Multinomial Distribution (Section 9.2)

Physical setup:

Similar to the binomial distribution, except that now we have $\boldsymbol{k}$ **types of outcomes** that can happen in a single repetition of an experiment.

The experiment is repeated **independently** $\boldsymbol{n}$ times, with one of $\boldsymbol{k}$ **distinct outcomes** possible each time.

Let the probabilities of these k types of outcomes be denoted by $\boldsymbol{p_{1,}p_2, \ldots, p_k}$ each time.

For $i = 1,2, \ldots, k$, let X_i count the number of times the i^{th} type of outcome occurs.

Note that

$$p_1 + p_2 + \cdots + p_k = 1.$$

Also, by construction, we would have

$$X_1 + X_2 + \cdots + X_k = n,$$

and hence we could "drop" one of the random variables (e.g. say X_k) and still recover

$$X_k = n - X_1 - X_2 - \cdots - X_{k-1}.$$

$(X_1, X_2, \ldots, X_k)$ is said to have a **multinomial distribution** with parameters n and $p_1, p_2, \ldots, p_k$.

Derivation of the Joint Probability Function

Similar to the binomial distribution, we first look at the number of ways we can arrange x_1 items of the 1st type, x_2 items of the 2nd type,…, x_k items of the kth type with a total of *n* trials.

That number is

$$\frac{n!}{x_1!\,x_2!\ldots x_k!}$$

Each of these arrangements has probability

$${p_1}^{x_1}{p_2}^{x_2}\ldots{p_k}^{x_k}$$

Putting it altogether, we obtain:

$$\boldsymbol{f(x_1, x_2, \dots, x_k) = \frac{n!}{x_1!x_2!\dots x_k!} p_1{}^{x_1} p_2{}^{x_2} \dots p_k{}^{x_k}}$$

such that each $x_i = 0,1,\dots,n$ and $\sum_{i=1}^{k} x_i = n$.

Of course, it is straightforward to verify

$$\sum_{\substack{x_i=0,1,\dots,n \\ with\ x_1+\cdots+x_k=n}} f(x_1, x_2, \dots, x_k) = 1.$$

using the multinomial theorem sum formula.

Shorthand notation:

$$(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{k}) \sim \mathrm{Multinomial}(\mathrm{n}; \mathrm{p}_1, \mathrm{p}_2, \dots, \mathrm{p}_\mathrm{k})$$

Marginal and Joint Probability Functions

If we are interested in finding the marginal distribution of one r.v., say X_2, in the multinomial distribution, we can use:

1. **Mathematical Approach:**

For this, we would fix the value of x_2 and then sum the joint p.f. over all the other $k-1$ variables, thereby leading to

$$f_2(x_2) = \sum_{all\ x_1, x_3, \dots, x_k} f(x_1, \dots, x_k)$$

for each $x_2 = 0,1, \dots, n$.

This can be algebraically challenging.

2. Intuitive and Simple Approach:

- Intuitively, if we are only interested in X_2 (i.e. the number of occurrences of the 2nd type among n trials), we take note that:
  i. The experiment is repeated n times and $X_2 \in \{0, 1, \ldots, n\}$.
  ii. The probability that the 2nd type of outcome occurs on any trial is p_2 and the probability it does not occur is $1 - p_2$.
  iii. All trials are assumed to be independent of each other.

Hence,

$$\boldsymbol{X_2 \sim Bin(n, p_2)}.$$

What about the distribution of, say,

$$T = X_1 + X_2?$$

Using a similar argument, we could view our *"success"* as an occurrence of one of the first two types of outcomes, whereas anything else is considered a *"failure"*.

Then, it follows that

$$\boldsymbol{T = X_1 + X_2 \sim Bin(n, p_1 + p_2).}$$

Example: The probabilities that a certain electronic component will last less than 50 hours, last between 50 and 90 hours, or last more than 90 hours, are 0.2, 0.5, and 0.3, respectively. The time to failure of eight such components is recorded.

a) What is the probability that one will last less than 50 hours, five will last between 50 and 90 hours, and two will last more than 90 hours?

b) What is the probability that at least 3 components will <u>last between 50 and 90 hours</u>?

