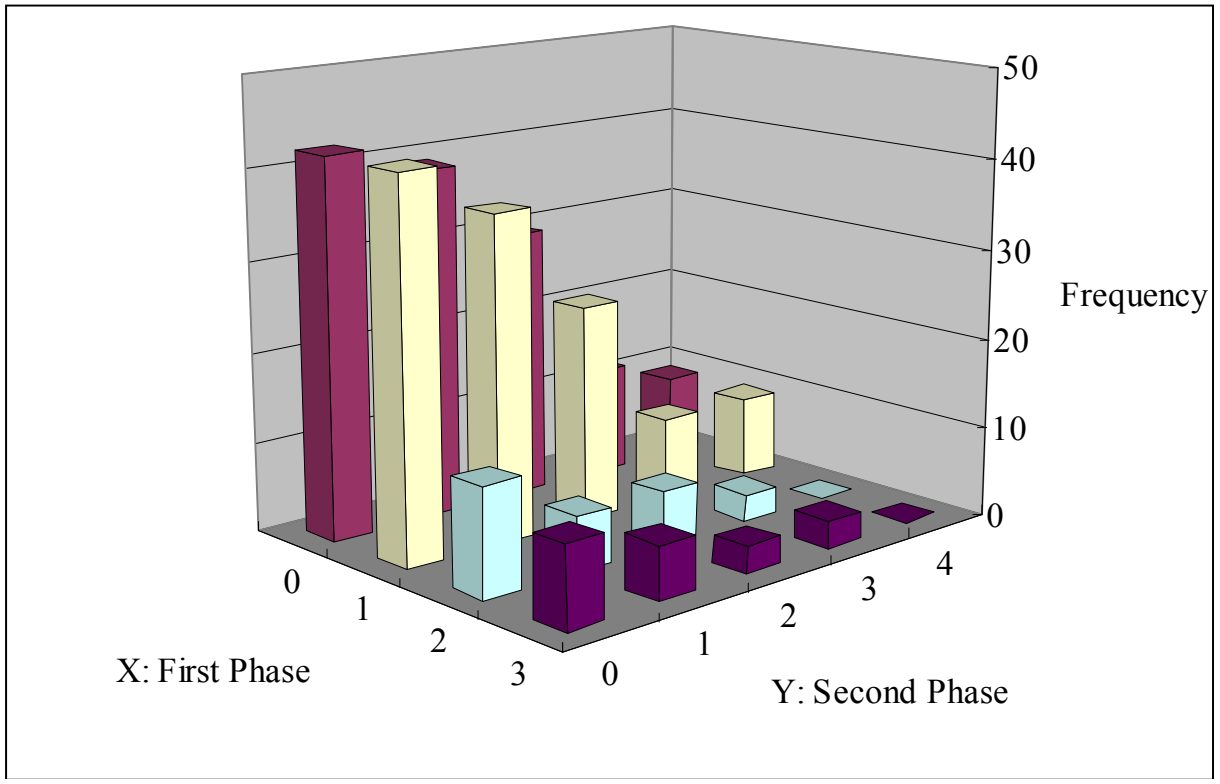
$$p_X(x) = \sum_{all\ y_i} p_{X,Y}(x, y_i) \tag{3.62}$$

$$p_Y(y) = \sum_{all\ x_i} p_{X,Y}(x_i, y) \tag{3.62a}$$

Ex. In a two phase production process, defects are attributed to either the first or second phase. Table below gives the number of times of x and y, where x and y are number of defects in the first and second phases, respectively.