Expectation for Multivariate Distributions (Section 9.4)

Definition:

$$E[g(X,Y)] = \sum_{all\ (x,y)} g(x,y)f(x,y)$$

and

$$E[g(X_1, X_2, \dots, X_n)]$$

$$= \sum_{all\ (x_1,x_2,\dots,x_n)} g(x_1, x_2, \dots, x_n)f(x_1, x_2, \dots, x_n)$$

Example: Let the joint p.f. $f(x, y)$ of (X,Y) be given by the following table:

		x		
	$f(x, y)$	0	1	2
y	1	0.1	0.2	0.3
	2	0.2	0.1	0.1

Calculate $E(XY)$.

Important Property of Multivariate Expectation:

$$E[ag_1(X,Y) + bg_2(X,Y)]$$
$$= aE[g_1(X,Y)] + bE[g_2(X,Y)]$$

where a and b are constants and g_1 and g_2 are arbitrary functions.

Note: This result can be naturally extended to more than 2 functions and more than 2 random variables.

Relationships Between Random Variables

We look at two ways of measuring the strength of the **relationship** between two random variables.

Definition: The **covariance** of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)].$$

The above formula is used mainly for definition / interpretation purposes, but not for computational purposes.

For computational purposes, we tend to use another formula we will derive below:

Example: Let the joint p.f. $f(x, y)$ of (X, Y) be given by the following table:

		x			$f_Y(y)$
	$f(x, y)$	0	1	2	
y	1	0.1	0.2	0.3	0.6
	2	0.2	0.1	0.1	0.4
	$f_X(x)$	0.3	0.3	0.4	1

Calculate $Cov(X, Y)$.

You Try:

Let the joint p.f. $f(x, y)$ of (X,Y) be given by the following table:

			x		
	$f(x,y)$	0	1	2	3
	0	0.15	0.1	0.0875	0.0375
y	1	0.1	0.175	0.1125	0
	2	0.0875	0.1125	0	0
	3	0.0375	0	0	0

Calculate $Cov(X, Y)$.

Cov(X,Y) = E(XY) – E(X)E(Y)

E(XY) = 0 + 0 + 0 + 0 + 0 + (1)(1)0.175 + (1)(2)0.1125 + 0 + 0 + (2)(1)0.1125 + 0 + 0 + 0 + 0 + 0 + 0 = 0.625

E(X) = (0)0.375 + (1)0.3875 + (2)0.2 + (3)0.0375 = 0.9

E(Y) = (0)0.375 + (1)0.3875 + (2)0.2 + (3)0.0375 = 0.9

So, Cov(X,Y) = 0.625 – 0.81 = -0.185

You Try:

Let the joint p.f. $f(x, y)$ of (X,Y) be given by the following table:

			x	
	$f(x,y)$	1	2	3
	1	0.09	0.12	0.13
y	2	0.12	0.11	0.11
	3	0.13	0.10	0.09

Calculate *Cov*(*X*,*Y*).

Cov(X,Y) = E(XY) – E(X)E(Y)

E(XY) = (1)(1)0.09 + (1)(2)0.12 + (1)(3)0.13 + (2)(1)0.12 + (2)(2)0.11 + (2)(3)0.11 + (3)(1)0.13 + (3)(2)0.1 + (3)(3)0.09 = 3.86

E(X) = (1)0.34 + (2)0.33 + (3)0.33 = 1.99

E(Y) = (1)0.34 + (2)0.34 + (3)0.32 = 1.98

So, Cov(X,Y) = 3.86 – (1.99)(1.98) = -0.0802

Interpretation of Covariance

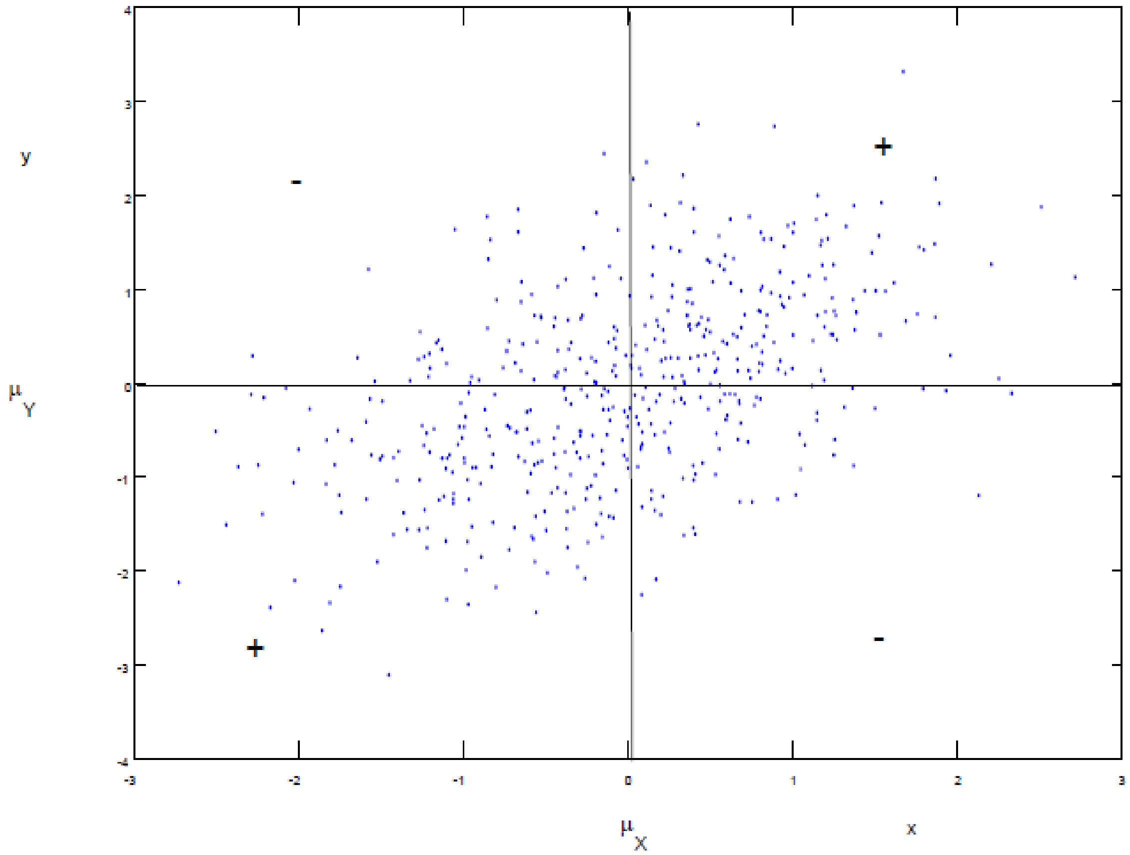


Figure 9.1: Random points (X, Y) with covariance 0.5, variances 1.