N (x,y)		x				
		0	1	2	3	
	0	42	42	12	9	105
	1	39	36	6	6	87
y	2	30	24	6	3	63
	3	12	9	3	3	27
	4	9	9	0	0	18
		132	120	27	21	300

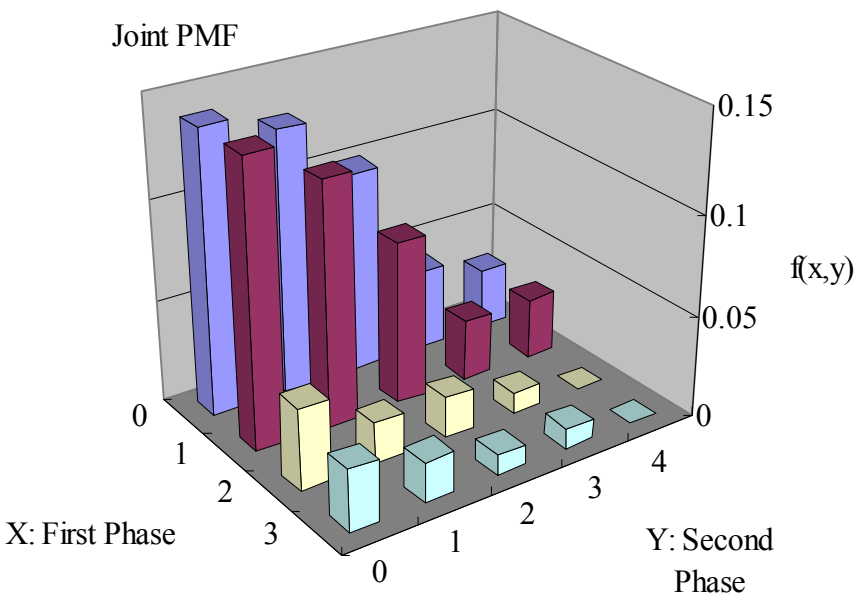
*Likelihood of more than 3 times in x, 4 times in y is negligible, respectively.

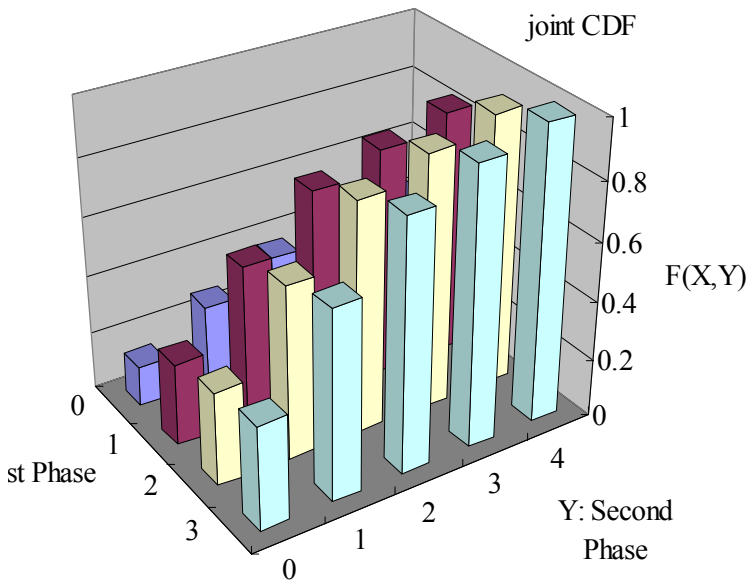


Joint PMF and Joint CDF

p(x,y)		x				
		0	1	2	3	
y	0	0.14	0.14	0.04	0.03	0.35
	1	0.13	0.12	0.02	0.02	0.29
	2	0.1	0.08	0.02	0.01	0.21
	3	0.04	0.03	0.01	0.01	0.09
	4	0.03	0.03	0	0	0.06
		0.44	0.4	0.09	0.07	1

F(x,y)		x			
		0	1	2	3
y	0	0.14	0.28	0.32	0.35
	1	0.27	0.53	0.59	0.64
	2	0.37	0.71	0.79	0.85
	3	0.41	0.78	0.87	0.94
	4	0.44	0.84	0.93	1