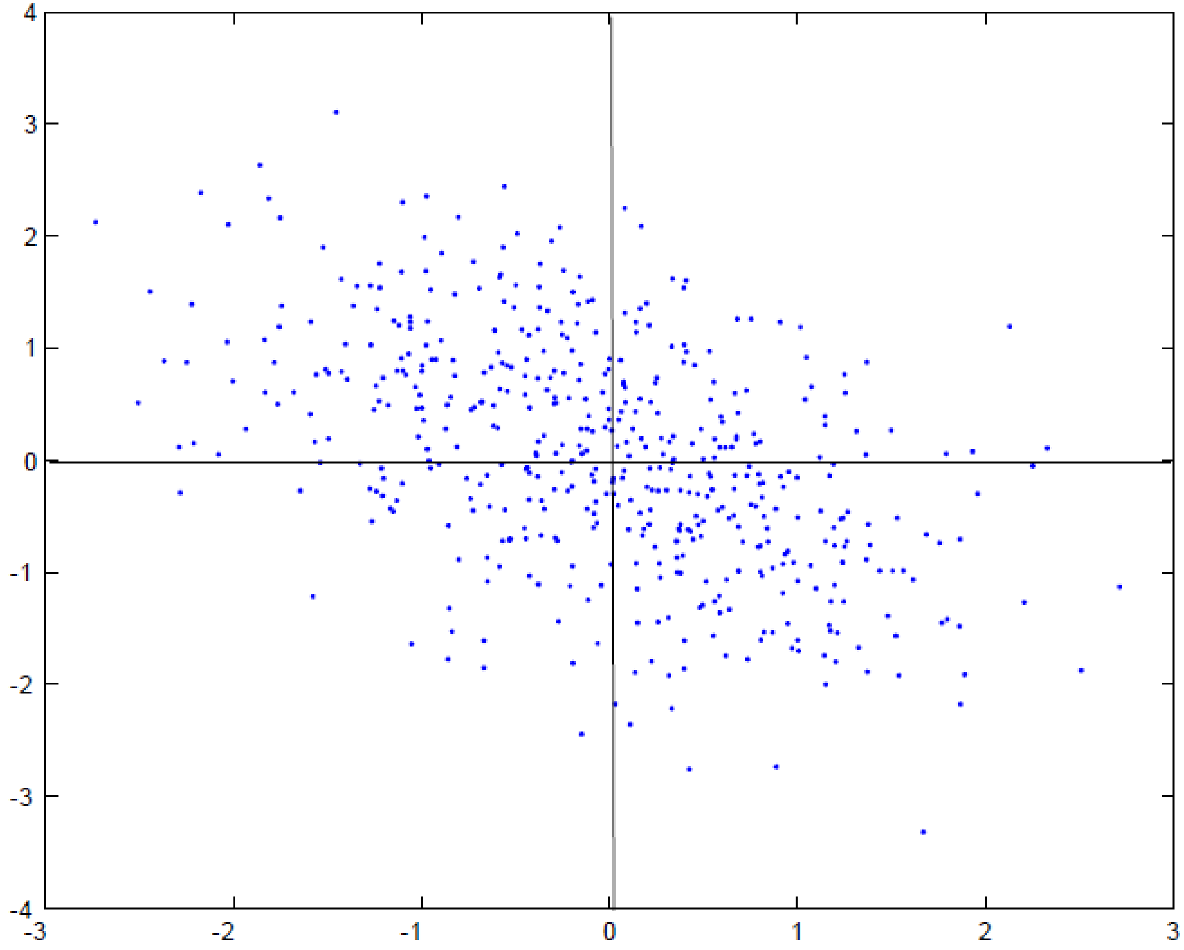


Figure 9.2: Covariance $= -0.5$, variances $= 1$

Theorem: If X and Y are independent random variables, then $E(XY) = E(X)E(Y)$.

Proof:

More generally…

Suppose X and Y are **independent random variables**. If $g_1(X)$ and $g_2(Y)$ are any two arbitrary functions, then

$$\boldsymbol{E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]}.$$

Theorem: If X and Y are independent random variables, then $Cov(X, Y) = 0$.

Proof:

Note: The above result is NOT reversible!

$Cov(X, Y) = 0$ does not imply X and Y are independent.

The Correlation Coefficient

Definition: The **correlation coefficient** between X and Y is

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

This measures the strength of the **LINEAR** relationship between X and Y.

It is essentially a rescaled version of the covariance, which now conveniently lies in the interval [-1, 1].

Properties of $\boldsymbol{\rho}$:

1. Since σ_X and σ_Y are always **positive, $\boldsymbol{\rho}$ will have the same sign as $\boldsymbol{Cov(X,Y)}$.** Hence, the interpretation of the sign of ρ is the same as for $Cov(X,Y)$. Moreover, if $\rho = 0$, then we say X and Y are <u>uncorrelated</u>.

2. $-1 \leq \rho \leq 1$ and as $\boldsymbol{\rho \rightarrow \pm 1}$, then the relationship between X and Y becomes closer to **linear**.