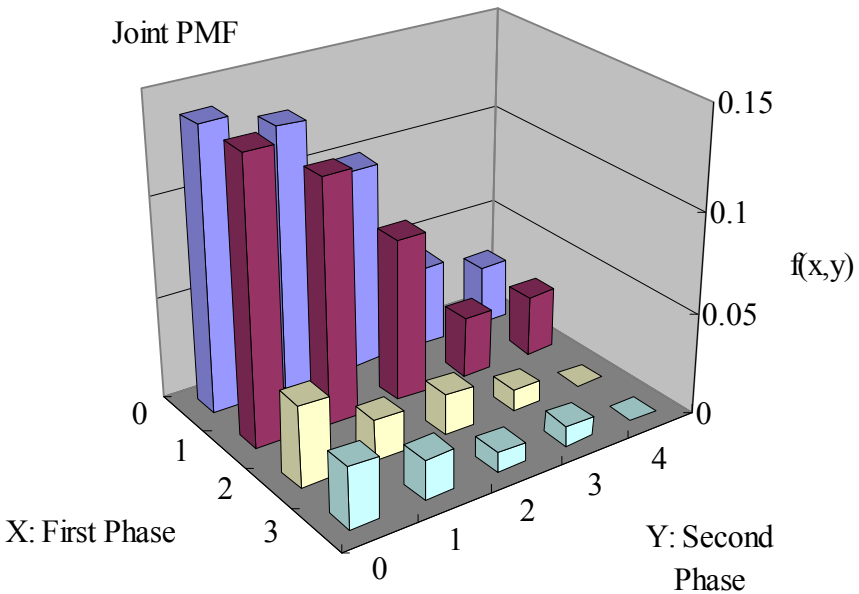




Marginal PMF and Conditional PMF

Marginal PMF

		x = 0	1	2	3	p(y)
y	0	0.14	0.14	0.04	0.03	0.35
	1	0.13	0.12	0.02	0.02	0.29
	2	0.1	0.08	0.02	0.01	0.21
	3	0.04	0.03	0.01	0.01	0.09
	4	0.03	0.03	0	0	0.06
p(x)		0.44	0.4	0.09	0.07	1



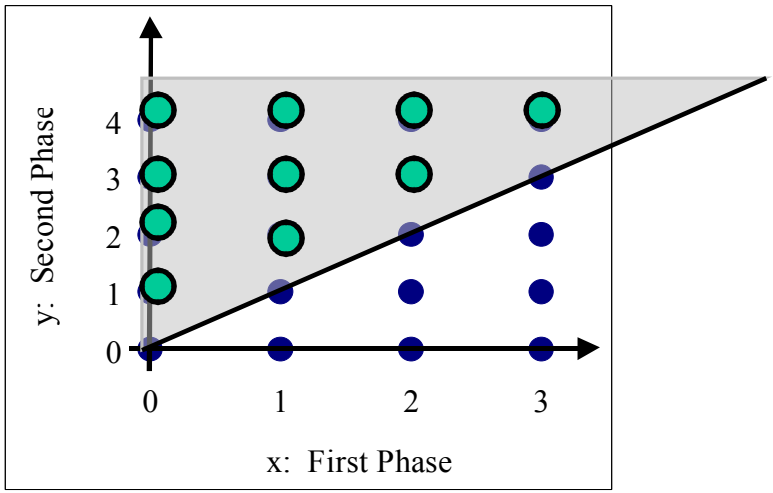
$p_{X|Y}(x|y) = p(x, y)/p(y)$

p(x|y)		x = 0	1	2	3	p(y)
y	0	0.4	0.4	0.114	0.086	1
	1	0.448	0.414	0.069	0.069	1
	2	0.476	0.381	0.095	0.048	1
	3	0.444	0.333	0.111	0.111	1
	4	0.5	0.5	0	0	1
		0.44	0.4	0.09	0.07	1

$p_{Y|X}(y|x) = p(x,y)/p(x)$

p(y|x)		x = 0	1	2	3	
y	0	0.318	0.35	0.444	0.429	0.35
	1	0.295	0.3	0.222	0.286	0.29
	2	0.227	0.2	0.222	0.143	0.21
	3	0.091	0.075	0.111	0.143	0.09
	4	0.068	0.075	0	0	0.06
p(x)		1	1	1	1	1

P(X<Y)