$\boldsymbol{Cov(X,Y)}$ ⇒ interpret the **SIGN**

$\boldsymbol{\rho}$ ⇒ interpret the **MAGNITUDE** and **SIGN**

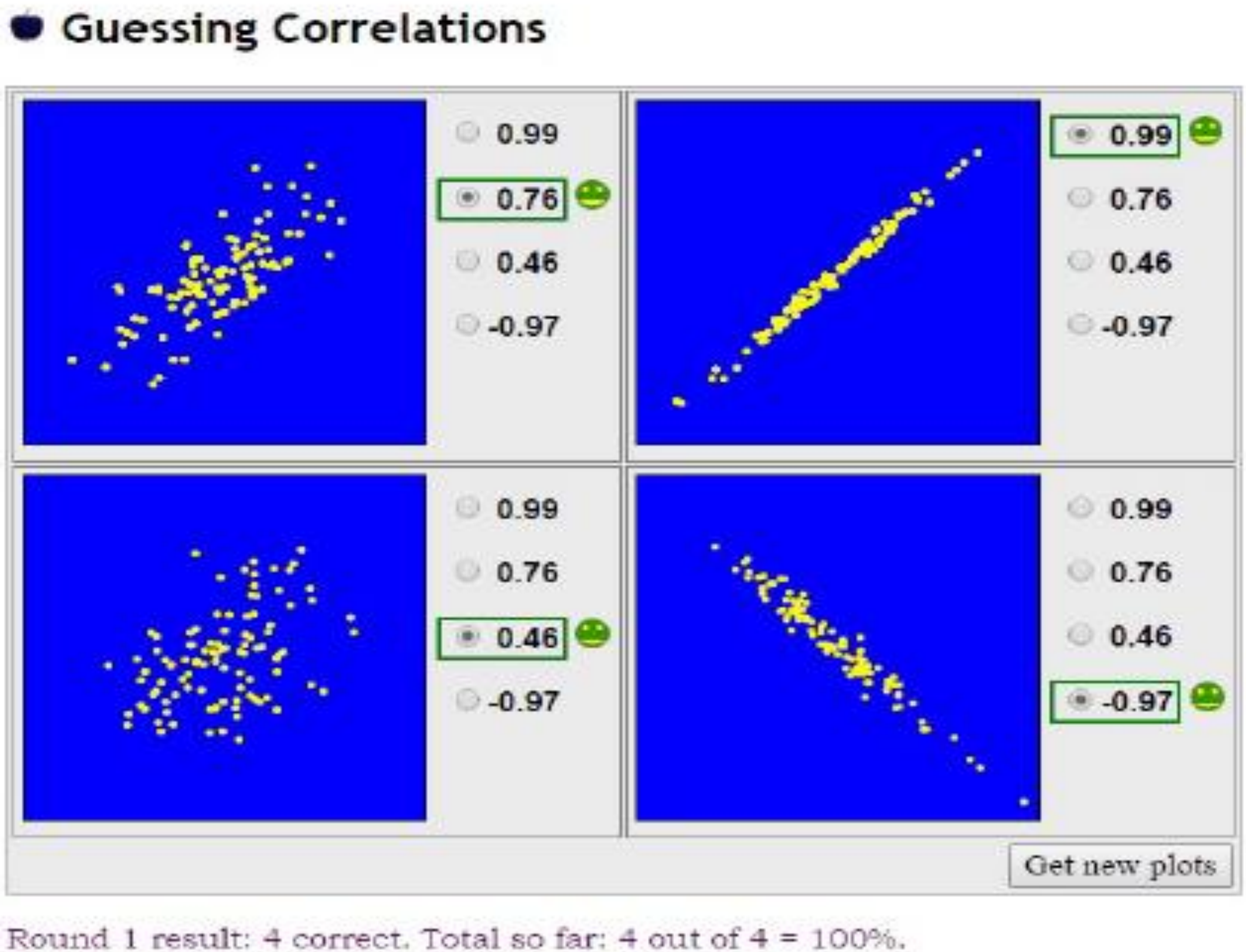
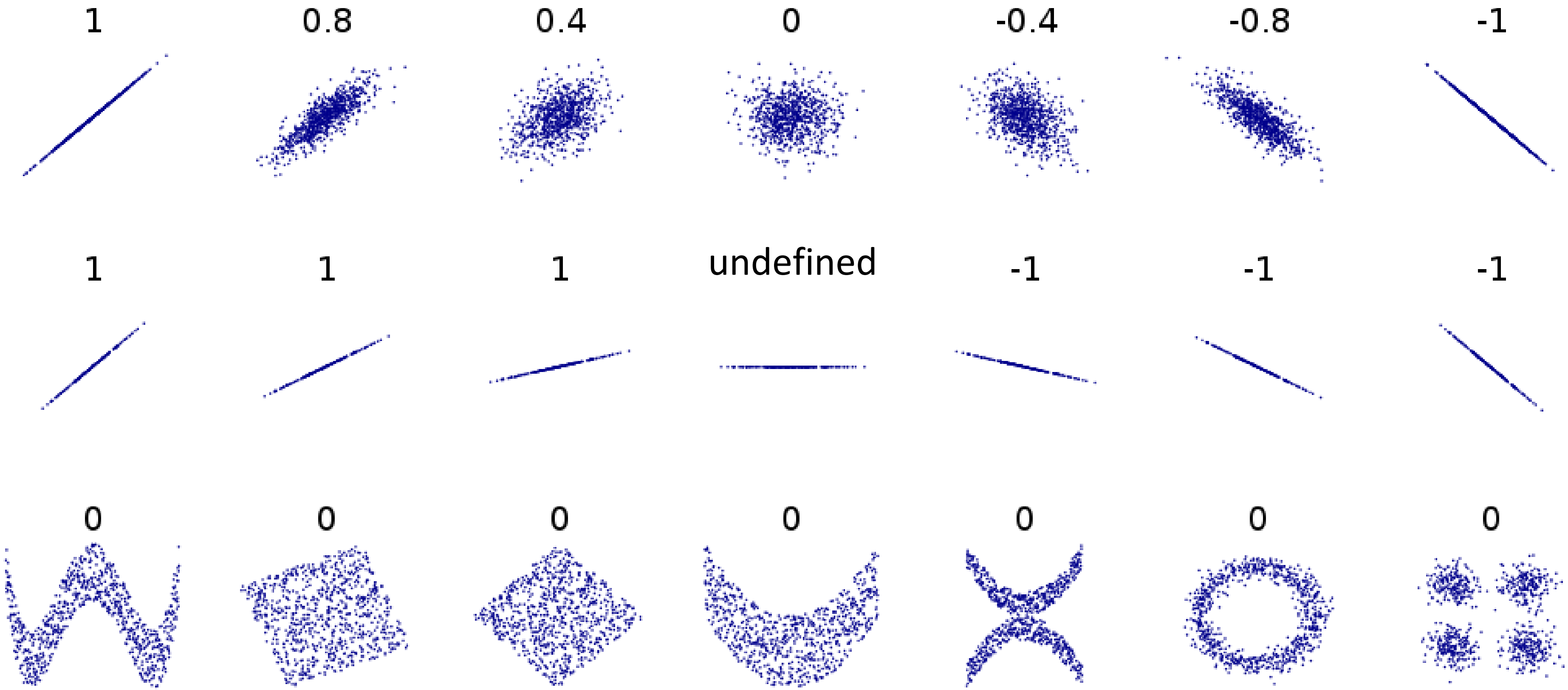


Figure 9.3: Guessing the correlation based on a scatter diagram of points

Correlation Examples



Example: The joint p.f. of (X, Y) is given by

		x			
$f(x,y)$		0	1	2	$f_Y(y)$
y	0	0.06	0.15	0.09	
	1	0.14	0.35	0.21	
	$f_X(x)$				

Calculate the correlation coefficient.

What does this indicate about the relationship between X and Y?

You try:

The joint p.f. of *X* and *Y* is given by:

			x	
	f(*x*,*y*)	1	2	3
	1	0.09	0.12	0.13
y	2	0.12	0.11	0.11
	3	0.13	0.10	0.09

Find the **Corr(X,Y) = ρ** and interpret its value.

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

We calculated Cov(X,Y) in an earlier You try exercise.

We found that **Cov(X,Y) = 3.86 – (1.99)(1.98) = -0.0802**

We will continue the solution on the next slide.

You try Solution (Cont'd)

The joint p.f. of X and Y is given by:

			x	
	$f(x,y)$	1	2	3
	1	0.09	0.12	0.13
y	2	0.12	0.11	0.11
	3	0.13	0.10	0.09

To find Corr(X,Y), we need the standard deviations of X and Y.

$Var(X) = E(X^2) - [E(X)]^2$ and $Var(Y) = E(Y^2) - [E(Y)]^2$

We found E(X) and E(Y) when we calculated the covariance, so we just need $E(X^2)$ and $E(Y^2)$.

$E(X^2) = (1)0.34 + (4)0.33 + (9)0.33 = 4.63$

$Var(X) = 4.63 - (1.99)^2 = 0.6699$

Similarly, $E(Y^2) = (1)0.34 + (4)0.34 + (9)0.32 = 4.58$

$Var(Y) = 4.58 - (1.98)^2 = 0.6596$

So, $Corr(X, Y) = \rho = \dfrac{-0.0802}{\sqrt{(0.6699)(0.6596)}} = -0.12065$

This implies that there is a weak linear relationship between X and Y.