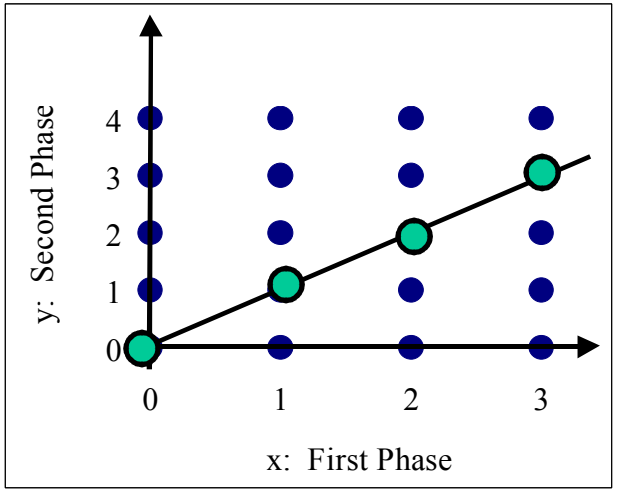


		x			
		0	1	2	3
y	0	0.14	0.14	0.04	0.03
	1	0.13	0.12	0.02	0.02
	2	0.1	0.08	0.02	0.01
	3	0.04	0.03	0.01	0.01
	4	0.03	0.03	0	0

P(X<Y)=0.13+0.10+0.04+0.03

+0.08+0.03+0.03+0.01=0.45

P(X=Y)



		x			
		0	1	2	3
y	0	0.14	0.14	0.04	0.03
	1	0.13	0.12	0.02	0.02
	2	0.1	0.08	0.02	0.01
	3	0.04	0.03	0.01	0.01
	4	0.03	0.03	0	0

P(X=Y)=0.14+0.12+0.02+0.01=0.29

X and Y are **<u>continuous</u>**:

Joint Probability Density Function (joint PDF)

$$f_{X,Y}(x,y)dxdy = P(x < X \le x + dx,\ \mathrm{y} < \mathrm{Y} \le \mathrm{y} + \mathrm{dy})$$

Joint (Cumulative) Distribution Function (joint CDF)

$$F_{X,Y}(x,y) = P(X \le x,\ Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,\ \mathrm{v})dvdu \qquad (3.59)$$

$$F_{X,Y}(-\infty,-\infty) = 0; \qquad F_{X,Y}(-\infty, y) = 0; \qquad F_{X,Y}(x,-\infty) = 0$$

$$F_{X,Y}(\infty, y) = F_y(y); \qquad F_{X,Y}(x,\ \infty) = F_x(x) \qquad F_{X,Y}(\infty,\infty) = 1$$

Conversely

$$f_{X,Y}(x,\ \mathrm{y}) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \qquad (3.60)$$

Also

$$P(a < X \le b,\ c < Y \le d) = \int_a^b \int_c^d f_{X,Y}(u,\ \mathrm{v})dvdu$$

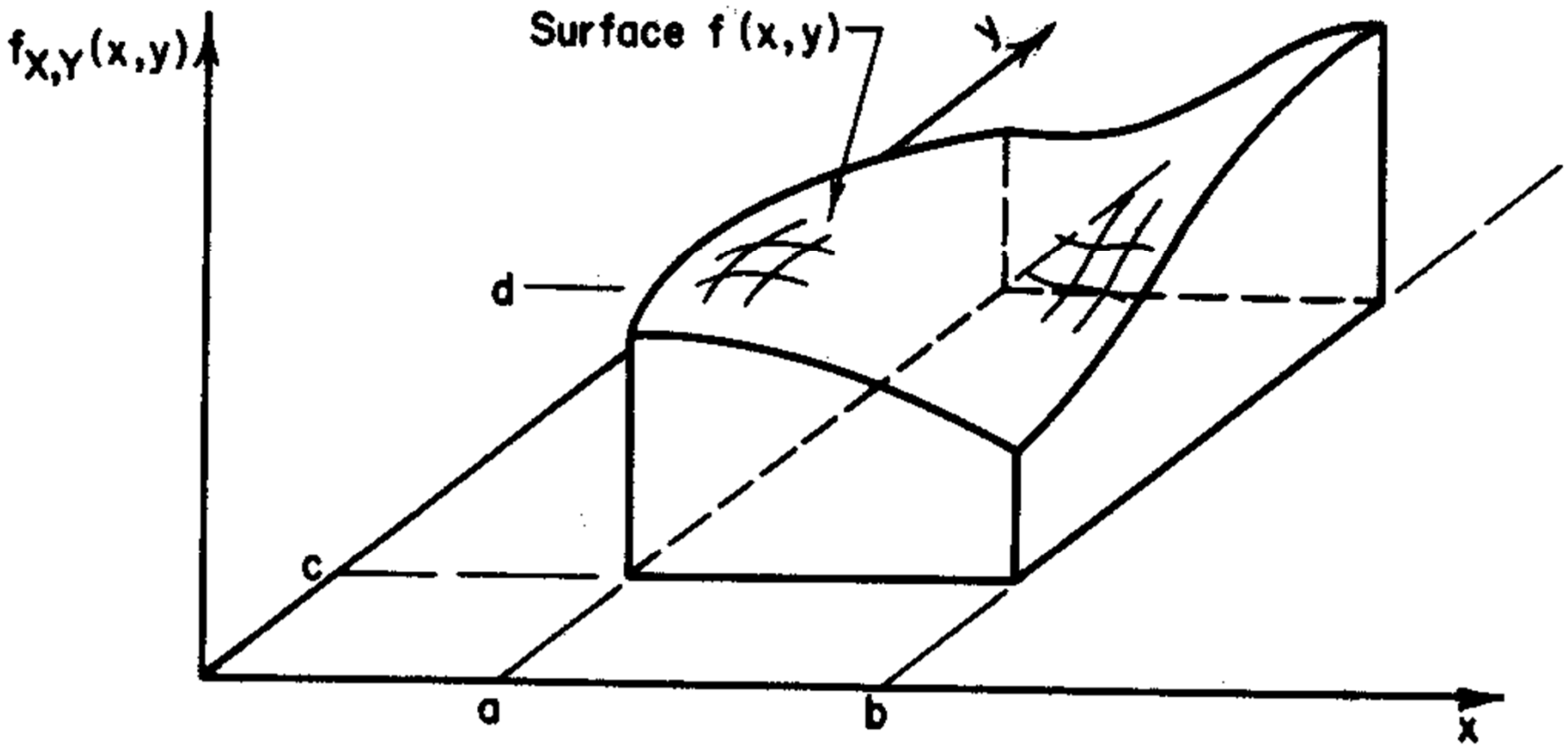


Figure 3.13 Joint PDF of X and Y

Conditional Probability Density Function

$$f_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

$$f_{Y|X}(y \mid x) = P(Y = y \mid X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \qquad (3.64)$$

<u>If X and Y are statistically independent</u>

$$f_{X|Y}(x \mid y) = f_X(x) \qquad \text{and} \qquad f_{Y|X}(y \mid x) = f_Y(y)$$

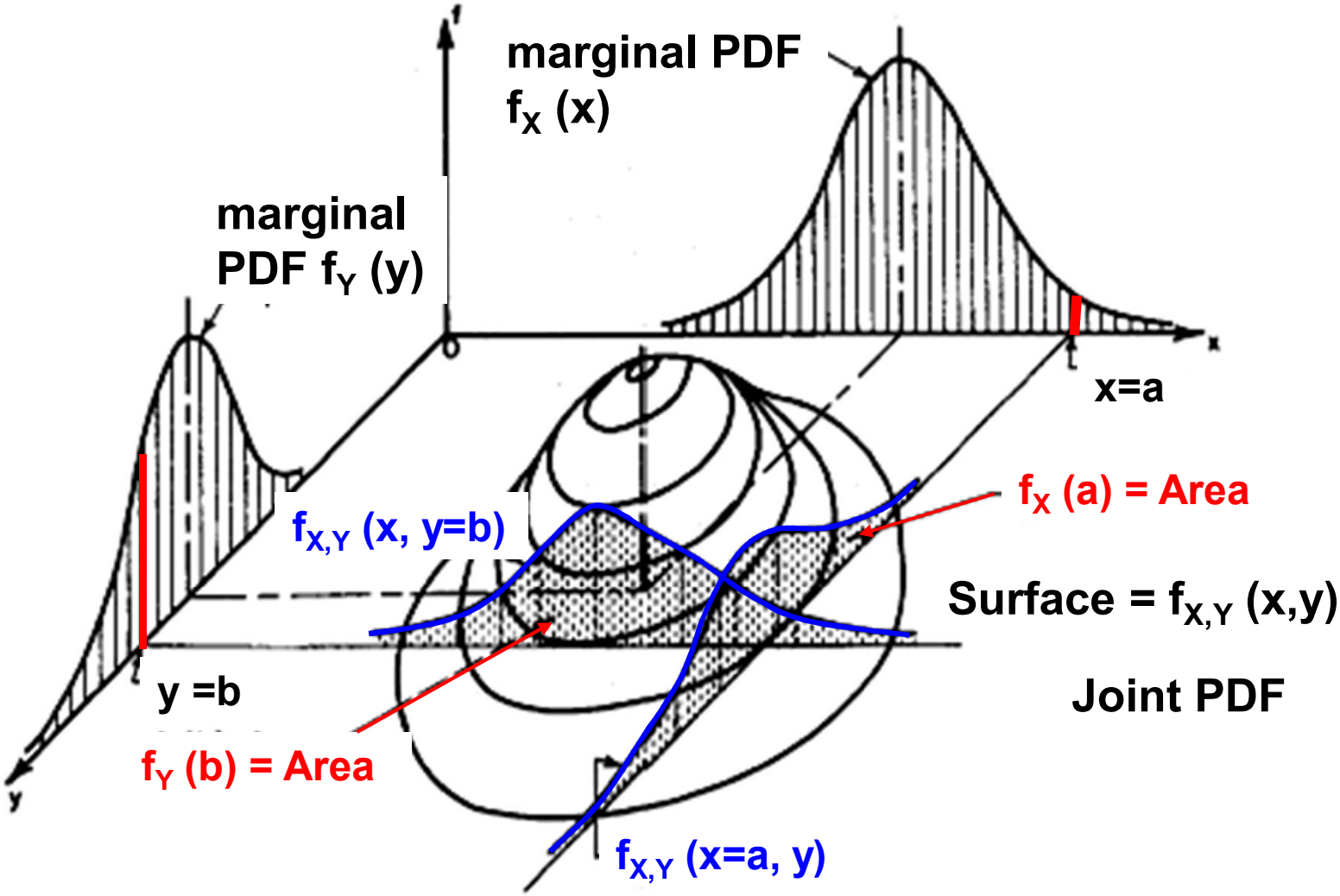
$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \qquad (3.66)$$

Marginal Density Function

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \qquad (3.67)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx \qquad (3.68)$$

Joint and Marginal PDF



Ex. 3.34

x: Work Duration
y: Average Productivity

		x				
		6	8	10	12	$p_Y(y)$
y	50	0.014	0.036	0.058	0.072	0.18
	70	0.036	0.216	0.18	0.043	0.475
	90	0.072	0.18	0.079	0.014	0.345
	$p_X(x)$	0.122	0.432	0.317	0.129	1

$$P_{Y|X}(90\ \%\ |\ 8\ \text{hrs}) = \frac{P_{X,Y}(8,90)}{P_X(8)} = 0.417$$

3.3.2 Covariance and Correlation

Covariance

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

$$= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y = E[XY] - \mu_X \mu_Y$$

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy \qquad (3.71)$$

in particular, X and Y are statistically independent

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \qquad (3.68)$$

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy \qquad (3.69)$$

$$= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = E[X]E[Y] = \mu_X \mu_Y \qquad (3.70)$$

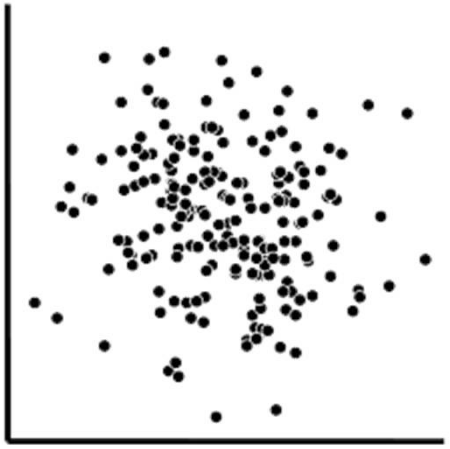
$$Cov(X,Y) = \mu_X \mu_Y - \mu_X \mu_Y = 0$$

Correlation Coefficient

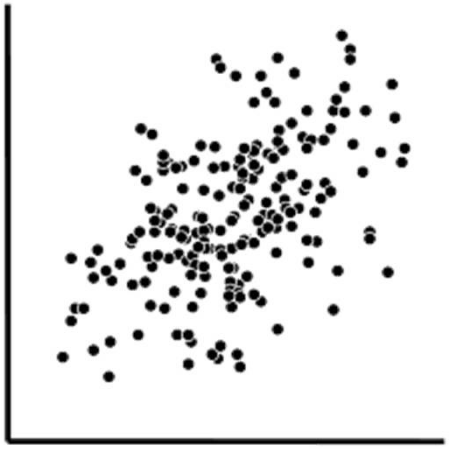
:Measure of the degree of linear interrelationship

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \qquad -1 \le \rho \le 1 \qquad (3.72)\ (3.73)$$

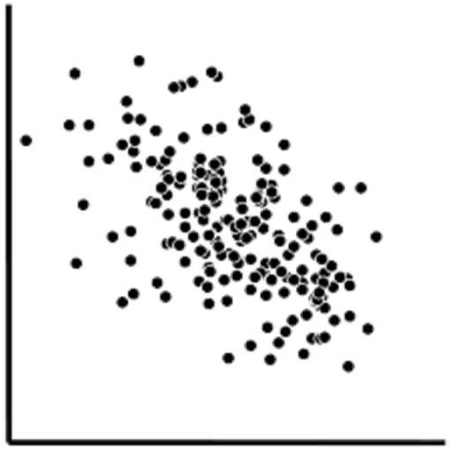
Significance of Correlation Coefficient



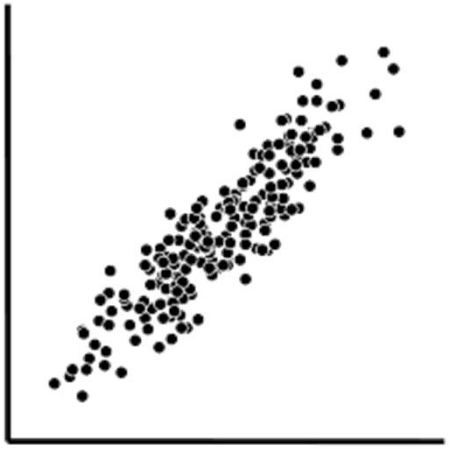
$\rho = 0$



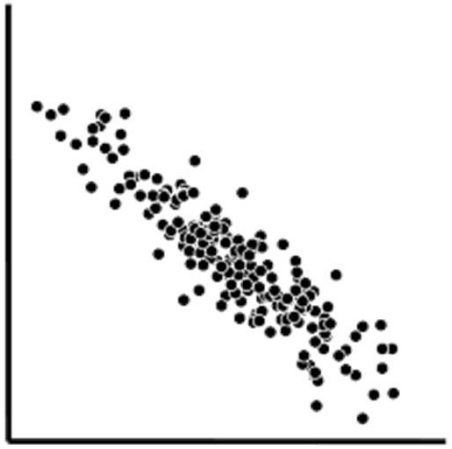
$\rho = 0.5$



$\rho = -0.5$



$\rho = 0.9$



$\rho = -0.9$

Covariance and Correlation Coefficient

p(x,y)		x				
		0	1	2	3	
	0	0.14	0.14	0.04	0.03	0.35
	1	0.13	0.12	0.02	0.02	0.29
y	2	0.1	0.08	0.02	0.01	0.21
	3	0.04	0.03	0.01	0.01	0.09
	4	0.03	0.03	0	0	0.06
		0.44	0.4	0.09	0.07	1

μ_X= 0*0.44+1*0.4+2*0.09+3*0.07=0.79

σ_X^2=0*0.44+1^2*0.4+2^2*0.09+ 3^2*0.07-0.79^2=0.7659 σ_X=0.8752

μ_Y= 0*0.35+1*0.29+2*0.21+3*0.09+4*0.06=1.22

σ_Y^2=0*0.35+1^2*0.29+2^2*0.21+ 3^2*0.09+4^2*0.06-1.22^2=1.4116
σ_Y=1.1881

Cov(X,Y)=0*0*0.14+0*1*0.14+0*2*0.04+0*3*0.03

+1*0*0.13+1*1*0.12+1*2*0.02+1*3*0.02

+2*0*0.10+2*1*0.08+2*2*0.02+2*3*0.01

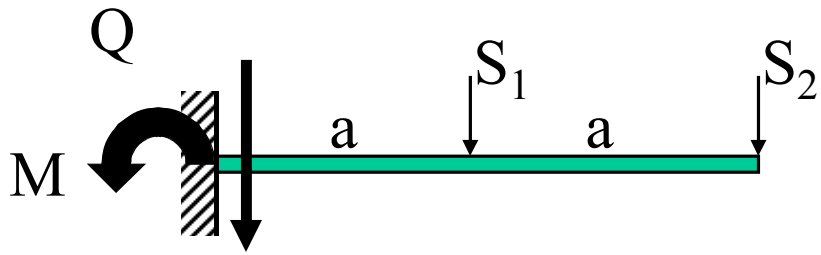
+3*0*0.03+3*1*0.03+3*2*0.00+3*3*0.00-0.79*1.22= -0.3538

ρ= -0.3538/(0.8752*1.1881)= -0.3402

Ex. 3.36

$Q=S_1+S_2$ $\quad S_1 : N(\mu_1, \sigma_1)$ $\quad S_2 : N(\mu_2, \sigma_2)$

$M=aS_1+2aS_2$



$$\mu_Q = \mu_1 + \mu_2 \qquad \sigma_Q{}^2 = \sigma_1^2 + \sigma_2^2$$

$$\mu_M = a\mu_1 + 2a\mu_2 \qquad \sigma_M{}^2 = a^2\sigma_1^2 + 4a^2\sigma_2^2$$

$$E[QM] = E[(S_1 + S_2)(aS_1 + 2aS_2)] = aE[S_1^2] + 3aE[S_1S_2] + 2aE[S_2^2]$$

$$E[S_1^2] = \sigma_1^2 + \mu_1{}^2, \quad E[S_2^2] = \sigma_2^2 + \mu_2{}^2, \quad E[S_1S_2] = E[S_1]E[S_2] = \mu_1\mu_2$$

$$E[QM] = a(\sigma_1^2 + \mu_1{}^2) + 3a\mu_1\mu_2 + 2a(\sigma_2^2 + \mu_2{}^2) = a(\sigma_1^2 + 2\sigma_2^2) + a(\mu_1{}^2 + 3\mu_1\mu_2 + 2\mu_2{}^2)$$

$$= a(\sigma_1^2 + 2\sigma_2^2) + a(\mu_1 + \mu_2)(\mu_1 + 2\mu_2) = a(\sigma_1^2 + 2\sigma_2^2) + \mu_Q\mu_M$$

$$\mathrm{Cov(Q, M)} = E[QM] - \mu_Q\mu_M = a(\sigma_1^2 + 2\sigma_2^2)$$

$$\rho_{QM} = \frac{\mathrm{Cov(Q, M)}}{\sigma_Q\sigma_M} = \frac{a(\sigma_1^2 + 2\sigma_2^2)}{\sqrt{\sigma_1^2 + \sigma_2^2}\sqrt{a^2\sigma_1^2 + 4a^2\sigma_2^2}} = \frac{(\sigma_1^2 + 2\sigma_2^2)}{\sqrt{\sigma_1^2 + \sigma_2^2}\sqrt{\sigma_1^2 + 4\sigma_2^2}}$$

In particular, $\sigma_1 = \sigma_2$

$$\rho_{QM} = \frac{3}{\sqrt{2}\sqrt{5}} = 0.948$$